ОБЩИЙ КУРС ФИЗИКИ

Д.В. Сивухин

МЕХАНИКА

ТОМ I

УДК 530.1(075.8)
С34
ББК 22.2

Сивухин Д. В. **Общий курс физики.** Учеб. пособие: Для вузов. В 5 т. Т. I. **Механика.** — 4-е изд., стереот. — М.: ФИЗМАТЛИТ; Изд-во МФТИ, 2005. — 560 с. — ISBN 5-9221-0225-7; 5-89155-078-4.

Первый том курса физики, широко известного у нас в стране и за рубежом. Книга написана на основе лекций, которые в течение ряда лет читались автором студентам Московского физико-технического института. Основное внимание уделено выяснению физического смысла и содержания основных законов и понятий механики, установлению границ применимости этих законов, развитию у студентов навыков физического мышления и умения ставить и решать конкретные задачи.

Третье издание первого тома вышло в 1989 г.

Для студентов физических и математических факультетов университетов, физико-технических и инженерно-физических институтов, а также вузов, где физика является основной дисциплиной.

Учебное издание

СИВУХИН Дмитрий Васильевич

ОБЩИЙ КУРС ФИЗИКИ

Том I
МЕХАНИКА

Редактор *Д. А. Миртова*
Оригинал-макет: *А. И. Чугуров*
Оформление переплета: *А. Ю. Алехина*

ЛР №071930 от 06.07.99. Подписано в печать 15.12.04. Формат 60×90/16. Бумага офсетная № 1. Печать офсетная. Усл. печ. л. 35. Уч.-изд. л. 36,45. Заказ №

Издательская фирма «Физико-математическая литература»
МАИК «Наука/Интерпериодика»
117864 Москва, Профсоюзная, 90
E-mail: fizmat@maik.ru; http://www.fml.ru

Издательство МФТИ
141700 Долгопрудный, Московская обл., Институтский пер., 9

При участии издательской и книготорговой фирмы «ФИЗМАТКНИГА»
141700 Долгопрудный, Московская обл., Институтский пер., 9
Тел.: (095) 408-76-81, 409-93-28

Интернет-магазин физико-математической литературы:
http://www.fizmatkniga.ru

Отпечатано с диапозитивов в РГУП «Чебоксарская типография № 1»
428019 Чебоксары, пр. И. Яковлева, 15

ISBN 5-9221-0225-7 (Т. I)
ISBN 5-9221-0229-X
ISBN 5-89155-078-4 (Т. I)
ISBN 5-89155-077-6

© ФИЗМАТЛИТ, 1989, 2002, 2005
© Издательство МФТИ, 2002, 2005

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ К ЧЕТВЕРТОМУ ИЗДАНИЮ 7
ПРЕДИСЛОВИЕ К ТРЕТЬЕМУ ИЗДАНИЮ 8
ПРЕДИСЛОВИЕ К ПЕРВОМУ ИЗДАНИЮ 9
ВВЕДЕНИЕ 12

ГЛАВА I

КИНЕМАТИКА

§ 1. Пространство и время 19
§ 2. Кинематическое описание движения. Материальная точка 32
§ 3. Скорость и ускорение при прямолинейном движении. Угловая скорость и угловое ускорение 34
§ 4. Скорость и ускорение при криволинейном движении 38
§ 5. Границы применимости классического способа описания движения 47
§ 6. О смысле производной и интеграла в приложениях к физическим вопросам 49
§ 7. О векторах и сложении движений 53
§ 8. Степени свободы и обобщенные координаты 65

ГЛАВА II

ЗАКОНЫ НЬЮТОНА

§ 9. Закон инерции. Инерциальная система отсчета 69
§ 10. Масса. Закон сохранения импульса 73
§ 11. Второй закон Ньютона. Сила 76
§ 12. Третий закон Ньютона и закон сохранения импульса 84
§ 13. Взаимодействие тел на расстоянии и полевое взаимодействие . . . 89
§ 14. Роль начальных условий 94
§ 15. Принцип относительности Галилея 97
§ 16. Аддитивность и закон сохранения массы 103
§ 17. О законах трения 105

ГЛАВА III

НЕКОТОРЫЕ СЛЕДСТВИЯ И ПРИМЕНЕНИЯ ЗАКОНОВ НЬЮТОНА

§ 18. Импульс силы и изменение импульса 112
§ 19. Теорема о движении центра масс 115

§ 20. Приведенная масса 117
§ 21. Движение тел с переменной массой. Реактивное движение 119

ГЛАВА IV

РАБОТА И ЭНЕРГИЯ

§ 22. Работа и кинетическая энергия 129
§ 23. Связь между кинетическими энергиями в различных системах отсчета. Теорема Кёнига 137
§ 24. Консервативные и неконсервативные силы 138
§ 25. Потенциальная энергия. Закон сохранения энергии в механике . 143
§ 26. Абсолютно неупругий удар 151
§ 27. Внутренняя энергия. Общефизический закон сохранения энергии 155
§ 28. Абсолютно упругий удар 157
§ 29. Силы и потенциальная энергия 168

ГЛАВА V

МОМЕНТ ИМПУЛЬСА

§ 30. Момент силы и момент импульса относительно неподвижного начала 176
§ 31. Связь момента импульса материальной точки с секториальной скоростью. Теорема площадей 181
§ 32. Момент импульса и момент сил относительно неподвижной оси . 182
§ 33. Уравнение момента импульса для вращения вокруг неподвижной оси. Момент инерции 184
§ 34. Примеры на закон сохранения вращательного импульса 186
§ 35. Теорема Гюйгенса—Штейнера 192
§ 36. Вычисление моментов инерции 193
§ 37. Уравнение моментов относительно движущегося начала и движущейся оси 200
§ 38. Законы сохранения и симметрия пространства и времени 210

ГЛАВА VI

ГАРМОНИЧЕСКИЕ КОЛЕБАНИЯ

§ 39. Кинематика гармонического колебательного движения 215
§ 40. Гармонические колебания груза на пружине 216
§ 41. Физический маятник 220
§ 42. Бифилярный и трифилярный подвесы 224
§ 43. Адиабатические инварианты 234

ГЛАВА VII

МЕХАНИКА ТВЕРДОГО ТЕЛА

§ 44. Твердое тело в механике. Уравнения движения и равновесия твердого тела 242
§ 45. Мгновенная ось вращения 245
§ 46. Угловая скорость как вектор. Сложение вращений 248
§ 47. Теорема Эйлера. Общее движение твердого тела 258
§ 48. Скатывание тел с наклонной плоскости 262
§ 49. Гироскопы. Движение свободного гироскопа 278

§ 81. Скорость распространения продольных упругих возмущений в стержнях 435
§ 82. Применения принципа суперпозиции 441
§ 83. Скорости распространения продольных и поперечных возмущений в неограниченной среде 447
§ 84. Скорость распространения поперечных возмущений в натянутом шнуре 450
§ 85. Скорость распространения звука в жидкостях и газах 452

ГЛАВА XI

МЕТОДЫ ПОДОБИЯ И РАЗМЕРНОСТИ

§ 86. Размерность и системы единиц 456
§ 87. Формула размерности 458
§ 88. Правило размерности 463

ГЛАВА XII

МЕХАНИКА ЖИДКОСТЕЙ И ГАЗОВ

§ 89. Общие свойства жидкостей и газов 468
§ 90. Основные уравнения равновесия и движения жидкостей 472
§ 91. Гидростатика несжимаемой жидкости 475
§ 92. Барометрическая формула 484
§ 93. Кинематическое описание движения жидкости 487
§ 94. Стационарное движение идеальной жидкости. Уравнение Бернулли 489
§ 95. Примеры на применение уравнения Бернулли. Формула Торричелли 496
§ 96. Вязкость 499
§ 97. Стационарное течение жидкости по прямолинейной трубе. Формула Пуазейля 506
§ 98. Законы гидродинамического подобия 511
§ 99. Турбулентность и гидродинамическая неустойчивость 516
§ 100. Парадокс Даламбера. Разрывные течения 520
§ 101. Применение теории размерности 525
§ 102. Потенциальные и вихревые движения 527
§ 103. Пограничный слой и явление отрыва 530
§ 104. Подъемная сила крыла самолета 536
§ 105. Эффект Магнуса 541
ПРИЛОЖЕНИЕ 468
ИМЕННОЙ УКАЗАТЕЛЬ 554
ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ 555

ПРЕДИСЛОВИЕ К ЧЕТВЕРТОМУ ИЗДАНИЮ

Четвертое издание первого тома выходит уже после смерти создателя фундаментального «Общего курса физики», выдающегося педагога и подвижника-ученого Д. В. Сивухина. За многие годы, прошедшие со дня опубликования первого издания, книга стала неотъемлемой частью учебного процесса в вузах, готовящих инженеров по специальностям, требующим глубокой подготовки по физике. За эти годы книга нисколько не устарела и не потеряла своей привлекательности, поскольку автор акцентирует внимание читателя не на описательной стороне физических демонстраций, а на физическом смысле рассматриваемых явлений и понятий физики.

Данное учебное пособие выгодно отличает стройность и ясность логических построений. Широкий охват рассматриваемых вопросов при их глубоком изучении делает книгу сравнимой с физической энциклопедией.

Четвертое издание ничем не отличается от третьего, вышедшего в издательстве «Наука» в 1989 г. В текст лишь внесены необходимые несущественные уточнения с целью исправления допущенных опечаток.

Издательство МФТИ

ПРЕДИСЛОВИЕ К ТРЕТЬЕМУ ИЗДАНИЮ

Третье издание первого тома выходит уже после написания всех пяти томов «Общего курса физики», поэтому автор имел возможность внимательно просмотреть весь первый том и внести в него изменения, которые, по его мнению, улучшили изложение.

Впрочем, план и построение курса остались прежними. Кое-что добавлено, исключено и заменено. Общее число и нумерация параграфов и пунктов остались без изменения.

Для третьего издания книга была тщательно прорецензирована профессором С. А. Ахмановым и доцентом С. С. Чесноковым, которые сделали ряд существенных замечаний, учтенных автором. Выражаю им искреннюю благодарность.

ПРЕДИСЛОВИЕ К ПЕРВОМУ ИЗДАНИЮ

Основное содержание предлагаемого курса составляет расширенное изложение лекций по физике, которые автор читал в течение многих лет (начиная с 1956 г.) в Московском физико-техническом институте. Общий план лекционного курса, а также основной подход к изложению принципиальных вопросов физики на протяжении всех лет менялись мало. Однако с каждым годом курс обновлялся включением новых вопросов частного порядка и примеров; многие ранее рассматривавшиеся вопросы при этом исключались. Делалось это не по принципиальным соображениям, а из-за недостатка времени.

В настоящий курс вошли практически все вопросы, излагавшиеся на лекциях в разные годы. Вошли также и такие вопросы, которые на лекциях не излагались. Они занимают около 10—15% текста. Кроме того, включено много задач с ответами или подробными решениями. Весь этот материал может оказать пользу студентам при углубленном изучении физики и преподавателям при проведении семинарских занятий. Он, как надеется автор, будет способствовать развитию у студентов навыков физического мышления и умения самостоятельно ставить и решать принципиальные вопросы и конкретные физические задачи, что и является главной целью предлагаемого руководства. Разумеется, не весь этот материал является обязательным. Для удобства читателя основной материал крупным шрифтом, вспомогательный — мелким.

В устном преподавании большое внимание уделялось лекционным демонстрациям. Лекционные демонстрации, а также лабораторные работы являются наилучшим средством для наглядного ознакомления студентов с физическими явлениями. Их не может заменить никакое самое образное и точное описание явления. Лекционные демонстрации, как очень важная и неотъемлемая часть лекционного курса, дают необходимый опытный материал, на основе которого вводятся физические понятия и обобщения, формулируются и уясняются физические законы и принципы. Кроме того, они оживляют читаемый курс и придают ему эстетическую привлекательность.

В книге изложение носит несколько иной характер. Много демонстраций опущено, а описание остальных дано схематично без указания технических и экспериментальных подробностей. На первый план выступает логическая сторона предмета. Главное внимание обращается на выяснение физического смысла и содержания основных положений и понятий физики. Много внимания уделяется установлению границ применимости физических законов, а также идеализированных моделей и схем, применяемых в физике.

В предлагаемом томе курса физики дано систематическое изложение физических основ классической нерелятивистской механики. Однако излагать физические основы механики без всякой связи с другими разделами физики невозможно. Поэтому уже в первом томе дается некоторое представление об идеях теории относительности и квантовой механики. Без этого невозможно точно установить границы применимости классической нерелятивистской механики. Необходимые сведения из квантовой механики даны кратко без какого бы то ни было обоснования и обсуждения. О релятивистской механике или механике теории относительности говорится более подробно. В основу изложения положена зависимость массы от скорости, рассматриваемая как экспериментальный факт. Этого достаточно не только для качественного рассмотрения, но и для количественного решения простейших задач, примеры которых приводятся в книге. Однако систематическое изложение вопросов релятивистской и квантовой механики будет дано в других томах курса — после изложения электрических и оптических явлений.

В следующих томах курса предполагается изложить термодинамику и молекулярную физику, электричество, оптику, атомную и ядерную физику.

Основной системой единиц, принятой в курсе, является система СГС. В механике она по существу не отличается от Международной системы единиц СИ. Различие между ними здесь чисто количественное и сводится к различному выбору масштабов единиц длины и массы (вместо сантиметра (в системе СГС) в системе СИ используется метр, вместо грамма — килограмм). Это не сказывается на смысле физических понятий и виде физических формул, а потому в механике обе системы одинаково удобны. Не так обстоит дело в электродинамике. В системе СИ электрическое состояние вакуума характеризуется четырьмя векторами: напряженностью электрического поля **E** и электрическим смещением **D**, напряженностью **H** и магнитной индукцией **B** магнитного поля. Это с физической точки зрения является ненужным и противоестественным усложнением. В системе СГС для характеристики вакуума достаточно двух векторов. Все четыре вектора **E**, **B**, **H**, **D** в системе СГС имеют одинаковую размерность, в системе СИ их размерности разные. Это также противоестественно, поскольку электрическое и магнитное поле тесно связаны между собой. В наиболее общем виде связь между этими векторами устанавливается в релятивистской электродинамике. Она объединяет векторы **E** и **B**, а также **H** и **D** в четырехмерные антисимметричные тензоры. При составлении таких тензоров в системе СИ надо вводить специальные размерные множители для уравнивания размерностей их компонентов. Это также ненужное усложнение. Можно указать только одно преимущество системы СИ по сравнению с системой СГС. В систему СИ органически входят ампер, вольт, ом и все производные этих единиц, получившие благодаря историческим случайностям широкое распространение в элек-

тротехнике. Это преимущество практическое, но не принципиальное. Однако чисто измерительные вопросы будут занимать в настоящем курсе ничтожное место. Ради них нет смысла приносить в жертву стройность и логичность физических понятий и формул, которыми характеризуется система СГС.

Появление настоящего курса было бы невозможно, если бы профессор Г. С. Горелик, возглавлявший кафедру физики Московского физико-технического института, не привлек автора к чтению лекций по общему курсу физики. Автор обсуждал с ним различные вопросы физики и ее преподавания.

Демонстрации для лекций готовили и осуществляли мои ассистенты М. И. Маклаков, В. А. Кузнецова, Е. Н. Морозов, В. П. Молчанов, Л. Д. Кудряшева, Г. Н. Фрейберг. Хотя далеко не все из этих демонстраций описаны в курсе, все они дали ценный материал, использованный при его написании.

Значительная часть задач, включенных в курс, предлагалась студентам на письменных экзаменах, а затем использовалась на семинарских занятиях. В их составлении принимали участие многие преподаватели кафедры физики МФТИ.

Рукопись настоящего тома была частично просмотрена академиком В. Л. Гинзбургом, Б. Б. Кадомцевым, М. А. Леонтовичем, Р. З. Сагдеевым; профессорами С. С. Герштейном и Н. А. Яковлевым. Она подверглась внимательному рецензированию и обсуждению на кафедре общей физики для механико-математического факультета Московского государственного университета им. М. В. Ломоносова, возглавлявшейся профессором С. П. Стрелковым.

Доброжелательная критика, советы и пожелания всех лиц, упомянутых в настоящем предисловии, немало способствовали улучшению книги. Всем им автор выражает глубокую благодарность.

ВВЕДЕНИЕ

1. При изучении всякого круга явлений очень важно установить основные законы или принципы, с помощью которых можно объяснить все известные явления из рассматриваемого круга, а также предсказать новые. Такой подход к изучению явлений природы получил название *метода принципов*. Основоположником его в физике является великий Исаак Ньютон (1643—1727). Непревзойденным мастером метода принципов был и великий физик Альберт Эйнштейн (1879—1955).

Сами основные законы или принципы не могут быть доказаны логически. Их доказательством является опыт. Основное значение имеют не столько опыты по непосредственной проверке самих принципов, сколько опыты, на которых проверяются вытекающие из них следствия. В этом смысле основные принципы являются *обобщениями опытных фактов.* Но никакие опыты никогда не охватывают все разнообразие условий, в которых могут протекать явления, а измерения всегда сопровождаются погрешностями. Поэтому *опытным путем* (а другого пути нет) *можно установить справедливость принципов лишь в ограниченных пределах и с ограниченной точностью.* При расширении круга изучаемых явлений и повышения точности измерений могут расшириться и эти пределы. Но может случиться, что вне определенных границ основные принципы перестанут быть справедливыми. Тогда возникнет необходимость в их обобщении или замене новыми принципами, имеющими более широкую область применимости. Старые принципы при этом не утратят своего значения. Но ими можно будет пользоваться только внутри установленной области применимости. Сила и привлекательность метода принципов в том, что весь материал, полученный из основных принципов логическим и математическим путем, является достоверным, конечно, в пределах тех границ, в которых основные принципы доказаны экспериментально, и с той точностью, с которой были выполнены эксперименты.

2. Раньше других разделов физики развилась *механика. Механика есть наука о движении и равновесии тел. В широком смысле слова движение материи есть всякое изменение ее. Однако под движением в механике понимается только простейшая форма его, а именно перемещение тела относительно других тел.* Принципы механики впервые были сформулированы Ньютоном в его основном сочинении «Математические начала натуральной философии», первое издание которого вышло в 1687 г. Ньютон имел, правда, много крупных предшественников: Архимеда (ок. 287—212 до н. э.), Иоганна Кеплера (1571—1630), Галилео Галилея (1564—1642), Христиана Гюйгенса (1629—1695) и др., решивших немало частных вопросов статики и отчасти динамики. Однако Ньютон был первым,

кто сформулировал полную систему принципов механики и на их основе воздвиг стройное здание этой науки. Громадные достижения механики Ньютона, а также его непререкаемый научный авторитет почти на 200 лет отвлекли внимание ученых от недостатков его системы механики. Серьезное критическое отношение к механике Ньютона возникло лишь во второй половине XIX века.

После Ньютона механика быстро развивалась. Однако до начала XX века это развитие шло в основном в направлении совершенствования математических методов механики и применения ее законов ко все новым и новым областям знания. Оно не затрагивало содержания основных принципов и физические представления механики Ньютона. Ничего принципиально нового в физические основы механики внесено не было вплоть до XX века, когда положение изменилось.

Хотя механика Ньютона и покоится на прочном фундаменте экспериментальных фактов, однако все они относятся к медленным движениям макроскопических тел. *Макроскопическими* называют обычные тела, окружающие нас, т. е. тела, состоящие из громадного количества молекул или атомов. Под *медленными* или *нерелятивистскими движениями* понимают движения, скорости которых очень малы по сравнению со скоростью света в вакууме $c = 300\,000$ км/с. Движения, скорости которых приближаются к скорости света в вакууме, называют *быстрыми* или *релятивистскими*. В этом смысле движение спутника или космического корабля со скоростью $v = 8$ км/с является еще очень медленным. В том же смысле очень медленными движениями являются движения планет Солнечной системы, их спутников и комет относительно Солнца. Применяя к таким телам принципы механики Ньютона, удалось объяснить и предсказать их движение в полном соответствии с наблюдениями. Это явилось первым и притом наиболее убедительным доказательством справедливости механики Ньютона. Движение искусственных спутников и космических кораблей также находится в полном соответствии с расчетами, производимыми на основе механики Ньютона.

3. Можно ли экстраполировать на случай быстрых движений принципы механики Ньютона, экспериментально установленные для медленных движений макроскопических тел? Можно ли применять основные понятия и принципы механики Ньютона к явлениям микромира, т. е. явлениям, происходящим с отдельными молекулами, атомами, электронами, протонами, нейтронами и прочими «элементарными частицами»? На эти вопросы логически ответить сразу нельзя. Ответ на них могут дать лишь опыты с быстро движущимися телами, а также опыты с отдельными атомами, электронами и пр. Такие опыты были поставлены только в XX веке. Они показали, что на оба вопроса в общем следует дать отрицательный ответ.

Теория относительности Эйнштейна предсказала, а опыт подтвердил это предсказание, что механика Ньютона не может быть

применима к движениям частиц, скорости которых близки к скорости света в вакууме. На основе теории относительности была создана новая механика, применимая не только к медленным, но и к сколь угодно быстрым движениям. Она называется *релятивистской механикой*, или *механикой теории относительности*. Согласно механике Ньютона скорость, до которой можно ускорить тело из состояния покоя, в принципе ничем не ограничена. По релятивистской механике значение скорости ускоряемого тела не может перейти через определенный предел, равный скорости света в вакууме c. В этом смысле скорость света c является *предельной*. Скорость тела не может ее достигнуть, но в принципе может подойти к ней сколь угодно близко. В современных ускорителях, например, получаются частицы с энергией порядка 500 ГэВ ($5 \cdot 10^{14}$ эВ). Скорость протонов таких энергий меньше скорости света всего на $5{,}25 \cdot 10$ см/с, а скорость электронов — на $1{,}5 \cdot 10^{-2}$ см/с. В космических лучах регистрировались протоны, скорость которых меньше скорости света всего на величину порядка 10^{-8} см/с. К движениям таких быстрых частиц нерелятивистская механика Ньютона совершенно не применима. Ускорители рассчитываются на основе релятивистской механики Эйнштейна, и то обстоятельство, что они работают в согласии с расчетами, является одним из наиболее убедительных и прямых экспериментальных доказательств правильности релятивистской механики.

4. Теория относительности установила границы применимости ньютоновской механики со стороны *больших скоростей*. Другое ограничение, и притом не только ньютоновской, но и релятивистской макроскопической механики, было получено в результате изучения *микромира* — мира атомов, молекул, электронов и пр.

При изучении микромира сначала применяли понятия и законы, введенные и установленные для макроскопических тел. Электрон, например, рассматривался как твердый или деформируемый шарик, по объему которого распределен электрический заряд. Считалось, что поведение электрона управляется теми же законами механики и электродинамики, которые были экспериментально установлены для макроскопических электрически заряженных тел; что все понятия и законы макроскопической физики применимы и имеют смысл для тел сколь угодно малых размеров и сколь угодно малых промежутков времени; что для понимания явлений микромира не требуется новых понятий и законов, помимо тех, которыми располагает макроскопическая физика, т.е. микромир рассматривался просто как уменьшенная копия макромира. Такой подход к изучению явлений природы и теории, основанные на нем, называются *классическими*.

Вопрос о применимости или неприменимости классического подхода к изучению микромира не может быть решен умозрительно. На этот вопрос может ответить только опыт. Опыты показали,

что *классический подход к изучению явлений микромира не применим*, или точнее, *его применимость к этому кругу явлений ограничена.* Адекватное описание явлений микромира (применимое, конечно, также в каких-то пределах) дает *квантовая механика*, существенно отличающаяся от механики классической. Квантовая механика вводит радикальные изменения в наши представления о движении. Так, классическая картина движения частицы вдоль траектории, в каждой точке которой частица имеет определенную скорость, в общем случае не применима при описании движения микрочастиц. Движение в микромире является более сложной формой движения, чем механическое перемещение тел в пространстве. Вообще, описание явлений в квантовой механике лишено наглядности в том смысле, что здесь требуются *принципиально новые* представления и понятия, не сводимые к привычным представлениям и понятиям, возникающим при изучении макроскопических объектов. Поскольку наш курс механики посвящен изучению движения макроскопических тел, нет необходимости останавливаться на дальнейшей характеристике квантовой механики. Достаточно указать границы применимости понятий и законов, которыми мы будем пользоваться (см. § 5.)

5. Таким образом, *механика Ньютона может быть охарактеризована как классическая нерелятивистская механика. Это значит, что она изучает медленные движения макроскопических тел.*

Релятивистская и квантовая механики являются более общими теориями, чем механика Ньютона. Последняя содержится в них как приближенный предельный случай. Релятивистская механика переходит в механику Ньютона в случае медленных движений. Квантовая механика переходит в механику Ньютона в случае медленных движений. Квантовая механика переходит в механику Ньютона в случае тел достаточно больших масс, движущихся в достаточно плавно меняющихся силовых полях. Это не означает, что механика Ньютона утратила свое значение. Во многих случаях фактические изменения, вносимые теорией относительности и квантовой механикой, сводятся к небольшим поправкам к ньютоновской механике. Они называются соответственно *релятивистскими* и *квантовыми.* Эти поправки в случае обычных медленных движений макроскопических тел столь ничтожны, что, как правило, далеко выходят за пределы точности самых тонких физических измерений. Кроме того, уже простейшие задачи на движение макроскопических тел, с которыми механика Ньютона легко справляется, привели бы к непреодолимым математическим трудностям при попытке найти их точные решения методами релятивистской и квантовой механик. Чтобы практически получить решение, надо было бы ввести упрощения и перейти к приближенным методам, а это по своему результату эквивалентно переходу к механике Ньютона.

Если, например, движение космического корабля относительно Земли рассчитывается по законам механики Ньютона, не учитывающей релятивистские эффекты, то при скорости корабля $v = 8$ км/с возникающая вследствие этого относительная погрешность будет порядка $(v/c)^2 = (8/300\,000)^2 \sim 10^{-9}$. Таким образом, здесь механика Ньютона обеспечивает погрешность вычислений до $10^{-7}\%$. Вводить в подобных случаях релятивистские поправки не только не нужно, но и иллюзорно, хотя бы уже потому, что входные параметры, необходимые при расчетах, могут быть определены с несравненно меньшей точностью. Кроме того, в этом нет практической необходимости.

Таким образом, механика Ньютона имеет очень широкую и практически важную область применимости. В пределах этой области она никогда не утратит своего научного и практического значения. Отказываться от механики Ньютона надо лишь вне области ее применимости, когда она приводит либо к неверным, либо к недостаточно точным результатам. Такова, например, задача о движении заряженных частиц в ускорителях, где надо пользоваться релятивистской механикой. Таковы задачи о движении электронов в атомах, которые надо решать с помощью квантовой механики.

6. Чтобы дать представление о величине поправок, вносимых общей теорией относительности в механику Ньютона, полезно рассмотреть движение планет Солнечной системы. Скорость планеты очень мала по сравнению со скоростью света в вакууме. В этом смысле движение планеты является медленным; к нему применима механика Ньютона. На основе применения закона всемирного тяготения (сила гравитационного притяжения двух точечных тел пропорциональна их массам и обратно пропорциональна квадрату расстояния между ними), ньютоновская механика позволяет прийти к заключению, что планета движется вокруг Солнца по замкнутой траектории, а именно по эллипсу. Астрономические наблюдения в этот результат вносят лишь маленькую поправку. Траектория планеты не совсем замкнута. Планета практически вращается по эллипсу, но его перигелий (ближайшая к Солнцу точка) медленно смещается в том же направлении, в котором вращается сама планета. Это явления называется *прецессией* орбиты. Наибольшее смещение перигелия наблюдается у Меркурия. Меркурий находится ближе всего к Солнцу, т. е. солнечное гравитационное поле для него громадно. Кроме того, его траектория наиболее вытянута. Поэтому в дальнейшем речь будет идти в основном о смещении перигелия Меркурия.

Замкнутая эллиптическая траектория Меркурия в ньютоновской механике получается в предположении, что Меркурий подвергается гравитационному воздействию одного лишь Солнца — воздействием остальных планет пренебрегается; возмущающее же действие всех остальных планет приводит к смещению перигелия Меркурия в

нужном направлении. Это и есть основная, но не единственная причина этого явления. Французский астроном Урбен Леверье (1811—1877) рассчитал орбиту Меркурия на основе механики Ньютона и закона всемирного тяготения. Он нашел, что наибольший вклад в смещение вносит ближайшая к Меркурию планета Венера (280,6″ в столетие), затем идет самая большая планета Юпитер (152,6″), Земля (83,6″), Сатурн (7,2″), Марс (2,6″), Уран (0,1″). Общий вклад всех планет в смещение составил 527″ в столетие. Действительное смещение в столетие — по Леверье — оказалось на 39″ больше.

Американский астроном Саймон Ньюком (1835—1909) в 1882 г. подтвердил и уточнил вычисления Леверье: вместо 39″ он получил число 42,95″, которое и применяется в настоящее время в расчетах. Это число получило название *аномального смещения* перигелия Меркурия. Таким образом, полное смещение перигелия Меркурия составляет 5,70″ в год, а период обращения — около 227 000 лет.

Было немало попыток объяснить аномальное смещение перигелия Меркурия, оставаясь в рамках ньютоновской механики. Леверье ввел гипотезу, что аномальное смещение вызывается неизвестной планетой (Вулканом), находящейся между Солнцем и орбитой Меркурия. Хуго Зелигер (1849—1924) пытался объяснить недостающее смещение влиянием околосолнечного вещества, вызывающего зодиакальный свет. Многие ученые (в их числе и Ньюком) вводили различные малые поправки в закон всемирного тяготения Ньютона. Так, делались попытки заменить в формуле $f \propto 1/r^2$ показатель на $2 + \delta$, где δ — малая поправка. Вводились предположения о зависимости гравитационной силы от скорости и ускорения взаимодействующих частиц. Но все это представляет только исторический интерес.

Аномальное смещение перигелия Меркурия без каких-либо специальных предположений автоматически получается в общей теории относительности, окончательно созданной Эйнштейном к концу 1925 г. Эта теория органически включает и теорию тяготения, переходящую в ньютоновскую в слабых гравитационных полях. В отношении смещения перигелия Меркурия получается полное количественное согласие теории с опытом. Аномальное смещение испытывает не только перигелий Меркурия, но и перигелии других планет. Однако для этих планет смещение значительно меньше. Так, аномальное смещение перигелия земной орбиты составляет около 4″ в столетие.

Прецессию совершают и орбиты компонентов двойных звезд. Это есть также эффект общей теории относительности. В 1981 г. такую прецессию обнаружил Х. Ф. Халиуллин (р. 1942) в Государственном астрономическом институте им. П. К. Штернберга в движении затменной двойной звезды EK Цефея. Орбита каждой компоненты

двойной звезды практически не отличается от эллиптической. Но ось эллипса медленно поворачивается в плоскости орбиты на 4,6′ в год, т. е. период полной процессии составляет 4700 лет (для Меркурия — 227 000 лет).

В настоящее время изучаются и другие, более тонкие, эффекты общей теории относительности при движении близких планет Солнечной системы и их спутников. Возможность такого изучения обусловлена двумя обстоятельствами. Во-первых, движение системы взаимодействующих тел в общей теории относительности описывается сложной системой нелинейных уравнений, которые аналитически решены быть не могут. Зато их можно решить численно с помощью ЭВМ. Во-вторых, вычисленные малые отступления от механики Ньютона могут быть обнаружены благодаря возросшей точности измерения расстояний с использованием лазерной техники.

ГЛАВА I

КИНЕМАТИКА

§ 1. ПРОСТРАНСТВО И ВРЕМЯ

1. Как уже было сказано во введении, в механике *движением* называют изменения положения тел в пространстве с течением времени. Под положением здесь понимается *относительное положение*, т. е. положение тела относительно других тел. Понятие абсолютного положения, т. е. положения тела в каком-то «абсолютном пространстве» безотносительно к другим телам, лишено смысла.

Тело или система тел, относительно которых определяется положение остальных тел, называется *пространственной системой отсчета.*

Утверждение, что два различных не одновременных события произошли в одном и том же месте пространства, не имеет смысла, пока не указана система отсчета, в которой события рассматриваются. Пассажир в движущемся железнодорожном вагоне взял из своего чемодана какую-то вещь и спустя некоторое время положил ее обратно. Можно сказать, что он взял и положил эту вещь в одном и том же месте, если за систему отсчета принять движущийся вагон. Но те же два события будут происходить в различных местах, если их рассматривать в системе отсчета, связанной с полотном железной дороги. Например, одно событие внутри вагона могло произойти в Москве, а другое — в Ленинграде.

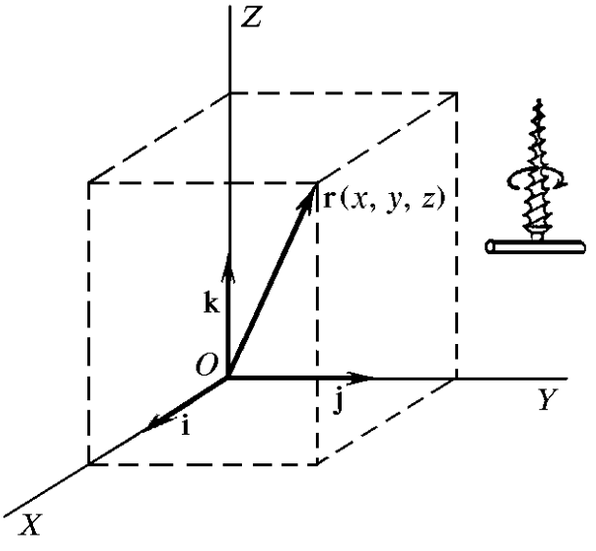


Рис. 1

2. В качестве пространственной системы отсчета можно взять произвольное твердое тело и связать с ним *координатные оси*, например, декартовой прямоугольной системы координат, реализованные в виде трех взаимно перпендикулярных твердых стержней. Положение каждой точки в избранной пространственной системе отсчета можно задавать тремя числами — *координатами точки* x, y, z, представляющими собой расстояния этой точки до координатных плоскостей YZ, ZX, XY соответственно (рис. 1). Три координаты x, y, z можно объединить в один *направленный отре-*

зок, называемый *радиусом-вектором* $\mathbf{r}$, проведенный из начала координат в рассматриваемую точку. Координаты x, y, z являются его проекциями на координатные оси, а потому

$$\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}, \tag{1.1}$$

где $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ — *координатные орты*, т. е. единичные векторы, направленные вдоль координатных осей X, Y, Z.

Существуют два вида координатных систем — *правая* и *левая* системы. Их различают с помощью *правила буравчика*. Будем ввинчивать буравчик с правой нарезкой, вращая его ручку в плоскости XY кратчайшим путем от положительного конца оси X к положительному концу оси Y. В правой системе координат (рис. 1) поступательное перемещение буравчика будет происходить в положительном, а в левой (рис. 2) — в отрицательном направлении оси Z. Правая система никакими вращениями не может быть совмещена с левой. Обе системы отличаются друг от друга примерно так же, как правая рука отличается от левой. Но правая система переходит в левую, если положительное направление одной из координатных осей изменить на противоположное. То же самое произойдет, если изменить на противоположные положительные направления всех трех осей. Последняя операция называется *инверсией координатных осей* или *отражением в начале координат*. Например, изображением правой системы в плоском зеркале будет левая система, и наоборот. В физике применяется исключительно правая система.

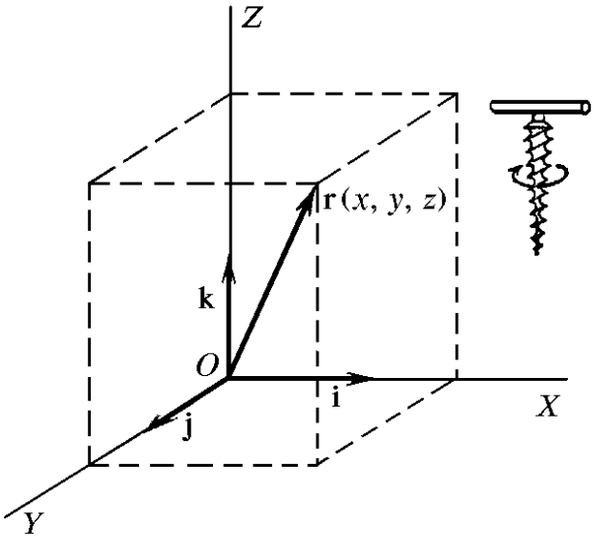


Рис. 2

3. Координаты x, y, z, которыми определяется положение точки в избранной системе отсчета, являются числами. *Количественное определение этих координат, равно как и количественное определение всякой физической величины, сводится в конце концов к указанию принципиального способа их измерения.* При этом имеются в виду *именно принципиальные, а не практические* способы измерения. Эти измерения должны лишь разъяснить смысл, точнее, принципиальный способ получения x, y, z, равно как и всяких чисел, с помощью которых количественно характеризуются все физические величины. Поэтому мы можем предполагать, что такие способы измерения являются *идеальными*, а сами измерения выполняются абсолютно точно. Координаты x, y, z являются длинами, а потому их нахождение сводится к измерению длин, т. е. к определению тех чисел, с помощью которых характеризуются длины. Когда мы говорим об измерении длин, то имеем в виду следующую из-

мерительную операцию. Некоторый твердый стержень условно принимается за эталон, а его длина — за единицу длины. При измерении длины тела в каком-либо направлении определяется число, показывающее, сколько раз в этом направлении в теле укладывается выбранный эталон. Это число и называется длиной тела в рассматриваемом направлении. Если оно не целое, то предварительно длину эталона следует разделить на более мелкие части: десятые, сотые и пр. Используя их наряду с самим эталоном, можно представить длину всякого тела в виде десятичной дроби или целого числа с десятичной дробью.

4. Измерение длины непосредственным прикладыванием эталона или его частей называется *прямым измерением.* Прямые измерения не всегда возможны. Так, они невозможны при измерении расстояний до удаленных тел, например планет, звезд и других небесных объектов. Они невозможны и при измерении очень малых длин, например таких, с которыми имеет дело физика атома, атомного ядра или элементарных частиц. Во всех этих случаях используют *косвенные методы.* Правильность таких методов должна контролироваться прямыми методами (разумеется, в тех случаях, когда последние применимы). За пределами же применимости прямых методов остаются одни только косвенные методы. Здесь прямые измерительные операции, с помощью которых первоначально было введено количественное понятие длины, становятся чисто умозрительными, а косвенные методы фактически играют роль основных принципиальных измерительных операций, которыми раскрывается смысл самих длин или, точнее, тех чисел, которыми длины характеризуются.

Примером косвенного метода может служить *триангуляция*, применяемая для измерения расстояний до удаленных предметов. Прямым методом измеряют длину «базы» AB (рис. 3), с концов которых делают «засечки» удаленного объекта C, т. е. измеряют углы α и β между базой AB и прямыми AC и BC. По этим данным искомое расстояние до объекта C может быть найдено геометрическим построением или вычислено по формулам геометрии. Если база AB настолько велика, что ее длина не может быть найдена прямым измерением, то можно выбрать более короткую базу и затем найти длину базы AB описанным косвенным методом. Принципиально это ничего не меняет. Более существенно уяснить теоретическую основу метода. В методе предполагается, что сторонами треугольника ABC являются *прямые линии, подчиняющиеся аксиомам геометрии Евклида.* Но какими материальными объектами реализуются эти стороны? Такими объектами являются световые лучи, приходящие от объекта C к точкам A и B. Следовательно, в основе рассматриваемого способа лежит гипотеза, что световые лучи прямолинейны, т. е. подчиняются

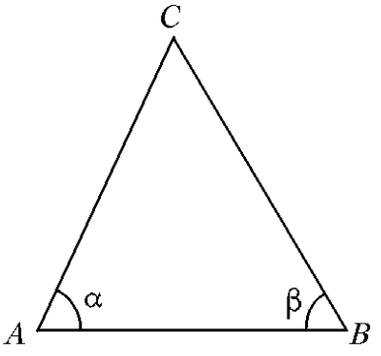


Рис. 3

тем же аксиомам геометрии Евклида, что и геометрические прямые линии. Но эта гипотеза не очевидна. Доказать или опровергнуть ее можно только опытным путем. При этом имеются в виду световые лучи в вакууме, а не лучи в атмосфере, где они действительно искривляются из-за изменения показателя преломления от точки к точке. Такое искривление лучей может быть учтено и действительно учитывается, когда точность измерений этого требует.

Как можно убедиться в применимости или неприменимости геометрии Евклида к реальному миру в указанном выше смысле? Прямой метод состоит в том, что надо подвергнуть экспериментальной проверке следствия, выводимые из аксиом геометрии Евклида. Одним из таких следствий является, например, теорема, утверждающая, что сумма внутренних углов треугольника равняется 180°. Великий немецкий математик, астроном и физик Карл Гаусс (1777—1855) измерял в 1821—1823 гг. со всей возможной тщательностью внутренние углы треугольника, образованного тремя удаленными горными вершинами. Длины сторон треугольника были порядка 100 км. Он нашел, что в пределах ошибок измерений не наблюдалось нарушений указанной теоремы. Этот метод не годится в масштабах Солнечной системы и бóльших, так как все измерения производятся с Земли, и мы не можем непосредственно измерить все три внутренних угла треугольника, вершинами которого помимо Земли являются, например, какие-либо две планеты или звезды. Здесь мы судим о применимости геометрии Евклида на основании косвенных данных — по согласованности различных результатов, полученных с использованием такой геометрии. Так, можно предвычислить движение планет Солнечной системы на много лет вперед и проверить полученные предсказания. Если бы они не оправдались, то одной из причин могла бы быть неприменимость геометрии Евклида к областям пространства порядка размеров Солнечной системы. Наоборот, согласие с опытом (что на самом деле имеет место) указывает на то, что сомневаться в применимости геометрии Евклида в областях такого размера нет оснований. Не вдаваясь в этот вопрос, ограничимся замечаниями, что, по-видимому, нет существенных нарушений геометрии Евклида в областях порядка размеров нашей *Галактики* ($\sim 10^{21}$ м) и даже *Метагалактики*, т. е. части Вселенной, доступной исследованию с помощью современных наиболее мощных оптических и радиотелескопов ($\sim 10^{26}$ м). Точно так же нет оснований ожидать существенных нарушений геометрии Евклида и в субатомных областях размером, скажем, порядка 10^{-15} м.

Световые лучи при определении положения удаленных тел выполняют и другую важную функцию. Они служат теми материальными объектами, с помощью которых конструируется сама система отсчета. Действительно, твердые стержни не могут быть неограниченно длинными, а потому они не пригодны в качестве координат-

ных осей во всем пространстве. Эту роль берут на себя световые лучи, являющиеся продолжениями в нужных направлениях координатных осей, первоначально реализованных твердыми стержнями.

5. В связи с изложенным целесообразно сделать одно замечание о связи физики с математикой. Математика играет исключительно важную роль в физике. Без нее современная физика немыслима. Однако необходимо правильно представлять себе истинную роль математики в физике, и к этому вопросу мы еще будем неоднократно возвращаться. Чистая математика имеет дело с абстрактными объектами и понятиями, подчиняющимися определенной системе аксиом. Единственное требование, предъявляемое в чистой математике к ее понятиям и аксиомам, сводится к их *логической непротиворечивости*. Все свои результаты чистая математика получает из этих аксиом путем логических рассуждений, основанных на правилах формальной логики. Содержание этих результатов, очевидно, не может выйти за пределы логической связи между различными объектами и понятиями чистой математики. В этом смысле чистая математика является логически *замкнутой дисциплиной*. Такая замкнутость и логическая согласованность придают чистой математике эстетическую привлекательность и доставляют чувство глубокого удовлетворения всякому уму, воспитанному в духе математической строгости.

Надо, однако, заметить, что строго замкнутая сама в себе математика оторвана от реальной действительности и не может быть использована в других науках и практической деятельности человека. Чтобы математика стала мощным средством при описании и изучении явлений природы, каким она в действительности является, необходимо установить связи между абстрактными математическими объектами и понятиями, с одной стороны, и реальными объектами и явлениями природы — с другой. Математические понятия и объекты должны появляться не как чисто логические категории, а как *абстракции каких-то реальных объектов или процессов природы*. Так, точка является абстракцией физического тела достаточно малых размеров, прямая линия — абстракцией достаточно тонкого прямого твердого стержня или светового пучка в однородной среде. Вопрос о справедливости математики сводится к справедливости ее аксиом. Справедливость же самих аксиом может быть установлена опытным и только опытным путем.

Правда, опыт с математическими объектами нельзя осуществить в чистом виде, поскольку эти объекты являются идеализациями и не встречаются в природе. Всякий опыт выполняется с реальными телами. Математическую строгость, которой, и не без оснований, так гордятся математики, надо понимать в смысле логической согласованности ее выводов, но не в смысле обоснования математических аксиом.

Одной математической строгости недостаточно для физики, как и для всякой другой опытной науки, имеющей дело с реальными объектами и явлениями природы. *Всякое теоретическое исследование, даже выполненное математически строго, никогда не мо-*

но высокой точности. Так, например, совсем не ясно, о чем идет речь, если поставить задачу об измерении длины твердого стержня с погрешностью до размеров электрона или даже атома. Принципиально неограниченная точность измерения длин имеет смысл для абстрактных прямолинейных отрезков геометрии, а не для реальных тел, имеющих атомистическую структуру.

6. Перейдем к вопросу об измерении *времени*. Как и всякая физическая величина, время количественно характеризуется некоторыми числами. Задача прежде всего состоит в том, чтобы выяснить, с помощью каких принципиальных измерительных операций эти числа могут быть получены. Тем самым устанавливается и точный смысл самих этих чисел.

Под временем в количественном смысле этого слова мы будем понимать показания каких-то часов. Точнее, надо говорить не о самом времени, а о *промежутке времени* между двумя событиями или моментами времени. Он характеризуется *разностью* показаний часов в рассматриваемые моменты времени. Когда говорят просто о времени, не указывая оба момента, являющихся границами рассматриваемого промежутка времени, то предполагают, что один из этих моментов фиксирован и условно принят за начальный. От него и ведется отсчет времени. Часы здесь понимаются в более широком смысле слова, чем в обыденной жизни. *Под часами понимают любое тело или систему тел, в которых совершается периодический процесс, служащий для измерения времени.* Примерами таких процессов могут служить колебания маятника с постоянной амплитудой, вращение Земли вокруг собственной оси относительно Солнца или звезд, колебания атома в кристаллической решетке, колебания электромагнитного поля, представляемого достаточно узкой спектральной линией, и пр. Так, если между двумя событиями Земля при вращении относительно звезд сделала один оборот, то говорят, что промежуток времени между этими двумя событиями составляет *звездные сутки*. Если при этом она совершила 10 оборотов, то соответствующий промежуток времени — 10 звездных суток, и т. д. Если в течение звездных суток маятник совершил приблизительно 86 164 колебания, то говорят, что период одного колебания составляет *одну секунду*, и т. д. От звездных суток следует отличать *солнечные сутки*. Так называется промежуток времени, в течение которого Земля делает один оборот при вращении вокруг собственной оси относительно Солнца. Ввиду того, что Земля движется вокруг Солнца не по круговой, а по эллиптической орбите, это ее движение не совсем равномерно (см. § 55). Это значит, что солнечные сутки изо дня в день несколько изменяются в течение года. Поэтому при измерении времени пользуются так называемыми *средними солнечными сутками*. Они составляют 24 ч = 1440 мин = 86 400 с.

К часам предъявляют требования, чтобы они шли «равномерно». Но что значит — часы идут равномерно? Говорят, что это — периодический процесс, служащий для отсчета времени; он должен по-

вторятся через строго одинаковые промежутки времени. Однако это не есть ответ на вопрос, так как убедиться в одинаковости следующих друг за другом промежутков времени можно только в том случае, когда мы уже располагаем равномерно идущими часами. Выйти из этого логического круга можно только путем определения, так как никакого априорного представления о равномерном течении времени не существует. Надо условиться считать какие-то часы *по определению* равномерно идущими. Такие часы должны рассматриваться как *эталонные* или основные часы, по которым должны градуироваться все остальные.

В принципе любые часы могут быть приняты за эталонные. Однако так поступать не целесообразно. Эталонные часы должны быть достаточно «хорошими» и прежде всего обладать *высокой воспроизводимостью*. Это означает, что если изготовить с возможной тщательностью много «одинаковых» эталонных часов, то они с большой точностью должны идти одинаково, независимо от того, изготовлены ли они одновременно или между моментами их изготовления прошло длительное время. Например, песочные часы дают несравненно худшую воспроизводимость, чем маятниковые часы.

Не так давно за основные или эталонные часы принимались *астрономические часы*. Долгое время основными часами служила Земля, вращающаяся вокруг собственной оси относительно звезд, а основной единицей времени — *сутки*. Недавно вместо осевого вращения Земли стали пользоваться ее орбитальным движением вокруг Солнца, принимая за основную единицу времени *тропический год*, т. е. промежуток времени между двумя последовательными прохождениями центра Солнца через *точку весеннего равноденствия*. При измерении времени таким путем достигалась лучшая воспроизводимость. Но еще лучшая воспроизводимость была достигнута после изобретения *кварцевых, молекулярных* и *атомных часов*.

Все эти часы представляют довольно сложные радиотехнические устройства. Здесь нет необходимости останавливаться на принципах действия и деталях устройства таких часов. Достаточно заметить, что роль маятника или балансира, регулирующих ход часов, выполняют в кварцевых часах колебания кристаллической решетки кварца, в молекулярных часах — колебания атомов в молекулах, в атомных часах — колебания электромагнитного поля в узких спектральных линиях атомов некоторых изотопов химических элементов, находящихся в точно определенных и строго контролируемых внешних условиях. Особой стабильностью обладают последние из отмеченных колебаний. Поэтому период именно таких колебаний в настоящее время и принимается в качестве основной единицы времени, с помощью которой воспроизводится секунда. Конкретно, *секунда — это промежуток времени, в течение которого совершается* 9 192 631 770 *колебаний электромагнитного излучения, соответствующего переходу между двумя определенными*

сверхтонкими уровнями основного состояния атома цезия-133 *в отсутствие внешних полей.*

С помощью кварцевых, молекулярных и атомных часов было показано, что Земля вокруг своей оси вращается «неравномерно».

7. С единицей длины дело обстоит так же, как с единицей времени. Идеально твердых тел не существует. Первоначальный эталон метра, реализованный в виде стержня из сплава платины и иридия, недостаточно надежен. Он подвержен внешним влияниям, его внутреннее молекулярное строение может измениться. Наконец, он может быть утерян или испорчен. Предпочтительнее в качестве основной единицы взять какую-либо естественную, точно воспроизводимую длину. В качестве такой длины до недавнего времени принималась длина световой волны определенной узкой спектральной линии, получаемой при определенных, строго контролируемых внешних условиях. С помощью такой естественной длины и воспроизводилась практическая единица длины — метр. В октябре 1983 г. Генеральная ассамблея мер и весов приняла новое определение метра:

Метр есть длина пути, проходимая светом в вакууме в течение времени интервала 1/299792458 *секунды.* Отсюда следует, что *скорость света по определению принимается равной* 299 792 458 *м/с* (*точно*).

Такое изменение эталона метра вызвано тем, что экспериментальная погрешность определения скорости света в вакууме в настоящее время составляет примерно 0,3 м/с, что превышает относительную погрешность воспроизведения метра по предыдущему определению, которая равна примерно $\pm 4 \cdot 10^{-9}$. Относительная точность измерения состояния будет непрерывно повышаться со временем, по крайней мере до достижения величины 10^{-12}, требующейся для астрономических измерений. Чтобы каждый раз не менять определение метра, поступили более естественным способом — зафиксировали значение скорости света.

Возможно, что современные эталоны времени и длины перестанут удовлетворять более жестким требованиям, которые будут предъявляться в будущем к точности измерений и воспроизводимости результатов. В таком случае старые эталоны будут заменены новыми, более стабильными. Принципиально это ничего не меняет.

8. Для описания движения, а также любых физических явлений, протекающих во времени, пространственных систем отсчета недостаточно. Надо превратить их в *пространственно-временные* системы отсчета. Принципиально это означает, что в системе отсчета должны быть достаточно часто расставлены неподвижные часы, обладающие одинаковой скоростью хода. Тогда каждое событие можно характеризовать местом, где оно произошло, и временем, когда оно произошло, т. е. показаниями часов, помещенных в этом месте. Однако таким путем будет определено только *местное время*, т. е. время в каждой точке пространства. Показания часов, находящихся

в различных точках пространства, еще никак не связаны между собой. Для описания физических процессов такая связь, в принципе, не обязательна. Можно было бы довольствоваться местным временем в каждой точке пространства, введя столько различных времен, сколько существует различных точек. Однако такое описание было бы крайне запутанным и абсолютно необозримым. Чтобы описание было простым и обозримым, необходимо пользоваться временем, *единым для всей пространственно-временной системы* отсчета. Для этой цели надо *«синхронизовать»* часы, расставленные в различных местах пространства, т. е. установить их так, чтобы они показывали «одинаковое» время.

Если часы неподвижны в рассматриваемой системе отсчета и показывают одно и то же время, то про такие часы говорят, что в этой системе отсчета они идут *синхронно*. Установление же «стрелок» часов так, чтобы часы всегда показывали одно и то же время, называется их *синхронизацией*. Синхронизация часов не вызывает затруднений, если часы расположены в одном и том же месте пространства. Но как синхронизовать часы, если они находятся в удаленных точках пространства? Более того: что вообще означает, что пространственно разделенные часы синхронизованы, т. е. всегда показывают «одинаковое» время? До теории относительности такой вопрос физика даже не ставила, считая ответ на него само собой разумеющимся. Между тем без постановки этого вопроса и ответа на него невозможно ясное понимание самой теории относительности. Это впервые (в 1905 г.) с полной ясностью было понято Эйнштейном.

На первый взгляд, в вопросе о синхронизации пространственно разделенных часов нет никаких трудностей. Поместим, например, двое одинаковых часов сначала в точку A, поставим их стрелки одинаково, а затем одни из часов перенесем в точку B. Тогда часы в A и часы в B будут синхронизованы между собой. Однако такой способ синхронизации только тогда имел бы смысл, когда показание переносимых часов не зависело от способа переноса их из начальной точки A в конечную точку B. Чтобы убедиться в этом, надо сравнить показания переносимых часов в этих двух точках с показаниями каких-то двух неподвижных часов 1 и 2, из которых одни находятся в точке A, а другие — в точке B. Но часы 1 и 2 должны быть заранее синхронизованы. Получился заколдованный круг. Поэтому способ синхронизации путем переноса не годится.

Остается *синхронизация часов с помощью сигналов*. Можно, например, поступить следующим образом. В произвольный момент времени t_A по часам A послать какой-то сигнал к часам B. В момент прихода сигнала часы B поставить так, чтобы они показывали время $t_B = t_A + \tau_{AB}$, где τ_{AB} — время, затрачиваемое сигналом на прохождение расстояния от A до B. Тогда часы A и B будут синхронизованы между собой. Однако для осуществления такой синхронизации надо знать время τ_{AB}. Но это время можно измерить только после

того, как часы A и B уже синхронизованы. Получился опять заколдованный круг. Эйнштейн указал, что выйти из этого круга можно *только путем определения* понятия одновременности. Дорелятивистская физика считала, что одновременность событий абсолютна и является чем-то само собой разумеющимся. Эйнштейн указал, что это заблуждение. Смысл одновременности событий не требует разъяснения только тогда, когда эти события происходят в одном и том же месте пространства. Если же события происходят в различных местах пространства, то надо условиться, такие события называть одновременными. Очевидно, вопрос об одновременности пространственно разделенных событий эквивалентен вопросу о синхронизации пространственно разделенных часов.

Решение всякого вопроса, сводящегося к определению, содержит некоторый произвол. Это полностью относится и к вопросу об определении одновременности пространственно разделенных событий. Объективное содержание законов природы, разумеется, не может зависеть ни от каких произвольных определений. Определения могут влиять только на форму законов. Надо стремиться вводить такие определения, чтобы форма законов была наиболее проста и легко обозрима. Это требование почти однозначно вынуждает принять определение одновременности, предложенное Эйнштейном в теории относительности.

Эйнштейн пользуется для синхронизации часов световыми сигналами в вакууме. Время, затрачиваемое сигналом на пути от A до B и обратно, очевидно, равно $\tau = \tau_{AB} + \tau_{BA}$, где τ_{BA} — время, которое он затрачивает на прохождение от B к A. Для измерения времени τ достаточны только одни часы A, и никакой предварительной синхронизации не требуется. Но время τ_{AB} и τ_{BA} можно измерить лишь после того, как часы A и B синхронизованы. Пока они еще не синхронизованы, можно потребовать, чтобы между τ_{AB} и τ_{BA} соблюдалось какое-то соотношение. Тогда из этого соотношения и соотношения $\tau = \tau_{AB} + \tau_{BA}$ найдутся времена τ_{AB} и τ_{BA} в отдельности. Тем самым часы A и B будут синхронизованы и установлено понятие одновременности пространственно разделенных событий. *Эйнштейн полагает по определению, что* $\tau_{AB} = \tau_{BA}$, *т. е. что свет в направлениях от A к B и от B к A располагается с одной и той же скоростью.* Тогда $\tau_{AB} = \tau_{BA} = \tau/2$, и следовательно, $t_B = t_A + \tau/2$. Для бесконечно быстрых сигналов $\tau = 0$, а потому $t_A = t_B$. Одновременность, устанавливаемая с помощью таких сигналов, была бы абсолютной. Однако бесконечно быстрых сигналов не существует.

Синхронизация часов по Эйнштейну не требует предварительного измерения скорости света. Ее можно осуществить также следующим способом. Пусть C — точка, находящаяся на середине отрезка между точками A и B (неподвижными друг относительно друга). Произведем в C световую вспышку. *По определению* свет от нее до-

стигнет точек A и B одновременно. Если в момент прихода света к часам A и B их показания сделать одинаковыми, то они будут синхронизованы между собой.

Необходимо подчеркнуть, что в основе обоих способов синхронизации часов (иначе определения одновременности) лежит предположение об *однородности* и *изотропии пространства* и об *однородности времени*. Однородность пространства означает, что в нем все пространственные точки эквивалентны, т. е. ничем не выделены одна от другой. Однородность времени означает то же самое по отношению к любым моментам времени. Наконец, изотропия пространства характеризуется отсутствием в нем выделенных направлений — все направления в пространстве эквивалентны. Поэтому-то при определении одновременности и использовались световые сигналы в вакууме и притом в отсутствие гравитационного поля, так как наличие последнего уже нарушает указанную однородность и изотропию пространства.

Теперь нетрудно убедиться в эквивалентности приведенных выше двух способов синхронизации часов. Сформулируем прежде всего основной постулат, являющийся следствием однородности и изотропии пространства и однородности времени. На прохождение отрезков равной длины свет затрачивает в вакууме одно и то же время, независимо от того, где эти отрезки расположены, как они ориентированы и в какое время в них распространяется свет. Следствием этого постулата является утверждение, что свет затрачивает в вакууме одно и то же время на прохождение отрезка в прямом и противоположном направлении.

Приведем теперь доказательство эквивалентности обоих способов определения одновременности, хотя оно почти и самоочевидно. По-прежнему будем предполагать такое же расположение точек A, B, C. Допустим сначала, что применяется первый способ синхронизации часов. Тогда $\vec{\tau}_{AB} = \tau_{BA}$, или $\vec{\tau}_{AC} + \tau_{CB} = \tau_{BC} + \tau_{CA}$. Для наглядности над временами поставлены стрелки, указывающие направление (слева направо и справа налево) распространения светового сигнала. Так как $AB = BC$, то на основании основного постулата $\vec{\tau}_{AC} = \tau_{BC}$. Поэтому $\tau_{CA} = \vec{\tau}_{CB}$. Но это есть второй способ синхронизации часов.

Допустим теперь, что синхронизация устанавливается вторым способом, так что $\tau_{CA} = \vec{\tau}_{CB}$. По основному постулату $\vec{\tau}_{CB} = \tau_{CA}$. Сложив это с предыдущим равенством и снова применив основной результат, получим $\vec{\tau}_{AC} + \vec{\tau}_{CB} = \tau_{BC} + \tau_{CA}$, или $\vec{\tau}_{AB} = \tau_{BA}$, т. е. первый способ синхронизации.

Подчеркнем еще раз, что эйнштейновский способ синхронизации часов предполагает отсутствие поля тяготения. При наличии последнего скорость света перестает быть универсальной постоянной. Этот вопрос был разработан Эйнштейном в его общей теории относительности.

Всякое определение должно быть непротиворечивым. Если синхронизовать между собой часы A и B, то с третьими часами C можно поступить двояко: синхронизовать их либо с часами A, либо с часами B. Оба способа должны приводить к одному и тому же результату. Иначе получилось бы внутреннее противоречие. Наша уверенность в непротиворечивости эйнштейновского определения одновременности основана на логической согласованности бесчисленного множества следствий, при выводе которых использовалось это определение.

9. *Одновременность пространственно разделенных событий в том смысле, какой придается ей определением Эйнштейна, относительна.* Это значит, что два события, одновременные в одной системе отсчета, могут оказаться не одновременными, если их рассматривать в другой системе отсчета, движущейся относительно первой. Поясним это следующим примером. Пусть система отсчета S связана с твердым стержнем AB, а система S' — с твердым стержнем $A'B'$, движущимся относительно AB (рис. 4). В момент, когда середины стержней C и C' совпадают между собой, производится световая вспышка в точке, где происходит это совпадение. Рассмотрим распространение света от этой вспышки сначала в системе отсчета S. Из определения одновременности следует, что сигнал достигнет одновременно концов стержня A и B. Значит, он достигнет раньше точки A' и позже — точки B', поскольку точка A' движется навстречу сигналу, а точка B' уходит от него. Итак, с точки зрения отсчета S световой сигнал приходит в точки A и B одновременно, а в точки A' и B' — не одновременно: в точку A' он приходит раньше, чем в B'. К другому заключению мы придем, если рассмотрим явление отсчета S'. По определению одновременными будут приходы сигнала в точки A' и B', а приходы в точки A и B — не одновременными: в точку B сигнал придет раньше, чем в точку A. Противоречия между обоими заключениями нет, поскольку «одновременность» понимается в них в разных смыслах. Этот пример и иллюстрирует относительность одновременности.

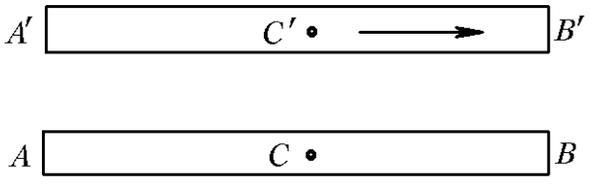


Рис. 4

10. *Длина тела также относительна, т. е. зависит от того, в какой системе отсчета она измеряется.* Что такое длина движущегося стержня — такой вопрос дорелятивистская физика не ставила. Теория относительности показала, что надо строго различать *длину покоящегося стержня* l_0, т. е. длину, измеренную в системе отсчета, в которой стержень покоится, и *длину движущегося стержня* l, т. е. длину, измеренную в системе отсчета, относительно которой он движется. Эти длины становятся равными только в пре-

дельном случае бесконечно медленных движений. Величина l_0 получается путем откладывания вдоль стержня единичного масштаба, покоящегося относительно этого стержня. Величина l сводится к измерению расстояния между неподвижными точками путем следующей измерительной операции. Надо отметить какими-либо неподвижными точками A и B положения концов движущегося стержня в рассматриваемой системе отсчета в один и тот же момент времени. Расстоянием между этими неподвижными точками и будет, по определению, длина движущегося стержня l. Если взять другую систему отсчета, то ввиду относительности одновременности концы стержня пройдут в этой системе отсчета мимо точек A и B, вообще говоря, не одновременно. Роль A и B будут играть другие точки A' и B', неподвижные в новой системе отсчета. Расстояние между этими точками l', вообще говоря, не будет совпадать с l. Таким образом, *как и промежутки времени, длины отрезков также относительны.*

11. Дорелятивистская физика считала длины тел и промежутки времени абсолютными. С ее точки зрения можно было говорить об одновременности событий, не указывая, в какой системе отсчета эти события рассматриваются. Можно было говорить о длине тела, не указывая, покоится оно или движется в рассматриваемой системе отсчета. Основанием, правда, явно не формулировавшимся, для такой точки зрения было убеждение в существовании сколь угодно быстрых сигналов. Но при рассмотрении медленных движений скорость световых сигналов в вакууме может считаться практически бесконечно большой. В этом приближении исчезают все релятивистские эффекты. Мы приведем в дальнейшем некоторые результаты релятивистской механики ввиду их важности в ядерной физике и физике элементарных частиц. Однако систематическое изложение релятивистской механики будет дано после того, как мы познакомимся с учением об электрических и оптических явлениях. Там же будут подробно разобраны с количественной стороны и вопросы пространства и времени, которые здесь были только поставлены или затронуты качественно.

§ 2. КИНЕМАТИЧЕСКОЕ ОПИСАНИЕ ДВИЖЕНИЯ. МАТЕРИАЛЬНАЯ ТОЧКА

1. *Кинематика* занимается описанием движения, отвлекаясь от его причин. Для описания движения можно выбирать различные системы отсчета. В различных системах отсчета движение одного и того же тела выглядит по-разному. В кинематике при выборе системы отсчета руководствуются лишь соображениями целесообразности, определяющимися конкретными условиями. Так, при рассмотрении тел на Земле естественно связать систему отсчета с Землей, что мы и будем делать. При рассмотрении движения самой Земли систему

отсчета удобнее связывать с Солнцем и т. п. Никаких принципиальных преимуществ одной системы отсчета по сравнению с другой в кинематике указать нельзя. Все системы отсчета *кинематически эквивалентны.* Только в *динамике*, изучающей движение в связи с силами, действующими на движущиеся тела, выявляются принципиальные преимущества определенной системы отсчета или, точнее, определенного класса системы отсчета.

2. Простейшим объектом, движение которого изучает классическая механика, является материальная точка. *Материальной точкой* называется макроскопическое тело, размеры которого настолько малы, что в рассматриваемом движении их можно не принимать во внимание и считать, что все вещество тела как бы сосредоточено в одной геометрической точке. Материальных точек в природе не существует. Материальная точка есть абстракция, идеализированный образ реально существующих тел. Можно или нельзя то или иное тело при изучении какого-либо движения принять за материальную точку — это зависит не столько от самого тела, сколько от характера движения, а также от содержания вопросов, на которые мы хотим получить ответ. Абсолютные размеры тела при этом не играют роли. Важны *относительные размеры*, т. е. отношения размеров тела к некоторым расстояниям, характерным для рассматриваемого движения. Например, Землю при рассмотрении ее орбитального движения вокруг Солнца с большой точностью можно принять за материальную точку. Характерной длиной здесь является радиус земной орбиты $R \approx 1{,}5 \cdot 10^8$ км. Он очень велик по сравнению с радиусом Земли $r \approx 6{,}4 \cdot 10^3$ км. Благодаря этому при орбитальном движении Земли все ее точки можно считать движущимися практически одинаково. Поэтому достаточно рассмотреть движение только одной точки, например центра Земли, и считать, что все вещество Земли как бы сосредоточено в этой геометрической точке. Такая идеализация сильно упрощает задачу об орбитальном движении Земли, сохраняя, однако, все существенные черты этого движения. Но эта идеализация не годится при рассмотрении вращения Земли вокруг собственной оси, ибо бессмысленно говорить о вращении геометрической точки вокруг оси, проходящей через эту точку.

В определение материальной точки мы включили условие, что точка должна быть *макроскопическим телом.* Это сделано для того, чтобы к ее движению можно было применять классическую механику. Однако в ряде случаев и движение микрочастиц может рассматриваться на основе классической механики. Сюда относится, например, движение электронов, протонов или ионов в ускорителях и электронно-ионных приборах. В этих случаях микрочастицы можно рассматривать как материальные точки классической механики.

3. Механика одной материальной точки или, короче, *механика точки* в классической физике является основой для изучения механики вообще. С классической точки зрения произвольное макроскопическое тело или систему тел можно мысленно разбить на малые макроскопические части, взаимодействующие между собой. Каждую из таких частей можно принять за материальную точку. Тем самым изучение движения произвольной системы тел сводится к изучению системы взаимодействующих материальных точек. Естественно поэтому начать изучение классической механики с механики одной материальной точки, а затем перейти к изучению системы материальных точек.

Выберем какую-либо произвольную систему отсчета и будем относить к ней движение материальной точки. Движение точки будет описано полностью, если будет известно ее положение в любой момент времени относительно выбранной системы отсчета. Положение точки мы условимся характеризовать ее прямоугольными координатами x, y, z, являющимися проекциями ее радиуса-вектора $\mathbf{r}$ на координатные оси. Полное описание движения сводится поэтому к нахождению трех координат x, y, z как функций времени t,

$$x = x(t), \quad y = y(t), \quad z = z(t), \tag{2.1}$$

или к нахождению одной векторной функции

$$\mathbf{r} = \mathbf{r}(t). \tag{2.2}$$

Однако для формулировки основных законов механики, с помощью которых теоретически могут быть найдены рассматриваемые функции, существенны два новых понятия — понятие *скорости* и в особенности понятие *ускорения*. К установлению этих понятий мы и перейдем.

§ 3. СКОРОСТЬ И УСКОРЕНИЕ ПРИ ПРЯМОЛИНЕЙНОМ ДВИЖЕНИИ. УГЛОВАЯ СКОРОСТЬ И УГЛОВОЕ УСКОРЕНИЕ

1. Рассмотрим сначала частный случай, когда материальная точка движется вдоль прямой линии. Примем эту прямую за координатную ось X, поместив начало координат O в какой-то произвольной точке ее (рис. 5). Положение материальной точки в рассматриваемом случае определяется одной координатой:

$$x = x(t). \tag{3.1}$$

Пусть в какой-то фиксированный момент времени t материальная точка находится в положении A_1. В этот момент ее координата равна $x_1 = x(t)$. В более поздний момент времени $t + \Delta t$ матери-

альная точка переместится в положение A_2 с координатой $x_2 = x(t + \Delta t)$. За время Δt материальная точка проходит путь $\Delta x = x_2 - x_1 = x(t + \Delta t) - x(t)$. Путь считается положительным, если перемещение совершается вправо, и отрицательным, если оно происходит влево. Отношение пройденного пути Δx к промежутку времени Δt называется *средней скоростью материальной точки за время* Δt, или, точнее, *за время между* t *и* $t + \Delta t$. Таким образом, по определению средняя скорость равна

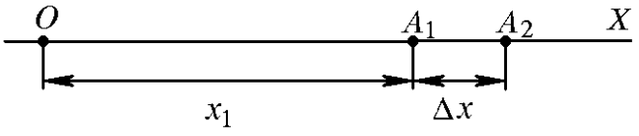


Рис. 5

$$v_{\text{ср}} = \frac{\Delta x}{\Delta t} = \frac{x(t+\Delta t) - x(t)}{\Delta t}. \tag{3.2}$$

Такое определение средней скорости имеет смысл для любых значений Δt. Надо исключить только значение $\Delta t = 0$, так как в этом случае для средней скорости мы получили бы выражение $\frac{0}{0}$, которое само по себе не имеет смысла. Однако ничто не мешает брать промежуток времени Δt как угодно малым, но отличным от нуля. Вообще говоря, средняя скорость зависит не только от t, но и от Δt. Будем теперь, оставляя момент времени t неизменным, брать промежуток времени Δt все меньше и меньше, устремляя его к нулю. Тогда будет стремиться к нулю и проходимый путь Δx. Отношение же $\frac{\Delta x}{\Delta t}$ при этом, как показывает опыт, будет стремиться к вполне определенному пределу, который может зависеть только от t, но уже не будет зависеть от Δt. Этот предел называется *истинной* или *мгновенной скоростью материальной точки в момент времени* t:

$$v = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} = \lim_{\Delta t \to 0} \frac{x(t+\Delta t) - x(t)}{\Delta t}. \tag{3.3}$$

Пределы типа (3.3) встречаются в самых разнообразных вопросах математики и ее приложениях. В математике предел, определяемый формулой (3.3), называется *производной* функции $x(t)$ по аргументу t. Производная во времени обозначается символом $\dot{x}(t)$ или $\frac{dx}{dt}$. Таким образом, по определению

$$\dot{x} \equiv \frac{dx}{dt} = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t}. \tag{3.4}$$

Понятие производной является основным понятием *дифференциального исчисления*. Используя это понятие, можно сказать, что *истинная* или *мгновенная скорость* v есть производная ко-

ординаты x по времени, или производная пройденного пути s по времени:

$$v = \dot{x} = \frac{dx}{dt} = \frac{ds}{dt}. \tag{3.5}$$

Скорость материальной точки, вообще говоря, является функцией времени: $v = v(t)$. Производная скорости по времени называется *ускорением материальной точки*. Ускорение мы обозначаем через a. Таким образом, по определению ускорения

$$a = \frac{dv}{dt} = \dot{v}(t), \tag{3.6}$$

или

$$a = \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t} = \lim_{\Delta t \to 0} \frac{v(t+\Delta t) - v(t)}{\Delta t}. \tag{3.7}$$

Производная (3.6) называется также *второй производной* координаты x по времени и обозначается символами

$$a = \ddot{x} \equiv \frac{d^2x}{dt^2}. \tag{3.8}$$

В существовании первой и второй производных координаты по времени в механике, как и во всех аналогичных вопросах физики, мы убеждаемся не путем логических рассуждений, а опытным путем.

2. Рассмотрим простейшие примеры.

Пример 1. $x = \text{const}$, т. е. координата x остается постоянной во времени. В этом случае материальная точка неподвижна, приращение координаты Δx равно нулю. Равны нулю также средняя и истинная скорости точки: $v = \dot{x} = 0$. Вообще, производная всякой постоянной величины равна нулю.

Пример 2. $x = Bt + C$, где B и C — постоянные коэффициенты. В этом случае говорят, что координата x является *линейной функцией времени* t. Очевидно,

$$x + \Delta x = B(t + \Delta t) + C = (Bt + C) + B\,\Delta t,$$

$$\Delta x = B\,\Delta t, \quad v_{\text{ср}} = \frac{\Delta x}{\Delta t} = B.$$

Средняя скорость постоянна и равна B. Поэтому истинная скорость также постоянна и равна средней скорости:

$$v = \frac{dx}{dt} = v_{\text{ср}} = B.$$

Движение с постоянной скоростью называется *равномерным*. Обозначим через x_0 значение координаты x в начальный момент времени $t = 0$. Величина x_0 называется *начальной координатой* и, очевидно, $x_0 = C$. Пройденный путь s определяется приращением координаты, $s = x - x_0 = Bt$, или

$$s = vt.$$

Пример 3. $x = At^2 + Bt + C$, где A, B и C — постоянные коэффициенты. В этом случае говорят, что координата x является *квадратичной функцией времени* t. Очевидно,

$$x + \Delta x = A(t + \Delta t)^2 + B(t + \Delta t) + C = \\ = (At^2 + Bt + C) + (2At + B)\,\Delta t + A(\Delta t)^2,$$

$$v_{\text{ср}} = \frac{\Delta x}{\Delta t} = (2At + B) + A\,\Delta t.$$

Здесь $v_{\text{ср}}$ зависит не только от t, но и от Δt. В пределе, когда $\Delta t \to 0$, член $A\,\Delta t$ обращается в нуль, и мы получаем следующее выражение для истинной скорости:

$$v = 2At + B.$$

Истинная скорость является линейной функцией времени t, а потому, дифференцируя ее, получаем постоянное значение для ускорения:

$$a = \frac{dv}{dt} = 2A.$$

Движение с постоянным ускорением называется *равноускоренным*. Постоянная A равна половине ускорения: $A = a/2$. Выясним теперь физический смысл постоянных B и C. При $t = 0$ наши формулы дают $v = B$. Скорость в момент времени $t = 0$ называется *начальной скоростью* и обозначается через v_0. Мы видим, что постоянная B равна начальной скорости: $B = v_0$. Аналогично доказывается, что постоянная C есть начальная координата движущейся точки: $C = x_0$. С введением этих величин можно написать

$$x = \frac{at^2}{2} + v_0 t + x_0, \qquad v = at + v_0.$$

Пройденный путь равен $s = x - x_0$, т. е.

$$s = \frac{at^2}{2} + v_0 t.$$

Примерами равноускоренного движения могут служить свободное падение тел и скатывание тела по наклонной плоскости без трения. С возрастанием высоты падения постоянство ускорения нарушается сопротивлением воздуха, а также неоднородностью поля тяготения.

3. По аналогии с *линейной* скоростью и ускорением вводятся *угловая скорость* и *угловое ускорение*. Эти понятия относятся к случаю движения материальной точки по окружности. Положение точки M на окружности можно задать углом α, который образует радиус-вектор OM с каким-либо неизменным направлением OX (рис. 6). Производная этого угла по времени

$$\omega = \frac{d\alpha}{dt}$$

называется *угловой скоростью*. Вращение называется *равномерным*, если угловая скорость ω постоянна. В этом случае $\alpha = \omega t + \text{const}$. При равномерном вращении величину ω называют также *угловой частотой вращения*. Величина $\nu = \omega/2\pi$ дает число оборотов в единицу времени и называется *частотой обращения*. Величина $T = 1/\nu$ есть продолжительность одного обращения и называется *периодом вращения*.

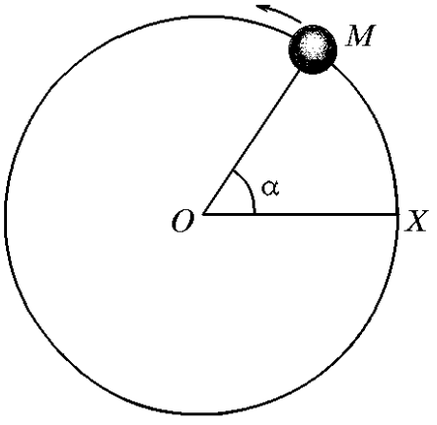


Рис. 6

Первая производная угловой скорости ω или вторая производная угла α по времени называется *угловым ускорением*:

$$\dot{\omega} = \frac{d\omega}{dt} = \frac{d^2\alpha}{dt^2}.$$

Если s означает длину дуги окружности XM, то ее производные $v = \frac{ds}{dt}$ и $a = \frac{d^2s}{dt^2}$ дают линейную скорость и линейное ускорение при движении точки по окружности. Если r — радиус окружности, то $s = r\alpha$. Дифференцируя это соотношение по времени, находим

$$v = \omega r, \quad a = \dot{\omega} r.$$

§ 4. СКОРОСТЬ И УСКОРЕНИЕ ПРИ КРИВОЛИНЕЙНОМ ДВИЖЕНИИ

1. Понятия скорости и ускорения естественным образом обобщаются на случай движения материальной точки по *криволинейной траектории*. Положение движущейся точки на траектории мы будем задавать радиусом-вектором $\mathbf{r}$, проведенным в эту точку из какой-либо неподвижной точки O, условно принимаемой за начало координат (рис. 7). Пусть в момент времени t материальная точка находится в положении M с радиусом-вектором $\mathbf{r} = \mathbf{r}(t)$. Спустя короткое время Δt она переместится в положение M_1 с радиусом-вектором $\mathbf{r}_1 = \mathbf{r}(t + \Delta t)$. Радиус-вектор материальной точки получит приращение, определяемое геометрической разностью

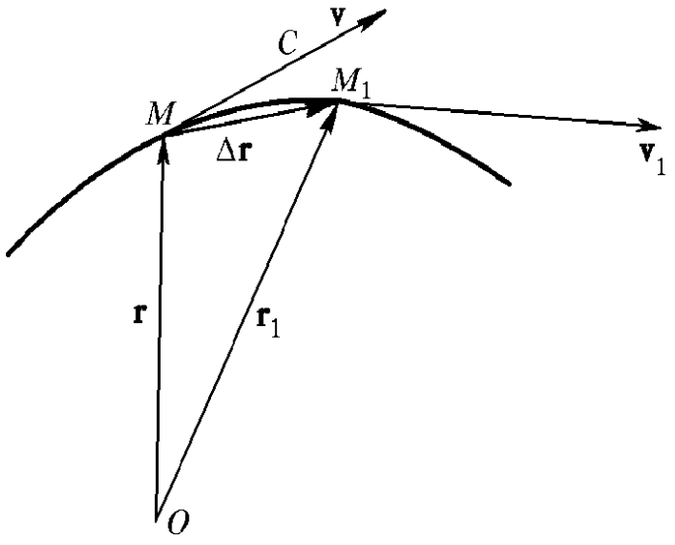


Рис. 7

$\Delta\mathbf{r} = \mathbf{r}_1 - \mathbf{r}$. Величина

$$\mathbf{v}_{\text{ср}} = \frac{\Delta\mathbf{r}}{\Delta t} = \frac{\mathbf{r}(t + \Delta t) - \mathbf{r}(t)}{\Delta t} \tag{4.1}$$

называется *средней скоростью движения за время* Δt или, точнее, *за время между* t *и* $t + \Delta t$. Она является величиной *векторной*, так как получается делением вектора $\Delta\mathbf{r}$ на скаляр Δt. Направление средней скорости $\mathbf{v}_{\text{ср}}$ совпадает с направлением хорды MM_1, т. е. с $\Delta\mathbf{r}$.

Предел средней скорости при $\Delta t \to 0$, т. е. производная радиуса-вектора $\mathbf{r}$ по времени

$$\mathbf{v} = \dot{\mathbf{r}} = \frac{d\mathbf{r}}{dt} = \lim_{\Delta t = 0} \frac{\Delta\mathbf{r}}{\Delta t} \tag{4.2}$$

называется *истинной* или *мгновенной скоростью материальной точки. Истинная скорость есть вектор, направленный по касательной к траектории движущейся точки.*

Совершенно аналогично определяется ускорение при криволинейном движении. *Ускорением* $\mathbf{a}$ *называется вектор, равный первой производной вектора скорости* $\mathbf{v}$ *или второй производной радиуса-вектора* $\mathbf{r}$ *по времени*:

$$\mathbf{a} = \dot{\mathbf{v}}(t) = \frac{d\mathbf{v}}{dt} = \lim_{\Delta t = 0} \frac{\Delta\mathbf{v}}{\Delta t}, \tag{4.3}$$

$$\mathbf{a} = \ddot{\mathbf{r}}(t) = \frac{d^2\mathbf{r}}{dt^2}. \tag{4.4}$$

2. Отметим следующую формальную аналогию между скоростью и ускорением. Из произвольной неподвижной точки O_1 будем откладывать вектор скорости $\mathbf{v}$ движущейся точки во всевозможные моменты времени (рис. 8). Конец вектора $\mathbf{v}$ назовем *скоростной точкой*. Геометрическое место скоростных точек есть кривая, называемая *годографом скорости*. Когда материальная точка описывает траекторию, соответствующая ей скоростная точка движется по годографу. Рисунок 8 отличается от рис. 7 только обозначениями. Радиус-вектор $\mathbf{r}$ заменен на вектор скорости $\mathbf{v}$, материальная точка — на скоростную точку, траектория — на годограф. Математические операции над вектором $\mathbf{r}$ при нахождении скорости и над вектором $\mathbf{v}$ при нахождении ускорения совершенно тождественны. Для математики безразлично, какой физический смысл имеют величины, над которыми выпол-

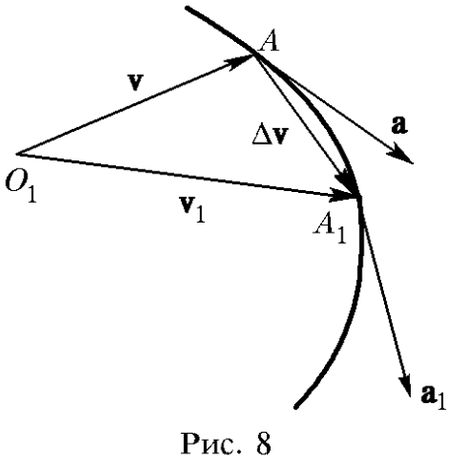


Рис. 8

няются математические операции. Не имеет значение также, какими символами эти величины обозначены. Для нахождения скорости **v** надо дифференцировать радиус-вектор **r**, для нахождения ускорения надо дифференцировать вектор скорости **v**. Скорость **v** направлена по касательной к траектории. Поэтому *ускорение* **a** *будет направлено по касательной к годографу скорости.* Можно сказать, что *ускорение есть скорость движения скоростной точки по годографу.* Следовательно, *все соотношения и теоремы, полученные для скорости, остаются справедливыми и для ускорения, если в них произвести замену величины и терминов согласно следующей таблице*:

Материальная точка	→ Скоростная точка
Радиус-вектор	→ Вектор скорости
Траектория	→ Годограф
Скорость	→ Ускорение

3. В качестве простейшего примера найдем ускорение точки, равномерно вращающейся по окружности радиуса r (рис. 9*а*). Скорость **v** направлена по касательной к окружности, ее модуль определяется выражением

$$v = \omega r = \frac{2\pi r}{T}. \tag{4.5}$$

Годографом будет окружность радиуса v (рис. 9*б*). Когда материальная точка M вращается по окружности радиуса r, соответствующая ей скоростная точка A вращается в том же направлении по окружности радиуса v, описывая эту окружность за то же время T. Положениям материальной точки на траектории M_1, M_2, M_3, M_4

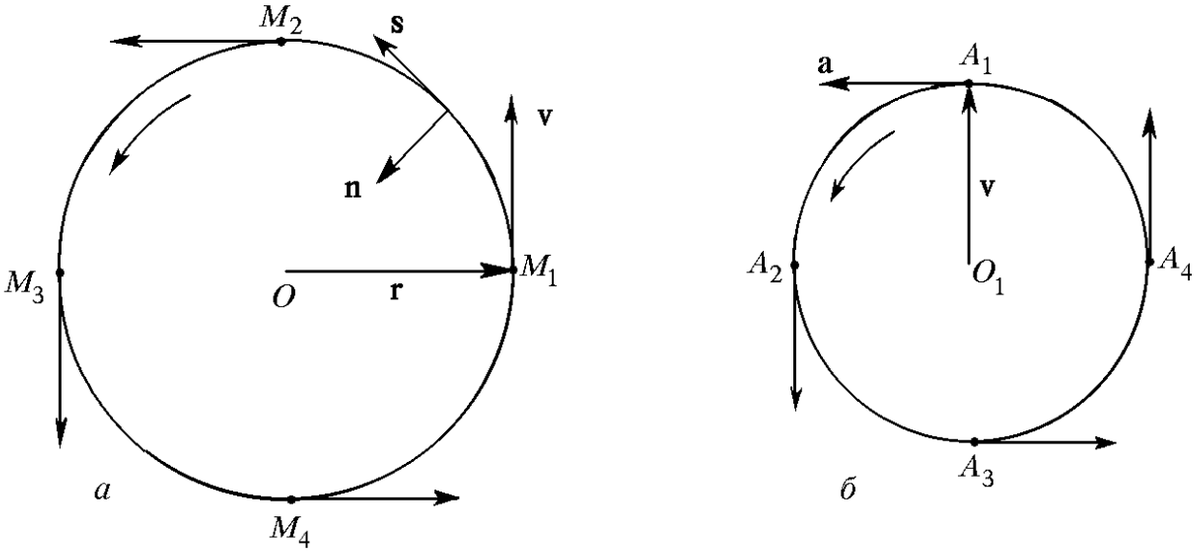


Рис. 9

соответствуют на годографе положения скоростной точки A_1, A_2, A_3, A_4. Ускорение **a** направлено по касательной к окружности — годогра-

фу и притом, как видно из рисунка, к центру O траектории вращающейся точки M. По аналогии с формулой (4.5) для модуля ускорения можно написать

$$a = \omega v = \frac{2\pi v}{T} = \frac{v^2}{r}. \tag{4.6}$$

Это известная формула для *центростремительного ускорения*. Ее можно записать в векторной форме:

$$\mathbf{a} = -\omega^2 \mathbf{r}. \tag{4.7}$$

Знак минус указывает на то, что направления векторов $\mathbf{a}$ и $\mathbf{r}$ взаимно противоположны, т. е. ускорение $\mathbf{a}$ направлено к центру круговой траектории, по которой вращается материальная точка. Можно также написать для любого положения движущейся точки

$$\mathbf{a} = \frac{v^2}{r}\mathbf{n}, \tag{4.8}$$

где $\mathbf{n}$ — единичный вектор нормали к круговой траектории движущейся точки, направленный к центру O (см. рис. 9a).

Имея в виду дальнейшие обобщения, представим вектор скорости в виде $\mathbf{v} = v\mathbf{s}$, где $\mathbf{s}$ — единичный вектор касательной к окружности. Первый множитель v есть модуль вектора скорости, второй множитель $\mathbf{s}$ указывает направление. При равномерном вращении модуль вектора скорости v остается неизменным, меняется только направление скорости, т. е. единичный вектор $\mathbf{s}$. Дифференцированию подлежит только этот вектор, а потому $\mathbf{a} = v\frac{d\mathbf{s}}{dt}$. Сравнивая это выражение с (4.8), получим

$$\frac{d\mathbf{s}}{dt} = \frac{v}{r}\mathbf{n}. \tag{4.9}$$

Обозначим через ds длину пути, проходимого материальной точкой за время dt при ее вращении по окружности. Эта положительная величина равна $ds = vdt$. Поэтому предыдущую формулу можно переписать в виде

$$\frac{d\mathbf{s}}{ds} = \frac{1}{r}\mathbf{n}. \tag{4.10}$$

В этом виде формула не содержит никаких кинематических величин. В нее входят только геометрические величины, характеризующие окружность. Поэтому она может быть получена чисто геометрически без привлечения кинематических понятий. Она определяет производную единичного вектора касательной $\mathbf{s}$ по длине дуги окружности. Взаимная перпендикулярность векторов $\mathbf{s}$ и $\frac{d\mathbf{s}}{dt}$ $\left(\text{или } \frac{d\mathbf{s}}{ds}\right)$ объясняется тем, что длина вектора $\mathbf{s}$ постоянна, меняется только направление этого вектора. Треугольник, составленный из векторов $\mathbf{s}$, $\mathbf{s} + \Delta\mathbf{s}$ и $\Delta\mathbf{s}$ (рис. 10), — равнобедренный. При стремлении элемента

дуги Δs к нулю стремится к нулю и угол α при его вершине. Поэтому направление вектора $\frac{\Delta \mathbf{s}}{\Delta s}$ в пределе оказывается перпендикулярным к вектору $\mathbf{s}$. Отмеченное свойство, разумеется, не является специфическим свойством единичного вектора $\mathbf{s}$. Производная любого вектора $\mathbf{A}$ постоянной длины по любому скалярному аргументу есть вектор, перпендикулярный к вектору $\mathbf{A}$.

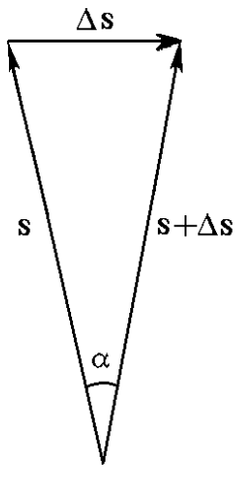


Рис. 10

4. Формула (4.10) допускает обобщение на случай произвольной гладкой кривой. Обозначим по-прежнему через $\mathbf{s}$ единичный вектор касательной к кривой, а через ds — длину элемента дуги этой кривой. Производная $\frac{d\mathbf{s}}{ds}$ есть вектор, направленный нормально к кривой в сторону ее вогнутости. Эту производную можно поэтому представить в виде (4.10), рассматривая величину $1/r$ как коэффициент пропорциональности между векторами $\frac{d\mathbf{s}}{ds}$ и $\mathbf{n}$. Фактическое содержание этой формулы сводится к тому, что производная $\frac{d\mathbf{s}}{ds}$ есть вектор, нормальный к кривой. В остальном на нее надо смотреть как на определение двух новых понятий: величины $1/r$ и единичного вектора $\mathbf{n}$. Величина $1/r$ называется *кривизной кривой*, r — *радиусом кривизны*, а $\mathbf{n}$ — *единичным вектором главной нормали* к кривой. При этом кривизна $1/r$ считается существенно положительной, а потому единичный вектор $\mathbf{n}$ всегда направлен в сторону вогнутости кривой. Оправданием такой терминологии служит интуитивное представление, что при рассмотрении кривизны малый элемент кривой приближено можно рассматривать как дугу окружности. Это приближение тем точнее, чем меньше длина дуги Δs. В случае окружности кривизна $1/r$ постоянна на протяжении всей кривой. В общем случае произвольной гладкой кривой кривизна непрерывно меняется от точки к точке. Непрерывно меняется и направление единичного вектора главной нормали $\mathbf{n}$. Кинематическая формула (4.9) также справедлива для движения вдоль произвольной кривой и притом независимо от того, постоянна величина v или меняется с течением времени. Действительно, формула (4.10) получается из формулы (4.9) с помощью соотношения $ds = v dt$.

Все геометрические кривые разделяются на плоские и кривые двоякой кривизны. *Плоской кривой* называется кривая, все точки которой лежат в одной плоскости. Примерами плоских кривых являются окружность, эллипс, гипербола, парабола, синусоида и пр. *Кривыми двоякой кривизны* называются такие кривые, которые не лежат в одной плоскости. Примером подобной кривой может служить винтовая линия — спираль. Плоскость, в которой лежат касательная и главная нормаль к кривой, называется *соприкасающейся плоскостью*. Для плоской кривой соприкасающаяся плоскость совпадает с плоскостью,

в которой лежит кривая. К понятию соприкасающейся плоскости приводит следующее интуитивное представление. Произвольную конечную дугу кривой двоякой кривизны, разумеется, нельзя уложить в плоскость. Но чем короче дуга кривой, тем точнее она приближается к элементу плоской кривой, тем с меньшей ошибкой ее можно уложить в плоскости. Такая плоскость приближенно и воспроизводит соприкасающуюся плоскость. Это интуитивное представление можно превратить в точное определение с помощью предельного перехода. Пусть M (см. рис. 7) — произвольная точка на кривой. Проведем в ней касательную MC и хорду MM_1. Этими двумя прямыми, вообще говоря, определится некоторая плоскость CMM_1. Будем неограниченно приближать точку M_1 к точке M. Тогда указанная плоскость, вообще говоря, будет стремиться к некоторому определенному предельному положению. Это предельное положение и называется соприкасающейся плоскостью. Перпендикуляр к соприкасающейся плоскости в точке M называется *бинормалью к кривой.*

5. При равномерном вращении точки по окружности ускорение направленно к ее центру, т. е. перпендикулярно к траектории. Ускорение перпендикулярно к траектории и при движении по любой кривой, если только скорость движущейся точки не меняется по модулю. Не так будет, когда меняется также и модуль скорости. Чтобы разобраться в этом вопросе, представим вектор скорости в виде $\mathbf{v} = v\mathbf{s}$. Применяя к этому выражению правило дифференцирования произведения, получим

$$\mathbf{a} = \frac{d}{dt}(v\mathbf{s}) = \frac{dv}{dt}\mathbf{s} + v\frac{d\mathbf{s}}{dt},$$

или ввиду формулы (4.9)

$$\mathbf{a} = \frac{dv}{dt}\mathbf{s} + \frac{v^2}{r}\mathbf{n}. \tag{4.11}$$

Отсюда следует, что *вектор ускорения* $\mathbf{a}$ *лежит в плоскости векторов* $\mathbf{s}$ *и* $\mathbf{n}$, *т. е. в соприкасающейся плоскости*; *вектор* $\mathbf{a}$ *не имеет составляющей по бинормали к траектории.* В общем случае ускорение $\mathbf{a}$ направлено под углом к траектории. Первое слагаемое в формуле (4.11)

$$\mathbf{a}_t = \frac{dv}{dt}\mathbf{s} \tag{4.12}$$

есть вектор, направленный по касательной к траектории. Этот вектор называется *касательным* или *тангенциальным ускорением.* Второе слагаемое

$$\mathbf{a}_n = \frac{v^2}{r}\mathbf{n} \tag{4.13}$$

есть вектор, направленный вдоль главной нормали в сторону вогнутости траектории. Он называется *нормальным ускорением.* Таким обра-

зом, в общем случае *ускорение* $\mathbf{a}$ *можно представить в виде геометрической суммы тангенциального и нормального ускорений*:

$$\mathbf{a} = \mathbf{a}_t + \mathbf{a}_n. \tag{4.14}$$

Тангенциальное ускорение меняет скорость только по модулю, нормальное ускорение меняет ее только по направлению.

Рисунок 11 поясняет разложение полного ускорения на тангенциальное и нормальное. Пусть $\mathbf{v}$ — скорость материальной точки в момент времени t, когда она находилась в положении M. Обозначим через $\mathbf{v}_1 = \mathbf{v} + \Delta\mathbf{v}$ скорость той же точки в момент $t + \Delta t$, когда она переместилась в положение M_1 (не обозначенное на рисунке). Отложим оба вектора $\mathbf{v}$ и $\mathbf{v}_1$ из одной и той же точки M и разложим приращение $\Delta\mathbf{v}$ скорости на две составляющие: составляющую $\Delta\mathbf{v}_t$ вдоль вектора $\mathbf{v}$ и составляющую $\Delta\mathbf{v}_n$, перпендикулярную к этому вектору. При уменьшении Δt оба отношения $\frac{\Delta v_t}{\Delta t}$ и $\frac{\Delta v_n}{\Delta t}$ будут стремиться к определенным пределам. Первый из них есть тангенциальное, а второй — нормальное ускорение.

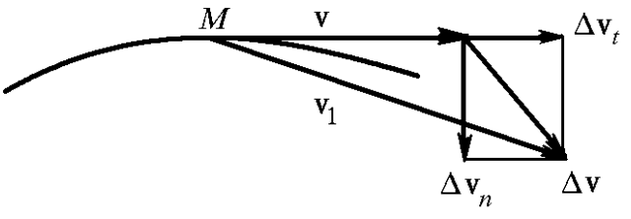


Рис. 11

При вычислении скорости точки бесконечно малую дугу траектории можно аппроксимировать бесконечно коротким прямолинейным отрезком, направление которого совпадает с направлением касательной к траектории. При определении ускорения такая аппроксимация уже не годится. Однако, как видно из рассуждений настоящего параграфа, при вычислении ускорения бесконечно малую дугу траектории можно аппроксимировать дугой окружности, плоскость которой совпадает с соприкасающейся плоскостью, а радиус равен радиусу кривизны траектории в рассматриваемой точке. Но и такая аппроксимация оказалась бы недостаточной, если бы потребовалось вычислить производные радиуса-вектора более высокого порядка: $\dddot{\mathbf{r}}$, $\ddddot{\mathbf{r}}$ и т. д.

ЗАДАЧИ

1. Шарик, которому сообщена горизонтальная скорость v, падает на горизонтальную плиту с высоты h. При каждом ударе о плиту теряется часть скорости (отношение вертикальной составляющей скорости после удара к ее значению до удара постоянна и равна α).

Определить, на каком расстоянии x от места бросания отскоки шарика прекратятся. Считать, что трение отсутствует, так что горизонтальная составляющая скорости шарика v не меняется.

Ответ: $x = v\sqrt{\frac{2h}{g}}\,\frac{1+\alpha}{1-\alpha}$.

2. Точка движется в плоскости, причем ее прямоугольные координаты определяются уравнениями

$$x = A\cos\omega t, \quad y = B\sin\omega t, \tag{4.15}$$

где A, B, ω — постоянные. По какой траектории движется точка? Вычислить ее ускорение.

Р е ш е н и е. Исключая время t из уравнений (4.15), находим

$$\frac{x^2}{A^2} + \frac{y^2}{B^2} = 1.$$

Точка движется по эллипсу. Ее радиус-вектор $\mathbf{r} = x\mathbf{i} + y\mathbf{i}$, а ускорение $\mathbf{a} = \ddot{x}\mathbf{i} + \ddot{y}\mathbf{j}$, Дифференцирование дает

$$\dot{x} = -\omega A\sin\omega t, \quad \ddot{x} = -\omega^2 A\cos\omega t = -\omega^2 x,$$

$$\dot{y} = \omega B\cos\omega t, \quad \ddot{y} = -\omega^2 B\sin\omega t = -\omega^2 y.$$

Следовательно,

$$\mathbf{a} = -\omega^2(x\mathbf{i} + y\mathbf{j}) = -\omega^2\mathbf{r}. \tag{4.16}$$

Ускорение направлено к центру эллипса и пропорционально r. В частном случае $A = B$ — эллипс вырождается в круг, а формула (4.16) переходит в известную формулу для центростремительного ускорения при равномерном вращении по кругу.

3. Точка с постоянной по модулю скоростью v движется по эллипсу с полуосями A и B. Найти ее ускорение в вершинах эллипса.

Р е ш е н и е. В вершинах эллипса ускорение только нормальное. Его можно найти по общей формуле (4.13) и притом в любой точке эллипса. Однако в вершинах эллипса поучительно получить решение, воспользовавшись тем обстоятельством, что эллипс может быть получен однородным растяжением или сжатием круга. Возьмем круг радиуса $r = B$ и равномерно растянем его параллельно оси X (рис. 11a) в A/B раз. Получится эллипс с полуосью A, параллельной оси X. Бесконечно малая дуга CED круга перейдет в дугу $C'E'D'$ эллипса. Время движения точки вдоль этих дуг будет одно и то же. Одинаковы будут и скорости точки в положениях E и E'. Однако нормальное смещение FE перейдет в $F'E'$, т. е. увеличится в A/B раз. В такое же число раз увеличится нормальное ускорение. В точке E' оно будет $a = (v^2/r)A/B = Av^2/B^2$. Сжимая круг радиуса A вдоль оси Y, таким же путем получаем ускорение в точке пересечения эллипса с малой осью. Оно равно $a = Bv^2/A^2$.

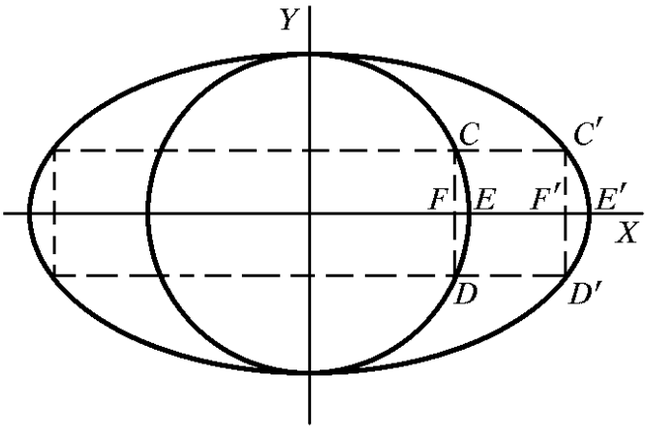


Рис. 11а

4. Установить связь между звездными и средними солнечными сутками. *Звездный год*, т. е. промежуток времени, в течение которого Солнце совершает свой видимый путь по небесной сфере относительно звезд, составляет 365,2564 средних солнечных суток. (Звездный год следует отличать от *тропи-*

ческого года, который соответствует периоду смены времен года и составляет 365,2422 средних солнечных суток.)

Р е ш е н и е. Пусть в положении *1* (рис. 12) плоскость земного меридиана *AB* проходит через центр Солнца *C* и какую-либо (бесконечно удаленную) звезду *D*. Когда Земля в своем орбитальном движении перейдет в положение *2*, плоскость того же меридиана повернется относительно направления на звезду на угол α, а относительно направления на центр Солнца — на угол β. Углы α и β могут превышать 2π, но они всегда связаны соотношением $\alpha = \beta + \gamma$, где γ — угол между направлениями на центр Солнца в положениях *1* и *2*. Спустя звездный год, когда Земля вернется в исходную плоскость *1CD*, угол γ примет значение 2π, а потому в этом положении $\alpha = \beta + 2\pi$. За это время пройдет $N_{зв} = \alpha/2\pi$ звездных и $N_{сол} = \beta/2\pi$ средних солнечных суток. Поэтому $N_{зв} = N_{сол} + 1$. Если $T_{зв}$ и $T_{сол}$ — продолжительность звездных и средних солнечных суток, то очевидно, что $N_{зв}T_{зв} = N_{сол}T_{сол}$, так как оба эти выражения представляют одно и то же время — звездный год. Используя соотношения $N_{зв} = N_{сол} + 1$, находим

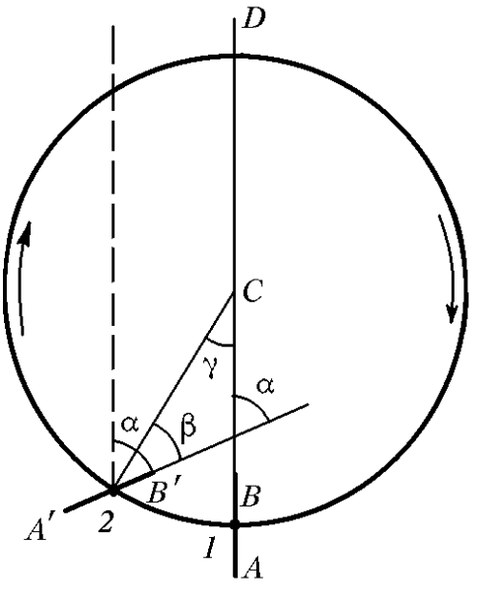


Рис. 12

$$T_{зв} = \frac{N_{сол}}{N_{сол}+1}\,T_{сол},$$

$$T_{сол} - T_{зв} = \frac{1}{N_{сол}+1}\,T_{сол}.$$

Подставив сюда $T_{сол} = 24 \cdot 60 \cdot 60 = 86\,400$ с, $N_{сол} = 365{,}2564$, получим

$$T_{сол} - T_{зв} = 325{,}9003 \text{ с} \approx 236 \text{ с}, \quad T_{зв} \approx 86\,164 \text{ с}.$$

Заметим, что при решении мы не вводили предположения, что Земля по своей орбите движется равномерно.

5. Определить скорость, с которой движется тень Луны по земной поверхности во время полного солнечного затмения.

Р е ш е н и е. Для простоты примем, что затмение наблюдается на экваторе и что земная ось перпендикулярна плоскостям солнечной и лунной орбит. Скорость света будем считать бесконечно большой по сравнению со всеми остальными скоростями, входящими в задачу. Пусть в рассматриваемый момент времени прямая Солнце—Луна перпендикулярна к земной поверхности в точке наблюдения *A* (рис. 13). Поверхность Земли в окрестности той же точки можно считать плоской. При решении выберем сначала систему отсчета, в которой Земля покоится. Пусть $\omega_С$ и $\omega_Л$ — угловые скорости вращения Солнца и Луны вокруг центра Земли, $R_С$ и $R_Л$ — расстояния их от того же центра, r — радиус Земли. За 1 с Солнце и Луна переместятся с востока на запад на расстояния $CC' = \omega_С R_С$ и $ЛЛ' = \omega_Л R_Л$. Соединив новые положения Солнца и Луны прямой линией, найдем, что за 1 с

граница лунной тени переместится по земной поверхности с запада на восток на расстояние $v = AA'$. Это расстояние и есть скорость движения тени Луны. Из рис. 13 видно, что

$$\frac{v}{\omega_C R_C} = \frac{x}{OC} \approx \frac{x}{R_C},$$

так как расстояние до Луны пренебрежимо мало по сравнению с расстоянием до Солнца, и можно принять $OC = R_C$. Таким образом, $v = \omega_C x$. Для нахождения x составляем пропорцию

$$\frac{\omega_C R_C}{\omega_Л R_Л} = \frac{CC'}{ЛЛ'} = \frac{OC}{OЛ}.$$

Полагая в ней $OC = R_C$, $OЛ = R_Л - x - r$, получим уравнение для нахождения x. Оно дает

$$x = \frac{\omega_C - \omega_Л}{\omega_C} R_Л - r.$$

Следовательно, скорость движения лунной тени с запада на восток будет

$$v = \omega_C x = (\omega_C - \omega_Л) R_Л - \omega_C r.$$

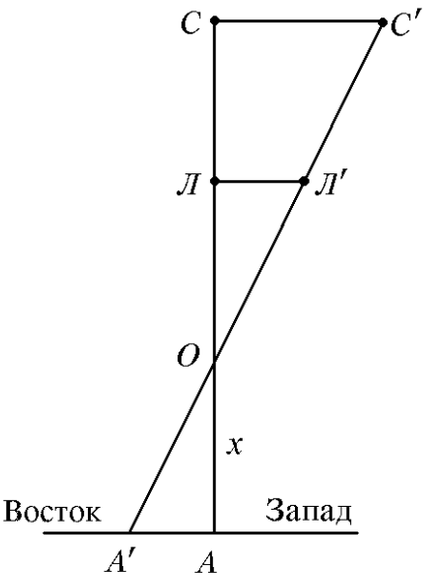


Рис. 13

Здесь $\omega_C = 2\pi/T_{сут}$, $\omega_C - \omega_Л = 2\pi/T_{мес}$, где $T_{сут} = 86\,400$ с — продолжительность солнечных суток, а $T_{мес} = 29{,}6\, T_{сут}$ — продолжительность месяца. Используя эти соотношения и подставляя числовые значения $R_Л = 3{,}8 \cdot 10^5$ км, $r = 6400$ км, получим

$$v = \frac{2\pi R_Л}{T_{мес}} - \frac{2\pi r}{T_{сут}} \approx 0{,}47 \text{ км/с}. \qquad (4.17)$$

Смысл последней формулы легко уяснить, перейдя в систему отсчета, в которой Солнце покоится. Считая Солнце бесконечно удаленным, можно отвлечься от движения центра Земли, приняв во внимание лишь вращение Земли вокруг своей оси, а также движение Луны по ее орбите вокруг Земли. Луна движется по орбите с запада на восток со скоростью $v_Л = 2\pi R_Л/T_{мес}$. Если бы Земля не вращалась, то с той же скоростью и в том же направлении по ее поверхности бежала бы лунная тень. Но из-за вращения Земли экваториальные точки последней движутся с запада на восток со скоростью $v_З = 2\pi r/T_{сут}$. Для нахождения скорости лунной тени эту величину надо вычесть из $v_Л$, что и сделано в формуле (4.17).

§ 5. ГРАНИЦЫ ПРИМЕНИМОСТИ КЛАССИЧЕСКОГО СПОСОБА ОПИСАНИЯ ДВИЖЕНИЯ

В классической механике состояние движения частицы в любой момент времени характеризуется положением (координатой x при одномерном движении) и *скоростью* **v**. Вместо скорости можно поль-

зоваться также *импульсом*, т. е. величиной $\mathbf{p} = m\mathbf{v}$, равной произведению *массы частиц* m на ее скорость *). Образом частицы является геометрическая точка, описывающая с течением времени непрерывную траекторию. В квантовой механике показано, что такой способ описания движения имеет принципиальные границы применимости.

Здесь преждевременно вдаваться в подробное обсуждение этого вопроса. Достаточно ограничиться предварительным сообщением основного результата, не касаясь его обоснования.

Согласно квантовой механике состояние частицы в каждый момент времени нельзя характеризовать точными значениями ее координаты и импульсом в этот момент времени. Если в каком-либо состоянии координата известна с неопределенностью δx, а импульс — с неопределенностью δp, то обе эти величины одновременно не могут быть сделаны сколь угодно малыми. Они связаны соотношением

$$\delta x \cdot \delta p \gtrsim h, \tag{5.1}$$

где h — универсальная постоянная, называемая *постоянной Планка* в честь немецкого физика-теоретика Макса Планка (1858—1947). Она играет основную роль во всех квантовых явлениях. Ее числовое значением равно

$$h = 6{,}626171 \cdot 10^{-27} \text{ эрг} \cdot \text{с}.$$

Соотношение (5.1) называется *принципом неопределенностей Гайзенберга* по имени немецкого физика-теоретика Вернера Гайзенберга (1901--1976). Это соотношение определяет принципиальный предел точности одновременного измерения координаты и импульса частицы, который не может быть превзойден никаким усовершенствованием приборов и методов измерения. Дело здесь не в ошибках измерений. Такова уж природа реальных частиц, что мгновенные состояния их движения не могут быть охарактеризованы классически — точными значениями координат и импульсов. Частицы ведут себя более сложно, чем материальные точки классической механики. Классическая картина движения по непрерывным траекториям лишь приближенно соответствует законам природы. Границы ее применимости определяются соотношением неопределенностей (5.1). Из него следует, что мгновенное состояние движения частицы нельзя также характеризовать абсолютно точными значениями координаты и скорости. Неопределенности этих величин должны удовлетворять условию

$$\delta x \cdot m\delta v \gtrsim h. \tag{5.2}$$

Для макроскопических тел практическая применимость классического способа движения не вызывает сомнений. Допустим, например, что речь идет о движении шарика с массой $m = 1$ г. Обычно положение шарика практически может быть определено с точностью

*) Понятия массы и импульса вводятся и подробно обсуждаются в § 10.

до десятой или сотой доли миллиметра. Во всяком случае, вряд ли имеет смысл говорить о погрешности в определении положения шарика, меньшей размеров атома. Положим поэтому $\delta x = 10^{-8}$ см. Тогда из соотношения неопределенностей (5.1) найдем

$$\delta v \gtrsim \frac{6{,}63 \cdot 10^{-27}}{10^{-8}} \approx 10^{-18} \text{ см/с}.$$

Одновременная малость величин δx и δv и является доказательством практической применимости классического способа описания движения для макроскопических тел. Не так обстоит дело, когда речь идет об атомных явлениях — явлениях, происходящих с частицами очень малой массы в очень малых объемах пространства. Рассмотрим, например, движение электрона в атоме водорода. Масса электрона $m = 9{,}11 \cdot 10^{-28}$ г. Погрешность в положении электрона δx во всяком случае не должна превышать размеры атома, т. е. погрешность δx должна быть $< 10^{-8}$ см. Но тогда из соотношения неопределенностей получаем

$$\delta v = \frac{h}{m \delta x} = \frac{6{,}62 \cdot 10^{-27}}{9{,}11 \cdot 10^{-28} \cdot 10^{-8}} \approx 7 \cdot 10^{8} \text{ см/с}.$$

Эта величина не меньше, а даже больше самой скорости электрона в атоме, которая по порядку величины равна 10^8 см/с. При таком положении классическая картина движения теряет всякий смысл.

§ 6. О СМЫСЛЕ ПРОИЗВОДНОЙ И ИНТЕГРАЛА В ПРИЛОЖЕНИЯХ К ФИЗИЧЕСКИМ ВОПРОСАМ

1. Процесс предельного перехода (3.4), с помощью которого определяется производная, называется *дифференцированием*. Понятие производной широко используется в механике и во всех других разделах физики. Именно задача об определении скорости произвольного движения привела к этому понятию Ньютона, который, наряду с Готфридом Лейбницем (1646—1716), является основоположником дифференциального и интегрального исчислений. Обозначение $\frac{dx}{dt}$ для производной принадлежит Лейбницу. На символ $\frac{dx}{dt}$ в математике следует смотреть как на единое целое, а не как на отношение двух «бесконечно малых» приращений dx и dt. Смысл производной $\dot{x} = \frac{dx}{dt}$ точно определен соотношением (3.4). Сначала надо образовать отношение конечных приращений $\frac{\Delta x}{\Delta t}$, предполагая, что Δt не равно нулю. Затем путем преобразований этого отношения или каким-либо иным способом следует совершить переход к пределу. Но ни в коем случае нельзя представлять себе, что сначала

совершен какой-то предельный переход от Δx и Δt к «бесконечно малым» величинам dx и dt, называемым *дифференциалами* функции x и аргумента t, а затем взято отношение этих дифференциалов $\frac{dx}{dt}$. Такой взгляд на производную существовал в начальной стадии развития дифференциального исчисления. Однако он несовместим с требованием математической ясности понятий, да и вообще лишен смысла. Правда, можно так определить дифференциалы dx и dt, что их отношение сделается равным производной $\dot{x}$. В математике дифференциал dt определяется как произвольное приращение аргумента t, а дифференциал dx — с помощью соотношения $dx = \dot{x}dt$. Но теперь в утверждении, что производная есть отношение двух конечных величин dx и dt, нет ничего удивительного, это — простая тавтология, иной способ выражения. Первичным по-прежнему является понятие производной, а не дифференциала.

Однако в приложениях математики к физике надо считаться с тем, что физические величины получаются в конце концов в результате конкретных измерений, а все измерения сопровождаются погрешностями и вносят искажения в естественный ход явлений. Это обстоятельство, строго говоря, делает невозможным предельный переход $\Delta t \to 0$, $\Delta x \to 0$, вводимый в математике при определении производной. Допустим, например, что измеряется скорость движущейся пули в воздухе. Задача сводится к измерению расстояния Δx и промежутка времени Δt, за который пуля проходит это расстояние. Если время Δt взять очень большим, то за это время скорость пули может заметно уменьшиться из-за сопротивления воздуха. Отношение $\frac{\Delta x}{\Delta t}$ в этом случае может оказаться заметно меньше скорости пули в рассматриваемый момент времени. Уменьшая время Δt, мы заметим, что начиная с определенного момента отношение $\frac{\Delta x}{\Delta t}$ в пределах доступной точности изменяется, если отвлечься от случайных погрешностей, сопровождающих каждое измерение. Дальнейшее уменьшение Δt бессмысленно. Оно может только ухудшить дело, так как при дальнейшем уменьшении Δt отношение $\frac{\Delta x}{\Delta t}$ начинает изменяться снова и притом все более и более нерегулярно. Оно принимает различные значения от очень больших до очень малых. Это обусловлено тем, что точность любого измерения тем меньше, чем меньше измеряемая величина. Не представляет, например, особо большого труда измерить длину в 1 м с погрешностью до 1 мм, т. е. с относительной точностью 1/1000. Но измерить с такой же относительной точностью длину в 1 мм значительно труднее. Чем меньше Δt, тем больше погрешность, с которой мы вычисляем отношение $\frac{\Delta x}{\Delta t}$. Если Δt уменьшать беспредельно, то вычисленные значения отношения $\frac{\Delta x}{\Delta t}$ не будут стремиться ни к какому определенному пределу. Это показывает, что в рассматриваемом

примере из-за погрешностей измерений предельный переход к $\Delta t \to 0$ не может быть осуществлен в строго математическом смысле. Вычислить истинную скорость или производную $v = \dot{x}$ из физических измерений можно лишь приближенно, отождествляя ее с отношением конечных приращений $\frac{\Delta x}{\Delta t}$. Оптимальное значение времени Δt, при котором точность вычисления истинной скорости максимальна, определяется конкретными условиями. Малые, но конечные приращения Δx и Δt, отношение которых с достаточной точностью аппроксимирует производную $\dot{x}$, физик называет *бесконечно малым* или, полнее, *физически бесконечно малыми величинами*. Он обозначает их через dx и dt и обращается с ними как с математическими дифференциалами. Таким образом, в *физике производная выступает как отношение конечных, но достаточно малых приращений функции и аргумента, а не как предел этого отношения.*

Однако не только погрешности измерений могут сделать невозможным практическое выполнение предельного перехода в строго математическом смысле. Такая невозможность может быть и принципиальной, обусловленной самой природой физических величин и физических законов. Так, точное выполнение предельного перехода невозможно из-за соотношения неопределенностей (5.1). Действительно, если промежуток времени Δt стремится к нулю, то при этом будет стремиться к нулю и проходимое расстояние Δx. Неопределенность δx в измерении проходимого расстояния не должна превосходить Δx. Иначе вычисление средней скорости по формуле $v_{\text{ср}} = \frac{\Delta x}{\Delta t}$ потеряло бы всякий смысл. Таким образом, при $\Delta t \to 0$ должна стремиться к нулю и неопределенность в координате δx. Но тогда, согласно соотношению (5.1), неопределенность скорости δv будет стремиться к бесконечности. Это значит, что погрешность, возникающая при вычислении скорости v по формуле (3.3), сколь угодно велика по сравнению с самой скоростью v.

2. Изложенные выводы относятся не только к производной координаты, но и к производным всяких физических величин. Допустим, например, что требуется определить плотность вещества в какой-либо точке пространства. С этой целью можно поступить следующим образом. Окружим рассматриваемую точку замкнутой поверхностью, ограничивающей объем ΔV. Обозначим через Δm массу вещества, содержащегося в этом объеме. Отношение

$$\rho_{\text{ср}} = \frac{\Delta m}{\Delta V}$$

называется *средней плотностью вещества в объеме* ΔV. Средняя плотность, вообще говоря, зависит от объема ΔV, внутри которого находится рассматриваемая точка, а также от его формы. Чтобы исключить эту зависимость, вводят понятие *истинной плотности* вещества, определяя ее путем предельного перехода $\Delta V \to 0$. Обычно

говорят, что при этом средняя плотность $\rho_{ср}$ стремится к определенному пределу ρ, который и называется истинной плотностью вещества в рассматриваемой точке пространства:

$$\rho = \lim_{\Delta V \to 0} \frac{\Delta m}{\Delta V} = \frac{dm}{dV}. \tag{6.1}$$

Истинная плотность определяется, таким образом, *как производная массы по объему.* Эта величина зависит только от положения точки, к которой она относится. Однако если в формуле (6.1) предельный переход понимать буквально в строго математическом смысле, то для реальных тел он выполнен быть не может из-за атомистической структуры вещества. При уменьшении объема в нем рано или поздно окажется лишь небольшое число молекул, например одна или даже ни одной молекулы. Кроме того, молекулы совершают беспорядочные тепловые движения, одни молекулы уходят из объема ΔV, другие вступают в него. Ввиду этого число молекул в фиксированном малом объеме ΔV весьма быстро и беспорядочно меняется во времени. При уменьшении ΔV отношение $\frac{\Delta m}{\Delta V}$ будет быстро и беспорядочно меняться от нуля, когда внутри объема ΔV нет молекул, до очень больших значений, когда в него попадает одна или несколько молекул. При бесконечном уменьшении ΔV отношение $\frac{\Delta m}{\Delta V}$ не будет стремиться к определенному пределу. Ввиду этого при определении истинной плотности вещества нельзя брать величины Δm и ΔV сколь угодно малыми. Объем ΔV должен иметь макроскопические размеры, т. е. содержать еще очень большое число молекул. Но он должен быть достаточно мал, чтобы содержащееся в нем вещество могло рассматриваться приближенно как макроскопически однородное. Если оба эти требования выполняются, то отношение $\frac{\Delta m}{\Delta V}$ будет иметь практически вполне определенное значение, не меняющееся при дальнейшем уменьшении макроскопического объема ΔV. Это отношение мы и принимаем в физике за производную массу m по объему V. Величины Δm и ΔV, удовлетворяющие указанным двум требованиям, в физике рассматриваются как *физически бесконечно малые*, и с ними физика обращается как с математическими дифференциалами. Математически этому соответствует замена реального тела идеализированной моделью с непрерывным распределением масс.

3. Совершенно так же обстоит дело с понятием интеграла. В математике интеграл определяется предельным переходом

$$\int_a^b f(x)\, dx = \lim_{\Delta x_i \to 0} \sum f(x_i)\, \Delta x_i.$$

Числовой промежуток (a, b) разбивается на n частичных промежутков Δx_1, Δx_2, ..., Δx_n. Длина каждого из них Δx_i умножается на

значение функции $f(x)$ в произвольной точке, лежащей внутри рассматриваемого частичного промежутка. Затем составляется сумма $\sum f(x_i)\,\Delta x_i$ и выполняется переход к пределу $n \to \infty$ в предположении, что длина каждого из частичных промежутков стремится к нулю. В физике, однако, из-за погрешностей измерений или по принципиальным соображениям (например, из-за атомистической структуры вещества) деление промежутка (a, b) на частичные промежутки меньше определенной длины (значение которой зависит от конкретных условий) теряет смысл. Ввиду этого предельный переход к $\Delta x_i \to 0$ не может быть выполнен до конца, а должен быть оборван в каком-то месте. Это означает, что *в физике интеграл выступает не как предел суммы, а как сумма большого числа достаточно малых слагаемых* $\sum f(x_i)\,\Delta x_i$.

§ 7. О ВЕКТОРАХ И СЛОЖЕНИИ ДВИЖЕНИЙ

1. Понятие вектора и основные операции векторной алгебры мы считаем известными. Остановимся только на разъяснении некоторых принципиальных моментов, представляющий особый интерес в физике. Среди физических величин встречаются величины, не имеющие направления, и величины, которым можно приписать определенное направление. Величины первого рода называются *скалярными*. К ним относятся, например, масса, энергия, температура, электрический заряд и пр. Величины второго рода называются *векторами*. Примерами векторов являются скорость, ускорение, сила, напряженности электрического и магнитного полей и пр. Векторы принято изображать направленными отрезками или стрелками и обозначать прямыми буквами полужирного шрифта ($\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, ...) или (реже) буквами, над которыми поставлены стрелки ($\vec{A}$, $\vec{B}$, $\vec{C}$, ...).

В качестве дополнения к приведенному определению иногда указывают, что не всякие направленные величины являются векторами, а только такие, которые складываются *геометрически*, т. е. по *правилу параллелограмма*. Однако это указание остается расплывчатым и бессодержательным, пока не сказано, что следует понимать под сложением рассматриваемых физических величин. Смысл сложения физических величин еще не определяется их физической природой. Сначала надо указать, что́ мы понимаем под сложением двух физических величин, а затем уже находить правила, по которым должно проводиться это сложение. Только тогда указание, о котором говорилось выше, приобретает определенное содержание. Нередко для решения вопроса, являются ли рассматриваемые физические величины векторами или нет, в их сложение вкладывают такой смысл, который к этому вопросу не имеет никакого отношения.

2. Например, сложение скоростей в механике понимают в следующем смысле. Пусть точка движется относительно системы отсчета

S_1 со скоростью $\mathbf{v}_1$ (например, пассажир идет по палубе корабля). Пусть далее система отсчета S_1 сама движется со скоростью $\mathbf{v}_2$ относительно другой системы отсчета S_2, условно принимаемой за неподвижную (например, корабль движется относительно берега). Под сложением движения понимают операцию, с помощью которой по этим данным можно найти скорость $\mathbf{v}$ точки (пассажира) относительно неподвижной системы S_2 (берега). В релятивистской кинематике это определение должно быть дополнено указанием, что каждая из скоростей $\mathbf{v}_1$ и $\mathbf{v}_2$ измеряется с помощью линеек и часов в той системе отсчета, относительно которой рассматривается движение. В нерелятивистской кинематике такое указание излишне, так как длины и промежутки времени в ней имеют абсолютный смысл, т. е. не зависят от системы отсчета. И вот оказывается, что сложение движений в указанном смысле в нерелятивистской кинематике производится по правилу параллелограмма, а в релятивистской кинематике это правило не справедливо. Тем не менее скорость точки считается вектором как в той, так и в другой кинематике. Это показывает, что правило параллелограмма скоростей для сложения движений в указанном смысле не имеет никакого отношения к вопросу о том, является скорость вектором или нет *).

Да и в самой нерелятивистской кинематике можно указать величины, которые считаются векторами, но тем не менее не всегда складываются по правилу параллелограмма, если в сложение этих величин вложить примерно такой же смысл, что и в сложение скоростей в вышеприведенном примере. К таким величинам относится, например, ускорение. Пусть точка движется относительно системы отсчета S_1 с ускорением $\mathbf{a}_1$, а система S_1 имеет ускорение $\mathbf{a}_2$ относительно «неподвижной» системы отсчета S_2. По этим данным можно найти ускорение $\mathbf{a}$ точки относительно системы S_2 только в том случае, когда складываемые движения поступательные. В этом случае вектор $\mathbf{a}$ находится по правилу параллелограмма. В остальных случаях для нахождения результирующего ускорения знания ускорений $\mathbf{a}_1$ и $\mathbf{a}_2$ недостаточно, и само нахождение вектора $\mathbf{a}$ производится по более сложному правилу, которое будет рассмотрено в § 64.

3. Приведенные примеры показывают, что определение вектора нуждается в уточнении. Необходимость этого диктуется также сле-

*) Если бы все скорости измерялись в одной и той же «неподвижной» системе отсчета S_2, то правило параллелограмма сохраняло бы силу и в релятивистской кинематике. Однако при этом изменился бы смысл скорости $\mathbf{v}_1$. Под $\mathbf{v}_1$ следовало бы понимать скорость точки относительно движущейся системы отсчета S_1, *измеренную в «неподвижной» системе* S_2. При сложении же скоростей в том смысле, в каком оно понимается в тексте, $\mathbf{v}_1$ есть скорость точки относительно движущейся системы S_1, *измеренная в той же системе*. А это существенно иная величина. Только в предельном случае бесконечно медленных движений обе скорости совпадают. При изложении теории относительности затронутые вопросы будут разобраны подробно.

дующими соображениями. Не всегда очевидно, какое направление следует приписать той или иной физической величине. Например, в случае геометрического отрезка AB не возникает вопроса, что следует считать его направлением. За таковое можно принять либо направление от точки A к точке B, либо противоположное направление — от точки B к точке A. Не возникает вопроса, что́ следует считать направлением смещения, скорости или ускорения точки, а также направлением силы, на нее действующей. Однако не очевидно, что́ следует считать за направление угловой скорости или геометрической поверхности, в особенности, когда последняя изогнута. Наконец, точное определение вектора необходимо дать для того, чтобы обобщить это понятие на случай *многомерных пространств*. Чтобы прийти к такому определению, рассмотрим сначала простейший вектор, а именно геометрический прямолинейный отрезок, на котором установлено определенное направление. Такой направленный отрезок будем изображать стрелкой **а**. Возьмем какую-либо произвольную прямоугольную или косоугольную систему координат и спроецируем отрезок **а** на координатные оси X, Y, Z. Проецирование будем производить плоскостями, параллельными координатным плоскостям. Например, чтобы получить проекцию на ось X, надо через концы отрезка **а** провести плоскости, параллельные координатной плоскости YZ. Эти плоскости и отсекут на оси X отрезок a_x, являющийся проекцией отрезка **а** на рассматриваемую ось. Аналогично получаются проекции a_y и a_z. Обычно рассматривают прямоугольные координатные системы. Тогда a_x, a_y, a_z будут *прямоугольными*, или *ортогональными*, *проекциями* отрезка **а**. Если проекции a_x, a_y, a_z известны в какой-либо системе координат S, то их можно найти и в любой другой координатной системе S', оси которой произвольным образом повернуты относительно системы S. Для этого по проекциям a_x, a_y, a_z в системе S надо восстановить отрезок **а** как диагональ параллелепипеда, построенного на отрезках a_x, a_y, a_z. Затем следует, спроецировать этот отрезок на оси X', Y', Z' новой системы координат S'. Получится тройка чисел $a_{x'}$, $a_{y'}$, $a_{z'}$, которые и являются проекциями отрезка **а** в новой системе координат. Теперь мы даем следующее определение вектора.

Вектором **а** называется упорядоченная тройка чисел a_x, a_y, a_z, заданная в каждой системе координат. (Упорядочение состоит в том, что первое число a_x приводится в соответствии оси X, второе число a_y — оси Y, третье число a_z — оси Z.) Эти числа называются *проекциями вектора* **а** на соответствующие координатные оси. Их называют также *составляющими*, или *компонентами*, *вектора. При переносе начала и повороте координатных осей составляющие* a_x, a_y, a_z *преобразуются по правилу преобразования проекций геометрических отрезков.*

Короче, *вектором* называется упорядоченная тройка чисел, заданная в каждой системе координат, которые при переносе начала

и повороте координатных осей преобразуются как разности координатных концов направленного геометрического отрезка.

Отложив эти числа вдоль координатных осей X, Y, Z, мы отсечем на них три отрезка. Если на таких трех отрезках как ребрах построить параллелепипед, то его диагональ можно рассматривать как направленный отрезок, служащий наглядным изображением вектора. Этот отрезок получится одним и тем же, какую бы систему координат мы не использовали при его построении. В этом проявляется *инвариантный* характер вектора, т. е. независимость его от системы координат, использованной для его представления. Компоненты вектора a_x, a_y, a_z в разных системах координат разные, но самый вектор $\mathbf{a}$ один и тот же. Векторное равенство $\mathbf{a} = \mathbf{b}$, записанное в координатной форме, равносильно трем равенствам $a_i = b_i$ $(i = x, y, z)$. При переходе к другой (штрихованной) системе координат обе части этих равенств преобразуются одинаково. Поэтому в новой системе координат они сохраняют прежний вид, т. е. $a_{i'} = b_{i'}$ $(i' = x', y', z')$. Уравнения, обе части которых при переходе к другой системе координат преобразуются одинаково и благодаря этому сохраняют свой вид во всех координатных системах, называются *ковариантными или инвариантными* по отношению к рассматриваемому преобразованию координатных систем. Мы видим, что *векторное уравнение* $\mathbf{a} = \mathbf{b}$ *инвариантно по отношению к переносу начала и повороту координатных осей. Ввиду этой инвариантности уравнения, выражающие физические законы в векторной форме, не зависят от выбора осей координат.* С помощью векторов физические законы формулируются в простой и обозримой форме, которая не сохраняется, если выразить их через проекции векторов в какой-либо системе координат.

Заметим, что координатные оси X, Y, Z не обязательно должны поворачиваться вместе подобно повороту твердого тела. Определение предусматривает и такие случаи, когда оси X, Y, Z поворачиваются независимо. Путем поворотов такого типа может быть совершен переход от любой прямолинейной системы координат к другой прямолинейной системе — правой или левой, оси которой ориентированы совершенно произвольно. В частности, такими поворотами может быть осуществлена инверсия осей, т. е. одновременное изменение на противоположные положительных направлений всех трех осей.

Если обе координатные системы прямоугольные, то формулы преобразования проекций имеют следующий вид:

$$\begin{aligned} a_{x'} &= \alpha_{x'x} a_x + \alpha_{x'y} a_y + \alpha_{x'z} a_z, \\ a_{y'} &= \alpha_{y'x} a_x + \alpha_{y'y} a_y + \alpha_{y'z} a_z, \\ a_{z'} &= \alpha_{z'x} a_x + \alpha_{z'y} a_y + \alpha_{z'z} a_z, \end{aligned} \tag{7.1}$$

где $\alpha_{x'x}$, $\alpha_{x'y}$, ... — косинусы углов между соответствующими координатными осями обеих систем координат. Например, $\alpha_{y'z}$ означает косинус угла между положительными направлениями осей Y' и Z.

4. Аналогично *скаляром* или *инвариантом* называется число, заданное в каждой системе координат, причем при переносе начала и повороте координатных осей это число остается неизменным. Таким образом, как и определение вектора, это определение предусматривает только перенос начала и поворот координатных осей. Оно предполагает, что обе координатные системы должны оставаться неподвижными одна относительно другой. Примерами скаляров являются время, масса, электрический заряд и пр. Абсцисса x неподвижной точки не является скаляром, так как ее числовое значение в разных системах координат разное. Скаляры можно образовывать из векторов. Например, скаляром является *длина вектора* или *ее квадрат*, который в прямоугольной системе координат представляется выражением $a_x^2 + a_y^2 + a_z^2$. *Скаляром является скалярное произведение двух векторов* $\mathbf{a}$ и $\mathbf{b}$, *т. е. величина* $(\mathbf{ab}) = ab\cos\vartheta$, где ϑ — угол между этими векторами. В прямоугольной системе координат, как известно, скалярное произведение представляется выражением $(\mathbf{ab}) = a_xb_x + a_yb_y + a_zb_z$ (см. задачи 1 и 3 к этому параграфу).

5. На основании изложенного ясно, что для доказательства векторного характера той или иной направленной физической величины надо только установить, как определяются ее составляющие вдоль координатных осей и как они преобразуются при переходе от одной координатной системы к любой другой, оси которой повернуты относительно осей первоначальной системы. При этом имеются в виду координатные системы, неподвижные одна относительно другой.

Например, двум векторам $\mathbf{a}$ и $\mathbf{b}$ с составляющими a_x, a_y, a_z и b_x, b_y, b_z можно сопоставить в каждой системе координат упорядоченную тройку чисел $c_x = a_x + b_x$, $c_y = a_y + b_y$, $c_z = a_z + b_z$. Легко видеть, что такая тройка чисел образует вектор, так как эти числа подчиняются тем же правилам преобразования, что и составляющие векторов $\mathbf{a}$ и $\mathbf{b}$. Вектор $\mathbf{c}(c_x, c_y, c_z)$ называется *суммой векторов* $\mathbf{a}$ *и* $\mathbf{b}$. Легко доказать, что он может быть получен из векторов $\mathbf{a}$ и $\mathbf{b}$ геометрическим построением по правилу параллелограмма. Аналогично определяется и *вычитание векторов. Разность двух векторов* $\mathbf{a}$ и $\mathbf{b}$ есть вектор $\mathbf{d}$, определяемый упорядоченной тройкой чисел $d_x = a_x - b_x$, $d_y = a_y - b_y$, $d_z = a_z - b_z$. Для его построения надо изменить на противоположное направление вектора $\mathbf{b}$ (получаемый таким путем вектор обозначают $-\mathbf{b}$), а затем на векторах $\mathbf{a}$ и $-\mathbf{b}$ построить параллелограмм.

В таком смысле сложение и вычитание векторов вводится путем *математического определения*. Над векторами можно производить и другие операции, вводимые таким же путем, например умножение вектора на скаляр или скалярное и векторное перемножение двух векторов. Все операции такого типа мы называем *математическими*. Их свойства устанавливаются соответствующими математическими теоремами. Не имеет смысла ставить вопрос об опытной

проверке результатов, получаемых с помощью таких математических операций. Например, о сложении векторов, как оно только что определено, мы будем говорить как о *математическом сложении* или *сложении в математическом смысле*. Но когда векторами изображают различные физические величины, часто в их сложение или вычитание вкладывается какой-то другой смысл. А именно для получения суммы или разности векторов над ними надо произвести какие-то (хотя бы мысленные) *физические операции*. Сложение и вычитание в таком смысле мы условимся называть *физическими*. Будет ли какое-либо конкретное физическое сложение совпадать с математическим (т. е. с правилом параллелограмма) и будет ли в результате такого сложения получаться вектор — это требует дополнительного исследования, в частности опытного.

6. Поставим, например, такой вопрос. Точка перешла из положения A в положение B вдоль прямолинейного отрезка $\overrightarrow{AB}$ (рис. 14). Затем из положения B она перешла в C вдоль отрезка $\overrightarrow{BC}$. Вдоль какого прямолинейного отрезка должна перемещаться точка, чтобы из A попасть в C? Ясно, что таким отрезком является отрезок $\overrightarrow{AC}$. Его можно рассматривать как геометрическую сумму отрезков $\overrightarrow{AB}$ и $\overrightarrow{BC}$. Сложение перемещений в таком понимании производится по правилу параллелограмма, т. е. совпадает с математическим сложением векторов. Тому же правилу подчиняется и сложение скоростей в следующем смысле. Точка в течение секунды перешла из A в B, двигаясь равномерно со скоростью $\mathbf{v}_1$. Затем также в течение секунды она перешла из B в C с постоянной скоростью $\mathbf{v}_2$. С какой постоянной скоростью v должна двигаться точка, чтобы в одну секунду перейти из A в C? Но в сложение скоростей обычно вкладывается другой смысл, разъяснимый на следующем примере. Точка перешла из A в B вдоль прямолинейного отрезка на палубе корабля, двигаясь равномерно со скоростью $\mathbf{v}_1$. За то же время сам корабль переместился относительно берега на отрезок BC, двигаясь с постоянной скоростью $\mathbf{v}_2$. С какой скоростью $\mathbf{v}$ двигалась точка относительно берега? Здесь сложение движений и их скоростей понимается в *другом смысле*. Оба движения рассматриваются в *разных* системах отсчета, движущихся одна относительно другой. Одной системой является корабль, и скорость $\mathbf{v}_1$ измеряется с помощью линеек и часов в этой системе. Другой системой является берег, и скорости $\mathbf{v}_2$ и $\mathbf{v}$ измеряются с помощью линеек и часов этой системы. На вопрос о результате сложения в таком смысле должен в конце концов ответить опыт. Дорелятивистская кинема-

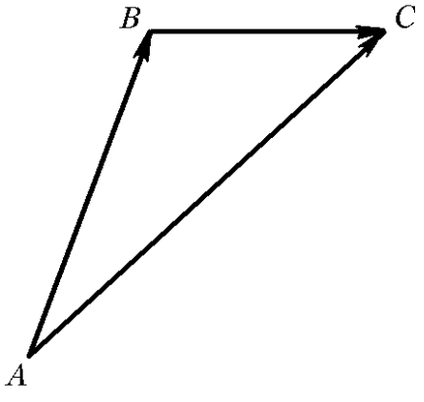


Рис. 14

тика утверждала, что по своему результату сложение движений во втором смысле не может отличаться от сложения в первом смысле. Это происходит потому, что в дорелятивистской физике длины отрезков и промежутков времени не зависят от того, в какой системе отсчета они измеряются. Сложение скоростей и во втором смысле в дорелятивистской кинематике происходило по правилу параллелограмма, т. е. совпадало с математическим сложением векторов. В релятивистской кинематике это уже не так. *Сложение скоростей во втором смысле не подчиняется правилу параллелограмма.* Это правило приближенно верно только в пределе, когда обе складываемые скорости очень малы по сравнению со скоростью света.

7. Каждому вектору $\mathbf{a}(a_x, a_y, a_z)$ и скаляру λ можно аксиоматически сопоставить объект $\lambda\mathbf{a}$, задаваемой упорядоченной тройкой чисел λa_x, λa_y, λa_z. Легко убедиться, что такой объект будет вектором. Он называется *произведением скаляра* λ *на вектор* $\mathbf{a}$. Бесконечно малое приращение вектора $d\mathbf{a}$ само является вектором. Бесконечно малое приращение любого скаляра t есть также скаляр dt. Этим двум величинам можно сопоставить вектор $\frac{1}{dt}\,d\mathbf{a} = \frac{d\mathbf{a}}{dt}$, называемый *производной вектора* $\mathbf{a}$ *по скаляру* t.

8. Теперь мы в состоянии доказать векторную природу многих физических величин, с которыми имеет дело механика. Прежде всего, *смещение точки из какого-либо положения A в другое положение B вдоль соединяющего их прямолинейного отрезка AB есть вектор.* Это очевидно, так как по самому определению при смещении начала и повороте координатных осей компоненты вектора должны преобразовываться так же, как проекции направленного отрезка. Обозначим рассматриваемый отрезок через $\mathbf{r}$. Продифференцируем этот отрезок по времени t в предположении, что начальная точка его закреплена. *Производная* $\frac{d\mathbf{r}}{dt}$ *будет вектором*, так как время — скаляр. Но такая производная есть скорость точки $\mathbf{v}$. Таким образом, *скорость* $\mathbf{v}$ также есть *вектор.* Дифференцируя $\mathbf{v}$ снова по t, найдем *другой вектор — ускорение точки* $\mathbf{a} = \frac{d\mathbf{v}}{dt}$. *Масса точки* m *является скаляром.* Умножая его на скорость $\mathbf{v}$, получаем вектор $\mathbf{p} = m\mathbf{v}$, называемый *импульсом точки.* Дифференцируя его по времени, получаем *силу* $\mathbf{F} = \frac{d\mathbf{p}}{dt}$, действующую на точку. Таким образом, *сила есть вектор.*

9. Приведем несколько более сложные примеры векторов. Возьмем в пространстве какой-либо *ориентированный контур* L, т. е. не самопересекающуюся замкнутую кривую, проходимую в каком-то опредленном направлении. Спроецируем этот контур на координатные плоскости прямоугольной системы координат XYZ. Получим три ориентированных плоских замкнутых контура L_x, L_y, L_z, лежа-

щих в координатных плоскостях YZ, ZX, XY соответственно (на рис. 15 контур L не изображен, изображены только его проекции). Обозначим через S_x, S_y, S_z площади, ограниченные замкнутыми контурами L_x, L_y, L_z. Эти величины будем считать положительными, если контуры L_x, L_y, L_z обходятся в *положительных направлениях*, и отрицательными — в противоположном случае. Положительные направления обхода контуров L_x, L_y, L_z задаются по-разному в зависимости от того, какая используется система координат — правая или левая. В правой системе координат направления обхода контуров L_x, L_y, L_z считаются положительными, если они находятся в *правовинтовом соотношении* с положительными направлениями координатных осей X, Y, Z соответственно, а в левой системе — в *левовинтовом*. Это значит, например, что в правой системе координат вращение ручки буравчика с правой нарезкой в положительном направлении контура L_z приводит к поступательному перемещению буравчика в положительном направлении оси Z. В левой системе будет то же самое, если взять буравчик с левой нарезкой. При таком соглашении о знаках площади S_x, S_y, S_z представляются интегралами

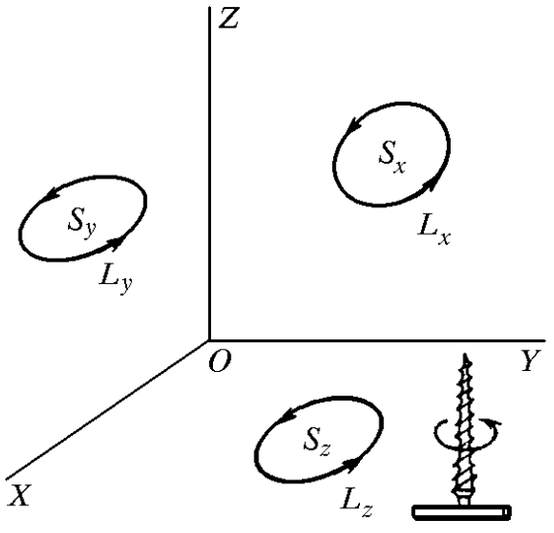


Рис. 15

$$S_x = \int\limits_{L_x} y\,dz, \quad S_y = \int\limits_{L_y} z\,dx, \quad S_z = \int\limits_{L_z} x\,dy, \tag{7.2}$$

взятыми по контурам L_x, L_y, L_z, независимо от того, применяется ли правая или левая система координат.

Мы утверждаем, что тройка чисел S_x, S_y, S_z образует вектор, с одной оговоркой, о которой будет сказано ниже. Для доказательства рассмотрим сначала частный случай, когда контур L плоский. Вдоль нормали к плоскости контура отложим направленный отрезок A, длина которого численно равна площади S, ограниченной контуром L, а направление находится в правовинтовом соотношении с направлением обхода по контуру, если используется правая система координат, и в левовинтовом соотношении, если используется левая система координат (рис. 16). Сначала будем пользоваться системами координат только какого-либо определенного типа: либо только одними правыми, либо только одними левыми. Построенный нами отрезок **A** совершенно не зави-

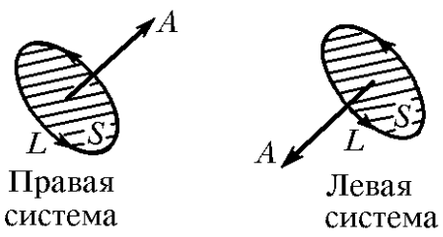


Рис. 16

сит от выбора координатных осей, а потому является вектором. В самом деле, его проекции на координатные оси равны

$$A_x = A\cos(\mathbf{A}, X), \quad A_y = A\cos(\mathbf{A}, Y), \quad A_z = A\cos(\mathbf{A}, Z).$$

С другой стороны, по известной геометрической теореме

$$S_x = S\cos(\mathbf{A}, X), \quad S_y = S\cos(\mathbf{A}, Y), \quad S_z = S\cos(\mathbf{A}, Z).$$

Так как длину $\mathbf{A}$ мы выбрали численно равной S, то в любой системе координат $S_x = A_x$, $S_y = A_y$, $S_z = A_z$. Отсюда следует, что при вращении координатной системы S_x, S_y, S_z преобразуются так же, как компоненты вектора $\mathbf{A}$. Поэтому S_x, S_y, S_z образуют вектор. Его мы будем обозначать буквой $\mathbf{S}$ и называть *вектором площади*, ограниченной ориентированным контуром L. В этом смысле говорят, что площадь является вектором. Это утверждение доказано нами для плоских контуров и плоских площадей.

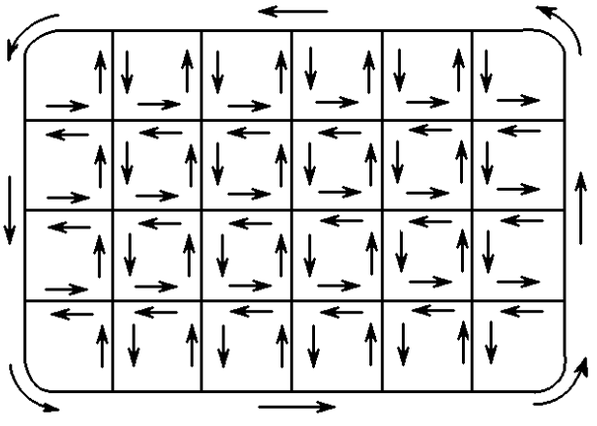


Рис. 17

Обобщение на случай неплоских контуров и площадей не представляет затруднений. Пусть L — такой контур. Натянем на него совершенно произвольную поверхность и разобъем ее на достаточно большое число n малых ориентированных областей, как указано на рис. 17. Проецируя их на координатные плоскости, получим

$$S_x = \sum_{i=1}^{n} S_{ix}, \quad S_y = \sum_{i=1}^{n} S_{iy}, \quad S_z = \sum_{i=1}^{n} S_{iz},$$

где S_{ix}, S_{iy}, S_{iz} — проекции на те же плоскости i-й элементарной области. Число n можно взять сколь угодно большим и рассматривать каждую малую область S_i как плоскую. Тогда на основании доказанного можно утверждать, что S_{ix}, S_{iy}, S_{iz} образуют вектор. Будет образовывать вектор и тройка чисел S_x, S_y, S_z, так как эти числа получаются путем сложения компонентов векторов $\mathbf{S}_i$.

10. В одном отношении, однако, тройка чисел S_x, S_y, S_z отличается от вектора. Эти числа преобразуются так же, как компоненты вектора при *вращении координатной системы как целого*, когда система координат все время остается либо правой, либо левой. Однако они ведут себя существенно иначе при *переходе от правой системы координат к левой или наоборот*, например при инверсии координатных осей. В этом случае для нахождения направления $\mathbf{S}$ надо перейти от одного винта к другому. Если в правой системе координат величину S изобразить стрелкой, то при переходе к левой

направление стрелки надо изменить на противоположное. Величины такого типа называются *псевдовекторами*, или *аксиальными векторами*, в отличие от *полярных векторов*, которые мы рассматривали до сих пор. *При повороте координатной системы как целого аксиальные векторы ведут себя в точности так же, как и полярные векторы. При инверсии координатных осей компоненты полярных векторов заменяют знаки, в то время как компоненты аксиальных векторов остаются неизменными.*

Можно было бы обойтись и без введения аксиальных векторов. Но тогда не все формулы имели бы один и тот же вид в правых и левых координатных системах. Например, если бы в правых системах координат мы определили тройку чисел S_x, S_y, S_z формулами (7.2), а в левых — теми же формулами, но с измененными знаками, то такая тройка чисел образовывала бы полярный вектор. Аксиальные векторы для того и вводятся, чтобы все формулы имели совершенно одинаковый вид в правых и левых системах координат.

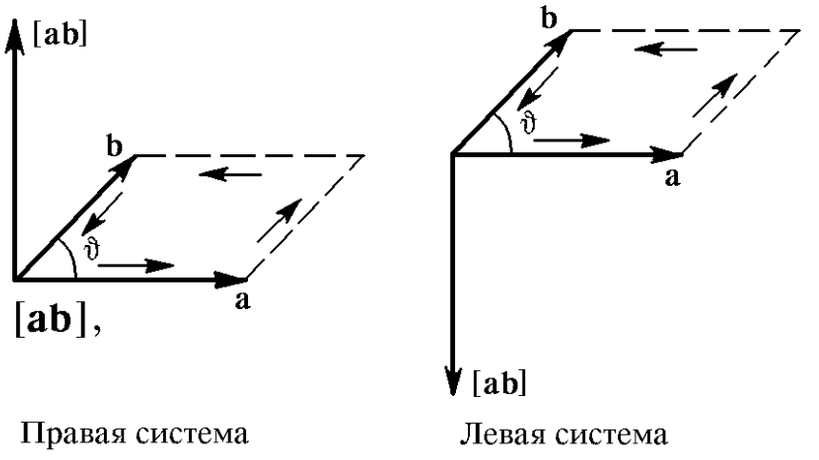
Аналогично наряду с *истинными скалярами* вводятся так называемые *псевдоскаляры. Скаляр, или инвариант*, есть число, остающееся неизменным во всех системах координат, как правых, так и левых. *Псевдоскаляр, или псевдоинвариант, остается неизменным при переходах от правых систем координат к правым же или от левых к левым же. При переходе от правой системы к левой или наоборот псевдоскаляр меняет знак, оставаясь неизменным по абсолютной величине. Произведение псевдоскаляра на полярный вектор есть вектор аксиальный. Произведение псевдоскаляра на аксиальный вектор есть вектор полярный.* Если пользоваться одними только правыми или одними только левыми системами координат (а в физике, как уже упоминалось, применяется почти исключительно правая система), то отпадает необходимость разделения векторов на полярные и аксиальные, а скаляров — на истинные скаляры и псевдоскаляры.

Операция сложения двух векторов имеет смысл только тогда, когда складываемые векторы оба полярные или оба аксиальные. Сумма $\mathbf{a} + \mathbf{b}$ не имеет смысла, если один из векторов полярный, а другой — аксиальный. Сумма такого рода не преобразовывалась бы по правилу преобразования полярного или аксиального вектора, а потому она не могла бы быть ни тем, ни другим.

11. Частным случаем вектора, представляющего площадку или поверхность, является так называемое *векторное произведение* двух векторов $\mathbf{a}$ и $\mathbf{b}$. Оно определяется как вектор площади параллелограмма, построенного на векторах $\mathbf{a}$ и $\mathbf{b}$. Чтобы ориентировать этот параллелограмм, надо обходить его периметр от начала вектора $\mathbf{a}$ к его концу, затем от конца вектора $\mathbf{a}$ параллельно вектору $\mathbf{b}$ и т. д., пока при таком обходе мы не вернемся в исходную точку (рис. 18). Короче говоря, первый вектор $\mathbf{a}$ надо проходить в прямом, а второй вектор $\mathbf{b}$ — в обратном направлениях. В согласии с изложенным выше векторное произведение можно изобразить стрелкой, направленной

перпендикулярно к плоскости параллелограмма и находящейся в нужном винтовом соотношении с направлением обхода периметра параллелограмма.

Длина стрелки численно равна площади параллелограмма, т. е. $ab \sin \vartheta$, где ϑ — угол между векторами **a** и **b**. Векторное произведение мы будем обозначать символом $\mathbf{c} = [\mathbf{ab}]$, т. е. будем заключать векторы **a** и **b** в квадратные скобки. Часто употребляется также косой крест: $\mathbf{c} = \mathbf{a} \times \mathbf{b}$.



Правая система Левая система

Рис. 18

Если векторы **a** *и* **b** *— полярные, то векторное произведение их будет вектором аксиальным. Векторное произведение полярного вектора на аксиальный есть вектор полярный. Векторное произведение двух аксиальных векторов есть также аксиальный вектор.*

ЗАДАЧИ

1. Доказать, что если **a** и **b** — два полярных или два аксиальных вектора, то в прямоугольных системах координат выражение $a_x b_x + a_y b_y + a_z b_z$ есть инвариант. (Это выражение называется *скалярным произведением* векторов **a** и **b** и обозначается символом (**ab**) или **ab**.)

Указание. Воспользоваться инвариантами $a_x^2 + a_y^2 + a_z^2$, $b_x^2 + b_y^2 + b_z^2$ и $(a_x + b_x)^2 + (a_y + b_y)^2 + (a_z + b_z)^2$.

2. Доказать, что скалярное произведение полярного вектора на аксиальный есть псевдоскаляр (псевдовариант).

3. Доказать, что скалярное произведение любых двух векторов **a** и **b** представляется выражением $(\mathbf{ab}) = ab \cos \vartheta$, где ϑ — угол между этими векторами.

Доказательство. Направим ось X вдоль вектора **a**. Тогда $a_y = a_z = 0$, $b_x = b \cos \vartheta$. Так как скалярное произведение $(\mathbf{ab}) \equiv a_x b_x + a_y b_y + a_z b_z$ есть инвариант, то $(\mathbf{ab}) = a_x b_x = ab \cos \vartheta$.

4. Скалярное произведение вектора **a** на векторное произведение других векторов [**bc**] называется смешанным произведением трех векторов **a**, **b**, **c** и обозначается как (**a**[**bc**]). Показать, что оно является псевдоскаляром, если один из этих векторов или все три полярные. Если же полярных векторов два или совсем нет, то смешанное произведение будет скаляром (инвариантом). Показать, что смешанное произведение численно равно объему параллелепипеда, построенного на векторах **a**, **b**, **c**. Пользуясь этим, доказать, что

$$(\mathbf{a}[\mathbf{bc}]) = (\mathbf{b}[\mathbf{ca}]) = (\mathbf{c}[\mathbf{ab}]) = -(\mathbf{a}[\mathbf{cb}]) = -(\mathbf{b}[\mathbf{ac}]) = -(\mathbf{c}[\mathbf{ba}]), \quad (7.3)$$

т. е. *смешанное произведение не меняется при любой циклической перестановке перемножаемых векторов*, а при нарушении цикличности меняет знак.

5. Доказать формулу

$$[\mathbf{a}[\mathbf{bc}]] = (\mathbf{ac})\mathbf{b} - (\mathbf{ab})\mathbf{c}. \tag{7.4}$$

Доказательство. Представим вектор $\mathbf{a}$ в виде $\mathbf{a} = \mathbf{a}_{\parallel} + \mathbf{a}_{\perp}$, где $\mathbf{a}_{\parallel}$ — составляющая вектора $\mathbf{a}$ вдоль вектора $\mathbf{d} \equiv [\mathbf{bc}]$, а $\mathbf{a}_{\perp}$ — составляющая, перпендикулярная к $\mathbf{d}$. Тогда

$$[\mathbf{a}[\mathbf{bc}]] \equiv [\mathbf{ad}] = [\mathbf{a}_{\perp}\mathbf{d}].$$

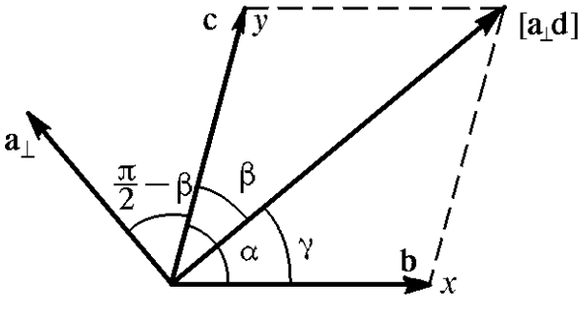


Рис. 19

Три вектора $\mathbf{a}_{\perp}$, $\mathbf{b}$, $\mathbf{c}$ лежат в одной плоскости. Примем ее за плоскость рисунка (рис. 19). Вектор $\mathbf{d}$ перпендикулярен к этой плоскости, его длина равна $bc \sin \alpha$, если α — угол между векторами $\mathbf{b}$ и $\mathbf{c}$. Поэтому длина вектора $[\mathbf{a}_{\perp}\mathbf{d}]$ будет $a_{\perp}bc \sin \alpha$. Поскольку этот вектор лежит в плоскости рисунка, его можно разложить по векторам $\mathbf{b}$ и $\mathbf{c}$, т. е. представить в виде

$$[\mathbf{a}_{\perp}\mathbf{d}] = x\mathbf{b} + y\mathbf{c}.$$

Неизвестные числа x и y найдутся с помощью теоремы синусов:

$$\frac{xb}{a_{\perp}bc \sin \alpha} = \frac{\sin \beta}{\sin \alpha}, \quad \frac{yc}{a_{\perp}bc \sin \alpha} = \frac{\sin \gamma}{\sin \alpha}.$$

Отсюда

$$x = a_{\perp}c \sin \beta = a_{\perp}c \cos (\mathbf{a}_{\perp}, \mathbf{c}) = (\mathbf{a}_{\perp}\mathbf{c}) = (\mathbf{ac}),$$

$$y = a_{\perp}b \sin \gamma = -a_{\perp}b \cos (\mathbf{a}_{\perp}, \mathbf{b}) = -(\mathbf{a}_{\perp}\mathbf{b}) = -(\mathbf{ab}).$$

6. Доказать формулу

$$([\mathbf{ab}]\ [\mathbf{cd}]) = (\mathbf{ac})\ (\mathbf{bd}) - (\mathbf{ad})\ (\mathbf{bc}).$$

7. Показать, что векторное произведение $[\mathbf{ab}]$ можно записать в виде символического определителя

$$[\mathbf{ab}] = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_x & a_y & a_z \\ b_x & b_y & b_z \end{vmatrix}, \tag{7.5}$$

если условиться разлагать его по элементам первой строки, состоящей из единичных векторов $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ вдоль координатных осей прямоугольной системы координат. Запись справедлива и в правых, и в левых системах координат. Компоненты векторного произведения определяются одними и теми же формулами, независимо от того, какие (прямоугольные) системы координат используются. С этим и связано то обстоятельство, что векторное произведение — аксиальный вектор.

8. Доказать, что в прямоугольной системе координат

$$(\mathbf{A}[\mathbf{BC}]) = \begin{vmatrix} A_x & A_y & A_z \\ B_x & B_y & B_z \\ C_x & C_y & C_z \end{vmatrix}. \tag{7.6}$$

9. Пусть $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ — произвольные векторы, не лежащие в одной плоскости. Векторы

$$\mathbf{e}_1^* = \frac{[\mathbf{e}_2\mathbf{e}_3]}{(\mathbf{e}_1[\mathbf{e}_2\mathbf{e}_3])}, \quad \mathbf{e}_2^* = \frac{[\mathbf{e}_3\mathbf{e}_1]}{(\mathbf{e}_1[\mathbf{e}_2\mathbf{e}_3])}, \quad \mathbf{e}_3^* = \frac{[\mathbf{e}_1\mathbf{e}_2]}{(\mathbf{e}_1[\mathbf{e}_2\mathbf{e}_3])} \tag{7.7}$$

называются по отношению к ним *взаимными*. Очевидно, что они также лежат в одной плоскости. Показать, что

$$\mathbf{e}_1 = \frac{[\mathbf{e}_2^*\mathbf{e}_3^*]}{(\mathbf{e}_1^*[\mathbf{e}_2^*\mathbf{e}_3^*])}, \quad \mathbf{e}_2 = \frac{[\mathbf{e}_3^*\mathbf{e}_1^*]}{(\mathbf{e}_1^*[\mathbf{e}_2^*\mathbf{e}_3^*])}, \quad \mathbf{e}_3 = \frac{[\mathbf{e}_1^*\mathbf{e}_2^*]}{(\mathbf{e}_1^*[\mathbf{e}_2^*\mathbf{e}_3^*])}. \tag{7.8}$$

Показать, далее, что

$$(\mathbf{e}_i\mathbf{e}_k^*) = \delta_{ik}, \tag{7.9}$$

где δ_{ik} — символ Кронекера, т. е. $\delta_{ik} = 1$ при $i = k$ и $\delta_{ik} = 0$ при $i \neq k$.

Пусть $\mathbf{A}$ и $\mathbf{B}$ — произвольные векторы. Представим их в виде

$$\mathbf{A} = A_1\mathbf{e}_1 + A_2\mathbf{e}_2 + A_3\mathbf{e}_3, \quad \mathbf{B} = B_1^*\mathbf{e}_1^* + B_2^*\mathbf{e}_2^* + B_3^*\mathbf{e}_3^*.$$

Показать, что

$$(\mathbf{AB}) = A_1B_1^* + A_2B_2^* + A_3B_3^*. \tag{7.10}$$

§ 8. СТЕПЕНИ СВОБОДЫ И ОБОБЩЕННЫЕ КООРДИНАТЫ

1. Положение точки в пространстве можно задать тремя прямоугольными координатами x, y, z. Но это можно сделать и иначе. Например, вместо прямоугольных можно взять *полярные* или какие-либо другие координаты. Существенно, однако, что при любом выборе число независимых координат, требующихся для однозначного определения положения точки, которая может перемещаться в пространстве как угодно, равно *трем*. Про такую точку говорят, что она обладает *тремя степенями свободы*.

Может случиться, что перемещение точки в заданных условиях не может быть каким угодно. Рассмотрим, например, маленький шарик, привязанный к концу нерастяжимой нити, другой коней которой закреплен (математический маятник). Если нить натянута, то шарик может перемещаться только по поверхности сферы с центром в точке закрепления. Можно привести много других примеров, в которых материальная точка все время вынуждена находиться на какой-либо заданной поверхности. В подобных случаях говорят, что на ее движение *наложены связи*. Координаты x, y, z такой точки должны удовлетворять соотношению вида $f(x, y, z) = 0$, которое является уравнением рассматриваемой поверхности. Ввиду этого независимыми остаются только две координаты, например x и y. Третья координата z может быть вычислена из уравнения связей $f(x, y, z) = 0$. В этих случаях говорят, что такая точка обладает *двумя степенями свободы*.

Если точка может перемещаться только вдоль какой-либо заданной кривой, то число независимых координат, требующихся для определения ее положения, снижается до одного. За координату можно принять, например, расстояние материальной точки от какой-либо точки рассматриваемой кривой, отсчитанное вдоль этой кривой. В таких случаях говорят, что точка обладает одной *степенью свободы.*

2. Все сказанное без труда обобщается на случай механической системы, состоящей из произвольного числа n материальных точек. Если эти точки могут перемещаться без всяких ограничений, то для определения мгновенного положения их надо задать $3n$ координат (по три координаты для каждой точки). В этом случае говорят, что система обладает $3n$ *степенями свободы.* В некоторых задачах, однако, свобода перемещения материальных точек ограничена. На $3n$ координат налагаются дополнительные условия, называемые *связями.* Для однозначного определения положения всех материальных точек системы достаточно знать меньшее число координат. Обозначим его через f. Остальные $3n - f$ координат могут быть вычислены из уравнений связи. Не обязательно в качестве независимых координат брать прямоугольные координаты. Для этой цели могут быть использованы любые f величины $q_1, q_2, \ldots, q_f$, заданием которых положение материальных точек системы определяется однозначно. Такие величины называются *обобщенными координатами.* Движение системы определится полностью, если обобщенные координаты будут найдены как функции времени. Производные обобщенных координат по времени $\dot{q}_1, \dot{q}_2, \ldots, \dot{q}_n$ называются *обобщенными скоростями.* Так, при вращении материальной точки по окружности ее положение можно задать значением центрального угла φ, который радиус-вектор вращающейся точки образует с положением его в некоторый момент времени (например, в момент $t = 0$). Обобщенная скорость в этом случае $\omega = \dot{\varphi}$ имеет смысл угловой скорости вращающейся точки.

Обобщенные координаты $q_1, q_2, \ldots, q_f$ могут быть выбраны как угодно, лишь бы они в любой момент времени полностью определяли положение механической системы. Однако число независимых обобщенных координат f во всех случаях будет одно и то же. Оно называется *числом степеней свободы системы.*

3. Определим, например, число степеней свободы *идеально твердого тела. Идеально твердым телом* в механике называют идеализированную систему материальных точек, все расстояния между которыми при движении системы не изменяются с течением времени. Докажем, что *идеально твердое тело, если на его движение не наложены никакие ограничения, обладает шестью степенями свободы.* Действительно, чтобы однозначно определить положение твердого тела, достаточно задать положение каких-либо трех его точек A, B, C, не лежащих на одной прямой (рис. 20). Для доказательства возьмем произвольную четвертую точку тела D. Расстояния AD, BD и CD для

рассматриваемого твердого тела могут считаться известными, так как при любых движениях эти расстояния не изменяются. Кроме того, следует учесть, что при любых движениях твердого тела точка D все время должна находиться по одну и ту же сторону плоскости треугольника ABC, никогда не пересекая ее. Чтобы определить положение в пространстве точки D, построим по заданным длинам AC, AD, CD треугольник ADC. Его основание AC в пространстве фиксировано. Чтобы найти положение вершины D, будем вращать треугольник ADC вокруг основания AC, пока вершина D не окажется на заданном расстоянии от третьей точки B. Этому условию удовлетворяют две точки D и D'. Но вторая из них не удовлетворяет условиям задачи, так как она находится не с той стороны от плоскости треугольника ABC. Таким образом, зная положение трех точек A, B, C, можно геометрическим построением найти положение любой другой точки твердого тела.

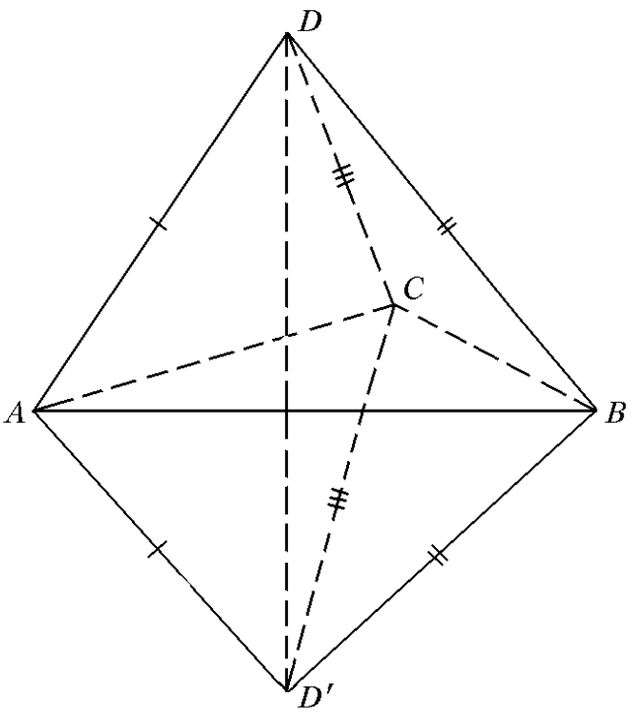


Рис. 20

Положение трех точек A, B, C можно задать их прямоугольными координатами x_A, y_A, z_A; x_B, y_B, z_B; x_C, y_C, z_C. Эти девять координат, однако, не независимы, а связаны тремя соотношениями

$$(x_A - x_B)^2 + (y_A - y_B)^2 + (z_A - z_B)^2 = AB^2 = \text{const},$$

$$(x_B - x_C)^2 + (y_B - y_{C)}^2 + (z_B - z_C)^2 = BC^2 = \text{const},$$

$$(x_C - x_A)^2 + (y_C - y_A)^2 + (z_C - z_A)^2 = CA^2 = \text{const},$$

поскольку длины AB, BC и CA не изменяются. Независимых координат остается только шесть — твердое тело имеет шесть степеней свободы.

При ограничении свободы движения число степеней свободы твердого тела уменьшается. Так, *твердое тело, одна из точек которого неподвижно закреплена, может только вращаться вокруг этой неподвижной точки и имеет три степени свободы. Твердое тело, которое может только вращаться вокруг закрепленной оси, имеет одну степень свободы. Если же твердое тело может скользить вдоль закрепленной оси и одновременно вращаться вокруг нее, то число степеней свободы становится равным двум и т. д.*

ГЛАВА II

ЗАКОНЫ НЬЮТОНА

В этой главе излагаются основные законы *динамики* — той части классической механики, которая занимается изучением движения тел в связи с действующими на них силами. Сила, действующая на тело, является мерой взаимодействия его с окружающими материальными объектами (другими телами, полями). Более полное определение приводится несколько ниже.

Законы динамики были установлены Ньютоном и носят его имя. Как и другие принципы, лежащие в основе физики, они являются *обобщением опытных фактов*. На них следует смотреть не как на изолированные независимые утверждения, а как на *систему взаимосвязанных законов*. Опытной проверке подвергается не каждый закон в отдельности, а вся система в целом.

Ввиду исключительной роли, которую играют законы Ньютона в механике, приведем их в том виде, в каком они были сформулированы самим Ньютоном (перевод акад. А. Н. Крылова (1863—1945)). Формулировке основных законов Ньютон предпосылает восемь определений, из которых для нас здесь важны первые четыре.

Определение 1. *Количество материи (масса) есть мера таковой, устанавливаемая пропорционально плотности и объему ее.*

Определение 2. *Количество движения есть мера такового, устанавливаемая пропорционально скорости и массе.*

Определение 3. *Врожденная сила материи есть присущая ей способность сопротивления, по которой всякое отдельно взятое тело, поскольку оно предоставлено самому себе, удерживает свое состояние покоя или равномерного прямолинейного движения.*

Определение 4. *Приложенная сила есть действие, производимое над телом, чтобы изменить его состояние покоя или равномерного прямолинейного движения.*

Закон 1. *Всякое тело продолжает удерживаться в своем состоянии покоя или равномерного и прямолинейного движения, пока и поскольку оно не понуждается приложенными силами изменить это состояние.*

Закон 2. *Изменение количества движения пропорционально приложенной движущей силе и происходит по направлению той прямой, по которой эта сила действует.*

Закон 3. *Действию всегда есть равное и противоположное противодействие, иначе — воздействие двух тел друг на друга между собою равны и направлены в противоположные стороны.*

Понятие массы у Ньютона отличается неясностью, поскольку им не было дано определения плотности. Кроме того, представление массы как произведения объема тела на плотность содержащегося в нем вещества возможно только для макроскопических тел, но не для элементарных и атомных частиц. Поэтому ньютоново определение массы не удержалось в науке и было заменено другими определениями.

Аристотель (384—322 до н. э.) и его последователи рассматривали силу как причину движения. Они считали, что с прекращением действия силы прекращается и движение тела. Сила необходима для поддержания движения. Установление первого закона Ньютона означало, что такое представление о силе является неправильным, так как для поддержания (равномерного) движения никаких «сил» не требуется. Силу стали рассматривать как *причину изменения количества движения тела.* А так как это изменение вызывается другими телами, то можно дать следующее определение силы. *Сила есть мера интенсивности взаимодействия тел, проявляющаяся в изменении их количества движения.*

Ниже мы подробно разберем содержание законов Ньютона и связанных с ними понятий, хотя и не будем следовать той системе изложения, которая была принята самим Ньютоном.

§ 9. ЗАКОН ИНЕРЦИИ. ИНЕРЦИАЛЬНАЯ СИСТЕМА ОТСЧЕТА

1. *В качестве первого закона движения* Ньютон принял *закон инерции*, высказанный в частной форме еще Галилеем. Согласно этому закону *тело* (*материальная точка*), *не подверженное внешним воздействиям, либо находится в покое, либо движется прямолинейно и равномерно.* Такое тело называется *свободным*, а его движение — *свободным движением* или *движением по инерции.*

Свободных тел, грубо говоря, не существует. Они являются физическими абстракциями. Однако можно поставить тело в такие условия, когда внешние воздействия на него по возможности устранены или практически компенсируют друг друга. Представив себе, что эти воздействия беспредельно уменьшаются, мы и приходим в пределе к представлению о свободном теле и свободном движении.

Здесь, однако, возникает следующая трудность. Как убедиться в том, что тело не подвержено внешним воздействиям? Об этом нельзя судить по отсутствию ускорений. Нужны какие-то другие независимые способы. Иначе закон инерции потерял бы всякое содержание. Вполне удовлетворительного ответа на этот вопрос не существует. В отсутствие внешних воздействий мы убеждаемся по отсутствию растягивающих пружин или веревок, которые тянут тело, по отсутствию тел, которые давят на него, т. д. Но тело может испытывать воздействия не только со стороны тел, с которыми оно

соприкасается. Оно может подвергаться воздействиям также со стороны различного рода *силовых полей*, возбуждаемых другими телами. Поэтому вопрос сводится к тому, как убедиться в том, что воздействиям со стороны силовых полей тело не подвергается.

Все силы, встречающиеся в природе, известные в настоящее время, сводятся к *силам гравитационного притяжения, электромагнитным силам, сильным и слабым взаимодействиям*. Сильные и слабые взаимодействия проявляются в атомных ядрах и в мире элементарных частиц. Они действуют на малых расстояниях: сильные — на расстояниях порядка 10^{-13} см, слабые — на расстояниях порядка 10^{-16} см. В макромире, который только и изучает классическая механика, от сильных и слабых взаимодействий можно отвлечься.

Электромагнитные и гравитационные силы, напротив, являются силами *дальнодействующими*. С расстоянием они убывают медленно. Если это статические силы, то они убывают обратно пропорционально квадрату расстояния. Если же они переменные (электромагнитные волны), то убывание происходит еще медленнее — обратно пропорционально расстоянию. Только благодаря электромагнитным волнам (свет, радиоизлучение, рентгеновское излучение), исходящим от планет, звезд, пульсаров, галактик и пр., мы и знаем о существовании этих небесных объектов. Поэтому нет оснований утверждать, что удаленные источники не возбуждают заметных электромагнитных и гравитационных полей в рассматриваемой нами области пространства. Однако в отсутствие электромагнитных полей всегда можно убедиться, так как они действуют *по-разному* на положительные и отрицательные заряды, из которых состоят тела. Под действием таких полей возникло бы некоторое разделение положительных и отрицательных зарядов, которое можно было бы обнаружить на опыте. Заряженный шарик, помещенный в одну и ту же точку пространства, двигался бы по-разному в зависимости от того, заряжен он положительно или отрицательно. Все имеющиеся факты не противоречат утверждению, что удаленные тела Вселенной не возбуждают сколько-нибудь заметных статических электромагнитных полей в малых областях пространства (порядка размеров Солнечной системы или Галактики).

О гравитационных полях этого нельзя сказать с той же уверенностью. Но если бы такие поля и существовали, то с ними можно было бы не считаться. Дело в том, что всем телам, независимо от их состава, одно и то же гравитационное поле сообщает в точности одинаковое ускорение. Статическое гравитационное поле удаленных тел Вселенной в малых областях пространства можно считать практически однородным. Можно ввести систему отсчета, свободно падающую в таком однородном гравитационном поле. На явлениях, происходящих в такой системе отсчета, наличие этого однородного гравитационного поля никак не сказывается. Здесь все происходит в точности так же, как в кабине космического корабля, свободно дви-

жущегося в космическом пространстве. В такой кабине космонавты не чувствуют наличия поля тяготения (невесомость). Переменные же гравитационные поля (гравитационные волны) слишком слабы. Попытки их экспериментального обнаружения стали предприниматься лишь в последнее время. Однако из-за малости ожидаемых эффектов гравитационные волны еще не обнаружены экспериментально. Ограничимся здесь этими замечаниями, откладывая более подробный разбор вопроса до гл. IX.

2. В кинематике выбор системы отсчета не был существенным. Все системы отсчета *кинематически эквивалентны.* Не так обстоит дело в динамике. Уже закон инерции с особой остротой ставит вопрос о выборе системы отсчета. Одно и то же движение выглядит *поразному* в разных системах отсчета. Если в какой-либо системе отсчета тело движется прямолинейно и равномерно, то в системе отсчета, движущейся относительно первой ускоренно, этого уже не будет. Отсюда следует, что закон инерции не может быть справедлив во всех системах отсчета. Без указания системы отсчета он просто теряет смысл. Классическая механика постулирует, что *существует система отсчета, в которой все свободные тела движутся прямолинейно и равномерно.* Такая система называется *инерциальной системой отсчета.* Содержание закона инерции, в сущности, сводится к утверждению, что существует по крайней мере одна инерциальная система отсчета.

Это утверждение является обобщением громадной совокупности опытных фактов. Точно так же, только опытным путем, можно установить, какие системы отсчета являются инерциальными, а какие — не инерциальными. Допустим, например, что речь идет о движении звезд и других астрономических объектов в доступной нашему наблюдению части Вселенной. Тогда можно утверждать, что система отсчета, в которой Земля принимается неподвижной (такую систему мы будем называть *земной*), не будет инерциальной. Действительно, в такой системе звезды совершают суточные вращения на небесном своде. Так как расстояния до звезд очень велики, то при этом развиваются очень большие центростремительные ускорения, направленные к Земле. Между тем каждая звезда, ввиду ее громадной удаленности от других небесных тел, практически является свободной. Свободное движение звезды в земной системе отсчета совершается по кругу, а не по прямой линии. Оно не подчиняется закону инерции, а потому земная система отсчета не будет инерциальной. Надо испытать на инерциальность другие системы отсчета. Попробуем взять *гелиоцентрическую систему отсчета,* иначе называемую *системой Коперника* (по имени польского астронома Николая Коперника (1473—1543)). Это есть координатная система, начало которой помещено в центре Солнца (точнее, в центре масс Солнечной системы), а координатные оси являются прямыми, направленными на три удаленные звезды и не лежащими в одной плоскости. Материальными объектами, с помощью которых реализуют-

ся эти оси, являются световые лучи, приходящие от звезд в Солнечную систему. Благодаря относительному движению звезд углы между координатными осями в системе Коперника не остаются постоянными, а медленно изменяются с течением времени. Однако ввиду колоссальности расстояний до звезд изменения направлений координатных осей происходят настолько медленно, что, как правило, их можно не принимать во внимание. Система Коперника практически является инерциальной системой, по крайней мере при изучении движений, происходящих в масштабе нашей планетной системы, а также всякой другой системы, размеры которой малы по сравнению с расстоянием до тех трех звезд, которые в системе Коперника выбраны в качестве опорных. Это доказывается опытами, большинство из которых являются косвенными. Некоторые прямые опыты (маятник Фуко и пр.) будут рассмотрены в гл. IX. Эти же опыты доказывают неинерциальность земной системы отсчета.

3. Неинерциальность земной системы отсчета объясняется тем, что Земля вращается вокруг собственной оси и вокруг Солнца, т. е. *движется ускоренно относительно системы Коперника.* Впрочем, оба эти вращения происходят медленно*). Поэтому *по отношению к громадному кругу явлений земная система отсчета ведет себя практически как инерциальная система.* Обычные, сравнительно грубые наблюдения и опыты над движением тел не позволяют обнаружить отступления от инерциальности земной системы отсчета. Для этого требуются более точные и тонкие опыты. Вот почему при установлении основных законов динамики можно начать с изучения движения тел относительно Земли, отвлекаясь от ее вращения, т. е. принять Землю за *приблизительно инерциальную систему отсчета.*

4. Если три звезды, используемые в системе Коперника для фиксирования направлений координатных осей, принадлежат нашей Галактике, то, разумеется, такая система может играть роль инерциальной или, точнее, приблизительно инерциальной системы отсчета только тогда, когда речь идет о движении объектов, малых по сравнению с размерами Галактики, например, о движении Солнечной системы или ее частей. Но при рассмотрении движений всей Галактики или нескольких галактик это будет уже не так. Тогда для построения (приблизительно) инерциальной системы отсчета можно использовать какие-либо другие четыре астрономических объекта, расстояния между которыми весьма велики по сравнению с размерами области пространства, внутри которой совершается движение рассматриваемых тел. Центр одного из этих астрономических объектов можно принять за начало координат, а остальные три объекта использовались для фиксирования направлений координатных осей.

При изучении движения тел мы будем сначала предполагать, что движение отнесено к инерциальной системе отсчета. После этого в

*) В каком смысле следует понимать медленность вращения выяснится в гл. IX.

гл. IX мы изучим, как изменится форма законов движения, когда оно рассматривается относительно неинерциальных систем отсчета.

§ 10. МАССА. ЗАКОН СОХРАНЕНИЯ ИМПУЛЬСА

1. Всякое тело оказывает сопротивление при попытках привести его в движение или изменить модуль или направление его скорости. Это свойство тел называется *инертностью*. У разных тел оно проявляется в разной степени. Так, сообщить одно и то же ускорение большому камню значительно труднее, чем маленькому мячику. *Мера инертности называется массой.*

Для точного количественного определения массы введем понятие *изолированной* или *замкнутой системы.* Так называют систему тел, настолько удаленных от всех остальных тел, что они практически не оказывают никакого действия на рассматриваемую систему. Тела системы могут взаимодействовать только между собой. Рассмотрим теперь изолированную систему, состоящую из двух материальных точек. Скорости точек должны быть малы по сравнению со скоростью света. В результате взаимодействия материальных точек их скорости меняются. Пусть $\mathbf{v}_1$ — скорость точки *1*, $\mathbf{v}_2$ — скорость точки *2*, а $\Delta\mathbf{v}_1$ и $\Delta\mathbf{v}_2$ — приращения этих скоростей за один и тот же промежуток времени Δt. Векторы $\Delta\mathbf{v}_1$ и $\Delta\mathbf{v}_2$ имеют противоположные направления и связаны между собой соотношением

$$m_1\Delta\mathbf{v}_1 = -m_2\Delta\mathbf{v}_2, \tag{10.1}$$

где коэффициенты m_1 и m_2 постоянны и имеют одинаковые знаки. Они совершенно не зависят от характера взаимодействия между материальными точками *1* и *2*. Например, взаимодействие может происходить путем столкновения материальных точек между собой. Его можно осуществить, сообщив материальным точкам электрические заряды или поместив между ними маленькую пружинку и т. д. Продолжительность времени Δt можно менять произвольным образом. Векторы $\Delta\mathbf{v}_1$ и $\Delta\mathbf{v}_2$ при этом будут меняться. Однако коэффициенты m_1 и m_2, точнее, их отношение, останутся одними и теми же. Эти результаты надо рассматривать как *опытные факты*, подтвержденные бесчисленным множеством примеров. Коэффициенты m_1 и m_2 могут зависеть только от *самих материальных точек системы.* Они называются *массами* или, точнее, *инертными массами* материальных точек *1* и *2*.

Таким образом, по определению *отношение масс двух материальных точек равно взятому с противоположным знаком отношению приращений скоростей этих точек в результате взаимодействия между ними.* При этом предполагается, что рассматриваемые

точки образуют изолированную систему и движутся с нерелятивистскими скоростями.

2. Чтобы от отношения масс перейти к самим массам, надо условиться массу какого-либо определенного тела считать равной единице. Такое тело называется *эталоном массы*. Тогда массы всех остальных тел определятся однозначно. В частности, все они окажутся *положительными*, так как знаки всех масс одинаковы, а масса эталонного тела положительна. В физике в качестве основной единицы массы принят *килограмм*. Килограмм есть масса эталонной гири из сплава иридия с платиной, хранящейся в Севре (Франция) в Международном бюро мер и весов. Приближенно килограмм равен массе кубического дециметра чистой воды при температуре 4 °С. Тысячная доля килограмма называется *граммом*. В отличие от длины и времени, для которых установлены *естественные единицы*, единица массы определена, таким образом, как масса некоторого случайно выбранного тела. И для массы было бы лучше установить естественную единицу. Можно было бы основным эталоном массы считать массу какой-либо элементарной частицы, например протона.

Отметим еще одно существенное обстоятельство, являющееся также результатом опыта. Отношение m_2/m_1 можно найти не только путем непосредственного сравнения масс рассматриваемых тел, но и следующим косвенным способом. Сначала измеряются отношения масс обоих тел к массе третьего тела, а затем эти отношения делятся одно на другое. Результат не зависит от массы третьего тела и совпадает с отношением m_2/m_1, полученным непосредственным сравнением масс m_1 и m_2.

Если соотношение (10.1) поделить на время взаимодействия Δt, то получится

$$m_1\mathbf{a}_{1\text{ср}} = -m_2\mathbf{a}_{2\text{ср}}, \tag{10.2}$$

а после перехода к пределу

$$m_1\mathbf{a}_1 = -m_2\mathbf{a}_2. \tag{10.3}$$

Этими соотношениями нахождение отношения масс двух тел сводится к сравнению *средних* или *истинных ускорений*, развивающихся во время их взаимодействия.

3. Придадим соотношению (10.1) другую форму. Пусть $\mathbf{v}_1$ и $\mathbf{v}_2$ — скорости тел до взаимодействия, $\mathbf{v}_1'$ и $\mathbf{v}_2'$ — после взаимодействия. Тогда $\Delta\mathbf{v}_1 = \mathbf{v}_1' - \mathbf{v}_1$, $\Delta\mathbf{v}_2 = \mathbf{v}_2' - \mathbf{v}_2$. Подставляя эти выражения в (10.1), получим

$$m_1\mathbf{v}_1 + m_2\mathbf{v}_2 = m_1\mathbf{v}_1' + m_2\mathbf{v}_2'. \tag{10.4}$$

Назовем *импульсом* или *количеством движения материальной точки* вектор, равный произведению массы точки на ее скорость:

$$\mathbf{p} = m\mathbf{v}. \tag{10.5}$$

Импульсом или *количеством движения системы материальных точек* назовем векторную сумму импульсов отдельных материальных точек, из которых эта система состоит. Для системы из двух материальных точек $\mathbf{p} = \mathbf{p}_1 + \mathbf{p}_2 = m_1\mathbf{v}_1 + m_2\mathbf{v}_2$. Равенству (10.4) можно придать вид

$$\mathbf{p} = \mathbf{p}', \tag{10.6}$$

где $\mathbf{p} = \mathbf{p}_1 + \mathbf{p}_2$, $\mathbf{p}' = \mathbf{p}_1' + \mathbf{p}_2'$ — импульсы системы до и после взаимодействия. Таким образом, *импульс изолированной системы двух материальных точек сохраняется, т. е. остается постоянным во времени, каково бы ни было взаимодействие между ними.* Это положение называется *законом сохранения импульса.* Оно является результатом опыта и введенного выше определения массы. То обстоятельство, что для величины $m\mathbf{v}$ имеет место «закон сохранения», и делает целесообразным дать этой величине специальное название и ввести для нее особое обозначение. Таким свойством не обладает, например, величина $m^2\mathbf{v}$, а потому она не играет никакой роли в механике. В дальнейшем закон сохранения импульса будет распространен на изолированные системы, состоящие из какого угодно числа материальных точек.

4. Закон сохранения импульса в приведенной выше форме есть закон нерелятивистской механики. Он справедлив только для медленных движений. В релятивистской механике этот закон обобщается на случай быстрых движений. Это обобщение будет подробно рассмотрено при изложении теории относительности. Сейчас же ограничимся предварительным сообщением основного результата. В релятивистской механике импульс частицы также определяется выражением (10.5), однако масса m зависит от скорости согласно формуле

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}}. \tag{10.7}$$

Здесь m_0 — постоянная для данной частицы величина, называемая ее *массой покоя.* Она совпадает с массой, рассматриваемой в нерелятивистской механике. Величина m, определяемая выражением (10.7), называется *массой движения* или *релятивистской массой.* Таким образом, в релятивистской механике закон сохранения импульса изолированной системы, состоящей их двух взаимодействующих частиц с массами покоя m_{01} и m_{02}, математически формулируется следующим образом:

$$\frac{m_{01}v_1}{\sqrt{1 - v_1^2/c^2}} + \frac{m_{02}v_2}{\sqrt{1 - v_2^2/c^2}} = \frac{m_{01}v_1'}{\sqrt{1 - v_1'^2/c^2}} + \frac{m_{02}v_2'}{\sqrt{1 - v_2'^2/c^2}}. \tag{10.8}$$

Следует, однако, сделать оговорку, что эта формула относится к тому случаю, когда весь импульс сосредоточен только в *движущихся частицах.* Это было бы так, если бы частицы не взаимодействовали,

а также в случае, когда взаимодействие между ними распространялось бы мгновенно. Но теория относительности отвергает мгновенные взаимодействия. Все взаимодействия осуществляются *полями* и распространяются с конечными скоростями, не превышающими скорость света в вакууме c. Поля же — это материальные объекты, обладающие импульсом и энергией. Поэтому в общем случае формулу (10.8) следует обобщить, включив в нее импульсы полей. Только в частной случае, когда изменениями импульсов полей можно пренебречь, можно пользоваться законом сохранения импульса в приближенной форме (10.8).

Для медленных движений, когда $v^2/c^2 \ll 1$, зависимостью массы от скорости можно пренебречь, полагая $m = m_0$. Тогда релятивистская механика переходит в нерелятивистскую как в свой предельный приближенный случай. Чтобы составить представление о погрешности, которая возникает при таком пренебрежении, рассмотрим космический корабль, движущийся со скоростью $v = 8$ км/с. В этом случае $(v/c)^2 = (8/300\,000)^2 \approx 7 \cdot 10^{-10}$. Если масса космического корабля $m = 5$ т $= 5 \cdot 10^6$ г, то релятивистская масса m будет превышать массу покоя всего на $m - m_0 = 3{,}5 \cdot 10^{-3}$ г. При всех расчетах движения космического корабля такой поправкой не только можно, но и нужно пренебречь, хотя бы потому, что входные данные, необходимые для расчетов, не могут быть измерены с такой высокой точностью.

§ 11. ВТОРОЙ ЗАКОН НЬЮТОНА. СИЛА

1. Общие законы движения Ньютон сформулировал с использованием понятия *силы*. Представление о силе возникает в связи с ощущением мускульного усилия, которое у нас возникает при поднятии тяжелых тел или при приведении их в движение. Но всякое ощущение слишком неопределенно и субъективно. Оно может дать только интуитивное представление о силе, но не может служить основанием для точного определения этой величины, равно как и для всякого другого научного понятия. *Под силой в механике понимают всякую причину, изменяющую импульс движущегося тела.* Это определение также недостаточно, так как оно чисто качественное. Количественное определение можно дать на основе следующего факта, являющегося обобщением опытных фактов.

В инерциальной системе отсчета производная импульса $\mathbf{p}$ материальной точки по времени представляется уравнением

$$\dot{\mathbf{p}} = \frac{d\mathbf{p}}{dt} \equiv \frac{d}{dt}(m\mathbf{v}) = \mathbf{F}, \tag{11.1}$$

или для медленных движений (когда массу m можно считать не зависящей от скорости $\mathbf{v}$)

$$m\dot{\mathbf{v}} = \mathbf{F}, \tag{11.2}$$

$$m\ddot{\mathbf{r}} = \mathbf{F}. \tag{11.3}$$

Здесь $\mathbf{F}$ однозначно определяется свойствами рассматриваемой материальной точки и окружающих ее тел, а также положениями и скоростями этих тел относительно материальной точки. Величина $\mathbf{F}$ называется силой, действующей на рассматриваемую материальную точку. В частных случаях сила $\mathbf{F}$ может определяться одним только положением материальной точки или только одной ее скоростью, но по самому смыслу в механике Ньютона она не может явно зависеть от ускорения этой точки. Очевидно, *сила* $\mathbf{F}$ *есть вектор*, поскольку она равна производной вектора $\mathbf{p}$ по времени. Отсюда следует, что сложение сил подчиняется правилу параллелограмма.

Таким образом, в *инерциальной системе отсчета производная импульса материальной точки по времени равна действующей на нее силе.* Это утверждение называется *вторым законом Ньютона*, а соответствующие ему уравнения — *уравнениями движения материальной точки.* В случае системы материальных точек уравнения движения должны быть написаны для каждой из них и учтено их взаимодействие между собой. Следует, однако, иметь в виду, что второй закон Ньютона только тогда имеет смысл закона, а не определения понятия силы, когда в нем сила $\mathbf{F}$ определена не по сообщаемому ей действию (т. е. не по ускорению или изменению импульса тела), а другим *независимым способом* (именно, по расположению и движению окружающих тел).

Конечно, если движение тела (материальной точки) известно, т. е. известны ее координаты как функции времени t, то простым дифференцированием по t можно определить величину действующей силы как функцию того же времени. Однако природа силы может считаться установленной только тогда, когда будет выяснено, как эта сила зависит от состояния (положения и скоростей) окружающих тел.

2. Приведем простейший пример, ограничиваясь ради простоты одномерным движением материальной точки.

Подвесим тело на спиральной пружине (рис. 21). Когда система успокоится, немного оттянем тело вниз из положения равновесия, а затем отпустим. Возникнут колебания вверх и вниз. При подходящих параметрах системы они будут затухать слабо. Тело успеет совершить несколько десятков колебаний, прежде чем колебания заметно затухнут. Мгновенное положение тела можно характеризовать одной координатой x — смещением тела из положения равновесия. Для определения функции $x = x(t)$ можно через малые промежутки вре-

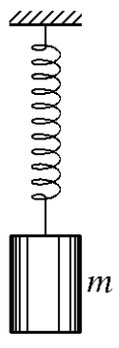


Рис. 21

мени фотографировать тело на кинопленку, а затем обработать фотографию и построить график $x = x(t)$. Можно поступить и как-нибудь иначе. Для слабо затухающих колебаний график почти не отличается от синусоиды (рис. 22) и представляется уравнением

$$x = A \cos \frac{2\pi t}{T}, \tag{11.4}$$

где A и T — постоянные, называемые *амплитудой* и *периодом колебаний*. Дважды дифференцируя это выражение, находим скорость и ускорение:

$$\dot{x} = -\frac{2\pi A}{T} \sin \frac{2\pi t}{T}, \quad \ddot{x} = -\left(\frac{2\pi}{T}\right)^2 A \cos \frac{2\pi t}{T}.$$

Сравнивая, последнее выражение с (11.4), получаем

$$\ddot{x} = -\left(\frac{2\pi}{T}\right)^2 x,$$

или после умножения на массу тела

$$m\ddot{x} = -kx, \tag{11.5}$$

где введено обозначение

$$k = \left(\frac{2\pi}{T}\right)^2 m. \tag{11.6}$$

Сравнивая (11.5) с (11.3), находим силу

$$F = -kx. \tag{11.7}$$

Мы видим, что величина F зависит только от удлинения пружины x — единственного переменного параметра, определяющего положения внешних тел, оказывающих действие на рассматриваемое тело. Если к пружине подвесить тело другой массы, то изменится и период колебаний T. Однако опыт показывает, что отношение m/T^2, а с ним и коэффициент k остаются без изменения. Значит, сила F определяется только растяжением пружины и совершенно не зависит от того, каким телом это растяжение вызвано. Эти опытные факты могут служить подтверждением второго закона Ньютона. Следовательно, можно ожидать, что если, кроме пружины, на тело больше ничто не действует, то его ускорение всегда будет равно kx/m и направлено вдоль оси пружины в сторону, противоположную ее удлинению x. Оно совершенно не зависит от того, как движется тело: прямолинейно, по кругу или как-нибудь иначе. Это предположение также подтверждается опытами.

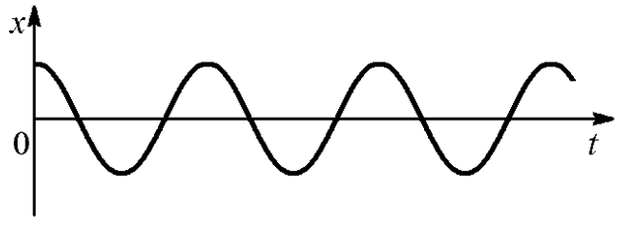


Рис. 22

Одновременно мы видим, что *сила натяжения пружины F пропорциональна ее удлинению x*. Как показали более точные исследования, этот результат является приближенным. Им можно пользоваться, когда удлинение пружины не очень велико. Он называется *законом Гука* (по имени английского физика Роберта Гука (1635—1703)). Величина k называется жесткостью пружины (коэффициентом жесткости). Для конкретной пружины жесткость постоянна, но может меняться от пружины к пружине.

Необходимо сделать два замечания относительно приведенного примера. Во-первых, силу F можно было бы найти исключением времени из двух уравнений (11.4). Таким путем величина F была бы выражена как функция координаты x и скорости $\dot{x}$. Уравнение (11.2), конечно, было бы удовлетворено, но так найденное выражение для F содержало бы в качестве параметра также амплитуду A, и чтобы уравнение (11.2) удовлетворялось, амплитуда A должна иметь уже определенное значение. При той же пружине и массе m, но другой амплитуде A уравнение (11.2) не удовлетворялось бы. Это значит, что при заданных k и m и найденном значении F уравнение (11.2) описывало бы только *вполне определенный частный вид движения*. Между тем требуется получить такое выражение для F, которое *не содержало бы параметра A*. Тогда уравнение (11.2) будет описывать *все движения, возможные при заданных k и m* (т. е. движения с произвольной амплитудой A).

Во-вторых, каждой силе F соответствует определенное удлинение пружины. Получается проградуированная пружина, которую можно использовать для измерения сил в различных случаях. Описанный способ градуировки, основанный на втором законе Ньютона, принципиально возможен, но практически для градуировки пружины к ней подвешивают грузы различного веса и измеряют соответствующие удлинения. Нет необходимости описывать, как этим способом градуируются различные *динамометры* и как ими пользуются. Важно подчеркнуть, что при наличии динамометра можно независимо измерить силу и ускорение и проверить справедливость второго закона Ньютона.

Опыт показывает, что колебания тела, подвешенного на пружине, постепенно затухают и в конце концов прекращаются. Отсюда следует, что уравнение движения (11.5) является приближенным. Оказывается, что тело, движущееся в газообразной или жидкой среде, встречает сопротивление, зависящее от скорости тела. Если скорость тела (относительно окружающей среды) не очень велика, то эта сила приблизительно пропорциональна первой степени скорости. Так, в случае шара на пружине затухание его колебаний в газе довольно точно описывается уравнением

$$m\ddot{x} = -kx - b\dot{x}, \tag{11.8}$$

где b — постоянный коэффициент, зависящей от размеров шара и рода газа, в котором он колеблется. Здесь мы имеем пример силы, которая зависит не только от положения, но и от скорости шара.

3. Для решения задач на движение материальных точек и их систем нужны *дифференциальные уравнения движения.* Способ получения таких уравнений не имеет значения. В частности, их можно было бы получать и строить всю механику без введения понятия силы.

При рассмотрении различных динамических задач механика ставит и решает два вопроса: 1) по заданному движению тел вычислить силы, действующие на них; 2) по заданным силам определить движение тел. Задачи первого типа сравнительно просты. Они сводятся к вычислению ускорений материальных точек, из которых состоит система. Примером таких задач может служить разобранная нами задача о силе, действующей на колеблющееся тело, подвешенное на пружине. Задачи второго типа много сложнее и являются основными в механике. Здесь прежде всего надо написать уравнение движения для каждой материальной точки, входящей в систему. Это сводится к отысканию сил как функций координат и скоростей взаимодействующих точек. В результате получится система дифференциальных уравнений, решение которой (при определенных начальных условиях) даст полное представление о всех деталях движения. Таким образом, при решении таких задач требуется *интегрирование дифференциальных уравнений*, а это значительно сложнее дифференцирования.

Могут быть и задачи смешанного типа. Сюда относятся, например, такие задачи, когда на движение системы наложены определенные ограничения, например движущаяся точка должна находиться на какой-то линии или поверхности. Такого рода ограничения называются *связями.* Действие таких линий или поверхностей, как и всяких связей, ограничивающих свободу движения, сводится к тому, что они воздействуют на движущиеся тела с определенными силами, называемыми *реакциями связей.* Во всех подобных случаях задача сводится не только к определению движения каждой материальной точки системы, но и к нахождению реакций связей.

4. Остановимся на вопросе о соотношении между первым и вторым законами Ньютона. Если в уравнении (11.1) положить $\mathbf{F} = 0$, то получится $\frac{d\mathbf{p}}{dt} = 0$. Отсюда следует, что $\mathbf{p} = \text{const}$, т. е. импульс, а с ним и скорость свободно движущейся материальной точки постоянны. Таким образом, формально первый закон Ньютона является следствием второго. Почему же тогда он выделяется в самостоятельный закон? Дело в том, что уравнение (11.1), выражающее второй закон Ньютона, только тогда имеет смысл, когда указана система отсчета, в которой оно справедливо. Выделить же такую систему (или такие системы) отсчета позволяет первый закон. Он утверждает, что существует система отсчета, в которой свободная мате-

риальная точка движется без ускорения. В такой системе отсчета (и в этом состоит второй закон) движение всякой материальной точки подчиняется уравнению (11.1). Таким образом, по существу, первый закон нельзя рассматривать как простое логическое следствие второго. Связь между этими законами более глубокая.

5. Уравнение (11.2) предопределяет выбор единицы силы. Поскольку единицы длины, массы и времени уже установлены, это уравнение вынуждает нас за единицу силы принять такую силу, которая единице массы сообщает ускорение, равное единице. В 1960 г. XI Генеральная конференция по мерам и весам приняла так называемую *Международную систему единиц* SI (в русской транскрипции СИ). В основу этой системы положены семь *независимых* единиц: единица длины *метр* (м), единица времени *секунда* (с), единица массы *килограмм* (кг), единица разности температур *кельвин* (К), единица силы тока *ампер* (А), единица силы света *кандела* (кд) и единица количества вещества *моль* (моль). Остальные единицы являются их *производными*. Смысл термина «производная единица» легко уясняется на примере единицы силы. В системе СИ за единицу силы принимается ньютон (Н). Ньютон — эта такая сила, которая массе в 1 кг сообщает ускорение в 1м/с^2. Наряду с системой СИ в физике широко употребляется система СГС. Основными единицами в этой системе являются: *сантиметр* (см) — единица длины, *секунда* (с) — единица времени, *грамм* (г) — единица массы. Единицей силы в системе СГС является *дина* (дин). Дина есть сила, сообщающая массе в 1 г ускорение 1 см/с^2. Очевидно,

$$1\ \text{Н} = 10^5\ \text{дин}.$$

В механике обе системы одинаково удобны. Ни одна из них не обладает преимуществом по сравнению с другой, так как между ними нет разницы по существу. Обе системы в механике отличаются друг от друга только *масштабами* основных единиц — единицы длины и единицы массы. Все понятия механики имеют один и тот же смысл, а формулы пишутся совершенно одинаково в обеих системах единиц. Не так обстоит дело в учении об электрических, оптических и атомных явлениях. Для изучения таких явлений система СГС значительно лучше приспособлена, чем система СИ. Поэтому в нашем курсе отдается предпочтение системе СГС.

6. В заключение этого параграфа остановимся на вопросе о *сложении сил*. Как уже было сказано выше, *сила является вектором*. Этим мы хотим сказать только то, что при повороте координатных осей составляющие силы преобразуются как составляющие вектора. Как и для всякого вектора, для силы можно ввести операцию сложения в математическом смысле (см. § 7). По определению каждым двум силам $\mathbf{F}_1$ и $\mathbf{F}_2$ приводится в соответствие новый объект, изображающийся диагональю параллелограмма, построенного на векторах $\mathbf{F}_1$ и $\mathbf{F}_2$. Этот объект, как легко доказать, является вектором. Он

называется *равнодействующей* или *результирующей* сил $\mathbf{F}_1$ и $\mathbf{F}_2$ или их *геометрической суммой*. Проверять на опыте результат такого сложения имеет столько же смысла, что и проверять на опыте правильность арифметического равенства $2 + 3 = 5$. Результат верен *по самому определению* сложения векторов. Однако сложение сил понимают иногда и в другом (*физическом*) смысле. И именно о нем идет речь, когда в элементарной физике впервые говорят о сложении сил. При этом самый вопрос формулируется недостаточно ясно. Говорят, что на материальную точку одновременно действуют две силы $\mathbf{F}_1$ и $\mathbf{F}_2$. После этого спрашивают, какой одной силой $\mathbf{F}$ их можно заменить, чтобы получить тот же результат? Неясность заключается в том, что не указывается, в каком смысле следует понимать выражение: «На материальную точку одновременно действуют две силы». На всякую материальную точку в данных конкретных условиях действует всегда только одна сила, величина и направление которой определяется расположением этой точки относительно всех окружающих тел. Какой же смысл вкладывается в содержание поставленного вопроса? Разъясним это на двух примерах.

Допустим, что к некоторой материальной точке A прикреплена растянутая пружина, которая тянет ее с некоторой силой $\mathbf{F}_1$. Уберем эту пружину и будем тянуть ту же материальную точку A другой растянутой пружиной с силой $\mathbf{F}_2$. О направлении и модуле сил $\mathbf{F}_1$ и $\mathbf{F}_2$ мы судим по направлениям осей пружин и степени их растяжения. Прикрепим теперь к материальной точке A обе пружины вместе, направив и растянув их по-прежнему. Вопрос заключается в том, чтобы определить силу $\mathbf{F}$, которая вызывает такое же движение точки, что и обе силы $\mathbf{F}_1$ и $\mathbf{F}_2$ вместе.

В качестве второго примера рассмотрим неподвижный точечный заряд q, помещенный в некоторой точке пространства A. Пусть в точках B и C находятся другие точечные заряды, q_1 и q_2. Пусть они вместе действуют на заряд q с силой $\mathbf{F}$. Уберем второй из них и обозначим через $\mathbf{F}_1$ силу, с которой на q будет действовать заряд q_1. Аналогично определяется сила $\mathbf{F}_2$, с которой заряд q_2 действует на q в отсутствие заряда q_1. Вопрос опять заключается в том, как по силам $\mathbf{F}_1$ и $\mathbf{F}_2$ найти силу $\mathbf{F}$.

Пусть $\mathbf{F}_i$ означает силу, действующую на рассматриваемую материальную точку со стороны какого-то другого i-го тела (источника силы $\mathbf{F}_i$), когда все остальные источники сил удалены ($i = 1, 2, \ldots, n$). Чему будет равна действующая сила $\mathbf{F}$, когда все n источников действуют одновременно? Это *физический вопрос,* на который нельзя дать ответ путем определения. Обычно говорят, что сила $\mathbf{F}$ равна геометрической сумме сил $\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_n$. Однако такой ответ не является логическим следствием законов Ньютона или каких-либо других законов. Он может быть верным, но может быть и

неверным. Это может решить только опыт. Опыт показывает, например, что для растянутых пружин или электрических сил, возбуждаемых точечными зарядами, ответ верен. Если это имеет место, то говорят, что силы $\mathbf{F}_1$, $\mathbf{F}_2$, ... подчиняются *принципу суперпозиции.* В основе принципа суперпозиции лежит представление о *независимости действия сил.* Говорят, что силы действуют *независимо*, если каждая сила $\mathbf{F}_i$ сообщает рассматриваемому телу одно и то же ускорение $\mathbf{a}_i$, независимо от того, действует только один i-й источник сил или все n источников одновременно. Так как ускорение является вектором, то результирующее ускорение найдется векторным сложением всех $\mathbf{a}_i$. Поэтому и результирующая сила $\mathbf{F} = m\mathbf{a}$ также найдется векторным сложением независимо действующих сил $\mathbf{F}_i = m\mathbf{a}_i$. Следовательно, применимость правила параллелограмма для сложения сил в рассматриваемом физическом смысле эквивалентна предположению о независимости действия сил. Но когда тела, являющиеся источниками сил, влияют друг на друга и вследствие этого меняют свое состояние, то результат вычисления силы $\mathbf{F}$ по указанной схеме может оказаться неверным. Это получится, например, когда во втором примере вместо точечных зарядов q_1 и q_2 взять протяженные тела, заряженные электричеством. При сближении таких тел распределение электричества на них изменится из-за индукции, а это отразится на значении действующей силы. Но и в этом случае можно воспользоваться принципом суперпозиции, если заряды на телах в их окончательных положениях мысленно разделить на достаточно малые части. Считая такие части точечными зарядами, можно вычислить создаваемые ими электрические поля по *закону Кулона,* а затем воспользоваться принципом суперпозиции. Такое утверждение следует рассматривать как обобщение опытных фактов.

ЗАДАЧИ

1. Лифт движется с ускорением $a = \alpha g$, причем $|\alpha| < 1$. Зная вес покоящегося лифта P (вместе с нагрузкой), определить во время ускоренного движения натяжение троса T, на котором он подвешен.

Ответ. $T = P(1 - \alpha)$. Дробь α следует считать положительной, когда ускорение a направлено вниз, и отрицательной, когда оно направлено вверх.

2. К пружине прикреплено тело, которое может смещаться вдоль определенной прямой (например, вдоль стержня, на который оно надето). Эта система может служить *акселерометром,* т. е. прибором для измерения ускорения тела, на котором такой прибор установлен (автомобиля, самолета, поезда и пр.). Опишите принцип действия такого акселерометра.

3. Самолет совершает вираж, двигаясь по окружности с постоянной скоростью v на одной и той же высоте. Определить радиус r этой окружности, если плоскость крыла самолета наклонена к горизонтальной плоскости под постоянным углом α.

Ответ. $r = \dfrac{v^2}{g \operatorname{tg} \alpha}$.

У к а з а н и е. Когда самолет летит прямолинейно, плоскость крыла горизонтальна. Подъемная сила в этом случае направлена вертикально вверх, т. е. перпендикулярна к плоскости крыла. При повороте корпуса самолета вокруг продольной оси подъемная сила поворачивается на тот же угол, т. е. продолжает оставаться перпендикулярной к плоскости крыла, так как силы взаимодействия самолета с окружающей средой зависят лишь от относительного движения самолета и среды.

§ 12. ТРЕТИЙ ЗАКОН НЬЮТОНА И ЗАКОН СОХРАНЕНИЯ ИМПУЛЬСА

1. Рассмотрим замкнутую систему, состоящую из двух взаимодействующих материальных точек. В этом случае справедлив закон сохранения импульса

$$\mathbf{p}_1 + \mathbf{p}_2 = \text{const}.$$

Дифференцируя это соотношение по времени, получим

$$\dot{\mathbf{p}}_1 + \dot{\mathbf{p}}_2 = 0,$$

или, на основании второго закона Ньютона (11.1),

$$\mathbf{F}_1 = -\mathbf{F}_2, \tag{12.1}$$

где $\mathbf{F}_1$ и $\mathbf{F}_2$ — силы, с которыми рассматриваемые материальные точки действуют друг на друга. Привлечем сюда опытный факт, согласно которому силы $\mathbf{F}_1$ и $\mathbf{F}_2$ направлены *вдоль прямой,* соединяющей взаимодействующие точки. Тогда мы придем к *третьему закону Ньютона*:

Силы взаимодействия двух материальных точек равны по модулю, противоположно направлены и действуют вдоль прямой, соединяющей эти материальные точки.

Одну из сил, $\mathbf{F}_1$ или $\mathbf{F}_2$, согласно Ньютону, иногда называют *действием,* а другую — *противодействием,* и формулируют третий закон следующим образом. *Всякому действию соответствует равное и противоположно направленное противодействие.* Следует, однако, заметить, что «действие» по своей физической природе ничем не отличается от «противодействия». Если действующая сила обусловлена деформацией, всемирным тяготением или наличием электрического поля, то и противодействующая сила обусловлена тем же самым. Так, тяжелое тело, лежащее на столе, давит на стол, испытывая со стороны стола противоположно направленное противодавление. Действие — давление камня на стол — обусловлено деформацией камня, противодействие — давление стола на камень — обусловлено деформацией стола. В основе подразделения сил на «действующие» и «противодействующие» лежит представление об *активных телах*, производящих действие, и *пассивных телах,* оказывающих противодействие.

Так, если лошадь тянет телегу, то активным телом, производящим действие, будет лошадь, а пассивным телом, оказывающим противодействие, — телега. Однако подразделение тел на активные и пассивные можно провести далеко не всегда. Например, когда Солнце и планета притягиваются друг к другу силами всемирного тяготения, то в этом взаимодействии они выступают совершенно равноправно, и нельзя указать, какое из этих взаимодействующих тел является активным, а какое пассивным. Какую из сил $\mathbf{F}_1$ и $\mathbf{F}_2$ назвать действием и какую противодействием — это в большинстве случаев вопрос соглашения.

2. Третий закон Ньютона мы сформулировали для замкнутой системы, состоящей из двух взаимодействующих материальных точек. Постулируем теперь его справедливость для системы из произвольного числа материальных точек. Мы исходим из представления, что и в этом случае взаимодействие сводится к силам *попарного взаимодействия* между материальными точками. Пусть $\mathbf{F}_{ik}$ — сила, с которой i-я материальная точка действует на k-ю, а $\mathbf{F}_{ki}$ — сила, с которой k-я точка действует на i-ю. Третий закон утверждает, что обе эти силы направлены вдоль прямой, соединяющей взаимодействующие точки, причем $\mathbf{F}_{ik} = -\mathbf{F}_{ki}$. В таком понимании третий закон Ньютона позволяет выполнить переход от *механики отдельной материальной точки к механике системы материальных точек.* В частности, он позволяет распространить закон сохранения импульса на случай системы произвольного числа n взаимодействующих материальных точек. Рассмотрим этот вопрос, а также другие связанные с ним важные вопросы.

Силы, действующие на материальные точки системы, можно разделить на *внутренние* и *внешние.* Внутренние силы — это силы взаимодействия между материальными точками самой системы. Выше мы обозначили их символами $\mathbf{F}_{ik}$ с двумя индексами i и k, которые указывают, какие точки взаимодействуют. Внешние силы — это такие силы, с которыми на материальные точки системы действуют внешние тела. Согласно третьему закону Ньютона $\mathbf{F}_{ik} = -\mathbf{F}_{ki}$, т. е. $\mathbf{F}_{ik} + \mathbf{F}_{ki} = 0$. Отсюда следует, что геометрическая сумма всех внутренних сил, действующих в системе, равна нулю. Запишем этот результат в виде соотношения

$$\mathbf{F}_1^{(i)} + \mathbf{F}_2^{(i)} + \ldots + \mathbf{F}_n^{(i)} = 0, \tag{12.2}$$

снабдив каждую силу верхним индексом (i), который указывает, что речь идет о внутренних силах. Нижний индекс обозначает номер материальной точки, на которую действует сила. Таким образом, $\mathbf{F}_1^{(i)}$, например, обозначает полную внутреннюю силу, действующую на первую материальную точку. Обозначим далее символами $\mathbf{F}_1^{(e)}$, $\mathbf{F}_2^{(e)}$, ... внешние силы, действующие на материальные

точки системы. Тогда на основании второго закона Ньютона можно написать

$$\frac{d\mathbf{p}_1}{dt} = \mathbf{F}_1^{(i)} + \mathbf{F}_1^{(e)},$$

$$\frac{d\mathbf{p}_2}{dt} = \mathbf{F}_2^{(i)} + \mathbf{F}_2^{(e)},$$

. . . . . .

Сложив почленно эти уравнения и приняв во внимание соотношение (12.2), найдем

$$\frac{d}{dt}(\mathbf{p}_1 + \mathbf{p}_2 + \ldots + \mathbf{p}_n) = \mathbf{F}_1^{(e)} + \mathbf{F}_2^{(e)} + \ldots + \mathbf{F}_n^{(e)},$$

или

$$\frac{d\mathbf{p}}{dt} = \mathbf{F}^{(e)}, \tag{12.3}$$

где $\mathbf{p}$ — импульс всей системы, $\mathbf{F}^{(e)}$ — равнодействующая всех внешних сил, действующих на нее. Таким образом, *производная по времени от импульса системы материальных точек равна геометрической сумме всех внешних сил, действующих на систему.* Внутренние силы исключаются третьим законом Ньютона. Уравнение (12.3) является обобщением соответствующего уравнения для одной материальной точки.

Допустим теперь, что геометрическая сумма всех внешних сил равна нулю (это имеет место, например, для замкнутой системы). Тогда $\frac{d\mathbf{p}}{dt} = 0$. Производная постоянной величины равна нулю. Справедливо и обратное утверждение: если производная некоторой величины равна нулю, то эта величина постоянна. Поэтому из последнего уравнения следует, что $\mathbf{p} = \text{const}$.

Итак, *если геометрическая сумма внешних сил, действующих на систему, равна нулю, то импульс системы сохраняется, т. е. не меняется со временем.* В частности, это имеет место, когда система замкнута.

Допустим теперь, что $\mathbf{F}^{(e)} \neq 0$, однако равна нулю проекция силы $\mathbf{F}^{(e)}$ на какое-либо направление, например на направление оси X. Тогда из уравнения (12.3) следует, что для этой проекции $\frac{d\mathbf{p}_x}{dt} = 0$, а потому $\mathbf{p}_x = \text{const}$. Таким образом, полный импульс системы не сохраняется, но сохраняется проекция импульса на направление оси X. Например, импульс свободно падающего тела не может сохраняться, так как на тело действует вниз сила тяжести. Под действием этой силы вертикальная составляющая импульса непрерывно изменяется. Однако горизонтальная составляющая импульса при свободном падении остается неизменной. (Мы учитываем действие только силы тяжести и отвлекаемся от силы сопротивления воздуха и прочих сил.)

3. Относительно приведенного вывода закона сохранения импульса надо сделать следующее замечание. Вывод предполагает, что материальные точки замкнутой системы взаимодействуют между собой попарно, и это взаимодействие подчиняется третьему закону Ньютона. Для справедливости результата достаточно потребовать выполнения более слабого условия (12.2). Достаточно, чтобы обращалась в нуль геометрическая сумма внутренних сил, действующих в системе. Соблюдение этого условия, как будет показано в § 38, является следствием весьма общего свойства пространства — его *однородности.* Возможно, что и это более слабое условие не является необходимым. Возможно, что закон сохранения импульса останется справедливым даже в тех случаях, когда теряет смысл разделение системы на части и нельзя пользоваться представлением о силах взаимодействия между ними, а также другими представлениями и понятиями классической механики. Возможно, что такая ситуация встречается внутри атомных ядер или при превращениях «элементарных» частиц. Опыт показывает, что *закон сохранения импульса,* надлежащим образом обобщенный, *является фундаментальным законом природы, не знающим никаких исключений.* Однако в таком широком понимании он уже не может рассматриваться как следствие законов Ньютона.

4. В нашем изложении закон сохранения импульса для замкнутой системы из двух взаимодействующих материальных точек был постулирован. Его доказательством служил опыт. Это было сделано для того, чтобы ввести понятие массы. Но можно ввести это понятие иначе, а именно определить отношение масс сравниваемых тел по обратному отношению ускорений, сообщаемым им равными силами. Этот способ *не требует предварительного измерения сил.* Достаточно лишь располагать критерием *равенства сил.* Например, если на два тела последовательно подействовать одной и той же пружиной, растянутой на одну и ту же длину, то можно утверждать, что действующие на них силы одинаковы. В сущности, способ определения массы, использованной нами в § 10, является частным случаем этого второго, более общего способа. Он использует то обстоятельство, что два тела, приведенные во взаимодействие, подвергаются согласно третьему закону Ньютона воздействию сил, равных по модулю. Понятно, что если при определении массы не опираться на третий закон Ньютона, а пользоваться каким-либо другим *независимым способом*, то при доказательстве закона сохранения импульса не потребуется особо выделять случай двух взаимодействующих материальных точек. И в этом случае закон сохранения импульса будет теоремой механики. Определение массы, принятое нами в § 10, обладает, однако, тем преимуществом, что оно не нуждается в указании дополнительного критерия, позволяющего судить о равенстве действующих сил. В общем случае такой критерий, не опирающийся на третий закон Ньютона, указать затруднительно.

5. Иногда взаимодействие двух тел A и B осуществляется посредством третьего тела. Тогда мы имеем дело с системой трех тел, и надо принимать во внимание уравнение движения этого третьего тела. Между тем во многих случаях рассуждают так, как если бы этого третьего тела совсем не было. Выясним, когда такой способ рассуждения допустим и не приводит к ошибкам. Для этого рассмотрим следующим пример.

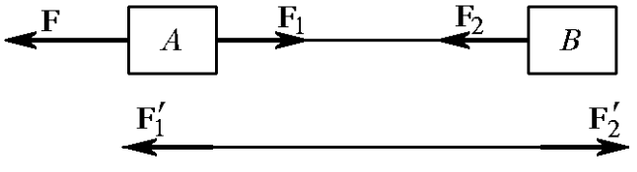


Рис. 23

Тела A и B связаны между собой нерастяжимой нитью (рис. 23). На тело A действует сила $\mathbf{F}$, натягивающая нить, вследствие чего оба тела A и B движутся с одним и тем же ускорением a. Обычно рассуждают следующим образом. Обозначим через F_1 силу, с которой тело B действует на тело A посредством натянутой нити, а через F_2 — противоположно направленную силу, с которой тело A действует на тело B. Тогда

$$m_A a = F - F_1, \quad m_B a = F_2, \tag{12.4}$$

где m_A и m_B — массы тел A и B. По третьему закону Ньютона $F_1 = -F_2$. Исключая F_1 и F_2, находим ускорение

$$a = \frac{F}{m_A + m_B},$$

а затем силы F_1 и $-F_2$:

$$F_2 = -F_1 = \frac{m_B}{m_A + m_B} F.$$

Это рассуждение неполно и может привести к неправильному результату. Из рассуждения выпало третье тело — нить, которая также движется ускоренно. Тела A и B не взаимодействуют непосредственно между собой. Они взаимодействуют с *нитью,* и третий закон Ньютона надо применять именно к таким взаимодействиям. Вот более подробное рассуждение, в котором учитывается ускорение, сообщаемое нити. В нем под F_1 и F_2 следует понимать силы, с которыми на тела A и B действует *натянутая нить.* Силы, с которыми на нить действуют тела A и B, обозначим через F_1' и F_2'. К уравнениям (12.4) надо присоединить уравнение движения нити: $ma = F_1' - F_2'$, где m — масса нити. Ввиду равенства действия и противодействия $F_1' = F_1$, $F_2' = F_2$, так что

$$ma = F_1 - F_2.$$

Решая это уравнение совместно с (12.4), получим

$$a = \frac{F}{m_A + m_B + m},$$

$$F_2 = m_B a, \quad F_1 = (m_B + m)\, a.$$

Теперь $F_1 \neq F_2$, поскольку $m \neq 0$. Допустим, однако, что масса нити пренебрежимо мала по сравнению с массами тел A и B. Тогда, отбрасывая член ma, получим приближенно $F_1 = F_2$. В этом приближении результат получается такой же, как если бы тела A и B *непосредственно* взаимодействовали между собой. Идеализируя задачу, говорят, что взаимодействие между телами A и B осуществляется посредством *«безмассового» тела* (нити). Подобные случаи встречаются очень часто. Безмассовые тела просто выбрасываются из рассмотрения. Однако безмассовых тел в действительности не существует, они являются идеализированными абстракциями. Надо отдавать себе отчет, когда можно и когда нельзя пользоваться такими идеализированными абстракциями. В приведенном примере было бы грубой ошибкой пользоваться соотношением $F_1 = F_2$ в тех случаях, когда масса нити сравнима с массами тела A и B.

§ 13. ВЗАИМОДЕЙСТВИЕ ТЕЛ НА РАССТОЯНИИ И ПОЛЕВОЕ ВЗАИМОДЕЙСТВИЕ

1. Взаимодействие тел может происходить либо при их *непосредственном* соприкосновении, либо *на расстоянии.* В первом случае взаимодействующие тела тянут или толкают друг друга. Возникающие при этом силы обычно вызываются *деформациями тел.* Деформации могут быть малы и не представлять непосредственного интереса в изучаемом явлении. Тогда от них можно отвлечься, учтя их влияние введением соответствующих сил *натяжения* и *давления.* Но если нас интересует происхождение и механизм действия сил, то надо подробно рассмотреть картину деформаций, возникающих в телах. Так, в примере, рассмотренном в конце предыдущего параграфа, нить действует на тела A и B (см. рис. 23) потому, что она растянута. В приведенных там расчетах мы отвлеклись от этого обстоятельства, так как предполагали, что растяжение нити мало, так что скорости связанных тел A и B можно было считать одинаковыми. В этом случае результаты расчета не зависят от степени растяжения нити. Это было бы, вообще говоря, не так, если вместо нити взять мягкую пружину. Мы не можем также отвлечься от растяжения нити, если хотим понять происхождение сил F_1 и F_2, с которыми нить действует на тела A и B. Точно также надо рассмотреть деформации тел A и B, если мы хотим понять, почему эти тела растягивают нить. Ведь тела A и B действуют на нить только потому, что они деформированы. Камень, привязанный к веревке, заставляют вращаться по окружности. При этом неизбежно возникнут деформации. Если бы камень не был деформирован, то он не мог бы двигаться с ускорением. Всякая мысленно вырезанная малая часть камня движется с ускорением потому, что на нее действуют окру-

жающие части камня. А это возможно только тогда, когда камень деформирован.

Помимо сил, зависящих от деформации тел, возможны и более сложные случаи. Например, силы взаимодействия могут зависеть не только от самих деформаций, но и от их *скоростей.* Примером могут служить силы трения. Но и эти силы возникают лишь при непосредственном соприкосновении взаимодействующих тел. Во всех этих случаях говорят, что силы взаимодействия являются *силами близкодействия.*

2. Помимо сил, действующих при соприкосновении тел, в природе существуют силы, которые, во всяком случае при непосредственном созерцании, воспринимаются нами как силы, *непосредственно действующие на расстоянии* без какого бы то ни было участия промежуточной среды. Они существуют даже тогда, когда взаимодействующие тела разделены «пустым» пространством. К силам такого рода относятся, например, гравитационные силы, а также силы взаимодействия наэлектризованных и намагниченных сил.

Согласно основным представлениям механики Ньютона силы, действующие на всякое тело в какой-либо момент времени, зависят от положения и скоростей остальных тел в тот же момент времени. Когда взаимодействующие тела не соприкасаются, такое представление предполагает либо непосредственное действие на расстоянии, либо передачу взаимодействий с *бесконечно большой скоростью.* Ньютоновская механика принципиально допускала взаимодействия, передающиеся с бесконечно большими скоростями. Логически против такого допущения возразить ничего нельзя. Вопрос должен быть решен опытом. Опытные факты привели к заключению, что мгновенных взаимодействий не существует. Скорость распространения взаимодействий ограничена (она не превосходит скорости света в вакууме). Отсюда следует, что описание взаимодействий, даваемое ньютоновской механикой, может привести и действительно приводит к практически верным результатам только тогда, когда скорости всех тел рассматриваемой системы пренебрежимо малы по сравнению со *скоростью распространения взаимодействия* (скоростью света). Но принципиально такое описание неверно. Для иллюстрации рассмотрим третий закон Ньютона. Из-за наличия конечной скорости распространения взаимодействий этот закон не может быть всегда верным для взаимодействий, осуществляющихся на расстоянии. Действительно, пусть две материальные точки длительно находятся в покое в положениях A и B (рис. 24). Допустим, что силы их взаимодействия в этих положениях подчиняются третьему закону Ньютона. Если первая точка перейдет в новое положение A' настолько быстро, что за время перехода взаимодействие не успеет распространиться до точ-

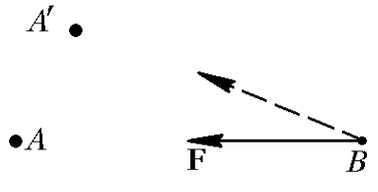


Рис. 24

ки B, то сила $\mathbf{F}$ при этом не изменится. Она будет определяться не новым, а прежним расположением материальных точек, т. е. будет направлена вдоль прямой BA, а не вдоль прямой BA', как должно было бы быть по третьему закону Ньютона. Получилось нарушение третьего закона, а с ним и механического закона сохранения импульса взаимодействующих сил. *Суммарный импульс тел A и B, вообще говоря, не может сохраняться из-за конечности скорости распространения взаимодействий.*

3. Физик XIX века сказал бы, что такое нарушение третьего закона Ньютона и механического закона сохранения импульса является кажущимся. В действительности нет непосредственного действия тел на расстоянии. Непосредственное действие тел на расстоянии предполагает, что тело может оказывать действие в тех местах, где оно находится и от которых оно отделено пустым пространством. По этой причине такое действие невозможно. Все взаимодействия осуществляются посредством промежуточной среды. Тело A, действуя на окружающую среду, вызывает в ней какие-то изменения, например, напряжения и давления, распространяющиеся с конечными скоростями. Когда эти изменения доходят до тела B, они проявляются в виде сил, действующих на него. Но как быть в тех случаях, когда взаимодействующие тела отдалены друг от друга совершенно пустым пространством? На это физик XIX века отвечал бы, что абсолютно пустого пространства не существует. Все пространство заполнено какой-то средой, например *мировым эфиром*, через которую и осуществляется взаимодействие. Такую гипотетическую среду физик XIX века наделял свойствами, аналогичными свойствам привычных нам твердых, жидких и газообразных тел. Третий закон Ньютона всегда справедлив — сказал бы он, — но его надо применять не к силам взаимодействия непосредственно между удаленными телами A и B (таких сил просто не существует, они кажущиеся), а к силам взаимодействия между этими телами и окружающей средой. Закон сохранения импульса также всегда справедлив. Однако полный импульс слагается не только из импульсов тел, входящих в систему, но и из импульса промежуточной среды, через которую передаются взаимодействия.

Физик нашего времени рассуждает аналогично, но более осторожно и абстрактно. Он также не признает непосредственного действия тел на расстоянии. Однако он не признает и промежуточной среды, через которую якобы передаются взаимодействия. Он говорит, что все взаимодействия осуществляются *полями* — гравитационными, электромагнитными и прочими. Тело A возбуждает в окружающем пространстве *силовое поле*, которое в месте нахождения тела B проявляется в виде действующих на него сил. В свою очередь, тело B возбуждает аналогичное силовое поле, действующее на тело A. Никаких других силовых взаимодействий, помимо полевых, современная физика не признает. Взаимодействия прикосновением являются частными случаями полевого взаимодействия. Они осуществляются *молекулярными полями.* Молекулярные поля быстро

убывают с расстоянием и проявляются, когда расстояние между взаимодействующими телами не превышает примерно 10^{-7} см. Вот почему такие *полевые взаимодействия* макроскопически воспринимаются как *«взаимодействия прикосновением»*.

4. Возникает вопрос, не является ли различие между точками зрения современного физика и физика прошлого века чисто терминологическим: раньше говорили о промежуточной среде, теперь говорят о поле. Дело, конечно, не в терминологии, а в существе, в реальных физических свойствах поля и гипотетической среды — мирового эфира, посредством которой, по воззрениям физиков XIX века, осуществляются взаимодействия тел. Мировому эфиру приписывались свойства, аналогичные свойствам обычных тел. Можно было говорить о покое и движении эфира, о его упругости, плотности, силах, действующих на эфир. Можно было говорить о движении тел относительно эфира, с эфиром можно было связать систему отсчета и т. д. Обо всем этом не имеет смысла говорить, когда речь идет о поле. На поле современная физика смотрит как на некоторую объективную реальность, посредством которой передаются взаимодействия. Поле может существовать и самостоятельно, независимо от возбудивших его тел. Таковы, например, электромагнитные волны, излучаемые радиостанциями. Радиостанция может прекратить работу, а излученное ею электромагнитное поле продолжает существовать и распространяться в пространстве. Достигнув удаленного приемника, оно доставляет информацию, которую несколько ранее передала радиостанция. Нет абсолютно пустого пространства, пространство заполнено полями. Поле, наряду с веществом, является *одним из видов материи.* Физикам XIX века казалось, что понять природу сил взаимодействия — это значит свести их к механическим силам, возникающим при соприкосновении тел, например, к силам упругости. Современная физика отрицает такую постановку вопроса.

Упругие силы, силы давления и натяжения, мускульная сила и прочие силы издавна были привычны человеку. Он считал эти силы чем-то простым и понятным. Электрические и магнитные силы, с которыми он столкнулся позднее, были восприняты им как нечто таинственное и непонятное. Возникло естественное стремление свести эти таинственные силы к более простым и привычным силам упругости, к силам давления и натяжения. Однако в действительности дело обстоит как раз наоборот. Более простыми и «понятными» являются силы электромагнитных взаимодействий, а упругие и прочие силы — более сложными. По современным представлениям силы упругости, силы трения, силы химического сродства, молекулярные силы, мускульная сила и прочие обычные силы, с которыми сталкивается человек, за исключением сил всемирного тяготения, являются не чем иным, как проявлением электромагнитных сил. Поэтому задача сведения электромагнитных сил к силам упругости просто бессмысленна.

Поле действует на тела с определенными силами. Однако не имеет смысла говорить о механических силах, действующих на поля. Поэтому с точки зрения полевого взаимодействия третий закон Ньютона может нарушаться: на тело действует сила, но нет силы противодействия, действующей на другое тело. Закон сохранения импульса, однако, остается верным, так как импульсом могут обладать не только тела, но и поля. Импульс поля проявляется в изменении импульса тела, излучившего или поглотившего энергию поля. При излучении тело теряет импульс, уносимый полем, при поглощении оно приобретает импульс за счет поглощения энергии поля.

5. Примером проявления импульса электромагнитного поля может служить *давление света*. Опытами русского физика-экспериментатора П. Н. Лебедева (1866—1912) было показано, что свет оказывает давление на тела, на которые он падает. Рассмотрим изолированную систему из двух тел A и B (рис. 25). Пусть тело A излучило кратковременный световой сигнал по направлению к телу B. Дойдя до тела B, свет поглощается, оказывая на него давление. В результате тело B приходит в движение, его импульс меняется. Закон сохранения импульса, очевидно, был бы нарушен, если бы свет, направлявшийся к телу B, не обладал импульсом. Мы должны приписать ему импульс, равный импульсу, приобретенному телом B после поглощения света. Но если свет обладает импульсом, то при его излучении излучающее тело A должно испытывать *отдачу*. Иначе процесс излучения света сопровождался бы нарушением закона сохранения импульса.

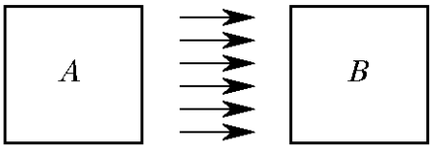


Рис. 25

Эти рассуждения можно облечь в количественную форму и прийти к важным соотношениям. Допустим, что на полностью поглощающее тело B перпендикулярно к его поверхности падает параллельный пучок света. Опытами Лебедева было показано, что давление π, оказываемое светом на единицу площади тела B, равно объемной плотности энергии падающего пучка. Обозначим через l длину, а через S — площадь поперечного сечения падающего пучка. Тогда $\pi = \frac{\varepsilon}{Sl}$, где ε — энергия пучка. Сила действующая на тело B, равна $F = \pi S = \varepsilon/l$. Она действует в течение времени $\tau = l/c$ (c — скорость света в вакууме), сообщая телу импульс $p = F\tau = \varepsilon/c$ (см. § 18). Это и есть импульс света, поглощенного телом B. Итак, свет, распространяющийся в определенном направлении, обладает импульсом

$$p = \frac{\varepsilon}{c}. \qquad (13.1)$$

Поскольку распространение идет со скоростью c, целесообразно представить импульс в виде $p = mc$, рассматривая величину m как *массу света.* Она равна

$$m = \frac{\varepsilon}{c^2}. \tag{13.2}$$

Соотношение (13.2) получено здесь для энергии света. Теория относительности показала, что оно справедливо для любых видов энергии. В таком расширенном понимании соотношение (13.2) выражает фундаментальный закон Эйнштейна о *взаимосвязи между массой и энергией.*

6. В механике нам не придется сталкиваться с явлениями, в которых проявляются импульсы полей. Мы ограничимся изучением только таких явлений, для которых третий закон Ньютона и закон сохранения импульса в их старом — ньютоновском — смысле выполняются.

§ 14. РОЛЬ НАЧАЛЬНЫХ УСЛОВИЙ

1. Векторное уравнение движения материальной точки (11.3) можно записать в *координатной форме*:

$$m\frac{d^2x}{dt^2} = F_x, \quad m\frac{d^2y}{dt^2} = F_y, \quad m\frac{d^2z}{dt^2} = F_z. \tag{14.1}$$

Одно векторное уравнение (11.3) эквивалентно трем числовым уравнениям (14.1). Все эти уравнения являются *дифференциальными,* а потому их недостаточно для однозначного определения движения материальной точки. Каждое из них есть уравнение *второго порядка.* (Порядок дифференциального уравнения определяется производной высшего порядка, входящей в это уравнение). По этой причине для однозначного определения движения точки к уравнениям движения надо присоединить дополнительные данные, определяющие значения двух векторных или шести числовых постоянных. В качестве таковых обычно берут значения радиуса-вектора $\mathbf{r}$ и скорости $\mathbf{v}$ или каких-либо двух функций их в момент времени $t = 0$. Эти значения называются *начальными условиями.* Выясним этот вопрос на примере свободного движения материальной точки в поле тяжести Земли.

2. Галилеем было установлено, что *все тела в пустоте падают с одинаковым ускорением.* Для качественного подтверждения этого положения может служить стеклянная трубка длиной около одного метра, из которой откачан воздух. В трубку помещаются различные тела, например дробинка, кусочек пробки, перышко, кусочек бумаги. Пока трубка не откачана, бумажки и перышки падают во много раз медленнее остальных тел, что объясняется сопротивлением воздуха. Но если воздух из трубки откачать, то все тела начнут падать

одинаково быстро. Более точное доказательство дают наблюдения за качаниями маятника: опыт показывает, что период качания маятника не зависит от материала, из которого он изготовлен. Ускорение свободного падения меняется с географической широтой: на полюсе оно максимально и составляет 9,83 м/с^2, на экваторе — минимально и равно 9,78 м/с^2. Ускорение g уменьшается с высотой над земной поверхностью: при поднятии на 1 м оно убывает приблизительно на $3 \cdot 10^{-6}$ м/с^2. Для средних широт можно принять, что вблизи земной поверхности $g = 9{,}80$ м/с^2. В расчетах, не требующих особой точности, ускорение свободного падения g может считаться одним и тем же для всей земной поверхности.

На тело в поле тяжести Земли действует сила $\mathbf{F} = m\mathbf{g}$, а потому уравнение движения (11.8) переходит в уравнение

$$\frac{d^2\mathbf{r}}{dt^2} = \mathbf{g}. \tag{14.2}$$

Мы пренебрегли всеми силами и учли только силу тяжести. Зависимостью g от географической широты и высоты над земной поверхностью также будем пренебрегать. Итак, ускорение g будем считать постоянным. Уравнение (14.2) эквивалентно двум уравнениям:

$$\frac{d\mathbf{v}}{dt} = \mathbf{g}, \quad \frac{d\mathbf{r}}{dt} = \mathbf{v}. \tag{14.3}$$

Простым дифференцированием нетрудно убедиться, что этим уравнениям удовлетворяют следующие решения:

$$\mathbf{v} = \mathbf{g}t + \mathbf{v}_0, \quad \mathbf{r} = \tfrac{1}{2}\,\mathbf{g}t^2 + \mathbf{v}_0 t + \mathbf{r}_0 \tag{14.4}$$

при произвольных значениях постоянных векторов $\mathbf{r}_0$ и $\mathbf{v}_0$. Решение (14.4) является *общим*. Это значит, что любое решение уравнения (14.2) может быть представлено в виде (14.4). Общее решение — это, в сущности, не одно решение, а целое *семейство решений*, зависящее от двух произвольных векторных постоянных $\mathbf{r}_0$ и $\mathbf{v}_0$. Придавая этим постоянным какие-либо конкретные значения, мы выделяем из этого семейства определенное *частное решение*. Постоянная $\mathbf{v}_0$ есть начальная скорость движущейся точки, $\mathbf{r}_0$ — радиус-вектор ее в начальный момент времени. В этом легко убедиться, если с помощью формул (14.4) найти значения $\mathbf{v}$ и $\mathbf{r}$ при $t = 0$. Постоянные $\mathbf{r}_0$ и $\mathbf{v}_0$ нельзя определить из дифференциального уравнения движения (14.2), так как при любых значениях этих постоянных выражения (14.4) являются решениями этого уравнения. Величины $\mathbf{r}_0$ и $\mathbf{v}_0$ определяются *начальными условиями*. В зависимости от значений $\mathbf{r}_0$ и $\mathbf{v}_0$ движения могут сильно отличаться друг от друга. Тело может подниматься вверх или опускаться вниз по прямой линии; оно может описывать параболу, достигая или не достигая ее вершины; дуга параболы может быть изогнута сильнее или

слабее и т. д. Получается довольно разнообразный и запутанный класс движений. Заслуга Ньютона, между прочим, и состоит в том, что он подметил, что вся эта сложность исчезает, а все многообразие движений может быть описано единой формулой, не содержащей никаких произвольных постоянных, если от положений и скоростей материальной точки перейти к ее ускорению.

3. Полученные результаты допускают обобщения. Допустим, что имеется система N материальных точек, взаимодействующих между собой и с внешними телами, положение которых предполагается заданным в любой момент времени. Записав математически второй закон Ньютона для каждой материальной точки, мы получим систему N векторных или $3N$ эквивалентных им числовых дифференциальных уравнений второго порядка. Можно показать, что для однозначного решения этих уравнений надо задать $2N$ векторных или $6N$ числовых величин, определяющих начальные значения координат и скоростей материальных точек системы.

ЗАДАЧА

Тело брошено вверх под углом α к горизонту с начальной скоростью $\mathbf{v}_0$. Исследовать его движения, пренебрегая сопротивлением воздуха. Найти уравнение траектории, дальность полета и максимальную высоту подъема, считая земную поверхность горизонтальной. При каком угле α дальность полета максимальна?

Р е ш е н и е. Точку земной поверхности, откуда брошено тело, примем за начало координат ($\mathbf{r}_0 = 0$). Тогда, как видно из (14.4), движение будет происходить в вертикальной плоскости, в которой лежат векторы $\mathbf{g}$ и $\mathbf{v}_0$. Примем ее за координатную плоскость XY, направив ось X горизонтально в сторону движения, а ось Y — вертикально вверх. Запишем уравнение (14.4) в проекциях на координатные оси,

$$v_x = v_0 \cos\alpha, \quad v_y = v_0 \sin\alpha - gt, \quad x = v_0 t \cos\alpha, \quad y = v_0 t \sin\alpha - \frac{1}{2} gt^2.$$

Исключая из последних двух уравнений время t, найдем уравнение траектории

$$y = x \operatorname{tg}\alpha \frac{gx^2}{2v_0^2 \cos^2\alpha}.$$

Это уравнение параболы. Отсюда находим дальность полета

$$x = \frac{v_0^2}{g} \sin 2\alpha$$

и максимальную высоту поднятия

$$y_{\text{макс}} = \frac{v_0^2 \sin^2\alpha}{2g}.$$

Максимальная дальность достигается при $\alpha = 45°$ и равна

$$x_{\text{макс}} = \frac{v_0^2}{g}.$$

§ 15. ПРИНЦИП ОТНОСИТЕЛЬНОСТИ ГАЛИЛЕЯ

1. Уравнение, выражающее закон Ньютона,

$$m\mathbf{a} = \mathbf{F}, \tag{15.1}$$

отчетливо показывает, что этот закон не может быть справедлив в любой системе отчета. Действительно, ускорение $\mathbf{a}$, вообще говоря, имеет разные значения в различных системах отсчета, движущихся относительно друг друга с ускорением. Сила же $\mathbf{F}$ не может зависеть от выбора системы отсчета, так как она определяется только взаимными расположениями и относительными скоростями материальных точек системы, а эти величины согласно нерелятивистской кинематике от выбора системы отсчета не зависят. Отсюда следует, что если второй закон Ньютона справедлив в какой-либо системе отсчета, то он не может оставаться справедливым в другой системе отсчета, движущейся относительно первой с ускорением.

2. Допустим, что система отсчета S инерциальна. Рассмотрим вторую систему отсчета S', движущуюся относительно первой поступательно с постоянной скоростью $\mathbf{V}$. Пусть известно движение материальной точки в одной из этих систем, например в системе S. Как найти движение той же точки в системе S'? Задача в дорелятивистской ее постановке сводится к нахождению формул, выражающих координаты x', y', z' движущейся точки в системе S' через ее координаты x, y, z в системе S в один и тот же момент времени.

Начало координат и направление координатных осей можно выбрать произвольно как в системе S, так и в системе S'. Если координатные системы неподвижны друг относительно друга и отличаются друг от друга только положениями начал и направлениями координатных осей, то преобразование координат есть чисто геометрическая задача. Ее решение известно из аналитической геометрии.

Остается только выяснить, что нового вносит в вопрос о преобразовании координат движение одной системы отсчета относительно другой? Для простоты можно принять, что координатные оси X', Y', Z' соответственно параллельны координатным осям X, Y, Z и что в начальный момент времени $t = 0$ начало O' совмещается с началом O. Кроме того, можно считать, что скорость $\mathbf{V}$ параллельна оси X. При этих условиях ось X' все время будет совпадать с осью X.

Такие упрощения в постановке задачи не лишают ее общности, поскольку переход к общим формулам может быть совершенен дополнительным переносом начал координат и поворотом координатных осей.

Пусть в момент времени t движущаяся точка находится в положении M (рис. 26). Тогда $\overrightarrow{OM} = \overrightarrow{OO'} + \overrightarrow{O'M}$. За время t начало ко-

ординат системы S' переходит из положения O в положение O', причем $\overrightarrow{OO'} = \mathbf{V}t$. Ввиду этого предыдущее соотношение принимает вид

$$\mathbf{r} = \mathbf{r}' + \mathbf{V}t', \quad t = t', \tag{15.2}$$

где $\mathbf{r} = \overrightarrow{OM}$, $\mathbf{r}' = \overrightarrow{O'M}$ — радиусы-векторы движущейся точки в системах S и S' соответственно. Запишем соотношение (15.2) в проекциях на координатные оси:

$$x = x' + \mathbf{V}t', \quad y = y', \quad z = z', \quad t = t'. \tag{15.3}$$

Формулы обратного преобразования имеют вид

$$\mathbf{r}' = \mathbf{r} - \mathbf{V}t, \quad t' = t, \tag{15.4}$$

или в координатной форме

$$x' = x - \mathbf{V}t, \quad y' = y, \quad z' = z, \quad t' = t. \tag{15.5}$$

Эти формулы и дают решение поставленной задачи. Они называются *преобразованием Галилея.* Мы присоединили к формулам преобразования координат дополнительную формулу $t' = t$, чтобы явно отметить, что в нерелятивистской кинематике время считается абсолютным, а потому не преобразуется.

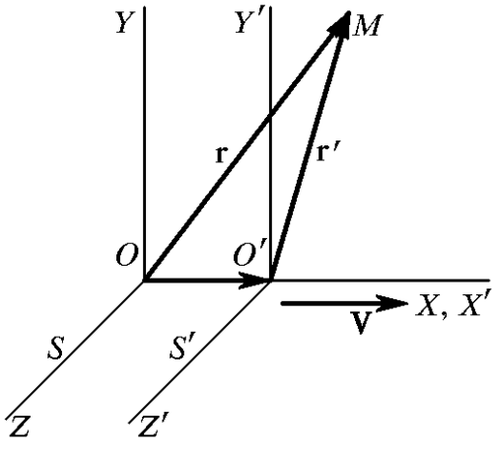


Рис. 26

С точки зрения «здравого смысла» преобразование Галилея кажется самоочевидным. Однако в основе его вывода лежит предположение дорелятивистской кинематики об абсолютности длин и промежутков времени. Абсолютность времени явно отмечена в уравнении $t = t'$. При выводе остальных формул использовано предположение об абсолютности длин. Действительно, формулы (15.2), (15.3) и (15.4) были бы самоочевидными, если бы радиусы-векторы $\mathbf{r}$ и $\mathbf{r}'$, а с ними и все координаты x, y, z, x', y', z' измерялись в одной и той же системе отсчета, например S. Но в действительности формулы предполагают, что «нештрихованные» величины $\mathbf{r}$, x, y, z измеряются в системе S, а «штрихованные» $\mathbf{r}'$, x', y', z' — в системе S'. По этой причине при выводе формул преобразования Галилея без предположения абсолютности расстояний и промежутков времени обойтись нельзя. Релятивистская физика отказалась от такой абсолютности. Преобразование Галилея она заменила *преобразованием Лоренца* (по имени ирландского физика-теоретика Лоренца (1853—1928)). Этот вопрос будет подробно рассмотрен при изложении теории относительности (см. т. IV). Сейчас достаточно отметить, что преобразование Галилея является предельным случаем преобразования Лоренца и получается из последнего, когда скорость V пренебрежимо мала по срав-

нению со скоростью света в вакууме. При изучении «медленных движений» ($V^2/c^2 \ll 1$) можно пользоваться преобразованием Галилея. В случае «быстрых движений» этого делать нельзя.

3. Дифференцируя соотношение (15.2) по времени t, получим

$$\frac{d\mathbf{r}}{dt} = \frac{d\mathbf{r}'}{dt} + \mathbf{V} = \frac{d\mathbf{r}'}{dt'} + \mathbf{V}',$$

или

$$\mathbf{v} = \mathbf{v}' + \mathbf{V}, \tag{15.6}$$

где $\mathbf{v}$ — скорость точки в системе S, а $\mathbf{v}'$ — в системе S'. Эта формула выражает *нерелятивистский закон сложения скоростей* (в физическом смысле). Она выведена здесь в предположении, что скорость $\mathbf{V}$ постоянна. Но формула верна и в случае, когда скорость $\mathbf{V}$ не постоянна. Однако для целей настоящего параграфа достаточно скорость $\mathbf{V}$ считать величиной постоянной.

Дифференцируя второй раз в предположении постоянства $\mathbf{V}$, получим

$$\frac{dv}{dt} = \frac{dv'}{dt} = \frac{dv'}{dt'},$$

или

$$\mathbf{a} = \mathbf{a}'. \tag{15.7}$$

Здесь $\mathbf{a}$ — ускорение точки в системе S, а $\mathbf{a}'$ — ускорение той же точки в системе S'. Таким образом, ускорение в обеих системах отсчета одно и то же. Говорят, что *ускорение инвариантно относительно преобразования Галилея.*

Свободная материальная точка движется в системе S без ускорения, так как по предположению система S инерциальна. Формула (15.7) показывает, что ее движение в системе S' будет также неускоренным. Следовательно, система S' — тоже инерциальная система отсчета. Таким образом, *система отсчета, движущаяся прямолинейно и равномерно относительно инерциальной системы, сама является инерциальной системой отсчета.* Если существует хотя бы одна инерциальная система отсчета, то существует и бесконечное множество инерциальных систем, движущихся друг относительно друга прямолинейно и равномерно. Сила является функцией только инвариантных величин: разностей координат и разностей скоростей взаимодействующих материальных точек. Поэтому она не меняется при переходе от одной системы отсчета к другой: $\mathbf{F} = \mathbf{F}'$. Иначе говоря, *сила инвариантна относительно преобразования Галилея.* Так как ускорение инвариантно: $\mathbf{a} = \mathbf{a}'$, то из уравнения (15.1) следует

$$m\mathbf{a}' = \mathbf{F}'.$$

Это уравнение выражает второй закон Ньютона в «штрихованной» системе отсчета S'. Оно имеет такой же вид, что и в «нештрихованной» системе S. Уравнения, остающиеся неизменными при переходе от одной системы отсчета к другой, называются *инвариантными.* Таким образом, *уравнения механики Ньютона инвариантны относительно преобразования Галилея.* Это утверждение называется *принципом относительности Галилея.*

4. Принцип относительности Галилея утверждает полное равноправие всех инерциальных систем отсчета. Значит ли это, что одно и то же движение выглядит одинаково во всех инерциальных системах отсчета? Конечно, нет. Движение тела, свалившегося с полки равномерно движущегося вагона, является прямолинейным, если его рассматривать относительно вагона. Но то же движение происходит по параболе в системе отсчета, связанной с полотном железной дороги, хотя законы механики Ньютона одинаковы в обеих системах отсчета. *Движение выглядит по-разному потому, что законы Ньютона выражаются дифференциальными уравнениями, а таких уравнений недостаточно, чтобы полностью определить движение.* Для этого к дифференциальным уравнениям надо присоединить начальные условия — задать начальное положение тела и его начальную скорость. В приведенном примере дифференциальные уравнения движения тела одни и те же в обеих системах отсчета, однако начальные условия разные. В вагоне тело падает с полки с начальной скоростью, равной нулю. В системе отсчета, связанной с полотном железной дороги, то же тело имеет начальную скорость в горизонтальном направлении, равную скорости поезда. Этим и объясняется различный характер движения в обеих системах отсчета. Для того чтобы движение получилось одинаковым, надо в обеих системах отсчета создать одинаковые начальные условия. Это надо понимать в следующем смысле.

Допустим, что имеются две замкнутые системы тел — две большие лаборатории, движущиеся относительно друг друга прямолинейно и равномерно. Каждая из лабораторий может служить системой отсчета. Пусть эти системы инерциальны. Предположим, что обе лаборатории совершенно тождественны, т. е. состоят из одного и того же набора одинаковых тел и оборудованы совершенно одинаково. Явления, происходящие внутри лабораторий, не зависят от того, что происходит в окружающем внешнем мире, так как по предположению лаборатории являются замкнутыми системами. Принцип относительности Галилея утверждает, что *основные механические законы, которыми определяются изменения состояния движения тел, в обеих лабораториях одни и те же.* Под основными механическими законами здесь понимаются законы, однозначно определяющие движение системы по начальным условиям, в которых она находилась, т. е. по значениям координат и скоростей всех материальных точек системы в произвольный момент времени, условно принимаемый за начальный. Если в обеих лабораториях создать

одинаковые начальные условия для всех без исключения тел, то все последующие движения их будут протекать совершенно одинаково в обеих лабораториях. Именно в таком смысле понимал принцип относительности сам Галилей. Он писал:

«Уединитесь с каким-нибудь приятелем в просторное помещение под палубой большого корабля и пустите туда мух, бабочек и других подобных мелких летающих насекомых. Пусть там находится также большой сосуд с водой и плавающими в нем рыбками. Подвесьте далее наверху ведерко, из которого капля за каплей вытекала бы вода в другой сосуд с узким горлышком, подставленный внизу. Пока корабль стоит неподвижно, наблюдайте старательно, как мелкие летающие живые существа с одной и той же скоростью летают во всех направлениях внутри помещения. Рыбки, как вы увидите, будут плавать безразлично во все стороны. Все падающие капли будут попадать в подставленный сосуд. Бросая приятелю какую-нибудь вещь, вам не придется применять бóльшую силу, чтобы бросить ее в одну сторону, чем в другую, если только вещь бросается на одни и те же расстояния. Прыгая двумя ногами, вы сделаете прыжок на одно и то же расстояние, независимо от его направления. Наблюдайте хорошенько за всем этим, хотя у нас не возникает никакого сомнения в том, что пока корабль остается неподвижным, все должно происходить именно так. Пусть теперь корабль приведен в движение с какой угодно скоростью. Если движение будет равномерным и без качки в ту и другую сторону, то во всех указанных явлениях вы не обнаружите ни малейшего изменения и ни по одному из них вы не сможете установить, движется ли корабль или стоит на месте».

Далее Галилей повторяет, как будут протекать на движущемся корабле все явления, описанные выше. Он замечает, в частности, что если бросить с одинаковой силой (надо было сказать — с одинаковой скоростью относительно корабля) один и тот же предмет сначала к корме, а затем к носу корабля, то в первом случае предмет пройдет относительно пола корабля не большее расстояние, чем во втором, хотя за время, пока предмет находится в воздухе, пол движущегося корабля успеет переместиться на значительное расстояние навстречу предмету. Аналогичные замечания делаются им и в отношении остальных явлений. Отмечая независимость всех явлений, наблюдаемых в закрытом помещении под палубою корабля, от равномерного движения последнего, Галилей приходит к следующему выводу:

«Причина согласованности всех этих явлений в том, что движение корабля обще всем находящимся в нем предметам, так же как и воздуху. Поэтому-то я и сказал, что вы должны находиться под палубой».

5. Было бы неправильно давать принципу относительности следующую формулировку: «Если в двух различных инерциальных системах отсчета в начальный момент времени *все без исключения тела и объекты Вселенной* поставить в совершенно одинаковые условия, то в дальнейшем в обеих системах отсчета все явления будут протекать совершенно одинаково». Такое утверждение бессодержательно и не может выражать никакого физического закона. Действительно, если две системы отсчета движутся одна относительно другой, то *все без исключения тела Вселенной* не могут в один и тот же момент времени находиться

в них в совершенно одинаковых условиях: скорости одних и тех же тел в этих двух системах отсчета будут разными. Поэтому предпосылка, о которой говорится в приведенной (неправильной) формулировке, не может быть выполнена, поскольку она имеет в виду все без исключения тела Вселенной. Принцип относительности (в правильной формулировке) является нетривиальным физическим законом потому, что в нем речь идет не о явлениях во всей Вселенной, а о явлениях внутри *конечных замкнутых систем* и *систем, находящихся в неизменных внешних условиях.* Примером может служить закрытое помещение на корабле, о котором говорил Галилей. Помещение должно быть закрыто. Иначе явления, в нем происходящие, зависели бы от скорости ветра, меняющейся с изменением скорости движения корабля. Такое помещение все же не является вполне замкнутой системой. Тела, в нем находящиеся, подвержены внешним явлениям: на них действует поле силы тяжести Земли. Однако это поле одно и то же, независимо от того, движется корабль равномерно или стоит на месте. Поэтому закрытое помещение с находящимися в нем телами ведет себя так же, как замкнутая система, хотя оно и находится во внешнем поле тяжести.

Принцип относительности иногда формулируют еще так: «Законы природы одинаковы во всех инерциальных системах отсчета». Недостаток этой формулировки состоит в том, что «одинаковость законов природы» может быть истолкована в смысле одинаковости протекания одного и того же явления во всех инерциальных системах отсчета. Это, как подробно разъяснено выше, неверно. Характер протекания физических явлений определяется не только основными законами природы, но и значениями параметров, определяющих начальные условия, в которых находилась система. Чтобы не возникло подобных неверных представлений, лучше говорить не просто о «законах природы», а по примеру Эйнштейна о «законах, по которым *происходят изменения состояний физических систем*», формулируя принцип относительности следующим образом:

Законы природы, по которым изменяются состояния физических систем, не зависят от того, к какой из инерциальных систем отсчета относятся эти изменения.

Эта формулировка является более общей, чем прежняя, в которой говорилось об инвариантности законов Ньютона относительно преобразования Галилея. Во-первых, здесь идет речь об инвариантности *всех физических законов,* а не только законов механики, каковыми являются законы Ньютона. Во-вторых, здесь не указан конкретный вид преобразований координат и времени, относительно которых законы природы инвариантны. Такие преобразования надо найти из *самого принципа относительности* и некоторых дополнительных соображений. Именно так в теории относительности получаются преобразования Лоренца, о которых было упомянуто выше. Законы природы инвариантны относительно преобразования Лорен-

ца. Принцип относительности в такой формулировке называется *принципом относительности Эйнштейна.* О нем будет идти речь в т. IV нашего курса.

§ 16. АДДИТИВНОСТЬ И ЗАКОН СОХРАНЕНИЯ МАССЫ

1. Пусть два тела с массами m_1 и m_2 сталкиваются между собой и соединяются в одно — составное — тело. Примером может служить слипание двух глиняных шаров при столкновении их между собой. Другим примером является *химическая реакция,* в которой два атома или ядра соединяются в молекулу. Требуется определить массу составного тела m, зная массы m_1 и m_2 соединяющихся тел. На первый взгляд ответ кажется тривиальным, а именно $m = m_1 + m_2$. Хотя это в какой-то мере и правильно, но требует обоснования. Обоснование в механике Ньютона можно дать на основе принципа относительности Галилея.

Рассмотрим процесс столкновения в какой-либо инерциальной системе отсчета S. Обозначим через $\mathbf{v}_1$ и $\mathbf{v}_2$ скорости тел до столкновения, а через $\mathbf{v}$ — скорость составного тела после столкновения. На основании закона сохранения импульса можно написать

$$m_1\mathbf{v}_1 + m_2\mathbf{v}_2 = m\mathbf{v}. \tag{16.1}$$

Рассмотрим теперь тот же процесс в системе отсчета S', движущейся относительно системы S прямолинейно и равномерно со скоростью $\mathbf{V}$. Согласно принципу относительности закон сохранения импульса справедлив также в системе S' и записывается в виде

$$m_1\mathbf{v}_1' + m_2\mathbf{v}_2' = m\mathbf{v}'. \tag{16.2}$$

Ввиду полного равноправия инерциальных систем отсчета массы всех тел в системе S' такие же, какими они были в системе S. В нерелятивистской физике скорости $\mathbf{v}_1'$, $\mathbf{v}_2'$ и $\mathbf{v}'$ в системе S' связаны с соответствующими скоростями в системе S соотношениями

$$\mathbf{v}_1' = \mathbf{v}_1 - \mathbf{V}, \quad \mathbf{v}_2' = \mathbf{v}_2 - \mathbf{V}, \quad \mathbf{v}' = \mathbf{v} - \mathbf{V}.$$

Поэтому (16.2) преобразуется в

$$m_1(\mathbf{v}_1 - \mathbf{V}) + m_2(\mathbf{v}_2 - \mathbf{V}) = m(\mathbf{v} - \mathbf{V}),$$

или на основании соотношения (16.1)

$$(m_1 + m_2)\mathbf{V} = m\mathbf{V}.$$

Отсюда ввиду произвольности $\mathbf{V}$

$$m = m_1 + m_2. \tag{16.3}$$

Масса составного тела, как мы и ожидали, *равна сумме масс составляющих тел. Это свойство называется аддитивностью массы.*

Доказательство может быть обобщено. Нет необходимости предполагать, что сталкиваются только два тела и что после столкновения они соединяются в одно тело. Можно взять, например, произвольную химическую реакцию, в которой реагирует несколько молекул или атомов, а в результате реакции получается несколько других молекул или атомов. Тогда, повторяя рассуждения, приведшие нас к соотношению (16.3), мы придем к более общему заключению, что *сумма масс веществ до реакции равна сумме масс веществ после реакции.* Это *закон сохранения вещества* или, точнее, *закон сохранения массы.* Масса пропорциональна весу. Поэтому этот закон называли также *законом сохранения веса.* Такое название применялось в старой литературе, но оно неудачно и теперь почти вышло из употребления. Однако Ломоносов (1711—1765) и Лавуазье (1743—1794), с именами которых обычно связывают утверждение в науке *закона сохранения вещества*, пришли к своему открытию на основании опытов по взвешиванию продуктов химических реакций. Поэтому то, что непосредственно доказали эти ученые, есть именно «закон сохранения веса».

2. Ньютон определил массу тела как *количество вещества,* содержащегося в нем. Такое определение бессодержательно, пока не указано, что следует понимать под «количеством вещества». Ньютон просто заменил термин «количество вещества» новым термином «масса», не дав ни одному из них точного определения. Однако с ньютоновским определением массы интуитивно связано представление об ее аддитивности. Если, например, две одинаковые гири с массами 1 кг каждая соединить в одну, то «интуитивно ясно», что получится гиря с массой 2 кг, ибо «количество вещества» в двух гирях вдвое больше, чем в одной из них. Интуиция является мощным источником научного творчества. Но в науке она не может служить доказательством. Доказательство должно основываться на точных определениях и законах природы. Точное определение понятия массы было дано в § 10. Если пользоваться только этим определением, то совсем не очевидно, что масса является величиной аддитивной. Необходимость доказательства этого свойства следует уже из того, что *аддитивность и закон сохранения массы вещества верны лишь приближенно.* В самом деле, этот закон мы получили как следствие галилеева принципа относительности. Но *галилеев принцип относительности не есть вполне точный закон природы — он является приближенным предельным случаем эйнштейновского принципа относительности.* Поэтому приведенное выше рассуждение должно быть пересмотрено — в основу его надо положить не галилеев, а эйнштейнов принцип относительности. Это будет сделано в т. IV при изложении теории относительности. Сейчас же мы ограничимся сообщением основного результата ввиду его фундаментального зна-

чения. *Закон сохранения массы в старом его понимании — сумма масс тел до реакции равна сумме масс тел после реакции — оказывается неверным. Законы сохранения массы и энергии, которые в дорелятивистской физике считались двумя независимыми точными законами природы, в релятивистской физике утратили свою независимость и были объединены в единый закон сохранения массы — энергии. Всякая энергия обладает массой, равной количеству энергии, деленному на квадрат скорости света в вакууме.* К такому заключению мы уже пришли в частном случае лучистой энергии (см. конец § 13). В балансе массы должна учитываться не только масса вещества, но и масса выделяющейся или поглощающейся энергии. Во всех процессах природы сумма массы вещества и энергии остается постоянной.

То обстоятельство, что в химических реакциях не было обнаружено изменение массы вещества, связано с их относительно очень малым энергетическим выходом. Изменение массы за счет выделения или поглощения энергии столь мало, что его обнаружение находится за пределами точности измерений. Так, при сгорании 12 г углерода с образованием углекислого газа CO_2 выделяется около 99 ккал тепла. В эргах величина равна $\varepsilon = 99 \cdot 4{,}19 \cdot 10^{10} \approx 4 \cdot 10^{12}$ эрг. По формуле (13.2) ей соответствует масса

$$\Delta m \approx \frac{4 \cdot 10^{12}}{9 \cdot 10^{20}} \approx 0{,}5 \cdot 10^{-8} \text{ г}.$$

Полная масса веществ, участвующих в реакциях, равна $m = 12 + 32 = 44$ г. Поэтому

$$\frac{\Delta m}{m} = \frac{0{,}5 \cdot 10^{-8}}{44} \approx 10^{-10}.$$

Чтобы обнаружить взвешиванием изменение массы вещества при химических реакциях, необходима относительная точность измерений не менее 10^{-10}.

Энергетический выход ядерных реакций в миллионы раз больше, чем химических реакций. На ядерных реакциях соотношение между массой и энергией (13.2) подтверждено экспериментально. В ядерной физике и физике элементарных частиц это соотношение играет исключительную роль.

§ 17. О ЗАКОНАХ ТРЕНИЯ

1. В механике обычно имеют дело с силами всемирного тяготения, упругими силами и силами трения. Иногда в механику включаются также задачи на движение электрически заряженных частиц в электрических и магнитных полях. Тогда к указанным силам добавляются еще электромагнитные силы, т. е. силы, которым подвер-

жены заряженные частицы со стороны таких полей. Силы всемирного тяготения и упругие силы будут подробно рассмотрены в дальнейшем. Сейчас же мы кратко рассмотрим *силы трения.* Заметим при этом, что механика не занимается изучением физической природы действующих сил — эти вопросы рассматриваются в других разделах физики. Поэтому мы здесь совсем не будем затрагивать вопрос о происхождении сил трения, а ограничимся описанием эмпирически найденных *законов трения.* Понятно, что эти вопросы при всей их важности не имеют все же того фундаментального значения, каким характеризуются, например, законы Ньютона, а сами законы трения являются приближениями, часто довольно грубыми.

Упругие силы, силы всемирного тяготения, а также силы притяжения и отталкивания электрически заряженных тел зависят только от *конфигурации тел,* т. е. от их взаимного расположения, но не от их скоростей. Силы трения, помимо конфигурации, зависят еще *от относительных скоростей тел,* между которыми они действуют.

Силы трения могут действовать между соприкасающимися телами или их частями как при их относительном движении, так и при их относительном покое. Трение называется *внешним,* если оно действует между различными соприкасающимися телами, не образующими единого тела (например, трение между бруском и наклонной плоскостью, на которой он лежит или с которой он соскальзывает). Если же трение проявляется между различными частями одного и того же тела, например между различными слоями жидкости или газа, скорости которых непрерывно меняются от слоя к слою, то трение называется *внутренним.* Впрочем, разделение трения на внутреннее и внешнее носит условный характер. Если соприкасающиеся тела объединить в одну механическую систему, то трение, которое ранее рассматривалось как внешнее, становится внутренним. Трение, испытываемое твердым телом при движении в жидкости (или газе), есть внутреннее трение в жидкости, а не внешнее трение между жидкостью и твердым телом. Действительно, опыт показывает, что слои жидкости или газа, непосредственно примыкающие к поверхности тела, прилипают к ней и движутся вместе с телом, а трение возникает в окружающей среде между различными слоями ее, соприкасающимися друг с другом.

Трение между поверхностью твердого тела и окружающей его жидкой или газообразной средой, в которой оно движется, а также трение между различными слоями такой среды, называется *вязким.*

Трение между поверхностями двух соприкасающихся твердых тел при отсутствии между ними жидкой или газообразной прослойки (смазки) называется *сухим.* Применительно к этому случаю, когда соприкасающиеся тела движутся друг относительно друга, различают *трение скольжения* и *трение качения.*

2. Рассмотрим сначала законы сухого трения. Такое трение возникает не только при скольжении одного тела по поверхности другого, но и при всякой попытке вызвать такое скольжение. В последнем слу-

чае трение называется *трением покоя* или *трением сцепления.* Наличие трения покоя — характерная особенность сухого трения. В более общем смысле, безотносительно к тому, между какими телами возникает трение, оно называется *сухим,* если силы трения не исчезают при обращении в нуль относительных скоростей соприкасающихся тел. В противоположном случае трение называется *жидким.* Положим тяжелый брусок на поверхность горизонтального стола (рис. 27). В состоянии покоя вес бруска P уравновешен силой нормального давления f_n, с которой на брусок действует стол ($P = f_n$). Приложим затем к бруску горизонтальную силу f, лежащую в вертикальной плоскости, проходящей через его центр масс, как можно ближе к поверхности стола, чтобы предотвратить опрокидывание бруска, когда он придет в движение. Опыт показывает, что если сила f не превосходит некоторой определенной величины f_0 ($f < f_0$), то брусок не приходит в движение. Отсюда следует сделать вывод, что на брусок со стороны стола действует равная и противоположно направленная сила $f_{тр}$, уравновешивающая силу f. Это и есть сила трения, а именно трения покоя. Такая же сила трения, но в противоположном направлении, действует на поверхность стола со стороны бруска. Сила трения покоя автоматически принимает значения, равные внешней силе f. Максимальное значение силы трения покоя равно f_0.

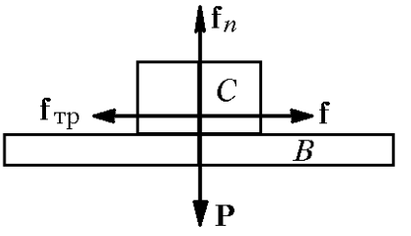


Рис. 27

Допустим теперь, что брусок скользит по поверхности стола со скоростью v. При равномерном движении действующая сила f по-прежнему уравновешивается силой $f_{тр}$. Если равновесия нет, то движение будет ускоренным. В обоих случаях сила трения $f_{тр}$, вообще говоря, зависит от скорости v. Характер этой зависимости графически изображен на схематическом рис. 28. Сила трения, приложенная к поверхности бруска, всегда действует против направления движения последнего. На графике это отражено тем, что знаки величин $f_{тр}$ и v всегда противоположны. При $v = 0$ график вырождается в отрезок вертикальной прямой. Этому соответствует тот факт, что сила трения покоя может принимать любое значение от $-f_0$ до $+f_0$. При увеличении модуля скорости модуль силы трения сначала убывает, проходит через минимум, а затем начинает возрастать. Вся кривая симметрична относительно начала координат [$f(+v) = -f(-v)$]. Как экспериментально установил Кулон, сила трения $f_{тр}$ не зависит от площади поверхности, вдоль которой тела

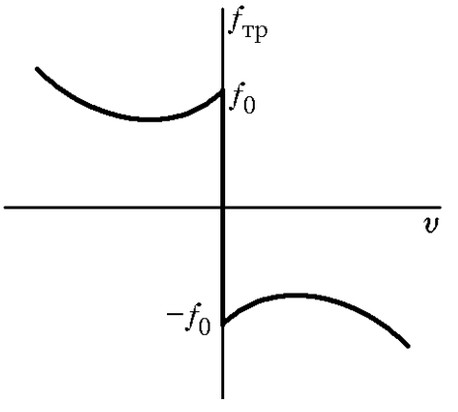


Рис. 28

соприкасаются, и пропорциональна силе нормального давления f_n, с которой одно тело действует на другое. Поэтому можно написать

$$f_{\text{тр}} = \mu f_n. \tag{17.1}$$

Постоянная μ называется *коэффициентом трения* и зависит от природы и состояния трущихся поверхностей. Если тело действительно скользит по поверхности другого тела, то μ называют *коэффициентом трения скольжения.* Если же тела покоятся друг относительно друга, то его называют *коэффициентом трения покоя.* В последнем случае предполагается, что в формуле $f_{\text{тр}}$ равно f_0, т. е. максимальному значению, которое может принимать сила трения покоя. В соответствии с рис. 28 коэффициент трения μ, вообще говоря, зависит от скорости v. Впрочем, как также установил Кулон, эта зависимость, как правило, выражена слабо, так что, когда не требуется большая точность, коэффициент μ можно считать не зависящим от скорости. Тогда кривая рис. 28 вырождается в кривую рис. 29. Сила трения и в этом идеализированном случае зависит от скорости v, поскольку при переходе скорости через нуль она меняет знак, а при $v = 0$ становится неопределенной. Во всех задачах на силы трения, приводимых в конце этого параграфа, предполагается, что μ не зависит от v.

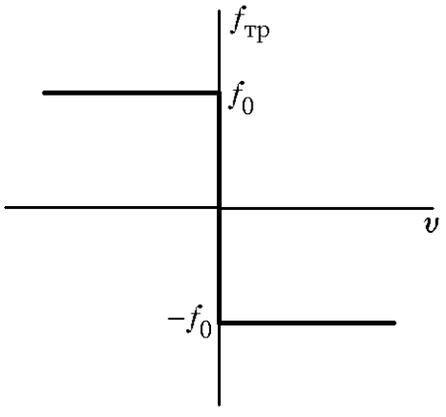


Рис. 29

Независимость силы трения покоя от площади соприкосновения тел можно демонстрировать с помощью следующего опыта. Брусок (например, кирпич), имеющий силу прямоугольного параллелепипеда, кладется различными гранями на наклонную плоскость. Увеличивая угол наклона наклонной плоскости, на опыте убеждаемся, что скольжение начинается при одном и том же угле, независимо от того, какой гранью брусок был положен на наклонную плоскость.

3. При наличии сухого трения тело может находиться в состоянии покоя, даже если на него подействовать какой-либо силой. Если приложенная сила f не превосходит максимального значения f_0 силы трения покоя, то тело не придет в движение. С этой особенностью сухого трения связано так называемое *явление застоя.* Допустим, например, что тело A, лежащее на поверхности горизонтального стола, находится в равновесии, когда обе прикрепленные к нему пружины не растянуты одинаково (рис. 30). В этом положении сила, действующая на тело A, равна нулю. Сместим тело A из положения равновесия в ту или другую сторону. Если сила f, действующая со стороны растянутых пружин, не превосходит f_0, то тело

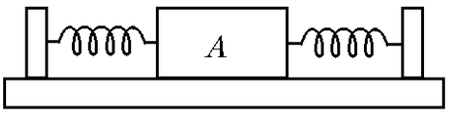


Рис. 30

A и в новом положении останется в равновесии. На поверхности стола не существует определенного положения равновесия тела. Напротив, существует область, при смещении в пределах которой тело остается в равновесии. Эта область называется *областью застоя.* Сухое трение, действующее, например, в подшипниках измерительных приборов со стрелками, ограничивает чувствительность таких приборов. Наличие же области застоя делает неопределенным положение равновесия, в котором устанавливается стрелка при измерениях, т. е. ограничивает точность измерения.

4. Во многих случаях силы трения оказываются полезными. Так, автомобиль приводится в движение силами трения, действующими между шинами колес и полотном дороги. Силы трения приводят в движение поезда. Силы трения, возникающие между приводным ремнем и шкивами, осуществляют передачу движения от одного маховика к другому. Подобных примеров можно привести неограничено много. Но сплошь и рядом силы трения являются вредными. Таковы, например, силы трения, возникающие между осью и втулкой, а также между другими деталями машины. Они приводя к преждевременному износу машин, и с ними приходится бороться. Для этой цели применяется *смазка.* Однако более радикальным способом уменьшения силы трения является замена трения скольжения *трением качения* (шарикоподшипники). Под трением качения понимают трение, возникающее, например, между шарообразным или цилиндрическим телом, катящимся без скольжения по плоской или изогнутой поверхности. Трение качения формально подчиняется тем же законам, что и трение скольжения. Однако коэффициент трения при качении значительно меньше, чем при скольжении. Наиболее радикальным способом уменьшения силы трения, который за последнее время начинает получать все большее и большее распространение, является создание «воздушной подушки» между соприкасающимися поверхностями.

5. Как уже говорилось выше, в отличие от сил сухого трения силы жидкого или вязкого трения обращаются в нуль вместе с относительными скоростями между соприкасающимися слоями среды. Более подробно вопрос о вязком трении будет разобран в механике жидкостей и газов, а также в кинетической теории газов. Здесь же мы очень кратко рассмотрим только силы жидкого трения, возникающие при движении твердого тела в жидкой или газообразной среде. Помимо сил, обусловленных собственно внутренним трением, на поверхность движущегося тела со стороны среды действуют также силы нормального давления. Результирующая этих нормальных давлений имеет составляющую, направленную против движения тела. Такая составляющая называется *силой сопротивления среды.* При больших скоростях она во много раз превосходит силы сопротивления, обусловленные собственно вязким трением. При рассмотрении движения тела в вязкой среде эти две силы целесообразно объединить вместе. Такую суммарную силу, направленную против

скорости движущегося тела, условно будем называть также *силой трения* и обозначать символом $\mathbf{f}_{тр}$.

При малых скоростях сила $\mathbf{f}_{тр}$ пропорциональна первой степени скорости тела:

$$\mathbf{f}_{тр} = -k_1\mathbf{v}. \tag{17.2}$$

При возрастании скорости зависимость становится более сложной, а затем сила трения начинает возрастать приблизительно пропорционально квадрату скорости:

$$\mathbf{f}_{тр} = -k_2 v^2 \frac{\mathbf{v}}{v} = -k_2 v\mathbf{v}. \tag{17.3}$$

Коэффициенты k_1 и k_2, а также область скоростей, в которой осуществляется переход от линейного закона (17.2) к квадратичному (17.3), в сильной степени зависят от формы и размеров тела, направления его движения, состояния поверхности тела и от свойств окружающей среды. Искусственно увеличивая поверхность тела и придавая ей надлежащую форму, можно сильно увеличить значения коэффициентов k_1 и k_2. На этом основано устройство и действие парашюта (см. задачу 8 к этому параграфу).

ЗАДАЧИ

1. Удобный метод измерения коэффициента трения покоя состоит в следующем. Тело кладется на наклонную плоскость. Измеряется минимальный угол наклона плоскости α, при котором начинается скольжение. Найти связь между углом α и коэффициентом трения μ.

Ответ: $\mu = \operatorname{tg}\alpha$.

2. Человек может, хотя и медленно, привести в движение тяжелую баржу на воде, если он будет тянуть за канат, привязанный к ней. Но он не в состоянии сделать это с тяжелым телом, лежащим на земле, если даже вес этого тела заметно меньше веса баржи. Почему?

3. Положите стержень в горизонтальном положении на указательные пальцы ваших рук. Вначале пальцы должны быть разведены, а стержень должен лежать на них своими концами. Затем приближайте пальцы друг к другу. Около какой точки стержня они сойдутся? После этого пальцы снова разведите так, чтобы стержень все время лежал на них в горизонтальном положении. В каком месте будет находиться один из пальцев, когда другой достигнет конца стержня? Произведите этот опыт и объясните наблюдаемые явления.

4. Шофер, едущий на автомобиле по горизонтальной площади в тумане, внезапно заметил недалеко впереди себя стену, перпендикулярную к направлению движения. Что выгоднее: затормозить или повернуть в сторону, чтобы предотвратить аварию?

Ответ. Затормозить.

5. Автомобиль движется «равномерно» вдоль извилистой горизонтальной дороги. Принимая дорогу за синусоиду, найти максимальную скорость, которую может развить автомобиль, чтобы не было заноса.

Р е ш е н и е. Если автомобиль движется по криволинейной траектории «равномерно», то его ускорение a будет только нормальным. Это ускорение создается силой трения покоя между колесами автомобиля и полотном дороги: $f_{тр} = ma$. Если скорость автомобиля превзойдет определенный предел, то для удержания автомобиля на требуемой траектории, где кривизна ее велика, максимальной силы трения f_0 будет недостаточно ($f_0 < ma$). Автомобиль начнет скользить в направлении нормали к траектории. При этом в соответствии с графиком рис. 28 сила трения скольжения уменьшится, что приведет к дальнейшему боковому смещению автомобиля с траектории. В этом и состоит явление заноса. При движении по синусоиде нормальное ускорение максимально в ее вершинах, где кривизна кривой максимальна. Если $y = y(x)$ — уравнение синусоиды, то в вершинах $y' = 0$, и радиус кривизны в этих точках можно вычислить по формуле $1/R = |y''|$. Имея все это в виду и записав уравнение синусоиды $y = A \sin \frac{2\pi x}{l}$ (амплитуда A и пространственный период l постоянны), нетрудно получить условие, при котором заноса не будет:

$$v < \frac{l}{2\pi} \sqrt{\frac{\mu g}{A}},$$

где μ — коэффициент трения, g — ускорение свободного падения.

6. Решить ту же задачу, предполагая, что автомобиль движется равномерно по эллипсу с полуосями A и B. В каких точках траектории нормальное ускорение автомобиля достигает максимального и минимального значений? Найти эти значения.

О т в е т. $a_{макс} = \frac{Av^2}{B^2}$, $a_{мин} = \frac{Bv^2}{A^2}$. Заноса не будет при условии $v < B\sqrt{\frac{\mu g}{A}}$.

7. Предполагая, что обе пружины на рис. 30 одинаковы, найти для тела A размеры области застоя.

О т в е т. Центр основания тела A может находиться в равновесии в любой точке в пределах области

$$-\frac{\mu P}{2k} < x < +\frac{\mu P}{2k},$$

где P — вес тела, μ — коэффициент трения, k — жесткость пружины (одной). За начало координат принят центр области застоя.

8. Парашютист совершает затяжной прыжок. Считая массу парашютиста m равной 70 кг, найти установившиеся скорости его падения без парашюта и с раскрытым парашютом. Для человеческого тела при падении без парашюта коэффициент $k_2 = 2$ г/см. При раскрытом парашюте этот коэффициент возрастает примерно в 100 раз, т. е. составляет приблизительно 200 г/см.

Р е ш е н и е. Установившаяся скорость падения найдется из условия, чтобы вес человека $P = mg$ уравновешивался силой трения. Это дает

$$v = \sqrt{\frac{mg}{k_2}} \approx \begin{cases} 60 \text{ м/с без парашюта;} \\ 6 \text{ м/с с раскрытым парашютом.} \end{cases}$$

ГЛАВА III

НЕКОТОРЫЕ СЛЕДСТВИЯ И ПРИМЕНЕНИЯ ЗАКОНОВ НЬЮТОНА

§ 18. ИМПУЛЬС СИЛЫ И ИЗМЕНЕНИЕ ИМПУЛЬСА

1. Как было показано в § 12, производная импульса $\mathbf{p}$ системы материальных точек по времени определяется уравнением

$$\frac{d\mathbf{p}}{dt} = \mathbf{F}^{(e)}, \tag{18.1}$$

где $\mathbf{F}^{(e)}$ — геометрическая сумма всех внешних сил, действующих на систему. Внутренние силы не входят в это уравнение из-за третьего закона Ньютона. В случае одной материальной точки уравнение (18.1) переходит в уравнение, выражающее второй закон Ньютона. Допустим, что сила $\mathbf{F}^{(e)}$ постоянна. Тогда из уравнения (18.1) следует

$$\mathbf{p} - \mathbf{p}_0 = \mathbf{F}^{(e)}(t - t_0), \tag{18.2}$$

где векторы $\mathbf{p}$ и $\mathbf{p}_0$ означают импульсы системы в моменты времени t и t_0 соответственно.

Произведение постоянной силы $\mathbf{F}^{(e)}$ на время ее действия называется *импульсом силы* за то же время. Это понятие нельзя смешивать с ранее введенной величиной $\mathbf{p} = m_1\mathbf{v}_1 + \ldots + m_n\mathbf{v}_n$, которая называется *импульсом системы материальных точек* или *импульсом тела.* Недоразумений возникнуть не может, так как слово «импульс» отдельно нигде встречаться не будет. Оно будет входить в комбинации либо со словом «сила», либо со словом «тело» (или «материальная точка» и «система материальных точек»). Поэтому всякий раз будет ясно, о каком импульсе идет речь. Впрочем, понятие импульса силы будет встречаться сравнительно редко.

2. Соотношение (18.2) означает, что приращение импульса тела или системы тел равно импульсу геометрической суммы всех внешних сил, действующих на систему. Этот результат получен нами в предположении, что сила $\mathbf{F}^{(e)}$ постоянна. Он может быть обобщен и на тот случай, когда эта сила меняется во времени. Разделим промежуток времени $t - t_0$ на более мелкие промежутки $(t_1 - t_0)$, $(t_2 - t_1)$, ..., $(t - t_{n-1})$ (рис. 31). Выберем эти промежутки

Рис. 31

настолько малыми, чтобы на каждом из них силу $\mathbf{F}^{(e)}$ без большой ошибки можно было считать приблизительно постоянной. Соответствующие значения силы $\mathbf{F}^{(e)}$ на таких промежутках обозначим через $\mathbf{F}_1^{(e)}, \ldots, \mathbf{F}_n^{(e)}$. Тогда на основании соотношения (18.2) можно написать приближенно

$$\mathbf{p}_1 - \mathbf{p}_0 = \mathbf{F}_1^{(e)}(t_1 - t_0),$$

$$\mathbf{p}_2 - \mathbf{p}_1 = \mathbf{F}_2^{(e)}(t_2 - t_1),$$

$$\ldots\ldots\ldots\ldots\ldots$$

$$\mathbf{p} - \mathbf{p}_{n-1} = \mathbf{F}_n^{(e)}(t - t_{n-1}),$$

где $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_{n-1}$ — импульсы системы в моменты времени $t_1, t_2, \ldots, t_{n-1}$ соответственно. Складывая эти равенства, получим

$$\mathbf{p} - \mathbf{p}_0 = \sum_i \mathbf{F}_i^{(e)} \Delta t_i,$$

где использовано стандартное обозначение $\Delta t_i = t_i - t_{i-1}$. Последнее равенство является приближенным и не совсем определенным, поскольку значения внешней силы $\mathbf{F}_1^{(e)}, \mathbf{F}_2^{(e)}, \ldots, \mathbf{F}_n^{(e)}$ не фиксированы точно. Однако эта неопределенность устраняется и указанное равенство переходит в точное соотношение, если перейти к пределу, устремляя к нулю наибольший из промежутков времени Δt_i при неизменной длине временного интервала $t - t_0$. В результате такого предельного перехода получится

$$\mathbf{p} - \mathbf{p}_0 = \lim_{\Delta t \to 0} \sum_i \mathbf{F}_i^{(e)} \Delta t_i.$$

Как известно, предел, стоящий в правой части этого равенства, называется определенным интегралом функции $\mathbf{F}^{(e)}(t)$ в пределах от t_0 до t:

$$\int_{t_0}^{t} \mathbf{F}^{(e)}(\tau)\, d\tau \equiv \lim_{\Delta t_i \to 0} \sum_i \mathbf{F}_i^{(e)} \Delta t_i.$$

Аргумент функции $\mathbf{F}^{(e)}$, по которому производится интегрирование, обозначен через τ, чтобы не смешивать его с верхним пределом интеграла t. Величина τ называется *переменной интегрирования.* Значение определенного интеграла при заданной подынтегральной функции определяется только значениями пределов интегрирования t_0 и t.

Таким образом, обобщением соотношения (18.2) является формула

$$\mathbf{p} - \mathbf{p}_0 = \int_{t_0}^{t} \mathbf{F}^{(e)}(\tau)\, d\tau. \qquad (18.3)$$

Интеграл, стоящий в правой части этого равенства называется *импульсом силы* $\mathbf{F}^{(e)}$ *за время от* t_0 *до* t. Следовательно, и в случае силы, меняющейся во времени, *приращение импульса системы материальных точек равно импульсу геометрической суммы всех действующих на нее внешних сил.* Внутренние силы, как уже подчеркивалось выше, не влияют на изменение полного импульса системы, поскольку они всегда входят попарно и удовлетворяют принципу равенства действия и противодействия.

3. Импульс, приобретаемый телом, зависит, таким образом, не только от силы, но и от продолжительности ее действия. Иллюстрацией этого может служить следующий простой опыт. Тяжелая гиря (рис. 32) подвешена на нити, снизу к ней прикреплена такая же нить. Если медленно тянуть за нижнюю нить, то рвется верхняя нить. Причина ясна. Так как гиря практически все время находится в покое, разность натяжений нити $T_1 - T_2$ должна уравновешивать вес груза P: $T_1 - T_2 = P$. Отсюда следует $T_1 > T_2$. Обозначим символом T_0 максимальное натяжение, которое может выдержать нить, не разрываясь. Когда мы медленно тянем за нижнюю нить, то в некоторый момент времени натяжение T_1 достигает предельной величины T_0. В этот момент натяжение нижней нити T_2 еще меньше T_0. Поэтому нижняя нить остается целой, а верхняя рвется. Однако если быстро дернуть за нижнюю нить, то верхняя нить остается целой, а нижняя рвется. Дело в том, что для разрыва верхней нити ее необходимо растянуть на определенную длину. А для этого надо привести в движение гирю. Чтобы сообщить гире необходимое смещение, требуется конечное время, даже когда на нее действует большая сила. Быстро дергая за нижнюю нить, мы не успеваем сообщить гире достаточное смещение. В нижней нити возникает натяжение, превосходящее предельное T_0, в то время как верхняя нить еще не успевает растянуться, и ее натяжение практически остается неизменным. Поэтому и рвется нижняя нить.

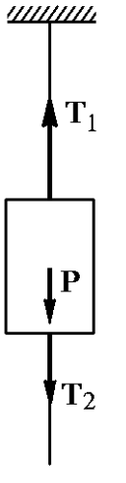


Рис. 32

Опишем второй опыт, иллюстрирующий влияние продолжительности действия силы. Из ватманской бумаги вырезаются два одинаковых кольца с наружным диаметром около 20 см и внутренним диаметром около 15 см. Кольца подвешиваются на двух горизонтальных металлических стержнях, зажатых в штативах. В кольца вставляется четырехугольная сосновая планка длиной 1 м с поперечным сечением 2—3 см^2. Расстояние между бумажными кольцами должно быть лишь немного меньше длины планки. Если плавно нажимать на середину планки, то одно из бумажных колец (или оба вместе), рвется, а планка остается целой. Нанесем теперь по середине планки резкий сильный удар тяжелым металлическим стерж-

нем. Планка ломается, а кольца остаются целыми. Поразительным в этом опыте является не то, что ломается планка, — она переломилась бы и при отсутствии колец, а то, что остаются целыми бумажные кольца.

§ 19. ТЕОРЕМА О ДВИЖЕНИИ ЦЕНТРА МАСС

В нерелятивистской механике, ввиду независимости массы от скорости, импульс системы $\mathbf{p} = m_1\mathbf{v}_1 + m_2\mathbf{v}_2 + \ldots$ может быть выражен через скорость ее *центра масс. Центром масс* или *центром инерции системы* называется такая воображаемая точка, радиус-вектор $\mathbf{R}$ которой выражается через радиусы-векторы $\mathbf{r}_1$, $\mathbf{r}_2$, ... материальных точек по формуле

$$\mathbf{R} = \frac{m_1\mathbf{r}_1 + m_2\mathbf{r}_2 + \ldots}{m}, \tag{19.1}$$

где $m = m_1 + m_2 + \ldots$ — общая масса всей системы. Эту точку мы обычно будем обозначать буквой C.

Если продифференцировать выражение (19.1) по времени и умножить на m, то получится

$$m\dot{\mathbf{R}} = m_1\dot{\mathbf{r}}_1 + m_2\dot{\mathbf{r}}_2 + \ldots,$$

или

$$m\mathbf{V} = m_1\mathbf{v}_1 + m_2\mathbf{v}_2 + \ldots,$$

где $\mathbf{V} = \dot{\mathbf{R}}$ — скорость центра масс системы. Таким образом,

$$\mathbf{p} = m\mathbf{V}. \tag{19.2}$$

Подставив это выражение в формулу (18.1), получим

$$m\frac{d\mathbf{V}}{dt} = \mathbf{F}^{(e)}. \tag{19.3}$$

Отсюда следует, что *центр масс системы движется как материальная точка, масса которой равна суммарной массе всей системы, а действующая сила — геометрической сумме всех внешних сил, действующих на систему.* Этот результат называется *теоремой о движении центра масс.*

Примером может служить движение снаряда по параболе в безвоздушном пространстве. Если в какой-либо момент времени снаряд разорвется на мелкие осколки, то эти осколки под действием внутренних сил будут разлетаться в разные стороны. Однако центр масс осколков и газов, образовавшихся при взрыве, будет продолжать свое движение по параболической траектории, как если бы никакого взрыва не было.

Центр масс системы совпадает с ее *центром тяжести,* т. е. с точкой приложения параллельных сил, действующих на матери-

альные точки системы в однородном поле тяжести. Поэтому вместо терминов «центр масс» и «центр инерции» употребляют также термин «центр тяжести». Однако в теореме о движении центра масс термином «центр тяжести» лучше не пользоваться, так как к этой теореме тяжесть не имеет прямого отношения. Термин «центр тяжести» распространен в курсах теоретической механики, особенно старых. В физике этот термин практически вышел из употребления.

Если система замкнута, то $\mathbf{F}^{(e)} = 0$. В этом случае уравнение (19.3) переходит в $\frac{d\mathbf{V}}{dt} = 0$, из которого следует $\mathbf{V} = \text{const}$. Центр масс замкнутой системы движется прямолинейно и равномерно.

ЗАДАЧИ

1. На дне маленькой запаянной пробирки, подвешенной над столом на нити, сидит муха, масса которой равна массе пробирки, а расстояние от дна до поверхности стола равно длине пробирки l. Нить пережигают, и за время падения муха перелетает со дна в самый верхний конец пробирки. Определить время, по истечении которого нижний конец пробирки стукнется о стол.

Ответ. $t = \sqrt{l/g}$.

2. Металлическое кольцо, подвешенное на нити к оси центробежной машины, как указано на рис. 33, равномерно вращается с угловой скоростью ω настолько быстро, что его плоскость вращения практически горизонтальна. Нить составляет угол α с осью. Найти расстояние от центра кольца до оси вращения.

Ответ. $x = \frac{g \operatorname{tg} \alpha}{\omega^2}$.

3. Однородный стержень длины l равномерно вращается вокруг свободной оси, перпендикулярной к стержню и проходящей через его центр. Какова должна быть угловая скорость вращения ω, при которой стержень еще не разрывается под действием внутренних напряжений, возникающих в нем

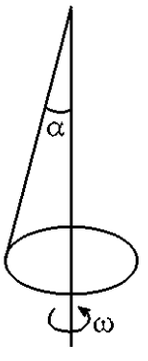


Рис. 33

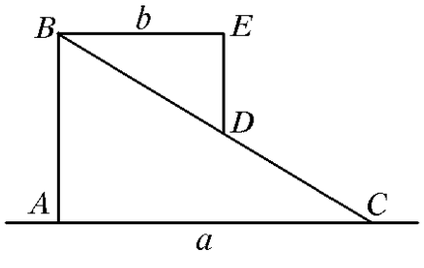


Рис. 34

при вращении? Максимальная сила натяжения, отнесенная к единице площади поперечного сечения стержня, равна T. Объемная плотность материала стержня равна ρ (см. также § 75, задача 4).

Ответ. $\rho l^2 \omega^2 < 8T$.

4. На прямоугольный трехгранный клин ABC массы M, лежащий на абсолютно гладкой горизонтальной плоскости, положен подобный же, но меньший клин BED массы m (рис. 34). Определить, на какое расстояние x сместится влево большой клин, когда малый клин соскользнет вниз и займет такое положение, что точка D совместится с C. Длины катетов AC и BE равны соответственно a и b.

Ответ. $x = \dfrac{m}{M+m}(a-b)$.

§ 20. ПРИВЕДЕННАЯ МАССА

1. Рассмотрим замкнутую систему, состоящую из двух взаимодействующих материальных точек с массами m_1 и m_2 (рис. 35). Уравнения движения этих точек можно записать в виде

$$\frac{d^2\mathbf{r}_1}{dt^2} = \frac{\mathbf{F}_1}{m_1}, \quad \frac{d^2\mathbf{r}_2}{dt^2} = \frac{\mathbf{F}_2}{m_2}, \tag{20.1}$$

причем по третьему закону Ньютона $\mathbf{F}_1 = -\mathbf{F}_2$. Вычитая из одного уравнения другое, находим

$$\frac{d^2}{dt^2}(\mathbf{r}_2 - \mathbf{r}_1) = \frac{\mathbf{F}_2}{m_2} - \frac{\mathbf{F}_1}{m_1} = \mathbf{F}_2\left(\frac{1}{m_1} + \frac{1}{m_2}\right).$$

Это уравнение описывает движение одной материальной точки относительно другой, так как разность $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$ есть радиус-вектор, проведенный от первой точки ко второй. Он однозначно определяет положение второй точки относительно первой. Введем обозначение

$$\frac{1}{\mu} = \frac{1}{m_1} + \frac{1}{m_2}, \quad \text{или} \quad \mu = \frac{m_1 m_2}{m_1 + m_2}. \tag{20.2}$$

Тогда предыдущее уравнение перейдет в следующее:

$$\mu \frac{d^2\mathbf{r}}{dt^2} = \mathbf{F}_2. \tag{20.3}$$

Это уравнение формально аналогично второму закону Ньютона. Роль силы играет сила $\mathbf{F}_2$, действующая на вторую материальную точку, а роль массы — вспомогательная величина μ, называемая *приведенной массой*.

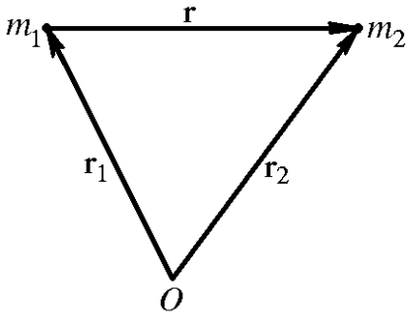


Рис. 35

Разумеется, одно уравнение (20.3) не может быть эквивалентно двум исходным уравнением (20.1). Однако такая эквивалентность может быть достигнута, если к уравнению (20.3) присоединить уравнение, выражающее теорему о движении центра масс системы. Последняя в рассматриваемом случае сводится к утверждению, что центр масс системы движется

прямолинейно и равномерно. Тем самым задача о движении двух материальных точек распадается на две независимые задачи: 1) определение равномерного движения центра масс; 2) определение относительного движения одной материальной точки относительно другой. Вторая задача формально сводится к задаче о движении одной материальной точки с массой μ в силовом поле другой точки. Этим и оправдывается введение понятия приведенной массы. Никакого глубокого физического смысла приведенная масса не имеет. На нее надо смотреть только как на целесообразное обозначение.

2. Рассмотрим пример, поясняющий пользу введения понятия приведенной массы. Пусть планета обращается вокруг Солнца по окружности радиуса r. Действующая на нее сила по закону всемирного тяготения равна $F = G\frac{Mm}{r^2}$, где M — масса Солнца, m — масса планеты, G — гравитационная постоянная. Так как сила направлена к Солнцу, то в векторной форме $\mathbf{F} = -G\frac{Mm}{r^2}\frac{\mathbf{r}}{r} = -G\frac{Mm}{r^3}r$. Вводя приведенную массу, запишем уравнение движения планеты относительно Солнца:

$$\mu\ddot{\mathbf{r}} \equiv \frac{Mm}{M+m}\ddot{\mathbf{r}} = \mathbf{F} = -G\frac{Mm}{r^3}\mathbf{r}.$$

Отсюда

$$\ddot{\mathbf{r}} = -G\frac{M+m}{r^3}\mathbf{r}.$$

Так как вращение планеты по орбите равномерно, то $\ddot{\mathbf{r}} = -\omega^2\mathbf{r}$, а потому

$$\omega^2 = \left(\frac{2\pi}{T}\right)^2 = G\frac{M+m}{r^3},$$

где ω — угловая скорость, а T — период обращения планеты. Если масса планеты пренебрежимо мала по сравнению с массой Солнца, то для угловой скорости ω_1 и периода обращения T_1 получаем

$$\omega_1^2 = \left(\frac{2\pi}{T_1}\right)^2 = G\frac{M}{r^3}.$$

Если бы масса планеты была бы равна массе Солнца (двойная звезда), то для угловой скорости ω_2 и периода обращения T_2 мы получили бы

$$\omega_2^2 = \left(\frac{2\pi}{T_2}\right)^2 = G\frac{2M}{r^3}.$$

При одном и том же расстоянии r

$$\left(\frac{\omega_2}{\omega_1}\right)^2 = \left(\frac{T_1}{T_2}\right)^2 = 2.$$

Период обращения во втором случае меньше, чем в первом, в $\sqrt{2}$ раз.

§ 21. ДВИЖЕНИЕ ТЕЛ С ПЕРЕМЕННОЙ МАССОЙ. РЕАКТИВНОЕ ДВИЖЕНИЕ

1. Термин *«переменная масса»* употребляется в этом параграфе в совершенно ином смысле, чем в теории относительности. В теории относительности масса движущегося тела изменяется за счет изменения его скорости, причем никакого вещества во время движения тело не получает и не теряет. Напротив, в настоящем параграфе говорится о медленном движении тел, масса которых меняется за счет *потери* или *приобретения вещества.* Например, масса автомобиля для поливки улиц уменьшается за счет вытекающих водяных струй; дождевая капля растет при падении в воздухе, пересыщенном водяными парами; масса ракеты или реактивного самолета уменьшается за счет истечения газов, образующихся при сгорании топлива. В таких случаях говорят о движении тел с переменной массой. Уравнения движения тел с переменной массой не содержат ничего принципиально нового по сравнению с законами Ньютона, а являются их следствиями. Тем не менее они представляют большой интерес, главным образом в связи с ракетной техникой.

2. Выведем уравнение движения материальной точки с переменной массой на примере движения ракеты. Принцип действия ракеты очень прост. Ракета с большой скоростью выбрасывает вещество (газы), воздействуя на него с большой силой. Выбрасываемое вещество с той же, но противоположно направленной силой в свою очередь действует на ракету и сообщает ей ускорение в противоположном направлении. Если нет внешних сил, то ракета вместе с выброшенным веществом является замкнутой системой. Импульс такой системы не может меняться во времени. На этом положении и основана теория движения ракет. Целесообразно, однако, обобщить задачу, предположив, что на ракету действуют внешние силы. Такими силами могут быть сила земной тяжести, гравитационное притяжение Солнца и планет, а также сила сопротивления среды, в которой движется ракета.

Пусть $m(t)$ — масса ракеты в произвольный момент времени t, а $\mathbf{v}(t)$ — ее скорость в тот же момент. Импульс ракеты в этот момент времени будет $m\mathbf{v}$. Спустя время dt масса и скорость ракеты получат приращения dm и $d\mathbf{v}$ (величина dm отрицательна!). Импульс ракеты станет равным $(m + dm) \times (\mathbf{v} + d\mathbf{v})$. Сюда надо добавить импульс движения газов, образовавшихся за время dt. Он равен $dm_{\text{газ}}\mathbf{v}_{\text{газ}}$, где $dm_{\text{газ}}$ — масса газов, образовавшихся за время dt, а $\mathbf{v}_{\text{газ}}$ — их скорость. Вычитая из суммарного импульса системы в момент $t + dt$ импульс системы в момент t, найдем приращение этой величины за время dt. Согласно известной теореме это приращение равно $\mathbf{F}\,dt$, где $\mathbf{F}$ — геометрическая сумма всех внешних сил, действующих на ракету. Таким образом,

$$(m + dm)\,(\mathbf{v} + d\mathbf{v}) + dm_{\text{газ}}\mathbf{v}_{\text{газ}} - m\mathbf{v} = \mathbf{F}\,dt.$$

Время dt, а с ним и приращения dm и $d\mathbf{v}$ мы должны устремить к нулю — нас интересуют предельные решения, или произвольные $\frac{dm}{dt}$ и $\frac{d\mathbf{v}}{dt}$. Поэтому, раскрывая скобки, можно отбросить произведение $dm \cdot d\mathbf{v}$, как бесконечно малую высшего порядка. Далее, ввиду сохранения массы, $dm + dm_{\text{газ}} = 0$. Пользуясь этим, можно исключить массу газов $dm_{\text{газ}}$. Наконец, разность $\mathbf{v}_{\text{отн}} = \mathbf{v}_{\text{газ}} - \mathbf{v}$ есть скорость истечения газов относительно ракеты. Мы будем называть ее *скоростью газовой струи.* С учетом этих замечаний предыдущее соотношение легко преобразуется к виду

$$m\, d\mathbf{v} = \mathbf{v}_{\text{отн}}\, dm + \mathbf{F}\, dt. \tag{21.1}$$

Отсюда делением на dt получаем

$$m \frac{d\mathbf{v}}{dt} = \mathbf{v}_{\text{отн}} \frac{dm}{dt} + \mathbf{F}. \tag{21.2}$$

По форме уравнение (21.2) совпадает с уравнением, выражающим второй закон Ньютона. Однако масса тела m здесь не постоянна, а меняется во времени из-за потери вещества. К внешней силе $\mathbf{F}$ добавляется дополнительный член $\mathbf{v}_{\text{отн}} \frac{dm}{dt}$, который может быть истолкован как *реактивная сила,* т. е. сила, с которой действуют на ракету вытекающие из нее газы. Уравнение (21.2) впервые было получено русским механиком И. В. Мещерским (1859—1935). Оно, так же как и эквивалентное ему уравнение (21.1), называется *уравнением Мещерского или уравнением движения точки с переменной массой.*

3. Применим уравнение (21.1) к движению ракеты, на которую не действуют никакие внешние силы. Полагая $\mathbf{F} = 0$, получим

$$m\, d\mathbf{v} = \mathbf{v}_{\text{отн}}\, dm.$$

Допустим, что ракета движется прямолинейно в направлении, противоположном скорости газовой струи $\mathbf{v}_{\text{отн}}$. Если направление полета принять за положительное, то проекция вектора $\mathbf{v}_{\text{отн}}$ на это направление будет отрицательной и равной $-v_{\text{отн}}$. Поэтому в скалярной форме предыдущее уравнение можно записать так: $m\, dv = -v_{\text{отн}}\, dm$, причем в соответствии с принятыми обозначениями величина $v_{\text{отн}}$ существенно положительна. Следовательно,

$$\frac{dv}{dm} = -\frac{v_{\text{отн}}}{m}. \tag{21.3}$$

Скорость газовой струи $v_{\text{отн}}$ может меняться во время полета. Однако простейшим и наиболее важным является случай, когда она постоянна. Предположение о постоянстве $v_{\text{отн}}$, очевидно, не затрагивает основные черты явления, но сильно облегчает решение уравнения (21.3). В этом случае

$$v = -v_{\text{отн}} \int \frac{dm}{m} = -v_{\text{отн}} \ln m + C.$$

Значение постоянной интегрирования C определяется начальными условиями. Допустим, что в начальный момент времени скорость ракеты равна нулю, а ее масса равна m_0. Тогда предыдущее уравнение дает $0 = -v_{\text{отн}} \ln m_0 + C$, откуда $C = v_{\text{отн}} \ln m_0$. Следовательно,

$$v = v_{\text{отн}} \ln \frac{m_0}{m}, \tag{21.4}$$

или

$$\frac{m_0}{m} = e^{v/v_{\text{отн}}}. \tag{21.5}$$

Последнее соотношение называется *формулой Циолковского* (по имени советского ученого К. Э. Циолковского (1857—1935)). Она получена нами для нерелятивистских движений, т. е. для тех случаев, когда обе скорости v и $v_{\text{отн}}$ малы по сравнению со скоростью света в вакууме c. Но ее можно обобщить на случай релятивистских движений. Если m_0 и m означают массы покоя ракеты в соответствующие моменты времени, то без вычислений ясно, что формула (21.5) дает заниженное значение для отношения m_0/m. Действительно, релятивистская масса возрастает со скоростью. Ввиду этого при одном и том расходе топлива «релятивистская» ракета достигнет меньшей скорости, чем получается по нерелятивистской формуле (21.5).

Релятивистская формула имеет вид

$$\frac{m_0}{m} = \left(\frac{1+\beta}{1-\beta}\right)^{c/2v_{\text{отн}}} \tag{21.6}$$

(см. задачу 2 к § 22). Здесь $\beta \equiv v/c$. При $\beta \ll 1$ и $v_{\text{отн}}/c \ll 1$ формула (21.6) переходит в формулу Циолковского. Действительно, в этом случае

$$\frac{1+\beta}{1-\beta} \approx 1 + 2\beta$$

и, следовательно

$$\frac{m_0}{m} \approx \left(1 + 2\beta\right)^{\frac{c}{2v}\frac{v}{v_{\text{отн}}}} = \left(1 + 2\beta\right)^{\frac{1}{2\beta}\frac{v}{v_{\text{отн}}}}.$$

Так как величина 2β мала, то

$$(1 + 2\beta)^{1/2\beta} \approx \lim_{\beta \to 0} (1 + 2\beta)^{1/2\beta} = e.$$

В результате в предельном случае медленных движений получаем

$$\frac{m_0}{m} = e^{v/v_{\text{отн}}},$$

т. е. формулу Циолковского.

4. Формула Циолковского позволяет рассчитать запас топлива, необходимый для сообщения ракете определенной скорости v. В табл. 1 приведены отношения начальной массы ракеты m_0 к ее конечной массе m при различных значениях отношения $v/v_{отн}$. Вычисления выполнены с помощью нерелятивистской формулы (21.5).

Допустим, например, что ракете надо сообщить *первую космическую скорость*, т. е. такую скорость, чтобы она начала двигаться вокруг Земли по окружности. Эта скорость $v \approx 8$ км/с. При скорости газовой струи $v_{отн} = 1$ км/с отношение масс $m_0/m = 2980$. Практически вся масса ракеты приходится на топливо. При $v_{отн} = 2$ км/с отношение $m_0/m = 54{,}6$, при $v_{отн} = 4$ км/с $m_0/m = 7{,}39$ и т. д. Отсюда видно, что относительная полезная масса ракеты очень быстро увеличивается с увеличением скорости газовой струи $v_{отн}$. Газы, выходящие из ракеты, должны иметь возможно меньшую молекулярную массу и быть нагреты до возможно более высокой температуры. Действительно, в молекулярной физике будет показано, что скорость газовой струи $v_{отн}$ пропорциональна $\sqrt{T/\mu}$, где T — абсолютная температура газа, а μ — его молекулярная масса.

Таблица 1

$v/v_{отн}$	m_0/m	$v/v_{отн}$	m_0/m	$v/v_{отн}$	m_0/m	$v/v_{отн}$	m_0/m
1	2,72	4	54,6	7	1100	10	22 000
2	7,39	5	148	8	2980	11	59 900
3	20,1	6	403	9	8100	12	163 000

В современных ракетах на химическом топливе скорость газовой струи порядка одного или нескольких километров в секунду. Вероятно, она не превосходит 4 км/с. Имея это в виду, оценим перспективы межпланетных и межзвездных полетов ракет на химическом топливе. Минимальная скорость, которую необходимо сообщить ракете относительно Земли, чтобы она вышла за пределы действия поля земного тяготения, называется *второй космической скоростью* и составляет 11,2 км/с. Практически такую скорость необходимо сообщить ракете, например, при отправке ее на Луну. Скорость ракеты, которую она должна приобрести относительно Земли, чтобы навсегда покинуть пределы Солнечной системы, называется *третьей космической скоростью*. Третья космическая скорость зависит от направления начальной скорости ракеты. Минимальное ее значение соответствует запуску ракеты по касательной к земной орбите в направлении орбитального вращения Земли. Эта скорость составляет около 16,7 км/с (см. § 61). Скорости такого порядка необходимы при межпланетных путешествиях. Допустим, что $v_{отн} = 4$ км/с. Тогда для достижения второй космической скорости отношение m_0/m должно составлять $m_0/m = e^{11{,}2/4} \approx 17$, а для достижения третьей $m_0/m = e^{16{,}7/4} \approx 64$. Оба отношения не очень велики. Однако надо принимать во внимание, что ракета должна иметь запас топлива для обратного возвращения на Землю, а также для ее торможения при посадке и для коррекции траектории. Поэтому отношение m_0/m (m — масса ракеты, вернувшейся обратно на Землю) должно быть значительно больше. Допустим, например, что поле тяготения и размеры второй планеты такие же, как у Земли. Тогда

при путешествии в прямом направлении в нашем примере должно быть $m_0/m' \approx 60$ (m' — масса ракеты, достигшей второй планеты). При обратном путешествии $m'/m \approx 60$, так что $m_0/m \approx 3600$. Таким образом, для осуществления межпланетных полетов запас топлива должен превышать массу космического корабля по меньшей мере в несколько тысяч раз. Технические трудности очень велики, но, по-видимому, все же преодолимы.

Но для межзвездных полетов ракеты на химическом топливе абсолютно непригодны. Возьмем, например, $v_{отн} = 10$ км/с, что для ракет на химическом топливе, по-видимому, превышает пределы возможного. (Если допустить, что газовая струя состоит из наиболее легкого вещества — атомарного водорода, то для достижения таких скоростей потребуется температура порядка 5000 °С.) Расстояния до звезд измеряются *световыми годами* — от ближайшей звезды свет идет до Земли около 4 лет. Поэтому для достижения даже ближайших звезд нужны космические корабли, скорости которых близки к скорости света c. В табл. 2 приведены значения отношения m/m_0 при различных значениях β, вычисленные по релятивистской формуле (21.6) и по формуле Циолковского (21.5) в предположении, что $v_{отн} = 10$ км/с. Таблица, между прочим, наглядно показывает, когда существенны релятивистские эффекты и формула Циолковского неприменима.

Таблица 2

$\beta = \frac{v}{c}$	m_0/m		$\beta = \frac{v}{c}$	m_0/m	
	по формуле (21.6)	по формуле (21.5)		по формуле (21.6)	по формуле (21.5)
0,001	$1{,}0690 \cdot 10^{13}$	$1{,}0686 \cdot 10^{13}$	0,25	$5{,}37 \cdot 10^{3327}$	$1{,}62 \cdot 10^{3257}$
0,01	$1{,}963 \cdot 10^{130}$	$1{,}942 \cdot 10^{130}$	1/3	$2{,}84 \cdot 10^{4515}$	$8{,}81 \cdot 10^{4342}$
0,1	$1{,}79 \cdot 10^{1307}$	$7{,}64 \cdot 10^{1302}$		,	

Допустим, что скорость космического корабля v должна составлять четверть скорости света ($\beta = 0{,}25$). Тогда должно быть $m_0/m \approx 5 \cdot 10^{3327}$. На каждую тонну полезного груза должно приходиться $5 \cdot 10^{3327}$ т топлива! Если полезная масса $m = 20$ т $= 2 \cdot 10^7$ г, то стартовая масса корабля должна быть $m_0 \approx 10^{3329}$ т $= 10^{3335}$ г! Обычно, когда имеют дело с очень большими величинами, их называют «астрономическими». В данном случае такое сравнение не годится — речь идет о величинах несравненно большего масштаба. Для сравнения приведем массы некоторых частиц и астрономических объектов:

масса электрона	$9{,}11 \cdot 10^{-25}$ г,	масса Солнца	$1{,}99 \cdot 10^{33}$ г,
масса протона	$1{,}67 \cdot 10^{-24}$ г,	масса Галактики	$3 \cdot 10^{44}$ г,
масса Земли	$5{,}98 \cdot 10^{27}$ г,	масса Метагалактики	10^{56} г.

Под *Метагалактикой понимают ту часть Вселенной*, которая доступна обнаружению методами современной оптической, радио- и гамма-астрономии. Масса Метагалактики превосходит массу электрона примерно в 10^{83} раза. Масса нашего фантастического корабля с топливом должна превосходить массу Метагалактики в 10^{3329} раз! Эти цифры превосходят всякое воображение. В масштабах нашего космического корабля Метага-

лактика выглядит несравненно более малым объектом, чем электрон в масштабах Метагалактики.

Вряд ли имеет смысл говорить о движении столь фантастически гигантского космического корабля относительно Метагалактики, имеющей по сравнению с ним ничтожные размеры. Вводить в рассмотрение объекты таких размеров и применять к ним обычные законы физики является недопустимой экстраполяцией. Кроме того, обычная теория движения ракет основана на предположении, что импульс, полученный от сгорания топлива в задней части ракеты, *практически мгновенно передается ракете в целом.* Это условие ввиду конечной скорости распространения взаимодействий не может выполняться для «ракет» достаточно больших размеров. Оно, разумеется, не выполняется для нашего фантастического «корабля». Наш пример доказывает только, что для межзвездных полетов ракеты на химическом топливе абсолютно непригодны.

Было бы неосторожным на основании изложенного сделать вывод, что звездные миры никогда не будут доступны земным космонавтам. Только отдаленное будущее покажет, возможно это или нет. Не собираясь входить в обсуждение этой фантастической проблемы, ограничимся следующими замечаниями. Для превращения ракеты в *звездолет* прежде всего необходимо повысить скорость струи $v_{\text{отн}}$, приблизив ее к скорости света. Идеальным был бы случай $v_{\text{отн}} = c$. Так было бы в *фотонной ракете,* в которой роль газовой струи должен играть световой пучок, излучаемый двигателем корабля в определенном направлении. Реактивная сила в фотонной ракете осуществлялась бы давлением света, оказываемым на корабль при излучении светового пуска. Превращение вещества в излучение постоянно происходит внутри звезд. Этот процесс осуществляется и на Земле и притом не только в лабораторных условиях, а в более крупном масштабе (взрывы атомных и водородных бомб). Возможно ли придать ему управляемый характер и использовать в фотонных ракетах — на этот вопрос отвечать преждевременно.

ЗАДАЧИ

1. Для лучшего уяснения закономерностей движения ракеты полезно рассмотреть мысленный случай, когда ракета выбрасывает вещество не непрерывно, а конечными дискретными порциями одной и той же массы Δm. Пусть при каждом выбрасывании порция вещества Δm получает одну и ту же скорость $v_{\text{отн}}$ относительно ракеты, направленную назад. Определить скорость ракеты v_N, которую она достигнет после N выбрасываний, если начальная масса ракеты равна m_0. Показать, что в предельном случае, когда $\Delta m \to 0$, $N \to \infty$, а произведение $N\Delta m$ остается постоянным, выражение для v_N переходит в формулу Циолковского. Ограничиться нерелятивистскими скоростями.

Р е ш е н и е. Пусть v_1, v_2, ... — скорость ракеты после 1-го, 2-го, ... выбрасываний. По закону сохранения импульса $(m_0 - \Delta m)v_1 + \Delta m \cdot w = 0$, где w — скорость выброшенной массы Δm после первого выбрасывания. Очевидно, $v_{\text{отн}} = v_1 - w$. Исключая w, получим

$$v_1 = \frac{\Delta m}{m_0} v_{\text{отн}}. \tag{21.7}$$

Найдем теперь v_2. В системе отсчета, движущейся со скоростью v_1, ракета перед вторым выбрасыванием неподвижна, а после второго выбрасывания приобретает скорость $v_2 - v_1$. Поэтому можно воспользоваться формулой (21.7), сделав в ней замену $m_0 \to m_0 - \Delta m$, $v_1 \to v_2 - v_1$. Это дает

$$v_2 - v_1 = \frac{\Delta m}{m_0 - \Delta m} v_{\text{отн}}.$$

Комбинируя это соотношение с (21.7), находим v_2. Продолжая этот процесс дальше, нетрудно получить

$$v_N = \left[\frac{\Delta m}{m_0} + \frac{\Delta m}{m_0 - \Delta m} + \ldots + \frac{\Delta m}{m_0 - (N-1)\,\Delta m}\right] v_{\text{отн}}.$$

В пределе, когда $\Delta m \to 0$, $N \to \infty$, $m_0 - (N-1)\,\Delta m \to m$, сумма, стоящая в квадратных скобках, переходит в интеграл, и мы получаем

$$v = v_{\text{отн}} \int_m^{m_0} \frac{dm'}{m'},$$

где m — конечная масса ракеты. После взятия интеграла получается формула Циолковского (21.5).

2. Найти связь между массой ракеты $m(t)$, достигнутой ею скоростью $v(t)$ и временем t, если ракета движется вертикально вверх в поле земного тяготения. Скорость газовой струи относительно ракеты $v_{\text{отн}}$ считать постоянной. Сопротивление воздуха и изменение ускорения свободного падения g с высотой не учитывать. Какую массу газов $\mu(t)$ должна ежесекундно выбрасывать ракета, чтобы оставаться неподвижной в поле тяготения?

Р е ш е н и е. Уравнение движения ракеты

$$m \frac{dv}{dt} = -v_{\text{отн}} \frac{dm}{dt} - mg$$

перепишем в форме

$$m \frac{d}{dt}(v + gt) = -v_{\text{отн}} \frac{dm}{dt}$$

или

$$\frac{d(v+gt)}{dm} = -\frac{v_{\text{отн}}}{m}.$$

Это уравнение имеет такой же вид, что и (21.3), если за неизвестное принять величину $v + gt$. Поэтому можно воспользоваться формулой (21.5), заменив в ней v на $v + gt$. Это дает

$$\frac{m_0}{m} = e^{(v+gt)/v_{\text{отн}}}, \quad v = v_{\text{отн}} \ln \frac{m_0}{m} - gt.$$

Величина μ, очевидно, равна $-dm/dt$. Она находится из условия, что для неподвижной ракеты $dv/dt = 0$, и равна

$$\mu = -\frac{dm}{dt} = \frac{m_0 g}{v_{\text{отн}}} e^{-gt/v_{\text{отн}}}.$$

3. Космический корабль движется с постоянной по модулю скоростью v. Для изменения направления его полета включается двигатель, выбрасывающий струю газа со скоростью $v_{\text{отн}}$ относительно корабля в направлении, перпендикулярном к его траектории. Определить угол α, на который повернется вектор скорости корабля, если начальная масса его m_0, а конечная m.

Решение. Ускорение корабля по модулю равно $\omega^2 r = \omega v$, причем $v = \text{const}$. Поэтому уравнение движения

$$m \frac{d\mathbf{v}}{dt} = \mathbf{v}_{\text{отн}} \frac{dm}{dt}$$

переходит в $mv\omega\, dt = -v_{\text{отн}}\, dm$. Замечая, что $d\alpha = \omega\, dt$ есть угол поворота за время dt, и интегрируя, получим

$$\alpha = \frac{v_{\text{отн}}}{v} \ln \frac{m_0}{m}.$$

4. Космический корабль, движущийся в пространстве, свободном от поля тяготения, должен изменить направление своего движения на противоположное, сохранив скорость по модулю. Для этого предполагаются два способа: 1) сначала затормозить корабль, а затем разогнать его до прежней скорости; 2) повернуть корабль, заставив его двигаться по дуге окружности, сообщая ему ускорение в поперечном направлении. При каком из этих двух способов потребуется меньшая затрата топлива? Скорость истечения газов относительно корабля считать постоянной и одинаковой в обоих случаях.

Ответ. Первый способ требует меньшей затраты топлива.

5. Определить коэффициент полезного действия ракеты, т. е. отношение кинетической энергии K, приобретенной ракетой, к количеству теплоты Q, выделившемуся при сгорании топлива. Скорость, достигнутая ракетой, $v = 9$ км/с. Теплота сгорания топлива $q = 4000$ ккал/кг, скорость выбрасываемых продуктов сгорания относительно ракеты $u = 3$ км/с.

Ответ. $\dfrac{K}{Q} = \dfrac{v^2}{2q(e^{v/u} - 1)} \approx 13\%$.

6. Продукты сгорания (газы) выбрасываются из ракеты со скоростью $u = 3$ км/с (относительно ракеты). Найти отношение ее кинетической энергии $K_{\text{рак}}$ к кинетической энергии продуктов сгорания $K_{\text{газ}}$ в момент достижения ракетой скорости $v_{\text{кон}} = 12$ км/с.

Решение. Приращение скорости ракеты $\mathbf{v}$ связано с изменением ее массы m соотношением $m\, d\mathbf{v} = \mathbf{v}_{\text{отн}}\, dm$. Переходя к скалярной форме и новым обозначениям, запишем его в виде $m\, dv = -u\, dm$, причем $dm = -dm_{\text{газ}}$, где $m_{\text{газ}}$ — масса выброшенных газов. Приращение кинетической энергии газов

$$dK_{\text{газ}} = -\frac{1}{2} v_{\text{газ}}^2\, dm = \frac{m v_{\text{газ}}^2}{2u}\, dv.$$

Подставив сюда $v_{\text{газ}} = v - u$ и воспользовавшись формулой Циолковского (21.5), получим

$$dK_{\text{газ}} = -\frac{m_0}{2u} (u - v)^2 e^{-v/u}\, dv,$$

или после интегрирования

$$K_{\text{газ}} = \frac{m_0 u^2}{2} (1 - e^{-x} - x^2 e^{-x}),$$

где для краткости введено обозначение $x = v_{\text{кон}}/u$. Кинетическая энергия ракеты

$$K_{\text{рак}} = \frac{1}{2} m v_{\text{кон}}^2 = \frac{1}{2} m_0 u^2 x^2 e^{-x}.$$

В результате находим

$$\eta \equiv \frac{K_{\text{рак}}}{K_{\text{газ}}} = \frac{x^2}{e^x - (1 + x^2)}.$$

При $x = 4$ отношение энергий $\eta = 45\%$.

7. С поверхности Луны стартует двухступенчатая ракета. При каком отношении масс первой (m_1) и второй (m_2) ступеней скорость контейнера с полезным грузом (массой m) получится максимальной? Скорость истечения газов u в двигателях обеих ступеней постоянна и одинакова. Отношения массы топлива к массе ступени равна для первой и второй ступеней соответственно α_1 и α_2. Отделение ступеней и контейнера производится без сообщения добавочных импульсов.

Р е ш е н и е. От действия силы тяжести Луны можно отвлечься. Сила тяжести уменьшает кинетическую энергию системы, но не влияет на условие максимума. Примем за единицу массы полную массу ракеты в момент старта. Тогда

$$m_1 + m_2 + m = 1. \tag{21.8}$$

После выгорания топлива в первой ступени масса системы уменьшится на $\alpha_1 m_1$. Если при этом будет достигнута скорость v_1, то по соотношению Циолковского

$$e^{v_1/u} = \frac{1}{(1 - \alpha_1) m_1 + m_2 + m}.$$

Масса $(1 - \alpha_1) m_1$ отделяется, и включается двигатель второй ступени. После выгорания топлива во второй ступени скорость ракеты возрастает еще на величину v_2, причем

$$e^{v_2/u} = \frac{m_2 + m}{(1 - \alpha_2) m_2 + m}.$$

В этом можно убедиться, если перейти в систему отсчета, в которой ракета в момент отделения первой ступени покоится. Полная достигнутая скорость найдется перемножением двух предыдущих соотношений и последующим логарифмированием. Исключая еще при этом массу m_2 с помощью соотношений (21.8), получим

$$\frac{v}{u} = \ln\,(1 - m_1) - \ln\,(1 - \alpha_1 m_1) - \ln\,[(1 - \alpha_2)\,(1 - m_1) + \alpha_2 m].$$

Здесь m и u играют роль постоянных параметров, а m_1 — роль аргумента, от которого зависит скорость v. Дифференцируя по m_1 и приравнивая производную нулю, получим условие максимума:

$$\frac{1}{m_1 - 1} + \frac{1}{\beta - m_1} + \frac{1}{\gamma + m} = 0, \tag{21.9}$$

где введены обозначения

$$\beta = \frac{1}{\alpha_1}, \quad \gamma = 1 + \frac{\alpha_2}{1-\alpha_2}\, m.$$

Условие (21.9) приводит к квадратному уравнению относительно m_1, решая которое, найдем

$$m_1 = 1 - \sqrt{1 + (\beta\gamma - \beta - \gamma)}\,.$$

Перед корнем взят минус, так как по смыслу задачи $0 < m_1 < 1$. С помощью (21.8) находим массу m_2, а затем искомое соотношение m_2/m_1. Возвращаясь при этом к прежним параметрам α_1 и α_2, получаем

$$\frac{m_2}{m_1} = \frac{\sqrt{\frac{\alpha_2}{\alpha_1}\frac{1-\alpha_1}{1-\alpha_2}} - \sqrt{m}}{1 - \sqrt{\frac{\alpha_2}{\alpha_1}\frac{1-\alpha_1}{1-\alpha_2}}\, m}\sqrt{m}\,. \tag{21.10}$$

Решение имеет смысл при выполнении условия

$$\frac{\alpha_2}{\alpha_1}\frac{1-\alpha_1}{1-\alpha_2}\, m < 1.$$

В реальных условиях, когда $m \ll 1$, а параметры α_1 и α_2 отличаются не очень сильно, это условие соблюдается. При $\alpha_1 = \alpha_2$ получается простая формула

$$\frac{m_2}{m_1} = \sqrt{m}\,. \tag{21.11}$$

ГЛАВА IV

РАБОТА И ЭНЕРГИЯ

§ 22. РАБОТА И КИНЕТИЧЕСКАЯ ЭНЕРГИЯ

1. *Работой силы* $\mathbf{F}$ *на перемещении* $d\mathbf{s}$ называется проекция F_s этой силы на направление перемещения, умноженная на само перемещение:

$$dA = F_s\, ds = F\, ds \cos \alpha, \tag{22.1}$$

где α — угол между векторами $\mathbf{F}$ и $d\mathbf{s}$ (рис. 36). Поскольку перемещение $d\mathbf{s}$ предполагается бесконечно малым, величина dA называется также *элементарной работой* в отличие от *работы на конечном перемещении.* Если воспользоваться понятием скалярного произведения, то можно сказать, что *элементарная работа* dA есть скалярное произведение силы $\mathbf{F}$ на перемещение $d\mathbf{s}$:

$$dA = (\mathbf{F}\, d\mathbf{s}). \tag{22.2}$$

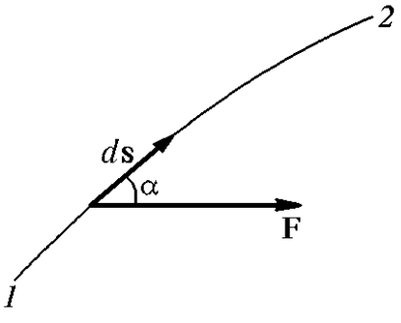


Рис. 36

В общем случае, когда материальная точка, двигаясь по криволинейной траектории, проходит путь конечной длины, можно мысленно разбить этот путь на бесконечно малые элементы, на каждом из которых сила $\mathbf{F}$ может считаться постоянной, а элементарная работа может быть вычислена по формуле (22.1) или (22.2). Если сложить все эти элементарные работы и перейти к пределу, устремив к нулю длины всех элементарных перемещений, а число их — к бесконечности, то такой предел обозначается символом

$$A = \int_L (\mathbf{F}\, d\mathbf{s}) \tag{22.3}$$

и называется *криволинейным интегралом* вектора $\mathbf{F}$ вдоль траектории L. Этот интеграл, по определению, и дает *работу силы* $\mathbf{F}$ *вдоль кривой* L.

Если $\mathbf{F} = \mathbf{F}_1 + \mathbf{F}_2$, то проецируя это векторное уравнение на направление элементарного перемещения $d\mathbf{s}$, получим $F_s = F_{1s} + F_{2s}$, а после умножения на ds: $F_s\, ds = F_{1s}\, ds + F_{2s}\, ds$, или

$$dA = dA_1 + dA_2. \tag{22.4}$$

Таким образом, *элементарная работа результирующей двух или нескольких сил равна сумме элементарных работ этих сил.* Очевидно, то же утверждение справедливо и для работ на конечных перемещениях:

$$A = A_1 + A_2. \tag{22.5}$$

Единицей работы в системе СИ является джоуль (Дж). *Джоуль* есть работа силы в один ньютон на перемещении в один метр при условии, что направление силы совпадает с направлением перемещения. В системе СГС единицей работы является эрг. *Эрг* есть работа силы в одну дину на перемещении в один сантиметр при том же условии, т. е. в предположении, что направления силы и перемещения совпадают. Очевидно,

$$1 \text{ Дж} = 10^7 \text{ эрг}.$$

Работа, отнесенная к единице времени, т. е. величина

$$P = \frac{dA}{dt}, \tag{22.6}$$

называется *мощностью.* Ее единицами являются эрг на секунду и джоуль на секунду, или ватт (Вт). Очевидно,

$$1 \text{ Вт} = 10^7 \text{ эрг/с}.$$

Подставив в формулу (22.3) $\mathbf{F} = \frac{d\mathbf{p}}{dt}$, $d\mathbf{s} = \mathbf{v}\,dt$, придадим этой формуле вид

$$A = \int (\mathbf{v}\,d\mathbf{p}). \tag{22.7}$$

2. Чтобы вычислить интеграл, надо знать связь между скоростью материальной точки $\mathbf{v}$ и ее импульсом $\mathbf{p}$. По определению импульса $\mathbf{p} = m\mathbf{v}$, причем в нерелятивистской механике масса m не зависит от скорости, так что $\mathbf{v}\,d\mathbf{p} = m\mathbf{v}\,d\mathbf{v}$. Здесь вектор $d\mathbf{v}$ означает элементарное приращение вектора $\mathbf{v}$, причем это приращение может и не совпадать по направлению с вектором $\mathbf{v}$ (рис. 37). Если мы условимся понимать под v длину вектора $\mathbf{v}$, то очевидно $v^2 = \mathbf{v}^2$. Действительно, справа стоит скалярное произведение вектора $\mathbf{v}$ на самого себя, а оно равно квадрату длины вектора, как это непосредственно следует из определения скалярного произведения. Дифференцируя теперь обе части соотношения $v^2 = \mathbf{v}^2$, получим $v\,dv = \mathbf{v}\,d\mathbf{v}$. Здесь dv есть элементарное приращение длины вектора $\mathbf{v}$. Его нельзя смешивать с длиной элементарного приращения вектора, т. е. с величиной $|d\mathbf{v}|$. Последняя величина по самому ее смыслу существенно положительна, в то время как приращение dv может быть как положительным, так и отрица-

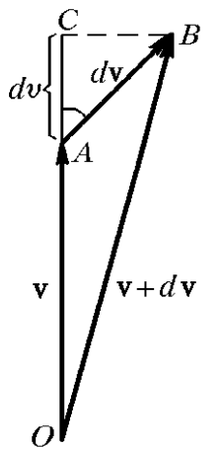


Рис. 37

тельным. На рис. 37 $d\mathbf{v} = \overrightarrow{AB}$, $dv = AC$. По определению скалярного произведения $\mathbf{v}\, d\mathbf{v} = v \cdot AB \cos \alpha = v \cdot AC = v\, dv$. Это дает другое доказательство соотношения $\mathbf{v}\, d\mathbf{v} = v\, dv$. Разумеется, такое соотношение справедливо не только для вектора $\mathbf{v}$, но и для любого другого вектора. Используя его в нашей задаче и вынося постоянный множитель m из-под знака интеграла, получим

$$A_{12} = m \int_{v_1}^{v_2} v\, dv = \frac{mv_2^2}{2} - \frac{mv_1^2}{2},$$

где v_1 — начальная, а v_2 — конечная скорости точки. Букву A мы снабдили индексами 1, 2, чтобы подчеркнуть, что речь идет о работе при перемещении материальной точки из начального положения *1* в конечное положение *2* (см. рис. 36). Величина

$$K = \frac{mv^2}{2} = \frac{p^2}{2m} \tag{22.8}$$

называется *кинетической энергией материальной точки.* С помощью этого понятия полученный результат запишется в виде

$$A_{12} = K_2 - K_1. \tag{22.9}$$

Таким образом, *работа силы при перемещении материальной точки равна приращению кинетической энергии этой точки.* Связь между работой и кинетической энергией, выражаемая этой теоремой, и оправдывает введение обоих этих понятий.

3. Полученный результат без труда обобщается на случай произвольной системы материальных точек. *Кинетической энергией системы* называется сумма кинетических энергий материальных точек, из которых эта система состоит или на которые ее можно мысленно разделить. Напишем соотношение (22.9) для каждой материальной точки системы, а затем все такие соотношения сложим. В результате снова получится формула (22.9), но уже не для одной материальной точки, а для системы материальных точек. Под A_{12} надо понимать сумму работ всех сил, как внутренних, так и внешних, действующих на материальные точки системы. Таким образом, *работа всех сил, действующих на систему материальных точек, равна приращению кинетической энергии этой системы.*

Имеется существенное отличие этой теоремы от аналогичной, в которой говорится о связи между импульсом силы и изменением импульса системы (§ 18). Внутренние силы вследствие равенства действия и противодействия не меняют импульс всей системы. Приращение импульса системы определяется только *внешними силами.* Не так обстоит дело в случае кинетической энергии. Работа внутренних сил, вообще говоря, не обращается в нуль. Представим себе, например, замкнутую систему, состоящую из двух материальных

точек, взаимодействующих между собой силами притяжения $\mathbf{F}_1$ и $\mathbf{F}_2$. Если точки придут в движение навстречу друг другу, то каждая из сил $\mathbf{F}_1$ и $\mathbf{F}_2$ совершит положительную работу. Будет положительной и работа обеих сил. Она пойдет на приращение кинетической энергии системы. Кинетическая энергия изменится под действием одних только внутренних сил. Следовательно, *приращение кинетической энергии определяется работой не только внешних, но и внутренних сил.*

4. Доказанная теорема для материальной точки имеет место и в релятивистской механике. Надо только изменить выражение для кинетической энергии. В релятивистской механике формула (22.7) также справедлива, однако при вычислении интеграла (22.7) надо учитывать зависимость массы от скорости. Масса определяется формулой

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}}.$$

Подставив в эту формулу $v = p/m$ и возведя в квадрат, получим

$$p^2 + (m_0 c)^2 = (mc)^2. \tag{22.10}$$

Дифференцированием этого соотношения находим

$$p\,dp = c^2 m\,dm.$$

А так как $p\,dp = \mathbf{p}\,d\mathbf{p}$ и $\mathbf{p} = m\mathbf{v}$, то

$$\mathbf{v}\,d\mathbf{p} = c^2\,dm.$$

Таким образом,

$$A_{12} = \int \mathbf{v}\,d\mathbf{p} = \int_{m_1}^{m_2} c^2\,dm.$$

Отсюда

$$A_{12} = c^2(m_2 - m_1) = c^2 \Delta m, \tag{22.11}$$

где m_1 и m_2 — массы материальной точки в начальном и конечном положениях.

Таким образом, *в релятивистской механике работа определяется только приращением массы материальной точки.* Этот результат проще соответствующего результата нерелятивистской механики. Введем обозначение

$$E = mc^2 \tag{22.12}$$

и назовем величину E *полной* или *релятивистской энергией частицы* (материальной точки). Тогда

$$A_{12} = E_2 - E_1. \tag{22.13}$$

В частном случае, когда частица покоится, ее релятивистская энергия определяется выражением

$$E_0 = m_0 c^2 \tag{22.14}$$

и называется *энергией покоя. Кинетическая энергия* есть часть релятивистской энергии, обусловленная движением частицы. Она представляется разностью

$$K = E - E_0 = (m - m_0)c^2, \tag{22.15}$$

или

$$K = m_0 c^2 \left(\frac{1}{\sqrt{1 - v^2/c^2}} - 1 \right). \tag{22.16}$$

Ясно, что работу A_{12} можно вычислить также по формуле

$$A_{12} = K_2 - K_1.$$

Если в формулу (22.10) ввести величины E и E_0, то получится

$$E^2 = E_0^2 + (pc)^2. \tag{22.18}$$

Эта формула выражает в релятивистской механике связь между импульсом частицы и ее полной энергией. Она справедлива не только для отдельных частиц, но и для систем, состоящих из нескольких частиц. Под m_0 и E_0 следует понимать массу и полную энергию такой системы в системе отсчета, относительно которой она покоится.

Формула (22.16) дает выражение для кинетической энергии в релятивистской механике. При медленных движениях она переходит в привычную формулу (22.8). Действительно, пользуясь формулой бинома Ньютона, может написать

$$\frac{1}{\sqrt{1 - v^2/c^2}} = \left(1 - \frac{v^2}{c^2}\right)^{-1/2} = 1 + \frac{1}{2}\frac{v^2}{c^2} + \frac{3}{8}\frac{v^2}{c^4} + \ldots$$

Когда $v^2/c^2 \ll 1$, можно оборвать это разложение на втором члене. Тогда формула (22.16) перейдет в формулу (22.8).

5. В атомной физике удобной единицей энергии является *электронвольт* (эВ). Это есть энергия, приобретаемая электроном в электрическом поле при прохождении разности потенциалов в один вольт:

$$1 \text{ эВ} = 1{,}602 \cdot 10^{-12} \text{ эрг}.$$

Употребляется также *килоэлектронвольт* (1 кэВ = 1000 эВ). В ядерной физике, а также в ускорительной технике употребляются более крупные единицы: *мегаэлектронвольт* (1 МэВ = 10^6 эВ), *гигаэлектронвольт* (1 ГэВ = 10^9 эВ) и *тераэлектронвольт* (1 ТэВ =

$= 10^{12}$ эВ). Энергия покоя для электрона и протона соответственно равна:

$$m_{0e}c^2 = 0{,}511 \text{ МэВ}, \quad m_{0p}c^2 = 938 \text{ МэВ}.$$

Если полная релятивистская энергия частицы E велика по сравнению с ее энергией покоя $E_0 = m_0c^2$, то говорят о движении с ультрарелятивистскими скоростями. Такие скорости получаются в ускорителях, они встречаются также в космических лучах.

Зная энергию ультрарелятивистской частицы, можно вычислить и ее скорость. Точнее, можно вычислить не саму скорость частицы (для этого недостаточна точность, с которой известна скорость света c), а разность между этой скоростью и скоростью света в вакууме. С этой целью перепишем формулу (22.12) в виде

$$E = \frac{m_0c^2}{\sqrt{1-v^2/c^2}}.$$

Отсюда получаем

$$E^2(c+v)\ (c-v) = m_0^2c^6 = E_0^2c^2.$$

Так как скорость v близка к c, то во втором множителе $c+v$ величину v можно заменить на c. В результате получится

$$\frac{c-v}{c} = \frac{E_0^2}{2E^2}. \tag{22.19}$$

Для протонов с энергией $E = 10$ ГэВ получаем

$$\frac{c-v}{c} = \frac{0{,}938^2}{2\cdot 10^2} = 0{,}0044.$$

Для электронов с энергией $E = 1$ ГэВ

$$\frac{c-v}{c} = \frac{0{,}511^2}{2\cdot 10^6} = 1{,}3\cdot 10^{-7}.$$

В космических лучах регистрировались протоны с энергией 10^{20} эВ $= 10^{11}$ ГэВ. В этом случае

$$\frac{c-v}{c} \sim 10^{-21},$$

т. е. скорость частицы отличается от скорости света всего на $3\cdot 10^{-11}$ см/с.

ЗАДАЧИ

1. Через неподвижный блок, массой которого можно пренебречь перекинута замкнутая тяжелая веревка массой M. В начальный момент времени за точку веревки, расположенную между блокам и нижним заворотом ее, цепляется обезьяна массой m и начинает карабкаться вверх так, чтобы удержаться на неизменной высоте. Какую мощность P должна для

этого развивать обезьяна? Через сколько времени она перестанет справляться со своей затеей, если максимальная мощность, которую она может развивать, равна $P_{\text{макс}}$?

Ответ. $P = \dfrac{(mg)^2}{M}\,t$; $t = \dfrac{M}{(mg)^2}\,P_{\text{макс}}$.

2. Вывести формулу (21.6), являющуюся релятивистским обобщением формулы Циолковского для движения ракеты. Считать, что скорость ракеты и газовой струи направлены вдоль одной прямой.

Решение. Решение основано на релятивистских законах импульса и энергии (релятивистской массы). Они нами были сформулированы. Кроме того, требуется знать *релятивистский закон сложения скоростей*, который нами не формулировался. Читатель, желающий разобрать решение, приводимое ниже, должен обратиться к руководствам по теории относительности или принять на веру формулу (22.22), приводимую ниже.

Пусть m и v — масса покоя и скорость ракеты в произвольный момент времени t, а $m_{\text{газ}}$ и $v_{\text{газ}}$ — те же величины для газов, образовавшихся из топлива ракеты к этому моменту времени. Так как газы, уже покинувшие ракету, не оказывают влияния на ее движение, то можно принять $m_{\text{газ}} = 0$. Однако газы непрерывно образуются, так что $dm_{\text{газ}} \neq 0$. На основании закона сохранения импульса и энергии (релятивистской массы)

$$\frac{mv}{\sqrt{1 - v^2/c^2}} + \frac{m_{\text{газ}} v_{\text{газ}}}{\sqrt{1 - v_{\text{газ}}^2/c^2}} = \text{const}, \tag{22.20}$$

$$\frac{m}{\sqrt{1 - v^2/c^2}} + \frac{m_{\text{газ}}}{\sqrt{1 - v_{\text{газ}}^2/c^2}} = \text{const}. \tag{22.21}$$

Дифференцируя уравнение (22.20) с учетом (22.21) и полагая в окончательном результате $m_{\text{газ}} = 0$, получим

$$\frac{m}{\sqrt{1 - v^2/c^2}}\,dv + (v - v_{\text{газ}})\,d\,\frac{m}{\sqrt{1 - v^2/c^2}} = 0.$$

По релятивистскому закону сложения скоростей

$$v_{\text{газ}} = \frac{v - v_{\text{отн}}}{1 - v v_{\text{отн}}/c^2}, \tag{22.22}$$

где $v_{\text{отн}}$ — скорость газовой струи относительно ракеты. Исключая $v_{\text{газ}}$, после несложных преобразований находим

$$\frac{dv}{v^2 - c^2} = \frac{v_{\text{отн}}}{c^2}\,\frac{dm}{m}.$$

Предполагая скорость $v_{\text{отн}}$ постоянной и интегрируя, получим

$$\frac{m_0}{m} = \left(\frac{1+\beta}{1-\beta}\right)^{c/(2v_{\text{отн}})}. \tag{22.23}$$

При выводе этой формулы предполагалось, что вся энергия и весь импульс сосредоточены в веществе. Энергия и импульс излучения все же учитываются, хотя и частично. Действительно, часть излучения поглоща-

ется газами, образующимися при сгорании топлива ракеты. За счет этого масса покоя газов несколько увеличивается. В результате некоторая часть энергии и импульса излучения переходит к веществу и учитывается при вычислении.

3. Импульс излучения $p_{\text{изл}}$ связан с его энергией $\mathcal{E}_{\text{изл}}$ соотношением

$$\mathcal{E}_{\text{изл}} = cp_{\text{изл}}.$$

Используя это соотношение, найти, как изменится формула (22.23) для фотонной ракеты.

Р е ш е н и е. Изменения полной энергии и импульса ракеты (с точностью до знака) равны энергии и импульсу образовавшегося излучения, т. е.

$$\left(m_0 - \frac{m}{\sqrt{1-\beta^2}}\right)c^2 = \mathcal{E}_{\text{изл}},$$

$$\frac{mv}{\sqrt{1-\beta^2}} = p_{\text{изл}}.$$

Значит,

$$\left(m_0 - \frac{m}{\sqrt{1-\beta^2}}\right)c^2 - \frac{mvc}{\sqrt{1-\beta^2}} = \mathcal{E}_{\text{изл}} - cp_{\text{изл}} = 0.$$

Отсюда

$$\frac{m_0}{m} = \sqrt{\frac{1+\beta}{1-\beta}}. \tag{22.24}$$

Формально эта формула является частным случаем формулы (22.23) при $v_{\text{отн}} = c$.

4. Определить мощность фотонной ракеты, которой сообщается ускорение $a = 10$ м/с^2.

Р е ш е н и е. Изменение импульса ракеты за время dt равно $d(mv) = m\,dv + v\,dm$, где $m = \dfrac{m_0}{\sqrt{1-\beta^2}}$ — релятивистская масса ракеты. Оно компенсируется изменением импульса излучения:

$$m\,dv + v\,dm = dp_{\text{изл}}.$$

По закону сохранения энергии

$$c^2\,dm + d\mathcal{E}_{\text{изл}} = 0.$$

Для излучения $d\mathcal{E}_{\text{изл}} = cp_{\text{изл}}$. Исключая $dp_{\text{изл}}$, получим

$$m\,dv + \frac{d\mathcal{E}_{\text{изл}}}{c}(1-\beta).$$

Мощность ракеты

$$P = \frac{d\mathcal{E}_{\text{изл}}}{dt} = \frac{mca}{1-\beta^2} = \frac{m_0 ca}{(1-\beta^2)^{3/2}}.$$

На старте $\beta = 0$,

$$P\ [\text{МВт}] = m_0 ca = 3000 m_0,$$

где m_0 — масса в килограммах. Удельная мощность излучения, т. е. мощность, требуемая для ускорения одного килограмма ракеты, равна мощности нескольких тысяч крупных электростанций.

§ 23. СВЯЗЬ МЕЖДУ КИНЕТИЧЕСКИМИ ЭНЕРГИЯМИ В РАЗЛИЧНЫХ СИСТЕМАХ ОТСЧЕТА. ТЕОРЕМА КЁНИГА

Как ясно из формулы (22.8), кинетическая энергия тела зависит от выбора системы отсчета, относительно которой рассматривается его движение. Можно поставить вопрос, как преобразуется кинетическая энергия при переходе от одной системе отсчета к другой. Приведем решение этого вопроса в нерелятивистской механике. Сначала рассмотрим частный случай, когда тело состоит всего из одной материальной точки. Обозначим через K кинетическую энергию материальной точки в какой-либо системе отсчета S, а через K' — в другой системе S', движущейся относительно S поступательно со скоростью $\mathbf{V}$. (Скорость $\mathbf{V}$ может быть постоянной, но может и меняться во времени.) В нерелятивистской механике скорости $\mathbf{v}$, $\mathbf{v}'$ и $\mathbf{V}$ связаны соотношением $\mathbf{v} = \mathbf{v}' + \mathbf{V}$. Поэтому

$$\tfrac{1}{2} mv^2 = \tfrac{1}{2} mv'^2 + \tfrac{1}{2} V^2 + m(\mathbf{v}'\mathbf{V}),$$

или

$$K = K' + \tfrac{1}{2} mV^2 + (\mathbf{p}'\mathbf{V}), \tag{23.1}$$

где $\mathbf{p}' = m\mathbf{v}'$ — импульс материальной точки в системе S'. Формула (23.1) справедлива и для произвольной системы материальных точек. Чтобы убедиться в этом, достаточно написать соотношение (23.1) для каждой материальной точки системы, а затем просуммировать по всем точкам. Тогда получится снова формула (23.1), в которой под $\mathbf{p}'$ надо понимать импульс *всей системы материальных точек* в системе отсчета S', т. е. $\mathbf{p}' = m_1\mathbf{v}_1' + m_2\mathbf{v}_2' + \ldots$ Его можно представить в виде $\mathbf{p}' = m\mathbf{v}_c'$, где $\mathbf{v}_c'$ — скорость центра масс системы материальных точек относительно S', а m — ее суммарная масса. Таким образом,

$$K = K' + \tfrac{1}{2} mV^2 + m(\mathbf{V}\mathbf{v}_c'). \tag{23.2}$$

Если центр масс покоится в системе отсчета S', т. е. $\mathbf{v}_c' = 0$, то

$$K = K' + \tfrac{1}{2} mV^2. \tag{23.3}$$

Это равенство выражает так называемую *теорему Кёнига*: *кинетическая энергия системы материальных точек равна сумме кинетической энергии всей массы системы, мысленно сосредоточен-*

ной в ее центре масс и движущейся вместе с ним, и кинетической энергии той же системы в ее относительном движении по отношению к поступательно движущейся системе координат с началом в центре масс.

§ 24. КОНСЕРВАТИВНЫЕ И НЕКОНСЕРВАТИВНЫЕ СИЛЫ

1. Все силы, встречающиеся в макроскопической механике, принято разделять на *консервативные* и *неконсервативные*. Прежде чем вводить эти понятия, рассмотрим некоторые примеры.

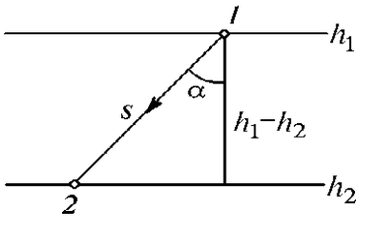


Рис. 38

Вычислим сначала работу силы тяжести, которую она совершает при переходе материальной точки из положения *1* в положение *2* вдоль прямолинейного отрезка *12* (рис. 38). Примером может служить скольжение без трения материальной точки по гладкой наклонной плоскости. Очевидно, эта работа $A_{12} = mgs\cos\alpha$, или

$$A_{12} = mg(h_1 - h_2) = mgh_1 - mgh_2, \qquad (24.1)$$

где h_1 и h_2 — высоты, на которых находилась материальная точка в начале и конце пути, отсчитанные от какого-либо произвольного уровня, например от земной поверхности или от уровня моря. Формула (24.1) остается справедливой и при перемещении вдоль произвольной кривой, например по пути *132* (рис. 39). Это станет очевидным, если разбить весь путь *132* горизонтальными плоскостями на малые участки, каждый из которых может быть принят за прямолинейный. Применив к каждому участку формулу (24.1) и сложив полученные работы, мы придем к прежнему результату (24.1). Если вместо пути *132* взять любой другой путь *142* между теми же начальным и конечными положениями *1* и *2*, то работа силы тяжести не изменится, так как она определяется только разностью высот $h_1 - h_2$, которая от формы пути не зависит. Таким образом, *работа тяжести не зависит от формы пути, а определяется только начальным и конечным положениями перемещающейся точки.*

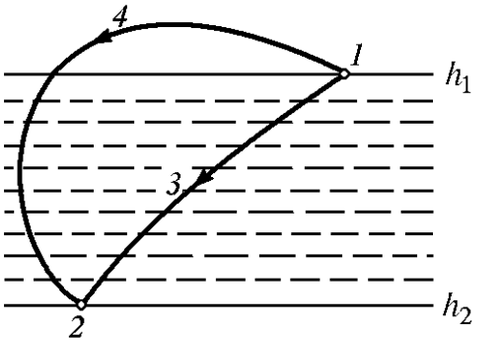


Рис. 39

2. В качестве второго примера рассмотрим работу при перемещении материальной точки в поле центральных сил. Сила называется *центральной,* если она направлена к одной и той же точке

(или от одной и той же точки) и зависит только от расстояния до этой точки, называемой *центром сил* или *силовым центром.* Примером может служить сила гравитационного притяжения, с которой Солнце действует на планету, или сила электростатического взаимодействия двух точечных зарядов. По определению элементарной работы $dA = F\,ds\cos(\mathbf{F}, d\mathbf{s})$. Величина $ds\cos(\mathbf{F}, d\mathbf{s})$ есть проекция элементарного перемещения $d\mathbf{s}$ на направление силы, или, что то же самое, на направление радиуса-вектора $\mathbf{r}$ (если за положительное направление силы принять направление от силового центра O). Следовательно, $ds\cos(\mathbf{F}, d\mathbf{s}) = dr$, где dr — элементарное приращение длины r, т. е. расстояние материальной точки от силового центра (рис. 40). Таким образом, $dA = F(r)\,dr$, причем по предположению сила F зависит только от расстояния r. Поэтому работа A_{12} выразится определенным интегралом

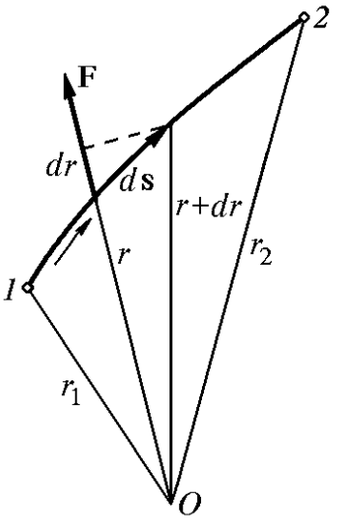


Рис. 40

$$A_{12} = \int_{r_1}^{r_2} F(r)\,dr, \tag{24.2}$$

значение которого зависит *только от расстояний* r_1 *и* r_2 точек *1* и *2* до силового центра O, но *не зависит от формы пути*, по которому материальная точка перешла из начального положения *1* в конечное положение *2*. В формулу (24.2) путь перехода вообще не входит, в нее входят только расстояния до силового центра.

3. Допустим, что в силовом центре помещено *физическое тело* (материальная точка), взаимодействующее с рассматриваемой материальной точкой (которая с тем же основанием может быть принята за силовой центр). При взаимодействии перемещаются как материальная точка, так и силовой центр. При выводе формулы (24.2) перемещение силового центра не принималось во внимание. Однако справедливость самой формулы не связана с этим ограничением. *Работа* A_{12} *зависит только от относительного перемещения материальных точек, но не может зависеть от абсолютных перемещений каждой из точек в отдельности. В этом можно убедиться простым* вычислением. Пусть взаимодействуют две материальные точки *1* и *2*, причем силы взаимодействия $\mathbf{F}_1$ и $\mathbf{F}_2$ подчиняются третьему закону Ньютона. Обозначим через $\mathbf{r}_1$ и $\mathbf{r}_2$ радиусы-векторы этих точек, проведенные из какого-либо неподвижного начала. Тогда для элементарной работы можно написать $dA = \mathbf{F}_1\,d\mathbf{r}_1 + \mathbf{F}_2\,d\mathbf{r}_2$. По третьему закону Ньютона $\mathbf{F}_1 = -\mathbf{F}_2$, а потому $dA = \mathbf{F}_2(d\mathbf{r}_2 - d\mathbf{r}_1) = \mathbf{F}_2\,d(\mathbf{r}_2 - \mathbf{r}_1)$. Но $\mathbf{r}_2 - \mathbf{r}_1$ есть ради-

ус-вектор точки *2* относительно точки *1*. Обозначим его через $\mathbf{r}_{12}$. Тогда

$$dA = \mathbf{F}_2\, d\mathbf{r}_{12}. \tag{24.3}$$

Значит, при вычислении элементарной, а с ней и полной работы точка *1* может считаться неподвижной, а точка *2* — перемещающейся относительно нее. Можно было бы, конечно, считать неподвижной точку *2*, а точку *1* — движущейся. Результат получился бы тот же самый. Вообще, как и раньше, выражение (24.3) может быть преобразовано к виду

$$dA = F(r)\, dr. \tag{24.4}$$

Сюда входят только расстояние между взаимодействующими точками r и его приращение dr. Отсюда немедленно получается формула (24.2), что и доказывает наше утверждение.

Отметим одно следствие формулы (24.2). Допустим, что материальные точки *1* и *2* соединены абсолютно жестким стержнем. При такой идеализации расстояние между взаимодействующими точками будет оставаться неизменным при любых их перемещениях: $dr = 0$. Поэтому всегда будет равен нулю интеграл в формуле (24.2), а с ним и работа сил взаимодействия материальных точек *1* и *2* на любом перемещении. Так называемые абсолютно твердые тела могут рассматриваться как системы материальных точек, расстояния между которыми не меняются при любых движениях. Такая неизменяемость обеспечивается внутренними силами или силами связей, действующими между материальными точками системы. Всю систему можно мысленно разбить на пары взаимодействующих точек и применить к ним доказанное выше следствие. Отсюда следует, что *работа внутренних сил, действующих в абсолютно твердых телах, равна нулю при любых движениях.* Реальные тела не являются абсолютно твердыми. Действующие в них силы обусловлены связями, которые могут быть очень жесткими, но не бесконечно жесткими. Работа таких сил, вообще говоря, отлична от нуля. Однако по мере увеличения жесткости работа становится все меньше и меньше и в пределе для бесконечно жестких связей обращается в нуль.

Результаты, полученные для двух материальных точек, обобщаются на случай произвольной системы материальных точек, между которыми действуют центральные силы. Если задать положение каждой материальной точки, то этим определится и положение всей системы или ее *конфигурация. Работа центральных сил не зависит от способа (или «пути») перехода системы из начальной конфигурации в конечную — она определяется исключительно самими конфигурациями.*

4. Если силы взаимодействия зависят только от конфигурации материальных точек системы (т. е. от их координат) и работа этих сил при перемещении системы из произвольного начального поло-

жения в произвольное конечное положение не зависит от пути перехода, а определяется только начальной и конечной конфигурациями системы, то такие силы называются *консервативными*. Рассмотренные нами примеры показывают, что *сила тяжести и все центральные силы являются силами консервативными.*

Можно дать другое определение консервативных сил, эквивалентное приведенному. Пусть система из положения *1* (рис. 41) перешла в положение *2* по пути *132*. (Мы символически изображаем положение системы точкой на плоскости, а путь перехода — линией, хотя буквально такой способ применим лишь для системы, состоящей всего из одной материальной точки.) При этом будет совершена работа A_{132}. Если бы система перешла в положение *2* по пути *142*, то совершенная работа была бы равна A_{142}. По определению консервативных сил $A_{132} = A_{142}$. Так как силы зависят только от конфигурации системы, то $A_{142} = -A_{241}$, где A_{241} — работа, которая была бы совершена при переходе системы из положения *2* в положение *1* по тому же пути, но в обратном порядке, т. е. по пути *241*. Таким образом, $A_{132} + A_{241} = 0$. Но сумма $A_{132} + A_{241}$ есть работа, совершенная силами, когда система вернулась в исходное положение *1*. В этом случае говорят о работе по «*замкнутому пути*». Итак, *работа консервативных сил по любому замкнутому пути равна нулю.* Проведя это рассуждение в обратном порядке, без труда докажем, что из обращения в нуль работы по любому пути следует независимость работы от пути перехода. Поэтому можно дать еще такое определение консервативных сил. *Консервативными* называются силы, зависящие только от конфигурации системы, и работа которых по любому замкнутому пути равна нулю.

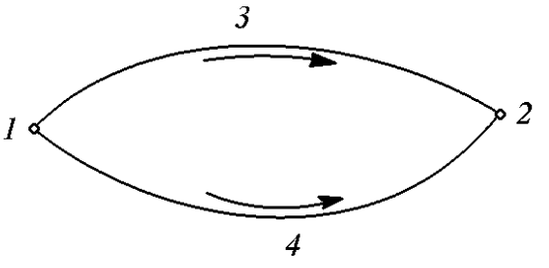


Рис. 41

5. Все силы, не являющиеся консервативными, называются *неконсервативными силами.* К ним относятся, прежде всего, так называемые *диссипативные силы*, например силы трения, возникающие при скольжении какого-либо тела по поверхности другого. Сюда же относятся силы сопротивления, испытываемые телом при движении в жидкой или газообразной среде. Их также иногда называют силами трения (см. § 17). Все эти силы зависят не только от конфигурации тел, но и от их *относительных скоростей.* Они направлены всегда против скорости тела (относительно поверхности, по которой оно скользит, или относительно сопротивляющейся среды, в которой оно движется). Поэтому, если тело скользит по неподвижной поверхности или движется в «неподвижной» сопротивляющейся среде, то при любом движении тела работа сил

трения, действующих на него, отрицательна. Но работа сил трения может быть и положительной, когда поверхность или среда сами движутся. Рассмотрим, например, тело B, по поверхности которого скользит тело C (рис. 42) с относительной скоростью $\mathbf{v}_{отн}$. Сила трения $\mathbf{F}_{тр}$, действующая на тело C, направлена против вектора $\mathbf{v}_{отн}$. Допустим, что само тело B движется в противоположном направлении со скоростью $\mathbf{v}$. Если $v > v_{отн}$, то в «неподвижной» системе отсчета тело C движется со скоростью $\mathbf{v} - \mathbf{v}_{отн}$ в том же направлении, в котором действует сила трения. Сила трения ежесекундно совершает над телом C положительную работу $A_1 = F_{тр}(v - v_{отн})$. Однако если система замкнута, то полная работа сил трения, действующих на все тела системы, всегда отрицательна. Так, в приведенном примере сила трения, действующая на тело B, совершает отрицательную работу $A_2 = -F_{тр}v$. Полная работа сил трения равна $A = A_1 + A_2 = -F_{тр}v_{отн}$, т. е. отрицательна. Поэтому мы даем следующее определение диссипативных сил. *Диссипативными* называют такие силы, полная работа которых при любых движениях в замкнутой системе всегда отрицательна.

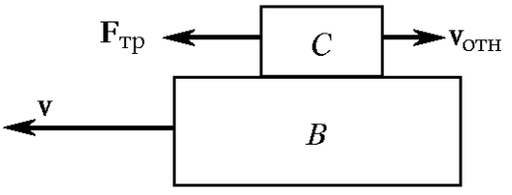


Рис. 42

6. Отметим, наконец, еще один вид неконсервативных сил, называемых *гироскопическими силами. Эти силы зависят от скорости материальной точки и действуют всегда перпендикулярно к этой скорости.* Работа таких сил равна нулю при любом перемещении материальной точки, в частности при ее движении по замкнутому пути. От консервативных гироскопические силы отличаются тем, что они определяются не только положением, но и скоростью движущейся материальной точки. Единственным примером гироскопических сил, известных в физике, является *сила Лоренца*, т. е. сила, действующая на заряженную частицу в магнитном поле. Она пропорциональна векторному произведению $[\mathbf{vB}]$, т. е. перпендикулярна как к направлению скорости $\mathbf{v}$, так и к вектору *напряженности магнитного поля* $\mathbf{B}$. Правда, в механике встречаются гироскопические силы и иного рода. Это так называемые *силы Кориолиса* (по имени французского физика Гюстава Кориолиса (1792—1843)). Однако эти силы не являются «настоящими силами» в смысле механики Ньютона. При рассмотрении движений относительно инерциальных систем отсчета (а только такие движения мы сейчас и рассматриваем) такие «силы» вообще не существуют. Они вводятся искусственно при рассмотрении движений в системах отсчета, вращающихся относительно инерциальных, чтобы придать уравнениям движения в таких системах формально такой же вид, что и в инерциальных системах отсчета (см. гл. IX).

§ 25. ПОТЕНЦИАЛЬНАЯ ЭНЕРГИЯ. ЗАКОН СОХРАНЕНИЯ ЭНЕРГИИ В МЕХАНИКЕ

1. Если на систему действуют только консервативные и гироскопические силы, то для такой системы можно ввести понятие *потенциальной энергии.* Примем какое-либо произвольное положение системы, характеризующееся заданием координат ее материальных точек, условно за *нулевое.* Работа, совершаемая консервативными силами при переходе системы из рассматриваемого положения в нулевое, называется потенциальной энергией системы в первом положении. Работа консервативных сил не зависит от пути перехода, а потому потенциальная энергия системы при фиксированном нулевом положении зависит только от координат материальных точек системы в рассматриваемом положении. Иными словами, *потенциальная энергия системы U является функцией только ее координат.*

Значение потенциальной энергии зависит от того, какое положение системы условно принято за нулевое. Если за нулевое принять положение 0 (рис. 43 *а*), то в положении *1* система будет обладать потенциальной энергией $U = A_{10}$, равной работе консервативных сил при переходе системы из положения *1* в положение 0. Если же за нулевое принять положение 0′, то потенциальная энергия будет равна $U' = A_{10'}$. Вследствие консервативности сил, действующих в системе, работа вдоль пути *1*0′ равна работе вдоль пути *1*00′: $A_{10'} = A_{10} + A_{00'}$, или $U' = U + A_{00'}$. Работа $A_{00'}$ постоянна, т. е. не зависит от координат системы в рассматриваемом состоянии *1*. Она полностью определяется выбором нулевых положений 0 и 0′. Мы видим, что при замене одного

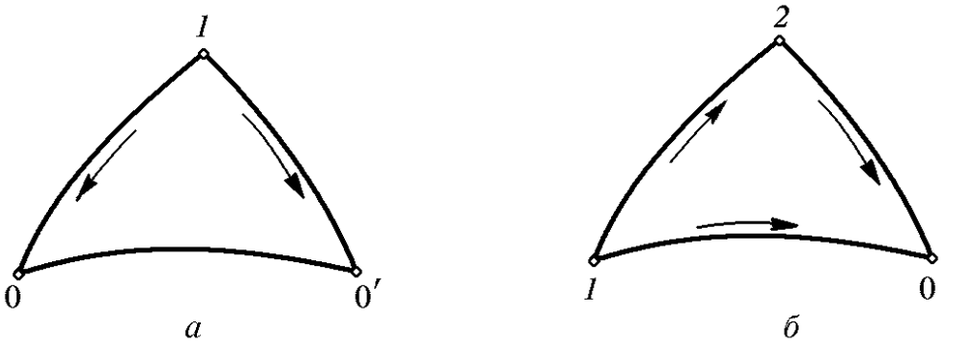


Рис. 43

нулевого положения другим потенциальная энергия системы меняется на постоянную величину. Неопределенность можно усилить еще больше, если условиться считать потенциальную энергию в нулевом положении равной не нулю, а какому-либо постоянному произвольному значению. Тогда в приведенном выше определении вместо потенциальной энергии следует говорить о ее *разности* в двух положениях. *Разностью потенциальных энергий в рассматриваемом и нулевом положениях* называется работа, совершаемая консервативными силами при переходе системы

из рассматриваемого в нулевое положение. Таким образом, *потенциальная энергия системы определена не однозначно*, а *с точностью до произвольной постоянной*. Этот произвол не может отразиться на физических выводах, так как ход физических явлений может зависеть не от абсолютных значений самой потенциальной энергии, а лишь от ее разностей в различных состояниях. Эти же разности от выбора произвольной постоянной не зависят.

Пусть система перешла из положения *1* в положение *2* по какому-либо пути *12* (рис. 43 *б*). Работу A_{12}, совершенную консервативными силами при таком переходе, можно выразить через потенциальные энергии U_1 и U_2 в состояниях *1* и *2*. С этой целью вообразим, что переход осуществлен через нулевое положение 0, т. е. по пути *102*. Так как силы консервативны, то $A_{12} = A_{102} = A_{10} + A_{02} = A_{10} - A_{20}$. По определению потенциальной энергии $U_1 = A_{10} + C$, $U_2 = A_{20} + C$, где C — одна и та же аддитивная постоянная. Таким образом,

$$A_{12} = U_1 - U_2, \tag{25.1}$$

т. е. *работа консервативных сил равна убыли потенциальной энергии системы.*

2. Та же работа A_{12}, как было показано, может быть выражена через приращение кинетической энергии по формуле (22.9). Приравнивая выражения (22.9) и (25.1), получим $K_2 - K_1 = U_1 - U_2$, откуда

$$K_1 + U_1 = K_2 + U_2.$$

Сумма кинетической и потенциальной энергий системы называется ее *полной энергией* E. Таким образом, $E_1 = E_2$, или

$$E \equiv K + U = \text{const}. \tag{25.2}$$

В системе с одними только консервативными (и гироскопическими) силами полная энергия остается неизменной. Могут происходить лишь превращения потенциальной энергии в кинетическую и обратно, но полный запас энергии системы измениться не может. Это положение называется *законом сохранения энергии в механике.*

3. Вычислим потенциальную энергию в некоторых простейших случаях.

а. Потенциальная энергия тела в однородном поле тяжести. Если материальная точка, находящаяся на высоте h, упадет на нулевой уровень (т. е. уровень, для которого $h = 0$), то сила тяжести совершит работу $A = mgh$. Поэтому на высоте h материальная точка обладает потенциальной энергией $U = mgh + C$. За нулевой можно принять произвольный уровень,

например, уровень пола (если опыт производится в лаборатории), уровень моря и т. д. Постоянная C равна потенциальной энергии на нулевом уровне. Полагая ее равной нулю, получим

$$U = mgh. \tag{25.3}$$

б. Потенциальная энергия растянутой пружины. Упругие силы, возникающие при растяжении или сжатии пружины, являются центральными силами. Поэтому они консервативны, и имеет смысл говорить о потенциальной энергии деформированной пружины. Ее называют *упругой энергией.* Обозначим через x *растяжение пружины,* т. е. разность $x = l - l_0$ длин пружины в деформированном и недеформированном состояниях. Упругая сила F зависит только от растяжения. Если растяжение x не очень велико, то сила пропорциональна ему: $F = kx$ (закон Гука, см. § 11). При возвращении пружины из деформированного в недеформированное состояние сила F совершает работу

$$A = \int_0^x F\,dx = k\int_0^x x\,dx = \frac{1}{2}kx^2.$$

Если упругую энергию пружины в недеформированном состоянии условиться считать равной нулю, то

$$U = \frac{1}{2}kx^2. \tag{25.4}$$

в. Потенциальная энергия гравитационного притяжения двух материальных точек. По закону всемирного тяготения Ньютона гравитационная сила притяжения двух точечных тел пропорциональна произведению их масс Mm и обратно пропорциональна квадрату расстояния между ними:

$$F = G\frac{Mm}{r^2}, \tag{25.5}$$

где G — *гравитационная постоянная.* Силы гравитационного притяжения, как силы центральные, являются консервативными. Для них имеет смысл говорить о потенциальной энергии. При вычислении этой энергии одну из масс, например M, можно считать неподвижной, а другую — перемещающейся в ее гравитационном поле. При перемещении массы m из бесконечности гравитационные поля совершают работу

$$A = \int_r^\infty G\frac{Mm}{r^2}\,dr = G\frac{Mm}{r},$$

где r — расстояние между массами M и m в конечном состоянии. Эта работа равна убыли потенциальной энергии:

$$A = U_\infty - U(r).$$

Обычно потенциальную энергию в бесконечности U_∞ принимают равной нулю. При таком соглашении

$$U = -G\frac{Mm}{r}. \tag{25.6}$$

Величина (25.6) отрицательна. Это имеет простое объяснение. Максимальной энергией притягивающиеся массы обладают при бесконечном расстоянии между ними. В этом положении потенциальная энергия считается равной нулю. Во всяком другом положении она меньше, т. е. отрицательна.

4. Допустим теперь, что в системе наряду с консервативными и гироскопическими силами действуют также диссипативные силы. Работа всех сил A_{12} при переходе системы из положения *1* в положение *2* по-прежнему равна приращению ее кинетической энергии $K_2 - K_1$. Но в рассматриваемом случае эту работу можно представить в виде суммы работы консервативных сил $A_{12}^{\text{кон}}$ и работы диссипативных сил $A_{12}^{\text{дис}}$. Первая работа может быть выражена через убыль потенциальной энергии системы: $A_{12}^{\text{кон}} = U_1 - U_2$. Поэтому

$$A_{12} = U_1 - U_2 + A_{12}^{\text{дис}}.$$

Приравнивая это выражение приращению кинетической энергии, получим

$$K_2 - K_1 = U_1 - U_2 + A_{12}^{\text{дис}},$$

или

$$E_2 - E_1 = A_{12}^{\text{дис}}, \tag{25.7}$$

где $E = K + U$ — полная энергия системы. Таким образом, в рассматриваемом случае механическая энергия E системы не остается постоянной, а уменьшается, так как работа диссипативных сил $A_{12}^{\text{дис}}$ отрицательна.

Уравнение (25.7) можно обобщить. Разделим все действующие силы на две группы. К первой группе отнесем силы, учитываемые посредством потенциальной энергии U, ко второй — все остальные силы, как внутренние, так и внешние, действующие в системе. Обозначим через A_{12} работу сил второй группы. Тогда, рассуждая так же, как и при выводе формулы (25.7), получим

$$E_2 - E_1 = A_{12}. \tag{25.8}$$

5. Допустим снова, что диссипативные силы в системе не действуют. Тогда справедлив закон сохранения энергии в форме (25.2). Поскольку кинетическая энергия K по своему смыслу не может быть отрицательной, из формулы (25.2) следует, что $E \geqslant U$. Этим соотношением определяется *область изменения всех координат системы*, в которой она может находиться при заданной полной энер-

гии E. В область, где $U > E$, система попасть не может, так как потенциальная энергия не может превышать полную.

Рассмотрим в качестве примера одномерное движение частицы, когда она движется вдоль определенной прямой линии. Примем эту линию за координатную ось X. На оси X величина U будет функцией только x: $U = U(x)$. Если E — полная энергия частицы, то частица может находиться только в тех местах оси X, где $U(x) \leqslant E$. Допустим, что график функции $U(x)$ имеет вид, изображенный на рис. 44. Проведем на этом рисунке горизонтальную прямую $U = E_1$, где E_1 — какая-то постоянная. Пусть эта прямая пересекает «*потенциальную кривую*» $U = U(x)$ в трех точках A, B, C с координатами x_A, x_B, x_C. Сразу видно, что частица с полной энергией E_1 не может находиться в областях I и III. Она может двигаться либо в области II, либо в области IV. Переходить из области II в область IV или обратно частица не может. Этому препятствует «*потенциальный барьер*» BNC на потенциальной кривой. В области II частица с полной энергией E_1 будет совершать так называемое *финитное движение*, т. е. движение, происходящее в ограниченной части пространства. Она окажется запертой в «*потенциальной яме*» AMB и будет совершать колебания между крайними точками x_A и x_B, называемыми *точками поворота.* Если же частица находится в области IV и движется налево, то она, достигнув точки x_C, повернет обратно и далее будет «уходить на бесконечность». Такое движение называется *инфинитным.* Пусть теперь частица обладает большей энергией $E_2 > E_1$, и горизонтальная прямая $U = E_2$ пересекает потенциальную кривую в единственной точке D с абсциссой x_D. Тогда для частицы окажется доступной вся область пространства правее точки x_D, и движение в этой области будет инфинитным.

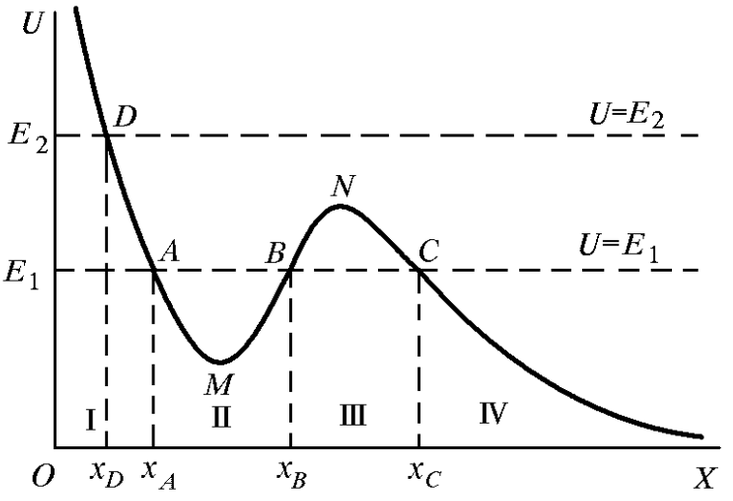


Рис. 44

Допустим, что потенциальная яма имеет характер кривой, изображенной на рис. 45. По обе стороны от точки M обе ветви потенциальной кривой монотонно поднимаются вверх. Пусть при $x = \pm\infty$ функция $U(x)$ обращается в нуль, т. е. ось абсцисс является для потенциальной кривой *асимптотой.* Тогда можно утверждать, что *движение частицы будет финитным, если ее полная энергия отрицательна, и инфинитным, если она положительна.*

Для уяснения качественного характера движения частицы в силовом поле с потенциальной энергией $U(x)$ полезна следующая иллюстрация. Изготовим идеально твердую и идеально гладкую дорожку, форма которой точно совпадает с профилем потенциальной кривой $U = U(x)$ (например, кривой, изображенной на рис. 44). Поместим такую дорожку в однородном поле тяжести и положим на нее на некоторой высоте маленький шарик. Тогда движение шарика под действие силы тяжести будет почти точно воспроизводить движение материальной точки в рассматриваемом силовом поле $U = U(x)$, если только надлежащим образом подобрать ее полную энергию. Некоторая неточность иллюстрации связана с тем, что при движении шарика по дорожке возникает вращение, на что расходуется часть энергии. Иллюстрация совершенно точно передавала бы все черты искомого движения, если бы шарик не катился по дорожке, а скользил по ней без трения. Если такой шарик поместить без начальной скорости в точку A (рис. 44), то он будет совершать колебания по дуге AMB между крайними точками A и B. Если его поместить в точку D, то он сможет преодолеть потенциальный барьер BNC и «уйти на бесконечность».

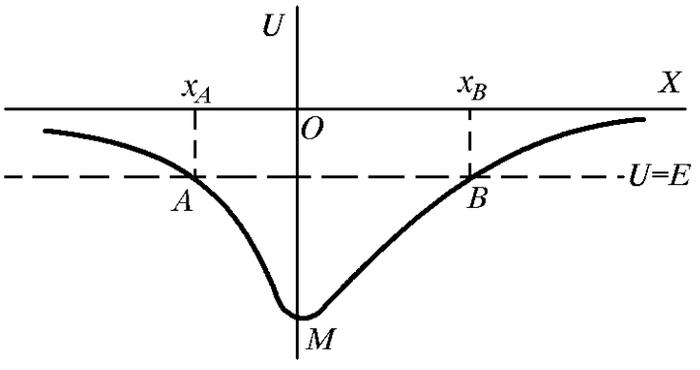


Рис. 45

6. Для финитных движений справедлива так называемая *теорема вириала*, имеющая многочисленные применения в различных отделах физики. Она была сформулирована и доказана немецким физиком Рудольфом Клаузиусом (1822—1888). Для произвольной системы материальных точек можно написать

$$\frac{d}{dt}\sum \mathbf{pr} = \sum \mathbf{rF} + \sum \mathbf{pv},$$

так как $\dot{\mathbf{p}} = \mathbf{F}$, $\dot{\mathbf{r}} = \mathbf{v}$ (суммирование ведется по всем материальным точкам системы). Последнее слагаемое в правой части есть удвоенная кинетическая энергия системы: $2K = \sum \mathbf{pv} = \sum mv^2$, и предыдущее соотношение можно переписать в виде

$$K = -\frac{1}{2}\sum \mathbf{rF} + \frac{d}{dt}\sum \frac{1}{2}(\mathbf{pr}). \qquad (25.9)$$

Величина $-\frac{1}{2}\sum \mathbf{rF}$ называется *вириалом сил*, действующих в системе.

Назовем *средним по времени значением функции* $f(t)$ на временном *интервале* $(t, t+T)$ величину, определяемую выражением

$$\overline{f(t)} = \frac{1}{T}\int_t^{t+T} f(t')\,dt'. \qquad (25.10)$$

Если функция $f(t)$ периодична, то в качестве времени T обычно берут ее период. Если же $f(t)$ не периодична, но ограничена, то время T берут достаточно большим и переходят к пределу

$$\overline{f} = \lim_{T\to\infty} \frac{1}{T} \int_{t}^{t+T} f(t')\, dt', \tag{25.11}$$

предполагая, конечно, что предел существует. Если $f(t)$ есть производная ограниченной функции по времени; $f = \frac{d\varphi}{dt}$, то $\overline{f} = 0$. Действительно,

$$\overline{f} = \lim_{T\to\infty} \frac{1}{T} \int_{t}^{t+T} \frac{d\varphi}{dt'}\, dt' = \lim_{T\to\infty} \frac{\varphi(t+T) - \varphi(t)}{T}.$$

Имея это в виду, усредним соотношение (25.9) по времени, устремляя T к бесконечности. Тогда для финитного движения последнее слагаемое в (25.9) даст нуль, и мы получим

$$\overline{K} = -\frac{1}{2} \overline{\sum \mathbf{rF}}. \tag{25.12}$$

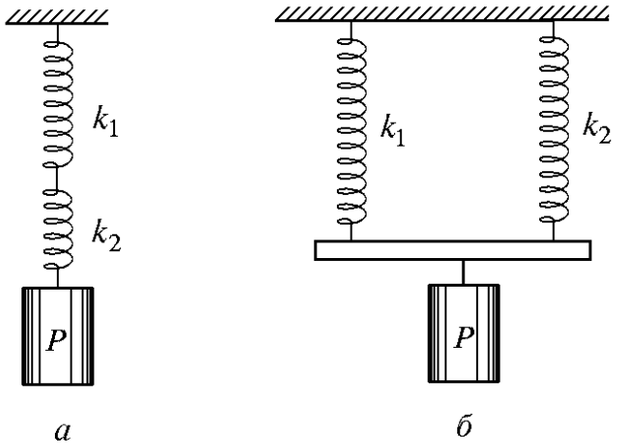
В случае финитных движений среднее по времени значение кинетической энергии системы равно среднему по времени значению вириала сил, действующих в системе. Это и есть теорема вириала Клаузиуса.

ЗАДАЧИ

1. Определить отношение потенциальных энергий деформации U_1 и U_2 двух пружин с жесткостями k_1 и k_2 в двух случаях: а) пружины соединены последовательно и растягиваются грузом P (рис. 46 *а*); б) пружины висят параллельно, причем груз P подвешен в такой точке, что обе пружины растягиваются на одну и ту же величину (рис. 46 *б*). Деформацией пружин под действием собственного веса пренебречь.

Ответ: а) $U_1/U_2 = k_2/k_1$; б) $U_1/U_2 = k_1/k_2$.

Когда одна из пружин очень жесткая по сравнению с другой, практически вся потенциальная энергия в случае а) будет запасена в более мягкой, а в случае б) — в более жесткой пружинах.



а *б*

Рис. 46

2. Два протона с энергией $E = 0{,}5$ МэВ каждый летят навстречу друг другу и испытывают лобовое столкновение. Как близко они могут сойтись, если учитывать только электростатическое взаимодействие между ними?

Ответ. $r = e^2/2E$, где e — заряд протона. Для вычислений формулу целесообразно преобразовать, положив $E = eV$. Тогда $r = e/2V = 1{,}4\cdot 10^{-13}$ см ($2V = 10^6$ В). Опыты по рассеянию ядерных частиц показали, что радиус действия ядерных сил по порядку величины равен 10^{-13} см. Поэтому при расчете столкновения протонов, энергии которых

превосходят примерно 0,5 МэВ, помимо электростатических сил надо учитывать также ядерные силы.

3. Три электрона в состоянии покоя находятся в вершинах правильного треугольника со стороной $a = 1$ см. После этого они начинают двигаться под действием взаимного отталкивания. Определить предельное значение их скоростей.

Ответ. $v = \sqrt{2e^2/ma} = 2{,}2 \cdot 10^4$ см/с.

4. Решить задачу 3 для релятивистских скоростей. При каких расстояниях a можно пользоваться нерелятивистским приближением?

Ответ. $v = c \dfrac{\sqrt{2m_0c^2e^2/a + e^4/a^2}}{m_0c^2 + e^2/a}$.

Нерелятивистское приближение справедливо при

$$a \gg \frac{e^2}{m_0c^2} = 2{,}8 \cdot 10^{-13} \text{ см}.$$

5. Исходя из того требования, что произведение импульса частицы mv на расстояние a велико по сравнению с L, определить, при каких расстояниях a в задаче 3 квантовые поправки не играют роли.

Ответ. При $a \gg \dfrac{h^2}{2me^2} \approx 10^{-7}$ см.

6. Четыре электрона в состоянии покоя находятся в вершине квадрата со стороной $a = 1$ см. После этого они начинают двигаться под действием взаимного отталкивания. Определить предельное значение их скоростей.

Ответ. $v = \sqrt{\left(2 + \dfrac{\sqrt{2}}{2}\right) \dfrac{e^2}{ma}} = 2{,}6 \cdot 10^4$ см/с.

7. Материальная точка совершает одномерное финитное движение в потенциальном силовом поле между точками поворота x_A и x_B (см. рис. 45). Показать, что время движения ее от точки x_A к точке x_B равно времени обратного движения от точки x_B к точке x_A.

8. Материальная точка (например, шарик на пружине) под действием квазиупругой силы $F = -kx$ совершает колебания вдоль оси X вокруг положения равновесия. Пользуясь теоремой вириала, показать, что средние по времени значения кинетической и потенциальной энергий при таком колебании одинаковы.

9. Идеально упругий шарик движется вверх и вниз в однородном поле тяжести, отражаясь от поля по законам упругого удара. Найти связь между средними по времени значениями его кинетической $\overline{K}$ и потенциальной $\overline{U}$ энергий.

Решение. Поместим начало координат в одной из точек пола, направив ось X вертикально вверх. Тогда сила давления пола на шарик не будет влиять на значение вириала, так как она действует только в таких положениях шарика, когда $x = 0$. Надо учитывать только силу тяжести $F = -mg$ (минус потому, что сила F действует вниз, т. е. в отрицательном направлении оси X). Вириал этой силы равен $-\frac{1}{2} Fx = \frac{1}{2} mgx = \frac{1}{2} U$. По теореме вириала находим

$$\overline{K} = \frac{1}{2} \overline{U}.$$

§ 26. АБСОЛЮТНО НЕУПРУГИЙ УДАР

1. Интересным примером, где имеет место потеря механической энергии под действием диссипативных сил, является *абсолютно неупругий удар.* Так называется столкновение двух тел, в результате которого они соединяются вместе и движутся дальше как одно тело. Примером может служить попадание ружейной пули в подвижную мишень, например в ящик с песком, подвешенный на веревках. Пуля, застряв в песке, остается в ящике и движется дальше вместе с ним. Шары из пластилина или глины при столкновении обычно слипаются и затем движутся вместе. Такое столкновение также может служить примером практически абсолютно неупругого удара. Точно также столкновение двух свинцовых шаров можно с хорошим приближением рассматривать как абсолютно неупругий удар. В атомной и ядерной физике неупругие удары сопровождаются внутренними превращениями сталкивающихся частиц.

Физические явления при столкновении тел довольно сложны. Сталкивающиеся тела деформируются, возникают упругие силы и силы трения, в телах возбуждаются колебания и волны и т. д. Однако если удар неупругий, то в конце концов все эти процессы прекращаются, и в дальнейшем оба тела, соединившись вместе, движутся как единое твердое тело. Его скорость можно найти, не вдаваясь в механизм явления, а используя только закон сохранения импульса.

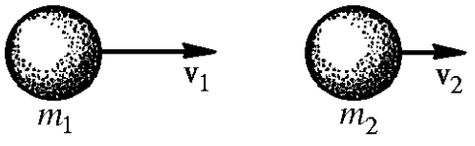


Рис. 47

Рассмотрим абсолютно неупругий удар на примере столкновения шаров. Пусть шары движутся вдоль прямой, соединяющей их центры, со скоростями $\mathbf{v}_1$ и $\mathbf{v}_2$ (рис. 47). В этом случае говорят, что удар является *центральным.* Обозначим через $\mathbf{v}$ общую скорость шаров после столкновения. Закон сохранения импульса дает

$$m_1 v_1 + m_2 v_2 = (m_1 + m_2) v,$$

где m_1 и m_2 — массы шаров. Отсюда получаем

$$v = \frac{m_1 v_1 + m_2 v_2}{m_1 + m_2}. \tag{26.1}$$

Кинетические энергии системы до удара и после удара равны соответственно

$$K_1 = \frac{1}{2} m_1 v_1^2 + \frac{1}{2} m_2 v_2^2, \quad K_2 = \frac{1}{2} (m_1 + m_2) v^2.$$

Пользуясь этими выражениями, нетрудно получить

$$K_1 - K_2 = \frac{1}{2} \mu (v_1 - v_2)^2, \tag{26.2}$$

где $\mu = \frac{m_1 m_2}{m_1 + m_2}$ — приведенная масса шаров. Таким образом, *при столкновении двух абсолютно неупругих шаров происходит потеря кинетической энергии макроскопического движения, равная половине произведения приведенной массы на квадрат относительной скорости.*

2. Неупругое столкновение тел всегда должно сопровождаться *потерей* кинетической энергии макроскопического движения. Действительно, согласно теореме Кёнига кинетическая энергия механической системы складывается из двух частей: 1) кинетической энергии движения системы как целого со скоростью ее центра масс; 2) кинетической энергии относительно движения материальных точек, на которые мысленно можно разбить систему, около ее центра масс. Обе части — как кинетические энергии — существенно положительны. Первая из них в результате столкновения тел не меняется в силу теоремы о движении центра масс. Вторая же после столкновения исчезает, так как в результате неупругого столкновения относительное движение частей системы прекращается, остается только общее движение их со скоростью центра масс. Поэтому столкновение приводит к уменьшению полной кинетической энергии макроскопического движения. Зато возрастает внутренняя энергия тела (см. § 27).

3. Нетрудно понять, почему в формулу (26.2) вошли приведенная масса и относительная скорость сталкивающихся шаров. Согласно общей формуле (25.7) потеря кинетической энергии по абсолютной величине равна работе диссипативных сил, действующих в системе во время столкновения. При вычислении этой работы, как было показано в § 24, можно одно из сталкивающихся тел считать неподвижным, а второе — движущимся относительно него. Относительное движение двух материальных точек описывается уравнением $\mu\ddot{\mathbf{r}} = \mathbf{F}$, аналогичным второму закону Ньютона. Ввиду этого работа диссипативной силы $\mathbf{F}$ за все время столкновения равна $\frac{1}{2}\mu(v_1 - v_2)^2$. Эта величина и дает убыль кинетической энергии системы за то же время.

Когда сталкиваются два тела, то разрушительное действие при столкновении зависит только от их относительной скорости $v_1 - v_2$. Кинетическая энергия, от которой зависит разрушительный эффект, равна $\frac{1}{2}\mu(v_1 - v_2)^2$. Остальная часть кинетической энергии связана с движением центра масс системы. Эта энергия при столкновении не изменяется, а потому она на разрушение не оказывает никакого влияния. Например, если сталкиваются два одинаковых автомобиля, движущихся навстречу друг другу с одной и той же скоростью v, то энергия, от которой зависит разрушение, равна

$$\frac{1}{2}\mu(v_1 - v_2)^2 = \frac{1}{2}\frac{mm}{m+m}(2v)^2 = mv^2,$$

т. е. вся кинетическая энергия тратится на разрушение. Это ясно без вычислений, так как после столкновения оба автомобиля, независимо от того, в какой мере они пострадали при аварии, должны остановиться. Тот же разрушительный эффект получится и в том случае, когда один из автомобилей неподвижен, а другой движется в направлении к нему со скоростью $2v$. Но в этом случае начальная кинетическая энергия системы составляет $\frac{1}{2}m(2v)^2 = 2mv^2$, т. е. она вдвое больше. Только половина энергии идет на разрушение.

Разрушительные эффекты при авариях, конечно, являются бедствием. Но в некоторых случаях, например при изучении превращений, претерпеваемых атомными ядрами и элементарными частицами во время столкновения, они являются целью исследования. В таких случаях стремятся к тому, чтобы разрушительные эффекты усилить. Из изложенного следует, что этого можно добиться, приводя в движение обе сталкивающиеся частицы. *При одной и той же затрате энергии наибольшее разрушение получится тогда, когда центр масс сталкивающихся частиц в лабораторной системе отсчета неподвижен.* Этот принцип используется в так называемых *ускорителях на встречных пучках.* Современные ускорители представляют дорогие и сложные технические сооружения, применяющиеся для сообщения высоких энергий заряженным частицам — электронам, протонам и пр. Они используются в ядерной физике и физике элементарных частиц для исследования различных процессов, происходящих при столкновении частиц высоких энергий. Обычно ускоренные частицы направляются на неподвижную мишень, при столкновении с которой и происходят процессы, подлежащие изучению. Тот же эффект, однако, может быть достигнут с меньшей затратой энергии, если привести в движение также саму мишень навстречу пучку. В качестве мишени используется встречный пучок ускоренных частиц. Если массы и скорости частиц в обоих пучках одинаковы, то согласно нерелятивистской механике должен получиться выигрыш в энергии в два раза. В действительности в ускорителях имеют дело с *релятивистскими пучками*, и при расчетах надо пользоваться релятивистской механикой. Оказывается, что в релятивистском случае можно получить принципиально ничем не ограниченный выигрыш в энергии, используя частицы, скорости которых приближаются к скорости света (см. т. IV).

4. Во время столкновения в системе действуют диссипативные силы, уменьшающие кинетическую энергию макроскопического движения. Поэтому применять закон сохранения энергии в его механической форме к процессам, происходящим во время удара, нельзя. Но после того как удар закончился и сталкивающиеся тела соединились в одно тело, законом сохранения энергии уже можно пользоваться (если, конечно, в дальнейшем не действуют диссипативные силы).

В качестве примера рассмотрим задачу о *баллистическом маятнике.* Он применяется для измерения скорости пуль или снарядов. Баллистический маятник обычно представляет собой подвешенный большой ящик с песком или землей, который может колебаться вокруг горизонтальной оси. Пуля или снаряд, попадая в маятник, останавливается в нем, и маятник отклоняется. Для простоты расчета будем считать маятник математическим. Процесс столкновения происходит настолько быстро, что за время столкновения маятник не успевает отклониться на заметный угол. В результате удара он только приходит в движение, и задача прежде всего заключается в том, чтобы найти скорость этого движения v непосредственно после того, как удар закончился. До удара, когда маятник находился в равновесии, внешние силы (сила веса и сила натяжения подвеса), действующие на него,

уравновешивались. Во время удара равновесие этих сил нарушается, и появляются новые силы, например силы трения. Однако во время самого удара все эти силы можно не принимать во внимание, так как они пренебрежимо малы по сравнению с силой, которая действует на маятник со стороны налетающих на него пули и снаряда. Иными словами, систему, состоящую из маятника и пули (снаряда) (рис. 48), во время удара можно считать замкнутой и применять к ней закон сохранения импульса. Из этого закона и найдется искомая скорость v, которую получит система непосредственно после удара:

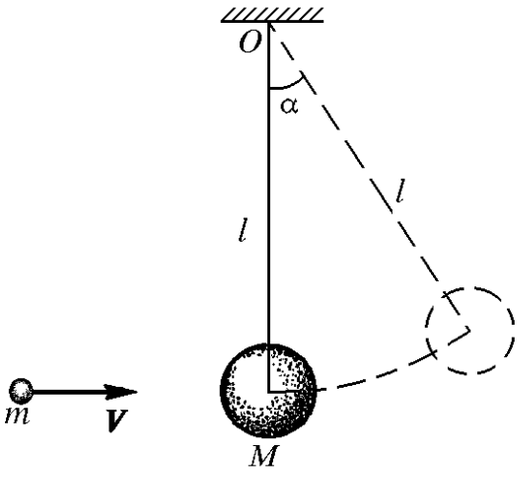


Рис. 48

$$v = \frac{m}{M+m} V,$$

где V — скорость пули до удара. После того как удар закончился, действие (внутренних) диссипативных сил прекращается. Поэтому к процессам после удара применим закон сохранения энергии. Скорость v надо рассматривать как начальную скорость, с которой начнет колебаться маятник в нижнем положении. В этом положении маятник и пуля (снаряд) обладают кинетической энергией $\frac{1}{2}(M+m)v^2$, которая при отклонении маятника переходит в потенциальную энергию $(M+m)gh$. Отсюда находится высота поднятия:

$$h = \frac{v^2}{2g} = \frac{1}{2g}\left(\frac{m}{M+m}\right)^2 V^2. \qquad (26.3)$$

Измерив высоту h, можно вычислить скорость пули V.

Было бы грубой ошибкой рассуждать следующим образом. В нижним положении (до удара) энергия системы равна кинетической энергии $mV^2/2$. При поднятии маятника эта энергия переходит в потенциальную энергию $(M+m)gh$. Такой способ рассуждения приводит к ошибочной формуле

$$h = \frac{1}{2g}\frac{m}{M+m} V^2.$$

Так как в случае баллистического маятника $m \ll M$, то эта формула дает совершенно неправильное (завышенное во много раз) значение для высоты h. Только в другом предельном случае, когда $m \gg M$, обе формулы фактически совпадают, как это ясно и без всяких вычислений. Ошибка приведенного рассуждения состоит в том, что оно не учитывает потери механической энергии при ударе.

При практических вычислениях удобно выразить высоту h через угол отклонения маятника из положения равновесия α, который легче поддается из-

мерению, чем высота h. Очевидно, $h = l(1 - \cos\alpha) = 2l\sin^2(\alpha/2)$, где l — длина маятника. Пользуясь этим выражением, формулу (26.3) нетрудно привести к виду

$$V = 2\frac{M+m}{m}\sqrt{gl}\sin\frac{\alpha}{2} \approx 2\frac{M}{m}\sqrt{gl}\sin\frac{\alpha}{2}. \qquad (26.4)$$

§ 27. ВНУТРЕННЯЯ ЭНЕРГИЯ. ОБЩЕФИЗИЧЕСКИЙ ЗАКОН СОХРАНЕНИЯ ЭНЕРГИИ

1. Потеря кинетической энергии без соответствующего увеличения потенциальной, о которой говорилось в предыдущем параграфе, происходит не только при неупругих ударах, но и во многих других процессах. Например, движения в замкнутой системе, где действуют силы трения, в конце концов прекращаются, так что запас кинетической энергии в системе уменьшается. Может происходить и потеря потенциальной энергии. Так, например, если растянуть пружину, перейдя при этом предел упругости, а затем предоставить ее самой себе, то она не возвращается в исходное состояние, в пружине сохраняется некоторое остаточное удлинение. При этом работа, которую в состоянии совершить растянутая пружина, меньше работы, затраченной на ее растяжение. Во всех подобных случаях наблюдаются потери механической энергии. Формальная макроскопическая механика объясняет эти потери тем, что энергия расходуется на работу против диссипативных сил, действующих в системе. Однако такое объяснение является чисто формальным и нефизическим, поскольку оно совсем не раскрывает физическую природу диссипативных сил.

2. Надо учесть, что всякий раз, когда наблюдается потеря механической энергии, в системе происходят какие-то внутренние изменения. Если, например, с помощью чувствительного термометра или термопары измерить температуру шаров до и после неупругого удара, то опыт покажет, что в результате удара шары немного нагрелись. То же самое происходит при трении и остаточной деформации. При положительном и интенсивном трении нагревание настолько сильное, что для его обнаружения не требуется никаких специальных приборов. Дикари добывали огонь трением одного куска дерева о другой. Если на ось мотора насадить диск из прочного картона (толщиной около 1 мм) и привести его в быстрое вращение, то можно перепилить деревянную доску, поднеся ее к краю этого вращающегося диска (картонная пила). Явление объясняется тем, что в месте контакта вращающегося картона с доской выделяется много тепла из-за трения. Дерево в этом месте сильно разогревается, обугливается и разрезается вращающимся диском. Картонный диск при этом не разрушается, так как он интенсивно охлаждается из-за быстрого вращения в окружающем воздухе. Разрез доски получается гладким и хорошо отполированным. Он имеет буроватую окраску из-за обугливания дерева при трении. Решающую роль в этом опыте играет натяжение картона, возникающее при вращении и придающее ему твердость. Вращением доски вокруг картонного диска ее распилить нельзя.

Могут быть и более сложные явления, сопровождающие потери механической энергии. Примером может служить следующая демонстрация. На вал небольшой динамомашины надет деревянный шкив, на который намотана длинная прочная нить. Нить перекинута через блок, укрепленный около потолка аудитории. К ее свободному концу подвешен груз в несколько килограммов. Вращая шкив, поднимают груз к потолку аудитории. Цепь динамомашины может замыкаться через ключ на небольшую электрическую цепочку. Если отпустить шкив, не замыкая цепи лампочки, то динамомашина не вырабатывает электрического тока. В этом случае груз падает ускоренно — потенциальная энергия груза переходит в кинетическую энергию. Если снова поднять груз и замкнуть цепь лампочки, когда он пройдет приблизительно половину пути до пола, то лампочка загорается, а движение груза и вращение динамомашины заметно затормозятся. После этого груз медленно опускается до пола с постоянной скоростью, а лампочка горит постоянным накалом во все время падения груза. Потенциальная энергия груза непрерывно уменьшается. Однако она не пропадает бесследно: динамомашина непрерывно вырабатывает электрический ток, выделяющий тепло в нити лампочки.

3. Макроскопическая механика учитывает только кинетическую энергию макроскопического движения тел и их макроскопических частей, а также их потенциальную энергию. Но она полностью отвлекается от внутреннего атомистического строения вещества. При ударе, трении и аналогичных процессах кинетическая энергия видимого движения тел не пропадает. Она только переходит в кинетическую энергию невидимого беспорядочного движения атомов и молекул вещества, а также в потенциальную энергию их взаимодействия. Эта часть энергии тела получила название *внутренней энергии.* Беспорядочное движение атомов и молекул воспринимается нашими органами чувств в виде *тепла.* Таково физическое объяснение кажущейся потери механической энергии при ударе, трении и пр.

Представление о теплоте как о беспорядочном движении атомов и молекул окончательно утвердилось во второй половине XIX века и составило эпоху в науке. Примерно тогда же в физике утвердился и взгляд на закон сохранения энергии как на *общефизический закон,* не знающий никаких исключений. Согласно этому закону *энергия никогда не создается и не уничтожается, она может только переходить из одной формы в другую.* Однако необходимо расширить понятие энергии, введя *новые формы ее*: энергию электромагнитного поля, ядерную энергию и пр. При этом необходимо заметить, что дать окончательную классификацию различных видов энергии не представляется возможным. Это можно было бы сделать, если бы окончательно были установлены все законы природы, и развитие науки, во всяком случае в ее основах, было бы окончательно завершено.

Деление энергии на кинетическую и потенциальную имеет смысл только в механике и не охватывает всех форм энергии. Кроме того, отнесение энергии к тому или иному виду часто зависит от точки зрения. Например, в макроскопической механике упругая

Так как одно из этих уравнений квадратное, а другое — линейное, то система (28.1) должна иметь два решения относительно неизвестных v_1' и v_2'. Одно из этих решений можно указать сразу, а именно $v_1' = v_1$, $v_2' = v_2$. Но это решение не удовлетворяет условию задачи. Ему соответствует случай, когда скорости шаров не изменились, т. е. шары не претерпели столкновения. Существование такого решения неизбежно. Действительно, законы сохранения импульса и энергии можно написать для двух любых состояний системы, разделенных каким-то промежутком времени Δt. Но в самих законах сохранения еще не заложено условие, что столкновение произошло. Это условие должно быть указано дополнительно. Если столкновение не произошло, то скорости шаров не могли измениться, и мы получаем решение $v_1' = v_1$, $v_2' = v_2$, указанное выше. Чтобы получить решение, относящееся к столкновению, очевидно, надо потребовать, чтобы скорости шаров *изменились*, т. е. чтобы $v_1' \neq v_1$, $v_2' \neq v_2$. Заметив это, перепишем уравнения (28.1) в виде

$$m_1(v_1' - v_1) = m_2(v_2 - v_2'), \quad m_1(v_1'^2 - v_1^2) = m_2(v_2^2 - v_2'^2).$$

Так как $v_1' - v_1$ и $v_2' - v_2$ не равны нулю, то уравнения можно поделить почленно. Это дает

$$v_1 + v_1' = v_2 + v_2'.$$

В результате задача сводится к решению системы двух линейных уравнений. Решая их, найдем единственное решение

$$v_1' = -v_1 + 2\frac{m_1v_1 + m_2v_2}{m_1 + m_2}, \quad v_2' = -v_2 + 2\frac{m_1v_1 + m_2v_2}{m_1 + m_2}, \tag{28.2}$$

удовлетворяющее условию задачи.

4. Полезно привести другой способ решения той же задачи. Он сокращает вычисления и лучше выявляет структуру окончательных формул. Рассмотрим процесс удара сначала в *системе центра масс*, т. е. в такой системе отсчета, в которой центр масс неподвижен. Относительно неподвижной системы отсчета (ее называют *лабораторной*) центр масс движется со скоростью

$$V = \frac{m_1v_1 + m_2v_2}{m_1 + m_2}. \tag{28.3}$$

Скорости в системе центра масс будем обозначать прежними буквами, но с индексом 0. Полный импульс в системе центра масс равен нулю, и законы сохранения импульса и энергии в такой системе запишутся в виде

$$\begin{gathered} m_1v_{10}' + m_2v_{20}' = m_1v_{10} + m_2v_{20} = 0, \\ \frac{1}{2}m_1v_{10}'^2 + \frac{1}{2}m_2v_{20}'^2 = \frac{1}{2}m_1v_{10}^2 + \frac{1}{2}m_2v_{20}^2. \end{gathered} \tag{28.4}$$

Эта система уравнений имеет два решения, которые могут быть указаны без вычислений. Первое решение

$$v'_{10} = v_{10}, \quad v'_{20} = v_{20}$$

не удовлетворяет условиям задачи. Годится только второе решение, а именно

$$v'_{10} = -v_{10}, \quad v'_{20} = -v_{20}.$$

Мы видим, что в системе центра масс столкновение приводит просто к изменению знака каждой из скоростей.

Перейдем теперь к лабораторной системе отсчета. Очевидно, $v_{01} = v_1 - V$, $v'_{01} = v'_1 - V$ и т. д. Поэтому

$$(v'_1 - V) = -(v_1 - V), \quad (v'_2 - V) = -(v_2 - V),$$

откуда

$$v'_1 = -v_1 + 2V, \quad v'_2 = -v_2 + 2V. \tag{28.5}$$

Подставив сюда значение для V из (28.3), придем к прежним формулам (28.2).

5. Допустим, что второй шар вначале был неподвижным ($v_2 = 0$). Тогда

$$v'_1 = \frac{m_1 - m_2}{m_1 + m_2} v_1, \quad v'_2 = \frac{2m_1}{m_1 + m_2} v_1.$$

Если $m_1 > m_2$, то первый шар будет двигаться в первоначальном направлении. При $m_1 < m_2$ он отскочит в противоположном направлении. При $m_1 = m_2$ первый шар остановится, а второй пойдет вперед со скоростью первого. Вообще, при $m_1 = m_2$ из формул (28.2) получаем

$$v'_1 = v_2, \quad v'_2 = v_1,$$

т. е. *при столкновении двух одинаковых абсолютно упругих шаров они просто обмениваются скоростями.*

Рассмотрим рад соприкасающихся одинаковых абсолютно упругих шаров, центры которых расположены вдоль одной и той же прямой линии (рис. 49). В соответствующем демонстративном опыте шары подвешиваются на нитях, а не располагаются на поверхности стола, чтобы не возникало вращение их из-за трения между шарами и поверхностью стола. Отклоним в сторону шар *1*. Ударившись о шар *2* со скоростью v, он передаст ему эту скорость, а сам остановится. С шаром *2* произойдет то же самое — при ударе о шар *3* он остановится, а шар *3* придет в движение со скоростью v. Этот процесс будет повторяться с каждым впереди находящимся шаром. В конце концов последний шар отскочит со скоростью v, а все прочие шары останутся в состоянии покоя.

Отклоним теперь два шара. При возвращении в нижнее положение они приобретут одну и ту же скорость v и, двигаясь с такой ско-

Рис. 49

ростью, ударят впереди находящийся шар (рис. 50). Оказывается, что в результате удара отскочат два последних шара со скоростью v, а все остальные шары останутся в покое. Явление можно объяснить следующим образом. Шар *2* ударяет в шар *3*. В результате этого шар *2* останавливается, а шар *3* приобретает скорость v. Однако шар *2* сразу же подвергается удару со стороны шара *1* и снова приобретает преж-

Рис. 50

нюю скорость v. Таким образом, шар *1* придет в состояние покоя, а шары *2* и *3* будут двигаться вместе со скоростью v. Повторяя это рассуждение, найдем, что затем остановится шар *2*, а начнут двигаться шары *3* и *4* и т. д. В конце концов скорость v приобретут два последних шара, а все остальные шары придут в состояние покоя. Вместо двух можно отклонить три, четыре и т. д. шара, сообщив им одну и ту же скорость v. После удара отскочит такое же количество шаров, остальные же шары останутся неподвижными.

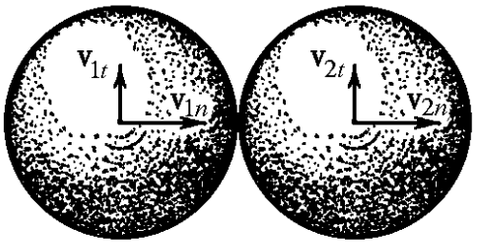


Рис. 51

6. Рассмотрим теперь *нецентральный удар* твердых упругих шаров. Так называется столкновение, когда в момент удара начальные скорости шаров не совпадают по направлению с линией центров. Разложим в момент столкновения начальную скорость каждого шара на нормальную v_n и тангенциальную v_t составляющие, т. е. составляющие вдоль линии центров и перпендикулярно к ней (рис. 51). Так же поступим с конечными скоростями шаров в момент начала их разлета. Тогда законы сохранения импульса тела и энергии запишутся в виде

$$\begin{aligned} m_1v'_{1n} + m_2v'_{2n} &= m_1v_{1n} + m_2v_{2n}, \\ m_1v'_{1t} + m_2v'_{2t} &= m_1v_{1t} + m_2v_{2t}, \\ \tfrac{1}{2}m_1(v'^2_{1n} + v'^2_{1t}) + \tfrac{1}{2}m_2(v'^2_{2n} + v'^2_{2t}) &= \tfrac{1}{2}m_1(v^2_{1n} + v^2_{1t}) + \tfrac{1}{2}m_2(v^2_{2n} + v^2_{2t}). \end{aligned} \tag{28.6}$$

Получилось всего три уравнения для определения четырех неизвестных v'_{1n}, v'_{1t}, v'_{2n}, v'_{2t}. Чтобы написать недостающее уравнение, введем предположение,

что *при столкновении шаров не возникают тангенциальные силы.* В сущности, ввести такое предположение вынуждает нас закон сохранения энергии, уже использованный при написании наших уравнений. Действительно, если бы тангенциальные скорости сталкивающихся шаров были одинаковы ($v_{1t} = v_{2t}$), то рассматриваемый случай сводился бы к случаю центрального удара, уже разобранному выше. Для этого достаточно было бы перейти в систему отсчета, в которой $v_{1t} = v_{2t}$. Поэтому без ущерба для общности мы будем предполагать, что $v_{1t} \neq v_{2t}$. Но тогда, если бы при столкновении развивались тангенциальные силы трения скольжения, механическая энергия не могла бы сохраняться. Поэтому, предполагая удар идеально упругим, мы должны считать сами шары *идеально гладкими.* При их столкновении тангенциальные силы не возникают. Если так, то не происходит также изменения тангенциальных скоростей, и к уравнениям (28.6) следует присоединить уравнения $v'_{1t} = v_{1t}$, $v'_{2t} = v_{2t}$. Тогда останутся только уравнения для нормальных скоростей, отличающиеся от уравнений (28.1) лишь обозначениями. В результате мы приходим к следующему заключению.

При столкновении гладких идеально упругих шаров их тангенциальные скорости не изменяются. Нормальные же скорости изменяются так же, как и скорости при центральном ударе. В частности, при столкновениях не изменяются состояния вращения шаров. Это было бы возможно только при наличии тангенциальных сил. *Если шары одинаковы, то при столкновении они обмениваются нормальными скоростями, тангенциальные скорости их остаются неизменными.*

7. Отметим случай, когда масса одного из шаров бесконечно велика. В этом случае скорость большего шара при столкновении вообще не изменится. Устремляя радиус этого шара к бесконечности, в пределе придем к задаче о столкновении гладкого упругого шара с *гладкой плоской стенкой.* Если связать систему отсчета с такой стенкой, то можно сказать, что при столкновении с ней тангенциальная скорость шара не меняется, а нормальная меняет знак. Это значит, что шар отражается от стенки «*зеркально*»: его скорость по модулю не изменяется, а угол падения равен углу отражения.

Удар под углом о массивную стенку сообщает ей конечный импульс, в то время как кинетическая энергия, приобретаемая стенкой, пренебрежимо мала. Например, если удар нормальный, то скорость шара v после удара становится равной $-v$, а импульс шара получает приращение $-2mv$. Чтобы общий импульс не изменился, стенка должна воспринять импульс $MV = 2mv$, проявляющийся в конечной силе удара, действующей на стенку. При этом стенка получает скорость $V = 2\frac{mv}{M}$, которая для массивной стенки бесконечно мала. Бесконечно мала и кинетическая энергия, приобретаемая стенкой. Действительно, ее можно представить в виде $\frac{MV^2}{2} = MV \cdot \frac{V}{2}$, т. е. в виде произведения конечного импульса MV на половину бесконечно малой скорости $V/2$.

ЗАДАЧИ

1. На гладком горизонтальном столе лежит шар массой m_1, соединенный с пружиной жесткости k. Второй конец пружины закреплен (рис. 52). Происходит лобовое упругое соударение этого шара с другим шаром, масса которого m_2 меньше m_1, а скорость равна v. В какую сторону будет двигаться второй шар после удара? Определить амплитуду колебаний первого шара A после соударения.

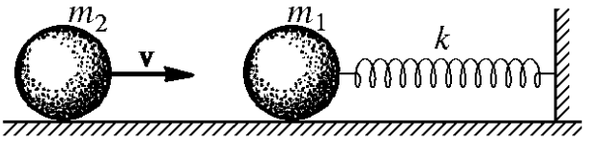


Рис. 52

Ответ. После соударения второй шар отскочит назад.

$$A = \frac{2m_2 v}{m_1 + m_2}\sqrt{\frac{m_1}{k}}.$$

2. Система состоит из двух шаров с массами m и M, соединенных между собой невесомой пружиной жесткости k (рис. 53). Третий шар массой m, движущейся вдоль оси пружины со скоростью v, претерпевает упругое столкновение с шаром m, как указано на рис. 53. Считая шары абсолютно жесткими, найти после столкновения: 1) кинетическую энергию K движения системы как целого; 2) внутреннюю энергию системы $E_{\text{вн}}$; 3) амплитуду колебаний A одного шара относительно другого . До удара система покоилась, а пружина не была деформирована. Какие шары могут рассматриваться как абсолютно жесткие?

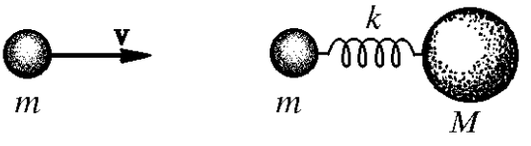


Рис. 53

Ответ. 1) $K = \dfrac{(mv)^2}{2(M+m)}$; 2) $E_{\text{вн}} = \dfrac{Mmv^2}{2(M+m)}$; 3) $A = v\sqrt{\dfrac{Mm}{k(M+m)}}$.

3. Ядра дейтерия D и трития T могут вступать в реакцию

$$\mathrm{D} + \mathrm{T} \rightarrow {}^4\mathrm{He} + \mathrm{n} + 17{,}6\ \text{МэВ},$$

в результате которой образуются нейтроны n и α-частицы, т. е. ядра гелия ^{4}He. При каждой реакции выделяется энергия 17,6 МэВ. Определить, какую энергию уносит нейтрон и какую — α-частица. Кинетические энергии, которыми обладали частицы до реакции, пренебрежимо малы.

Пояснение. Дейтерий — изотоп водорода с атомной массой 2, тритий — изотоп водорода с атомной массой 3, ^{4}He — обычный гелий с атомной массой 4.

Ответ. α-частица уносит 3,5 МэВ, нейтрон — 14,1 МэВ.

4. Ядра дейтерия D могут вступать друг с другом в реакцию, в результате которой образуется протон и ядра трития T. Каждый протон уносит кинетическую энергию 3 МэВ. Какую кинетическую энергию уносит ядро атома трития и каков общий энергетический выход реакции? Кинетические энергии, которыми обладали частицы до реакции, пренебрежимо малы.

Ответ. Ядро трития уносит энергию 1 МэВ; общий энергетический выход реакции 4 МэВ.

5. Ядра дейтерия могут вступать также в реакцию

$$\mathrm{D} + \mathrm{D} \rightarrow {}^3\mathrm{He} + \mathrm{n} + 3{,}25\ \text{МэВ}.$$

Какую энергию уносит нейтрон и какую — ядро гелия ^{3}He с атомной массой 3? Кинетические энергии, которыми обладали частицы до реакции, пренебрежимо малы.

Ответ. Нейтрон уносит энергию 2,44 МэВ, ядро ^{3}He — 0,81 МэВ.

6. В реакции

$$^3\mathrm{He} + \mathrm{D} \rightarrow {}^4\mathrm{He} + \mathrm{p}$$

получаются протоны с энергией 14,6 МэВ. Какую энергию уносит ядро гелия-4 (^{4}He) и какой общий энергетический выход реакции? Кинетические энергии, которыми обладали частицы до реакции, пренебрежимо малы.

Ответ. Ядро ^{4}He уносит энергию 3,7 МэВ, общий энергетический выход реакции 18,3 МэВ.

7. Движущаяся частица претерпевает упругое столкновение с покоящейся частицей такой же массы. Доказать, что после столкновения, если оно не было лобовым, частицы разлетятся под прямым углом друг к другу. Как будут двигаться частицы после лобового столкновения? При столкновении вращения не возникает.

Решение. Пусть $\mathbf{v}$ — скорость первой частицы до столкновения, $\mathbf{v}_1$ и $\mathbf{v}_2$ — скорости частиц после столкновения. Законы сохранения импульса и энергии дают

$$\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2, \quad v^2 = v_1^2 + v_2^2.$$

Возводя первое соотношение в квадрат и вычитая из него второе, получим $(\mathbf{v}_1\mathbf{v}_2) = 0$. Если оба вектора $\mathbf{v}_1$ и $\mathbf{v}_2$ не равны нулю, что будет при не лобовом ударе, то угол между ними равен 90°. При лобовом столкновении $v_1 = 0$, $v_2 = v$, т. е. частицы просто обмениваются скоростями.

8. При бомбардировке гелия α-частицами с энергией 1 МэВ найдено, что налетающая частица отклонилась на 60° от первоначального направления полета. Считая удар упругим, определить энергию частицы и энергию ядра отдачи.

Ответ. 1/4 МэВ и 3/4 МэВ.

9. Определить долю энергии, теряемую частицей массой m_1 при упругом столкновении ее с неподвижной частицей массой m_2, если после столкновения частица продолжает двигаться в прежнем (когда $m_1 > m_2$) или прямо противоположном (когда $m_1 < m_2$) направлениях. Показать, что доля теряемой энергии не зависит от того, какая частица движется, а какая покоится. При каком соотношении масс m_1/m_2 потеря энергии максимальна? Используя полученные результаты, объяснить, почему в ядерных реакторах для замедления нейтронов используется рассеяние их на ядрах легких (дейтерий, углерод), а не тяжелых атомов.

Ответ. $\dfrac{\Delta E}{E} = 4\dfrac{m_1 m_2}{(m_1 + m_2)^2}$. Потеря энергии максимальна при $m_1 = m_2$.

10. Определить долю энергии α, теряемую протоном при упругом рассеянии под углом 180° на протоне, дейтроне, ядре гелия и ядре углерода.

Ответ. $\alpha = 4A/(1 + A)^2$, где A — атомная масса частицы, с которой сталкивается протон:

A	1	2	3	12,
α	1	0,89	0,64	0,284,

11. Каков максимальный угол ϑ рассеяния α-частицы и дейтрона при упругом рассеянии в водороде?

Р е ш е н и е. Пусть m_1 — масса рассеиваемой частицы (α-частицы или дейтрона), v — ее скорость до рассеяния; m_2 — масса рассеивающей частицы (атома водорода); v_1 и v_2 — скорости частицы после рассеяния (рис. 54). Законы сохранения импульса и энергии дают

$$m_1 v = m_1 v_1 \cos\alpha + m_2 v_2 \cos\beta,$$

$$m_1 v_1 \sin\alpha = m_2 v_2 \sin\beta,$$

$$m_1 v^2 = m_1 v_1^2 + m_2 v_2^2.$$

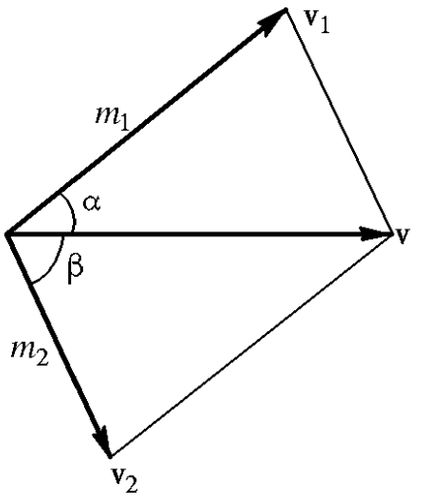


Рис. 54

Исключив отсюда угол β и скорость v_2, получим для v_1 квадратное уравнение

$$(m_1 + m_2)v_1^2 - 2m_1 v v_1 \cos\alpha + (m_1 - m_2)v^2 = 0.$$

Условие вещественности корней его, как легко видеть, имеет вид $\sin\alpha \leqslant m_2/m_1$. Максимальный угол α, удовлетворяющий этому условию, и будет равен углу ϑ. Таким образом, $\sin\vartheta = m_2/m_1$. Отсюда находим для α-частицы $\vartheta = 14°\,30'$, для дейтрона $\vartheta = 30°$.

12. Альфа-частица, летящая со скоростью v_0, испытывает упругое столкновение с неподвижным ядром и летит под углом 90° к первоначальному направлению движения. При каком соотношении масса α-частицы m и ядра M это возможно? Определить скорость α-частицы и ядра V после столкновения. Определить также угол ϑ между направлением скорости вылетающего ядра и первоначальным направлением движения α-частицы.

О т в е т. Масса α-частицы должна быть меньше массы ядра: $m < M$;

$$v = v_0\sqrt{\frac{M-m}{M+m}},$$

$$V = \frac{mv_0}{M}\sqrt{\frac{2M}{M+m}},$$

$$\operatorname{tg}\vartheta = \sqrt{\frac{M-m}{M+m}}.$$

13. Частица массой m, летящая горизонтально со скоростью V, сталкивается с невозбужденным неподвижным атомом массой M, после чего она отскакивает и летит в прямо противоположном направлении с половинной скоростью $V/2$, а атом переходит в возбужденное состояние, т. е. в состояние с более высокой внутренней энергией. Определить скорость атома v после столкновения и энергию E, которая пошла на возбуждение атома. Для каких невозбужденных атомов описанный процесс невозможен?

О т в е т. $v = \frac{3}{2}\,mV/M$, $E = \frac{3}{8}\,mV^2\left(1 - \frac{3m}{M}\right)$. Процесс невозможен, если $M < 3m$.

14. Ядра дейтерия и трития летят навстречу друг другу таким образом, что центр масс этих частиц остается неподвижным. Суммарная кинетическая энергия обеих частиц равна $E = 15$ кэВ. До какой энергии E_D надо ус-

корить ядро дейтерия, оставляя тритий неподвижным, чтобы получить тот же выход реакции? Какая энергия E_T потребуется для той же цели, если ускорять тритий?

Рассматриваемая реакция, а также реакция, о которой говорится в следующей задаче, являются основными реакциями, с помощью которых предполагается осуществить управляемую термоядерную реакцию синтеза для использования в мирных целях.

Ответ. $E_D = \dfrac{m_D + m_T}{m_T} E = \dfrac{5}{3} E = 25$ кэВ,

$$E_T = \frac{m_D + m_T}{m_D} E = \frac{5}{2} E = 37{,}5 \text{ кэВ}.$$

15. Ядро дейтерия сталкивается и вступает в реакцию с ядром трития. Предполагается осуществить этот процесс, ускорив перед столкновением лишь одну частицу до энергии $E = 20$ кэВ, оставив вторую неподвижной. Что выгоднее для осуществления реакции: ускорить легкую или тяжелую частицу? Предполагается, что удар между частицами центральный.

Ответ. Если ускорить дейтерий, то энергия

$$\frac{E}{1 + m_T/m_D} = 8 \text{ кэВ},$$

связанная с движением центра масс, не может принимать участие в реакции. В случае ускорения трития эта энергия равна

$$\frac{E}{1 + m_D/m_T} = 12 \text{ кэВ}.$$

Выгоднее ускорять дейтерий. Выигрыш в энергии в этом случае по сравнению с другим будет

$$\frac{m_T - m_D}{m_T + m_D} E = \frac{1}{5} E = 4 \text{ кэВ}.$$

16. Первая искусственная ядерная реакция

$$^{14}\mathrm{N} + {}^{4}\mathrm{He} = {}^{17}\mathrm{O} + \mathrm{p}$$

наблюдалась Резерфордом в 1919 г. Она идет с поглощением энергии $E = 1{,}13$ МэВ. Какую минимальную энергию E_0 надо сообщить в лабораторной системе α-частице (т. е. ядру атома гелия), чтобы при бомбардировке неподвижной мишени из $^{14}\mathrm{N}$ указанная реакция могла пойти?

Решение. Обозначим через p_0 импульс α-частицы до столкновения. В результате столкновения импульс не изменяется. С ним связана кинетическая энергия движения центра масс

$$K_{\text{ц. м}} = \frac{p_0^2}{2(m_{He} + m_N)} = \frac{m_{He}}{m_{He} + m_N} E_0,$$

которая также не изменяется, а потому никак не участвует в ядерных превращениях.

Следовательно, искомая энергия найдется из условия

$$E_0 = E + K_{\text{ц. м}} = E + \frac{m_{He}}{m_{He} + m_N} E_0,$$

откуда

$$E_0 = \frac{m_{He} + m_N}{m_N} E = 1{,}45 \text{ МэВ}.$$

17. *Пороговой энергией* $E_{пор}$ или *порогом* ядерной реакции называется такая энергия бомбардирующей частицы, что ядерная реакция при неподвижной мишени может идти только тогда, когда энергия E бомбардирующей частицы равна или превосходит $E_{пор}$, а при $E < E_{пор}$ реакция невозможна. Пороговая энергия ядерной реакции $^7\text{Li} + \text{p} \rightarrow {}^7\text{Be} + \text{n}$ (литий неподвижен) равна $E_{пор} = 1{,}88$ МэВ. При каких энергиях бомбардирующих протонов E_p нейтроны в такой реакции могут лететь назад от литиевой мишени?

Р е ш е н и е. Минимальное значение искомой энергии протона E_p соответствует лобовому столкновению, когда все частицы до и после столкновения движутся вдоль одной и той же прямой. Поэтому можно ограничиться только такими столкновениями.

Допустим сначала, что энергия бомбардирующего протона равна пороговой $E_{пор}$. Тогда получающиеся в результате реакции ядро Be и нейтрон в системе центра масс должны находиться в состоянии покоя, а потому в лабораторной системе двигаться вперед с одинаковыми скоростями. При таком движении они уносят кинетическую энергию

$$E_0 = \frac{P_{пор}^2}{2(m_{Be} + m_n)},$$

где $P_{пор}$ — импульс протона, соответствующий пороговой энергии

$$E_{пор} = \frac{P_{пор}^2}{2m_p}.$$

Разность этих двух энергий

$$E_{пор} - E_0 = \frac{m_{Be} + m_n - m_p}{m_{Be} + m_n} E_{пор} \tag{28.7}$$

затрачивается на ядерную реакцию.

Найдем теперь энергию бомбардирующего протона E_p, при которой получаются нейтроны в состоянии покоя, а ядра бериллия летят вперед. Если P_p — импульс протона до реакции, то

$$E_p = \frac{P_p^2}{2m_p},$$

а кинетическая энергия образовавшегося ядра бериллия $E_{Be} = P_p^2/2m_{Be}$. Разность этих энергий

$$E_p - E_{Be} = \frac{1}{2} P_p^2 \left(\frac{1}{m_p} - \frac{1}{m_{Be}} \right) = \frac{m_{Be} - m_p}{m_{Be}} E_p \tag{28.8}$$

идет на ядерную реакцию, а потому равна величине (28.7). Приравнивая выражения (28.7) и (28.8), находим

$$E_p = \frac{m_{Be}(m_{Be} + m_n - m_p)}{m_{Be}^2 - m_p^2} E_{пор},$$

или, пренебрегая различием масс протона и нейтрона,

$$E_{\text{р}} = \frac{m_{\text{Be}}^2}{m_{\text{Be}}^2 - m_{\text{р}}^2} E_{\text{пор}} = \frac{49}{48} E_{\text{пор}} = 1{,}92 \text{ МэВ}.$$

При больших энергиях появятся нейтроны, летящие назад.

§ 29. СИЛЫ И ПОТЕНЦИАЛЬНАЯ ЭНЕРГИЯ

1. *Взаимодействие тел можно описывать либо с помощью сил, либо с помощью потенциальной энергии как функции координат взаимодействующих частиц.* В макроскопической механике применимы оба способа. Первый способ обладает несколько большей общностью, так как он применим и к таким силам (например, силам трения), для которых нельзя ввести потенциальную энергию. Второй же способ применим только в случае консервативных сил. Но в квантовой механике, имеющей дело с явлениями микромира, диссипативных сил нет, и в ней для описания взаимодействия частиц применяется исключительно второй способ. В уравнения движения квантовой механики силы не входят, а входит лишь потенциальная энергия взаимодействующих частиц. Разумеется, в этом параграфе вопрос рассматривается только в рамках макроскопической механики.

2. Зная действующие силы как функции координат материальных точек системы, можно вычислить ее потенциальную энергию. Эта задача решается *интегрированием.* Простейшие примеры на такое вычисление были приведены в § 25. Можно поставить и обратную задачу: вычислить действующие силы по заданной потенциальной энергии как функции координат взаимодействующих материальных точек. Эта задача решается с помощью более простой математической операции — *дифференцирования.* Рассмотрим сначала отдельную материальную точку, находящуюся в силовом поле каких-то неподвижных тел. Если силы консервативны, то можно ввести потенциальную энергию U, которой обладает материальная точка в рассматриваемом силовом поле. Величина U будет функцией радиуса-вектора $\mathbf{r}$ этой точки или ее координат x, y, z. Пусть точка претерпела произвольное бесконечно малое перемещение $d\mathbf{r}$. Если $\mathbf{F}$ — сила, действующая на нее, то работа этой силы при таком перемещении будет равна убыли потенциальной энергии:

$$\mathbf{F}\, d\mathbf{r} = -dU. \tag{29.1}$$

Это равенство справедливо, каково бы ни было перемещение $d\mathbf{r}$. Поэтому если функция $U(\mathbf{r})$ известна, то оно полностью определяет силу $\mathbf{F}$. В самом деле, чтобы найти вектор $\mathbf{F}$, достаточно определить его проекции F_x, F_y, F_z на координатные оси прямо-

угольной системы координат. В этих проекциях уравнение (29.1) запишется так:

$$F_x\,dx + F_y\,dy + F_z\,dz = -dU. \tag{29.2}$$

Допустим, что смещение происходит вдоль какой-либо одной координатной оси, например оси X. Тогда

$$F_x\,dx = -(dU)_{y,z},$$

и, следовательно,

$$F_x = -\left(\frac{dU}{dx}\right)_{y,z}.$$

Индексы y, z означают, что при смещении, а следовательно, и при дифференцировании координаты y и z должны оставаться постоянными. Иными словами, $U(x, y, z)$ при дифференцировании должна рассматриваться как функция одного аргумента x; остальные два аргумента, y и z, являются *параметрами,* которые при дифференцировании по x должны оставаться постоянными. Величины, получающиеся в результате такого дифференцирования, называются *частными производными* функции U. Они обозначаются символом ∂, в отличие от символа d, применяемого при дифференцировании функций одного независимого переменного. Аналогичные соображения справедливы и для проекций силы вдоль остальных двух осей Y и Z. Таким образом,

$$F_x = -\frac{\partial U}{\partial x}, \quad F_y = -\frac{\partial U}{\partial y}, \quad F_z = -\frac{\partial U}{\partial z}. \tag{29.3}$$

Если функция $U(x, y, z)$ известна, то нахождение составляющих F_x, F_y, F_z сводится к вычислению ее частных производных по координатам. Разумеется, формулы (29.3) относятся только к случаю консервативных сил.

Приведем пример. Измеряя потенциальную энергию растянутой спиральной пружины, нашли, что она определяется выражением $U = {}^1/_2\,kx^2$, где x — удлинение пружины, а k — постоянная. Направим ось X вдоль оси пружины, закрепив один конец ее, а другой будем удерживать рукой. Тогда U будет функцией только одной координаты x. Растянутая пружина действует на руку с силой

$$F = -\frac{\partial U}{\partial x} = -\frac{d}{dx}\left(\frac{1}{2}kx^2\right) = -kx.$$

Знак минус указывает, что сила F направлена в сторону, противоположную смещению, т. е. является *силой притяжения.*

3. Три формулы (29.3) можно объединить в одну векторную формулу. С этой целью умножим эти формулы на единичные век-

торы координатных осей $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ и сложим. В результате получим

$$\mathbf{F} = -\operatorname{grad} U, \tag{29.4}$$

где символом grad U обозначена сумма

$$\operatorname{grad} U \equiv \frac{\partial U}{\partial x}\mathbf{i} + \frac{\partial U}{\partial y}\mathbf{j} + \frac{\partial U}{\partial z}\mathbf{k}. \tag{29.5}$$

Согласно соотношению (29.4) эта сумма является вектором. Вектор, определяемый соотношением (29.5), называется *градиентом скаляра* U. Для него, наряду с обозначением grad U, применяется также обозначение ∇U. Здесь ∇ («набла») означает символический вектор или *оператор*

$$\nabla = \frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial y}\mathbf{j} + \frac{\partial}{\partial z}\mathbf{k}, \tag{29.6}$$

называемый *оператором Гамильтона* (по имени ирландского математика и физика Уильяма Гамильтона (1805—1865)) или набла-оператором. Таким образом, ∇U формально может рассматриваться как произведение символического вектора ∇ на скаляр U. Понятно, что можно говорить о градиенте не только функции U, но и любой скалярной функции координат. Понятие градиента широко применяется в самых разнообразных вопросах физики и математики.

Для уяснения геометрического смысла градиента полезно ввести *поверхности уровня,* т. е. такие поверхности, на которых скаляр U остается постоянным. Пусть S — одна из таких поверхностей и пусть она проходит через точку пространства *1*, в которой ищется grad U (рис. 55). Поместим в этой точке начало координат. Ось X направим по нормали к поверхности уровня $U = \text{const}$, проведя единичный вектор $\mathbf{i}$ в сторону возрастания U. Координатные оси Y и Z расположатся в плоскости, касательной к поверхности уровня $U = \text{const}$. Ясно, что при таком выборе координатных осей частные производные $\frac{\partial U}{\partial y}$ и $\frac{\partial U}{\partial z}$ в рассматриваемой точке пространства обратятся в нуль, так что в формуле (29.5) остается одно только первое слагаемое: $\operatorname{grad} U = \frac{\partial U}{\partial t}\mathbf{i}$. Изменим теперь обозначения. Единичный вектор нормали к поверхности уровня $U = \text{const}$ обозначим символом $\mathbf{n}$, а расстояние между двумя бесконечно близкими поверхностями уровня U и $U + dU$, измеренное вдоль нормали, т. е. расстояние между точками *1* и *2*, — символом dn. Тогда, очевидно, $\frac{\partial U}{\partial x} = \frac{\partial U}{\partial n}$. Эту величину чаще обозна-

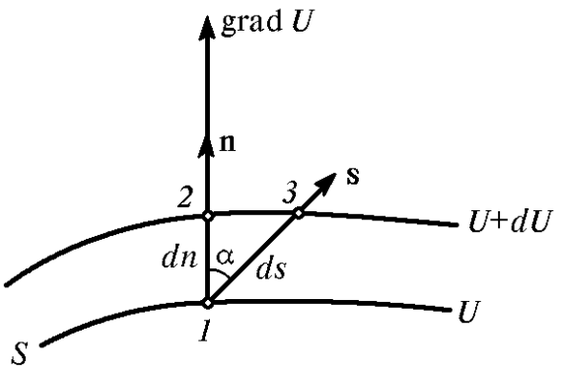


Рис. 55

чают через $\frac{\partial U}{\partial n}$ и называют *производной скаляра U в направлении нормали к поверхности уровня.* В этом направлении величина U, очевидно, изменяется наиболее быстро. Таким образом, в новых обозначениях формула (29.5) примет вид

$$\operatorname{grad} U = \frac{\partial U}{\partial n}\,\mathbf{n}. \tag{29.7}$$

Отсюда видно, что *градиент функции U есть вектор, направленный по нормали к поверхности уровня $U = \text{const}$ в сторону возрастания U; его длина численно равна производной по нормали функции U к той же поверхности.* Преимущество такого определения по сравнению с определением (29.5) состоит в том, что оно *инвариантно*, т. е. содержит только величины и понятия, имеющие непосредственный геометрический смысл, и не содержит ничего такого, что вносится случайным выбором координатной системы.

Отметим еще одну простую, но важную формулу. Проведем через точку *1* (см. рис. 55) отрезок прямой *13* под углом α к нормали n. Точку *3* возьмем на поверхности уровня $U + dU$. Длину этого отрезка обозначим через ds. Так как точка *3* лежит на той же поверхности уровня, что и точка *2*, то приращения функции U на отрезках *12* и *13* будут одни и те же. Так как сами отрезки бесконечно малы, то участки поверхностей уровня, через которые они проходят, могут считаться плоскими, а потому $ds = \frac{dn}{\cos\alpha}$. На этом основании

$$\frac{dU}{ds} = \frac{dU}{dn}\cos\alpha,$$

или в иных обозначениях

$$\frac{\partial U}{\partial s} = \frac{\partial U}{\partial n}\cos\alpha = \frac{\partial U}{\partial n}\,(\mathbf{ns}),$$

где $\mathbf{s}$ — единичный вектор в направлении отрезка *13*. Величина $\frac{\partial U}{\partial s}$ называется производной функции U в том же направлении. Учтя определение градиента (29.7), получим

$$\frac{\partial U}{\partial s} = (\mathbf{s}\operatorname{grad} U). \tag{29.8}$$

Формула эта справедлива независимо от конкретного смысла функции U. Если U является потенциальной энергией материальной точки, то с учетом (29.4) формула принимает вид

$$\frac{\partial U}{\partial s} = -\,(\mathbf{Fs}),$$

или

$$F_s = -\frac{\partial U}{\partial s}, \tag{29.9}$$

что, конечно, легко получить и прямо из (29.1).

4. Формулы (29.3) тривиальным образом обобщаются на случай произвольной системы материальных точек с одними только консервативными силами. В этом случае потенциальная энергия U является функцией координат всех взаимодействующих точек. Вместо (29.3) следует писать

$$F_{ix} = -\frac{\partial U}{\partial x_i}, \quad F_{iy} = -\frac{\partial U}{\partial y_i}, \quad F_{iz} = -\frac{\partial U}{\partial z_i}. \tag{29.10}$$

Здесь x_i, y_i, z_i — координаты i-й материальной точки системы, а F_{ix}, F_{iy}, F_{iz} — компоненты действующей на нее силы. Номер i может пробегать все возможные значения, так что формулы (29.10) справедливы для каждой точки системы.

5. Закон сохранения энергии в механике является следствием уравнения движения Ньютона. Можно ли, наоборот, вывести уравнение движения Ньютона из механического закона сохранения энергии? На этот вопрос следует ответить отрицательно. Уравнение, выражающее сохранение энергии, является *скалярным,* в то время как уравнение движения есть *векторное* и эквивалентно *трем независимым числовым уравнениям.* Ясно, что одного скалярного уравнения недостаточно для вывода из него трех независимых числовых уравнений. Но если движение одномерное, то при некоторых дополнительных предположениях из закона сохранения энергии можно вывести уравнение движения Ньютона. Допустим, что материальная точка движется вдоль какой-то фиксированной линии под действием одних только консервативных сил. По закону сохранения энергии $\frac{1}{2} mv^2 + U = \text{const}$. Потенциальная энергия U при таком движении может рассматриваться как функция только расстояния s, измеренного вдоль траектории. Дифференцируя последнее соотношение по времени, получим

$$mv\dot{v} + \frac{dU}{ds}\dot{s} = 0,$$

или, учитывая соотношения $v = \dot{s}$ и $F_s = -\frac{dU}{ds}$,

$$v(m\dot{v} - F_s) = 0. \tag{29.11}$$

Отсюда после сокращения на v получается уравнение движения Ньютона.

Необходимо, однако, отметить, что уравнение Ньютона в механике, обладает большей общностью, чем закон сохранения энергии. Во-первых, приведенный вывод справедлив только для консервативных сил. Во-вторых, при выводе в уравнении (29.11) производилось сокращение на v. Поэтому необходимо ввести *дополнительное предположение,* не содержащееся в самом законе сохранения энергии, что $v \neq 0$. Уравнение (29.11) имеет два решения, которые удовлетворяют условию сохранения энергии. Одно из них, а именно $v = 0$, было отброшено. Закон сохранения энергии для этого не дает осно-

ваний. Однако решение $v = 0$ не согласуется с уравнением Ньютона, если только сила F не обращается в нуль.

6. Используя понятие потенциальной энергии, можно выразить *условие равновесия* механической системы и его *устойчивости.* Рассмотрим сначала систему взаимодействующих материальных точек, на которую не наложены никакие связи. Пусть все действующие силы консервативны. Тогда их составляющие можно представить формулами (29.10). В состоянии равновесия все силы, а с ними и все первые производные потенциальной энергии U по координатам должны обращаться в нуль. Отсюда следует, что *для равновесия необходимо, чтобы потенциальная энергия была стационарна.* Стационарность означает, что при всяком выводе системы из расстояния равновесия, когда координаты материальных точек получают бесконечно малые приращения δx_1, δy_1, ..., δz_n, функция U остается почти постоянной. Точнее, приращения функции U при таких бесконечно малых изменениях координат являются бесконечно малыми *более высокого порядка*, чем приращения самих координат. В частности, *система будет находиться в равновесии, если потенциальная энергия U экстремальна, т. е. минимальна или максимальна.*

Если потенциальная энергия минимальна, то равновесие будет устойчивым. Действительно, пусть U_0 — значение потенциальной энергии в состоянии равновесия. По условию теоремы можно найти малую окрестность вблизи состояния равновесия, в которой разность $U - U_0$ положительна. Выберем эту окрестность так, чтобы было $0 < U - U_0 < \varepsilon$, где ε — некоторое положительное число, которое может быть взято сколь угодно малым. Выведем теперь систему из состояния равновесия, сообщив ей кинетическую энергию $K_0 < \varepsilon$. Затем предоставим систему самой себе. Свободное движение системы будет подчиняться закону сохранения энергии $K + U = K_0 + U_0$ или $U - U_0 = K_0 - K$. Отсюда видно, что $U - U_0 < \varepsilon$, так как кинетическая энергия K не может быть отрицательной.

Следовательно, система без внешних воздействий не может выйти за пределы области $0 < U - U_0 < \varepsilon$ и будет совершать в ней финитное движение. Это означает, что равновесие системы при минимуме потенциальной *энергии устойчиво*, точнее, *устойчиво по отношению к бесконечно малым возмущениям.*

Изложенное остается справедливым и при наличии диссипативных сил типа жидкого трения, а также гироскопических сил. Действительно, в состоянии равновесия, когда все материальные точки покоятся, такие силы равны нулю. Поэтому необходимое условие равновесия, требующее стационарности потенциальной энергии U, остается в силе. Сохраняет силу и доказательство устойчивости равновесия при минимуме U. Только равенство, выражающее закон сохранения энергии, при наличии диссипативных сил в доказательст-

ве следует заменить неравенством $(K+U)-(K_0+U_0)<0$ или $U-U_0<K_0-K$. Это только усилит дальнейшие заключения. *Диссипативные силы делают равновесие еще более устойчивым.* Если систему вывести из состояния равновесия и затем предоставить самой себе, то диссипативные силы в конце снова вернут систему в состояние равновесия.

Причина устойчивости равновесия при минимуме U выявится особенно наглядно, если рассмотреть всего одну материальную точку, могущую совершать одномерное движение. В этом случае график функции U имеет вид потенциальной ямы (аналогичной той, которая представлена на рис. 45). В состоянии равновесия материальная точка «лежит на дне потенциальной ямы». Никакие силы на нее в этом положении не действуют. При смещении точки в сторону, как легко видеть, появляется сила, направленная к положению равновесия и стремящаяся вернуть точку в это положение. Если же точка находится в равновесии там, где потенциальная энергия максимальна (т. е. «лежит на вершине потенциальной горы», например в точке N на рис. 44), то при ее смещении в сторону появляется сила, направленная от положения равновесия. Такая сила еще дальше уведет точку от этого положения. Равновесие будет неустойчивым. *Равновесие всякой механической системы, вообще говоря, неустойчиво, если потенциальная энергия максимальна.*

Изложенные результаты можно распространить и на системы, свобода перемещения которых ограничена *наложенными связями.* Надо только потребовать, чтобы связи были *идеальными*, т. е. такими, которые не производят работы при любых возможных перемещениях системы. Примером может служить идеально гладкий шарик, надетый на идеально твердую и гладкую спицу, которая задает направление возможного перемещения шарика. Сила, действующая на шарик со стороны спицы, перпендикулярна к направлению возможного перемещения и работы не производит.

При наличии связей условия равновесия материальной точки принимают вид

$$\begin{aligned} -\frac{\partial U}{\partial x}+R_x&=0,\\ -\frac{\partial U}{\partial y}+R_y&=0,\\ -\frac{\partial U}{\partial z}+R_z&=0, \end{aligned} \tag{29.12}$$

где $\mathbf{R}$ — *реакция связей*, т. е. сила, с которой связи действуют на рассматриваемую материальную точку. В целях краткости мы провели рассуждения для одной материальной точки. В случае системы изменится только число уравнений, но сами рассуждения останутся без изменений. Пусть δ_x, δ_y, δ_z — *возможные перемещения* материальной точки вдоль координатных осей. Умножая на них уравнения

(29.12), складывая и принимая во внимание, что реакции связей работы не производят, получим

$$\delta U \equiv \frac{\partial U}{\partial x}\,\delta x + \frac{\partial U}{\partial y}\,\delta y + \frac{\partial U}{\partial z}\,\delta z = 0.$$

Таково необходимое условие равновесия. Оно означает, что *в состоянии равновесия потенциальная энергия U стационарна.* Не изменятся и рассуждения относительно устойчивости равновесия, которые были приведены выше. Иллюстрацией может служить тяжелый шарик, помещенный на дно сферической чаши (устойчивое равновесие), или в вершину выпуклой поверхности (неустойчивое равновесие). *При наличии сил сухого трения стационарность потенциальной энергии U для равновесия не необходима.* Примером может служить равновесие бруска, лежащего на наклонной плоскости.

ГЛАВА V

МОМЕНТ ИМПУЛЬСА

§ 30. МОМЕНТ СИЛЫ И МОМЕНТ ИМПУЛЬСА ОТНОСИТЕЛЬНО НЕПОДВИЖНОГО НАЧАЛА

1. Важные законы механики связаны с понятиями *момента импульса* и *момента силы.* Следует различать и никоим образом не смешивать друг с другом моменты этих векторов *относительно точки* и *относительно оси.* Момент вектора относительно точки и относительно оси — разные понятия, хотя и связанные между собой. Момент вектора относительно точки сам есть *вектор.* Момент того же вектора относительно оси есть *проекция* на эту ось его момента относительно точки, лежащей на той же оси. Таким образом, момент вектора относительно оси уже не является вектором. Начнем с рассмотрения моментов относительно точки.

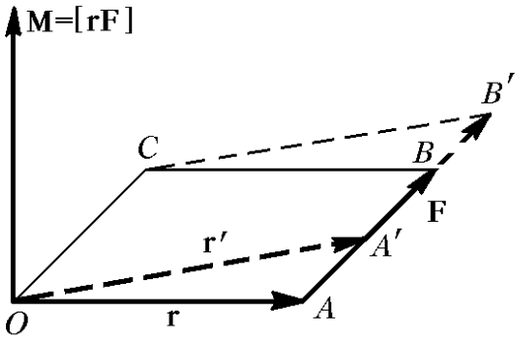


Рис. 56

Пусть O — какая-либо точка, относительно которой рассматривается момент вектора силы или вектора импульса. Ее называют *началом* или *полюсом.* Обозначим буквой $\mathbf{r}$ радиус-вектор, проведенный из этой точки к точке приложения силы $\mathbf{F}$ (рис. 56). *Моментом силы* $\mathbf{F}$ *относительно точки* O называется векторное произведение радиуса-вектора $\mathbf{r}$ на силу $\mathbf{F}$:

$$\mathbf{M} = [\mathbf{rF}]. \tag{30.1}$$

Из этого определения непосредственно следует, что *момент* $\mathbf{M}$ *не изменится, если точку приложения силы* $\mathbf{F}$ *перенести в любую другую точку, расположенную на линии действия силы.* Действительно, если точку приложения силы перенести из A в A', то параллелограмм $OABC$ перейдет в параллелограмм $OA'B'C$. Оба параллелограмма имеют общее основание OC и общую высоту. Поэтому их площади равны, что и доказывает наше утверждение.

Моментом $\mathbf{M}$ *нескольких сил относительно точки* называется сумма моментов этих сил относительно той же точки. Если линии действия сил $\mathbf{F}_1$, $\mathbf{F}_2$, ... пересекаются в одной точке A, то по доказанному при вычислении момента $\mathbf{M}$ этих сил точки приложения

их можно перенести в точку A. Если $\mathbf{r}$ — радиус-вектор точки A, то, следовательно,

$$\mathbf{M} = [\mathbf{r}\mathbf{F}_1] + [\mathbf{r}\mathbf{F}_2] + \ldots = [\mathbf{r}\mathbf{F}], \qquad (30.2)$$

где $\mathbf{F} = \mathbf{F}_1 + \mathbf{F}_2 + \ldots$ — геометрическая сумма сил $\mathbf{F}_1$, $\mathbf{F}_2$, ..., называемая их *равнодействующей.* Таким образом, в рассматриваемом случае *момент всех действующих сил относительно некоторой точки равен моменту их равнодействующей относительно той же точки.* Это справедливо и для параллельных сил, так как последние можно рассматривать как предельный случай пересекающихся сил (в бесконечно удаленной точке).

Доказанная теорема распространяется и на случай любой системы сил $\mathbf{F}_1$, $\mathbf{F}_2$, ..., линии действия которых лежат в одной плоскости. Действительно, при вычислении момента $\mathbf{M}$ силы $\mathbf{F}_1$ и $\mathbf{F}_2$ можно заменить их равнодействующей $\mathbf{F}_{12} = \mathbf{F}_1 + \mathbf{F}_2$; силы $\mathbf{F}_{12}$ и $\mathbf{F}_3$ можно заменить их равнодействующей $\mathbf{F}_{12} + \mathbf{F}_3 = \mathbf{F}_{123} = \mathbf{F}_1 + \mathbf{F}_2 + \mathbf{F}_3$ и т. д. В результате задача сведется к вычислению момента двух сил, лежащих в одной плоскости, а линии действия таких сил пересекаются в одной точке. Следовательно, *если линии действия сил лежат в одной плоскости, то момент этих сил относительно точки равен моменту их равнодействующей относительно той же точки.*

Отметим частный случай двух равных параллельных сил $\mathbf{F}_1$ и $\mathbf{F}_2$, направленных в противоположные стороны. Такие силы образуют так называемую *пару сил.* В этом случае

$$\mathbf{M} = [\mathbf{r}_1\mathbf{F}_1] + [\mathbf{r}_2\mathbf{F}_2] = [(\mathbf{r}_1 - \mathbf{r}_2)\,\mathbf{F}_1] = [(\mathbf{r}_2 - \mathbf{r}_1)\,\mathbf{F}_2],$$

т. е. *момент пары сил равен моменту одной из этих сил относительно точки приложения другой.* Момент пары сил, очевидно, не зависит от выбора точки, относительно которой он берется. В частности, если равные и противоположно направленные силы $\mathbf{F}_1$ и $\mathbf{F}_2$ действуют вдоль одной и той же прямой, то они коллинеарны с вектором $(\mathbf{r}_2 - \mathbf{r}_1)$, а потому момент $\mathbf{M}$ обращается в нуль.

Отметим, еще, что *момент произвольной системы сил может быть сведен к сумме момента их равнодействующей и момента какой-то пары сил.* Этим результатом мы нигде не будем пользоваться. При желании читатель сам может доказать его без особого труда или обратиться к курсам теоретической механики. Более существенно всегда иметь в виду, что перенос точки приложения силы вдоль линии ее действия, равно как и любая замена исходной системы сил на другую с тем же моментом и равнодействующей, приводит к эквивалентной системе сил только в том случае, когда рассматриваемое тело можно (приближенно) рассматривать как *идеально твердое.* Действительно, твердое тело обладает шестью степенями свободы. Для описания его движе-

ния достаточно двух векторных уравнений: уравнения движения центра масс (19.3) и выводимого ниже уравнения моментов (30.5). Во всех остальных случаях, например при рассмотрении движения деформированных тел, системы сил с одинаковыми равнодействующими и моментами не эквивалентны.

Аналогично определяется *момент импульса* p материальной точки относительно точки или *полюса* O. Так называется векторное произведение

$$\mathbf{L} = [\mathbf{rp}]. \tag{30.3}$$

Для системы материальных точек моментом импульса относительно некоторого начала O называется сумма моментов этих точек относительно того же начала.

2. Целесообразность введения моментов импульса и силы оправдывается тем, что они связаны между собой важным соотношением, которое называется *уравнением моментов*. Предположим сначала, что начало O неподвижно. В случае одной материальной точки, дифференцируя выражение (30.3) по времени, получим $\dot{\mathbf{L}} = [\dot{\mathbf{r}}\mathbf{p}] + [\mathbf{r}\dot{\mathbf{p}}]$. Но при неподвижном начале O импульс частицы $\mathbf{p}$ коллинеарен с ее скоростью $\mathbf{v} = \dot{\mathbf{r}}$. Кроме того, $\dot{\mathbf{p}} = \mathbf{F}$. Значит, $[\mathbf{r}\dot{\mathbf{p}}] = [\mathbf{rF}] = \mathbf{M}$. В результате имеем

$$\dot{\mathbf{L}} = \mathbf{M}. \tag{30.4}$$

Это и есть уравнение моментов для одной материальной точки. Оно справедливо как в ньютоновской, так и в релятивистской механике, так как при его выводе постоянство массы материальной точки не использовалось.

Распространим теперь уравнение (30.4) на систему материальных точек. Для этого напишем уравнение (30.4) для каждой точки, понимая под $\mathbf{M}$ момент всех действующих на нее сил, как внутренних, так и внешних. Затем сложим все эти уравнения. При таком сложении моменты внутренних сил исключатся. Действительно, внутренние силы входят в систему попарно: сила $\mathbf{F}_{ik}$, с которой k-я точка действует на i-ю, равна силе $\mathbf{F}_{ki}$, с которой i-я точка действует на k-ю. Эти силы направлены противоположно и действуют вдоль одной и той же прямой. Момент таких двух сил, а значит и моменты всех внутренних сил равны нулю. В результате получается уравнение моментов для системы материальных точек:

$$\dot{\mathbf{L}} = \mathbf{M}_{\text{внеш}}, \tag{30.5}$$

т. е. *производная по времени от момента импульса системы материальных точек относительно произвольного неподвижного начала равна геометрической сумме моментов всех внешних сил относительно того же начала* O.

3. *Если момент внешних сил относительно неподвижного начала* O *равен нулю, то момент импульса системы относительно того же начала остается постоянным во времени.* Это положение называет-

$\mathbf{F}_1 = -\mathbf{F}_2$, а из закона сохранения момента импульса

$$[\mathbf{r}_1\mathbf{p}_1] + [\mathbf{r}_2\mathbf{p}_2] = \text{const}.$$

Дифференцируя по времени это уравнение, получим

$$[\mathbf{r}_1\dot{\mathbf{p}}_1] + [\mathbf{r}_2\dot{\mathbf{p}}_2] = 0,$$

или

$$[\mathbf{r}_1\mathbf{F}_1] + [\mathbf{r}_2\mathbf{F}_2] = 0.$$

Так как $\mathbf{F}_1 = -\mathbf{F}_2$, то

$$[(\mathbf{r}_1 - \mathbf{r}_2)\mathbf{F}_1] = 0.$$

Отсюда следует, что векторы $\mathbf{r}_1 - \mathbf{r}_2$ и $\mathbf{F}_1$ коллинеарны. Коллинеарны также векторы $\mathbf{r}_1 - \mathbf{r}_2$ и $\mathbf{F}_2$. Это значит, что силы $\mathbf{F}_1$ и $\mathbf{F}_2$ направлены вдоль прямой, соединяющей взаимодействующие материальные точки.

4. Момент сил и момент импульса зависят не только от модуля и направления этих векторов, но и от положения начала. Оба момента, вообще говоря, изменятся, если перейти к новому началу. Пусть O и O' — два неподвижных начала. Радиусы-векторы $\mathbf{r}$ и $\mathbf{r}'$ одной и той же точки относительно этих начал связаны соотношением

$$\mathbf{r} = \mathbf{r}' - \mathbf{R},$$

где $\mathbf{R} = \overrightarrow{O'O}$ — радиус-вектор начала O относительно O'. Написав выражения для моментов импульсов каждой материальной точки системы и просуммировав эти выражения по всем материальным точкам, получим

$$\sum [\mathbf{r}m\mathbf{v}] = \sum [\mathbf{r}'m\mathbf{v}] - [\mathbf{R}\sum m\mathbf{v}],$$

или

$$\mathbf{L} = \mathbf{L}' - [\mathbf{R}\mathbf{p}], \tag{30.6}$$

где $\mathbf{p}$ — полный импульс системы, $\mathbf{L}$ и $\mathbf{L}'$ — моменты ее импульса относительно начала O и O' соответственно. Если импульс $\mathbf{p}$ равен нулю, то $\mathbf{L} = \mathbf{L}'$. В этом случае вектор момента импульса системы не зависит от выбора начала.

Аналогично

$$\mathbf{M} = \mathbf{M}' - [\mathbf{R}\mathbf{F}], \tag{30.7}$$

где $\mathbf{M}$ и $\mathbf{M}'$ — моменты сил, действующих на систему, относительно начал O и O', а $\mathbf{F}$ — геометрическая сумма этих сил. Если результирующая сила $\mathbf{F}$ равна нулю, то $\mathbf{M} = \mathbf{M}'$. Это имеет место, например, для *пары сил*. Вот почему можно говорить о моменте пары сил, не указывая начала, относительно которого этот момент берется.

§ 31. СВЯЗЬ МОМЕНТА ИМПУЛЬСА МАТЕРИАЛЬНОЙ ТОЧКИ С СЕКТОРИАЛЬНОЙ СКОРОСТЬЮ. ТЕОРЕМА ПЛОЩАДЕЙ

1. Если система состоит из одной материальной точки, то момент импульса имеет простой геометрический смысл. Пусть в момент времени t положение материальной точки определяется радиусом-вектором $\mathbf{r}$ (рис. 57). За время dt радиус-вектор получает приращение $\mathbf{v}\,dt$, описывая площадь бесконечно малого треугольника, заштрихованного на рис. 57. Площадь этого треугольника можно изобразить вектором

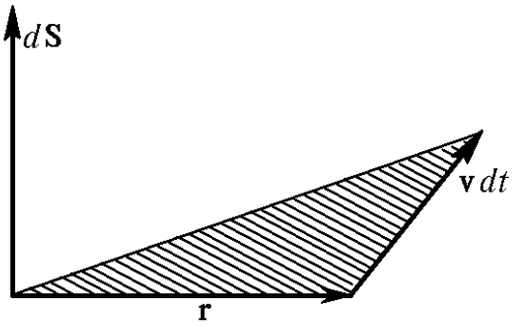


Рис. 57

$$d\mathbf{S} = \frac{1}{2}\,[\mathbf{rv}]\,dt,$$

длина которого представляет рассматриваемую площадь, а направление перпендикулярно к плоскости треугольника. Производная

$$\dot{\mathbf{S}} = \frac{d\mathbf{S}}{dt} = \frac{1}{2}\,[\mathbf{rv}] \tag{31.1}$$

определяет площадь, описываемую радиусом-вектором в единицу времени. Она называется *секториальной* скоростью. Так как по определению $\mathbf{L} = m[\mathbf{rv}]$, то

$$\mathbf{L} = 2m\dot{\mathbf{S}}. \tag{31.2}$$

При нерелятивистских движениях масса m постоянна, а потому момент импульса $\mathbf{L}$ пропорционален секториальной скорости $\dot{\mathbf{S}}$.

2. Если сила, действующая на материальную точку, центральная и ее направление проходит через полюс O, то вектор $\mathbf{L}$ не будет меняться во времени. В случае нерелятивистских движений не будет меняться и секториальная скорость $\dot{\mathbf{S}}$. В этом случае закон сохранения момента импульса переходит в *закон площадей*:

$$\dot{\mathbf{S}} = \text{const}. \tag{31.3}$$

Из этого уравнения вытекают два следствия. Во-первых, плоскость, в которой лежат векторы $\mathbf{r}$ и $\mathbf{v}$, перпендикулярна к направлению вектора $\dot{\mathbf{S}}$. А так как последнее направление остается неизменным, то будет неизменной и указанная плоскость. Это значит, что *траектория материальной точки в поле центральных сил есть плоская кривая.* Во-вторых, из постоянства длины вектора $\dot{\mathbf{S}}$ следует, что в *равные времена радиус-вектор материальной точки описывает одинаковые по размеру площади.* Это положение часто называют *законом площадей.* Мы предпочитаем, однако, придавать закону площадей более широкий смысл, характеризуя площадь не только размером, но и ее ориентацией в пространстве.

Справедливо и обратное утверждение. *Если траектория материальной точки — плоская кривая и радиус-вектор, проведенный из неподвижного полюса O, в равные времена описывает одинаковые площади, то направление действующей силы все время проходит через полюс O.* Действительно, условие теоремы эквивалентно утверждению, что секториальная скорость $\dot{\mathbf{S}}$ есть постоянный вектор. Будет постоянен и момент импульса $\mathbf{L}$.

Поэтому на основании (30.4) $\dot{\mathbf{L}} = \mathbf{M} = [\mathbf{rF}] = 0$. Отсюда следует, что вектор $\mathbf{F}$ коллинеарен радиусу-вектору $\mathbf{r}$, а следовательно, его направление все время проходит через точку O. Последняя является, таким образом, *силовым центром*, из которого должны исходить силы притяжения или отталкивания, действующие на материальную точку.

3. Теорема площадей справедлива не только в случае неподвижного силового центра. Пусть две материальные точки взаимодействуют между собой центральными силами. Применяя понятие приведенной массы, можно свести задачу об их относительном движении к задаче о движении одной точки в силовом поле неподвижного силового центра (см. § 20). В качестве такого силового центра можно принять любую из рассматриваемых материальных точек, относительно которой движется другая точка. Тогда радиус-вектор, проведенный от первой точки ко второй, будет в относительном движении описывать в равные времена равные площади.

§ 32. МОМЕНТ ИМПУЛЬСА И МОМЕНТ СИЛ ОТНОСИТЕЛЬНО НЕПОДВИЖНОЙ ОСИ

1. Векторное уравнение (30.5) эквивалентно трем скалярным уравнениям

$$\frac{dL_x}{dt} = M_x^{\text{внеш}}, \quad \frac{dL_y}{dt} = M_y^{\text{внеш}}, \quad \frac{dL_z}{dt} = M_z^{\text{внеш}}, \tag{32.1}$$

которые получаются из уравнения (30.5) путем проецирования на неподвижные оси декартовой системы координат. Индекс «внеш», указывающий на то, что при вычислении момента сил внутренние силы могут не приниматься во внимание, в дальнейшем обычно будет опускаться. Таким образом, под $\mathbf{M}$ в уравнении моментов всегда будет подразумеваться момент внешних сил. Величины L_x и M_x называются соответственно *моментами импульса* и *сил относительно оси X*. Аналогично говорят о моментах импульса и сил относительно координатных осей Y и Z.

Вообще, *моментами L_x и M_x импульса и сил относительно произвольной оси X* называют проекции векторов $\mathbf{L}$ и $\mathbf{M}$ на эту ось в предположении, что начало O лежит на рассматриваемой оси.

Уравнение

$$\frac{dL_x}{dt} = M_x \tag{32.2}$$

называется *уравнением моментов относительно неподвижной оси* X. Когда момент внешних сил относительно какой-либо неподвижной оси равен нулю, то момент импульса системы относительно той же оси остается постоянным. Это — *закон сохранения импульса относительно неподвижной оси.*

2. Чтобы выяснить геометрический смысл момента M_x, представим векторы $\mathbf{r}$ и $\mathbf{F}$ в виде

$$\mathbf{r} = \mathbf{r}_\perp + \mathbf{r}_\parallel, \quad \mathbf{F} = \mathbf{F}_\perp + \mathbf{F}_\parallel.$$

Здесь $\mathbf{r}_\perp$ — составляющая вектора $\mathbf{r}$, перпендикулярная к оси X, а $\mathbf{r}_\parallel$ — составляющая того же вектора, параллельная этой оси. Аналогичный смысл имеют векторы $\mathbf{F}_\perp$ и $\mathbf{F}_\parallel$. Используя эти разложения, можно написать

$$\mathbf{M} = [\mathbf{rF}] = [\mathbf{r}_\perp \mathbf{F}_\perp] + \{[\mathbf{r}_\perp \mathbf{F}_\parallel] + [\mathbf{r}_\parallel \mathbf{F}_\perp]\} + [\mathbf{r}_\parallel \mathbf{F}_\parallel].$$

Последний член как векторное произведение параллельных векторов равен нулю. Сумма, заключенная в фигурные скобки, есть вектор, перпендикулярный к оси X. При проецировании на эту ось он даст нуль. Таким образом, составляющая вектора $\mathbf{M}$, параллельная оси X, равна

$$\mathbf{M}_\parallel = [\mathbf{r}_\perp \mathbf{F}_\perp].$$

Только эта составляющая и играет роль при нахождении момента M_x относительно оси X. Аналогично при нахождении проекции L_x достаточно проецировать только параллельную слагаемую вектора $\mathbf{L}$:

$$\mathbf{L}_\parallel = [\mathbf{r}_\perp \mathbf{p}_\perp].$$

Изложенное тривиальным образом обобщается на случай системы нескольких сил и системы нескольких материальных точек.

Назовем *плечом силы относительно некоторой оси* кратчайшее расстояние между осью и линией действия силы. Тогда *момент силы относительно той же оси* может быть определен как взятое с надлежащим знаком произведение перпендикулярной составляющей силы на соответствующее плечо. Такое определение момента дается в элементарной физике. Так как точку приложения силы можно перемещать произвольно вдоль линии ее действия, то это определение согласуется с определением, которое было приведено выше. Это видно из рис. 58, где предполагается, что ось перпендикулярна к плоскости рисунка и проходит через полюс O.

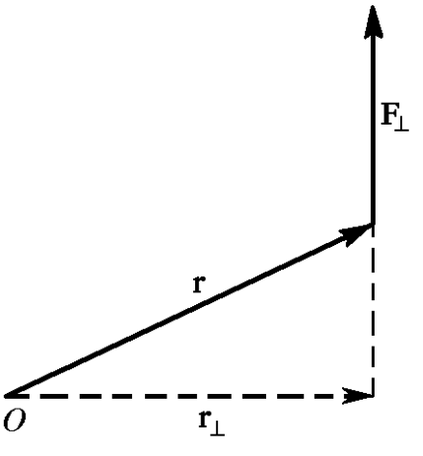


Рис. 58

Аналогично *момент импульса материальной точки относительно оси* можно определить как взятое с надлежащим знаком произведение слагающей импульса, перпендикулярной к этой оси, на соответствующее плечо.

§ 33. УРАВНЕНИЕ МОМЕНТА ИМПУЛЬСА ДЛЯ ВРАЩЕНИЯ ВОКРУГ НЕПОДВИЖНОЙ ОСИ. МОМЕНТ ИНЕРЦИИ

1. Применим уравнение моментов относительно оси к рассмотрению *вращательного движения.* За неподвижную ось моментов удобно выбрать *ось вращения.* Если материальная точка вращается по окружности радиуса r (рис. 59), то момент ее импульса относительно оси вращения O равен $L = mvr$. Пусть ω — угловая скорость вращения, тогда $v = \omega r$, и, следовательно, $L = mr^2\omega$. Если вокруг оси O вращается система материальных точек с одной и той же угловой скоростью ω, то $L = \sum mr^2\omega$, где суммирование производится по всем материальным точкам системы. Величину ω как одинаковую для всех материальных точек можно вынести из-под знака суммы. Тогда получится

$$L = I\omega, \tag{33.1}$$

где

$$I = \sum mr^2. \tag{33.2}$$

Величина I, равная сумме произведений масс материальных точек на квадраты расстояний их до оси вращения, называется *моментом инерции системы относительно этой оси.* Уравнение (33.1) показывает, что *при вращении системы момент ее импульса относительно оси вращения равен произведению момента инерции относительно той же оси на угловую скорость.*

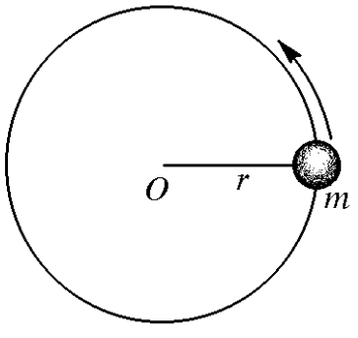


Рис. 59

Если на вращательное движение системы материальных точек накладывается еще *радиальное движение* их, а также *движение параллельно оси,* то наличие таких движений не отразится на справедливости формулы (33.1). Это следует из того, что *момент импульса материальной точки зависит от ее скорости* $\mathbf{v}$ *линейно.* Когда же скорость $\mathbf{v}$ направлена по радиусу или параллельно оси вращения, то момент импульса относительно этой оси равен нулю. Поэтому такие движения непосредственно не сказываются на виде связи между моментом импульса системы относительно оси вращения и ее угловой скоростью. Их влияние косвенное и состоит в том, что момент инерции I перестает быть постоянной величиной, а меняется во времени в соответствии

с изменением мгновенной конфигурации системы. В этом случае уравнение (32.2) принимает вид

$$\frac{d}{dt}(I\omega) = M, \tag{33.3}$$

где M — момент внешних сил относительно оси вращения. Это *основное уравнение динамики вращательного движения вокруг неподвижной оси.* Она напоминает уравнение Ньютона для движения материальной точки. Роль массы играет момент инерции I, роль скорости — угловая скорость ω, роль силы — момент силы M, роль импульса — момент импульса L. Момент импульса L часто называют *вращательным импульсом системы.* Пользуясь этой терминологией, можно сказать, что *производная вращательного импульса системы по времени равна моменту внешних сил относительно оси вращения.*

Если момент внешних сил M относительно оси вращения равен нулю, то вращательный импульс $I\omega$ сохраняется.

2. Важным частным случаем является вращение неизменяемой системы материальных точек или твердого тела вокруг неподвижной оси. В этом случае момент инерции I при вращении остается постоянным, и уравнение (33.3) переходит в

$$I\frac{d\omega}{dt} = M. \tag{33.4}$$

Произведение момента инерции твердого тела относительно неподвижной оси вращения на угловое ускорение $\frac{d\omega}{dt}$ равно моменту внешних сил относительно той же оси.

Для лучшего уяснения уравнения (33.4) приведем другой ее вывод, основанный непосредственно на уравнении движения материальной точки. Последнее в случае вращения материальной точки вокруг неподвижной оси имеет вид $m\frac{dv}{dt} = F_\tau$, где F_τ — тангенциальная слагающая действующей силы. Так как $v = \omega r$, то, умножая предыдущее уравнение на r, получим $mr^2\frac{d\omega}{dt} = rF_\tau$. Напишем такие соотношения для каждой материальной точки, а затем сложим их. Тогда мы снова придем к уравнению (33.4). При этом все внутренние силы исключаются, так что под M в уравнении (33.4) следует понимать момент одних только внешних сил. Этот элементарный вывод обладает, однако, тем недостатком, что он дает уравнение вращательного движения не в общей форме (33.3), а только в частной форме (33.4).

3. Аналогия между движением материальной точки и вращением твердого тела относительно неподвижной оси может быть прослежена дальше. Если материальная точка вращается по окружности, то элементарная работа при повороте на угол $d\varphi$ равна $dA = F\,ds = Fr\,d\varphi = M\,d\varphi$. Такое же выражение получится и для твердого тела, так как его можно рассматривать как систему материальных точек, вращающихся с общей угловой скоростью ω. Внут-

ренние силы исключатся, так как в случае твердого тела, как было показано в § 24, они работы не совершают. Итак, для твердого тела

$$dA = M\,d\varphi. \tag{33.5}$$

Роль силы играет момент внешних сил, роль линейного перемещения — угловое перемещение.

Кинетическая энергия вращающегося твердого тела представляется в виде

$$K = \frac{1}{2}\sum mv^2 = \frac{1}{2}\sum m(\omega r)^2 = \frac{\omega^2}{2}\sum mr^2,$$

или

$$K = \frac{1}{2}I\omega^2 = \frac{L^2}{2I}. \tag{33.6}$$

Эти выражения напоминают соответствующие выражения для кинетической энергии материальной точки. Они получаются из последних формальной заменой $m \to I$, $v \to \omega$, $p \to L$.

§ 34. ПРИМЕРЫ НА ЗАКОН СОХРАНЕНИЯ ВРАЩАТЕЛЬНОГО ИМПУЛЬСА

1. Поучительные демонстрационные опыты на закон сохранения момента импульса можно осуществить с помощью *скамьи Жуковского* (по имени русского ученого Н. Е. Жуковского (1847—1921)). Скамья Жуковского представляет собой угол, сидение которого имеет форму диска. Диск может свободно вращаться вокруг вертикальной оси на шариковых подшипниках. Во время опыта демонстратор садится или становится на скамью Жуковского и, отталкиваясь от пола, может приводить ее во вращение. После прекращения толчка единственными внешними силами, которые могут создавать момент относительно оси вращения, являются силы трения и сопротивления воздуха. Силы трения благодаря шариковым подшипникам очень малы, а сопротивление воздуха может не приниматься во внимание, пока число оборотов скамьи невелико. Поэтому момент импульса системы, состоящей из скамьи и демонстратора, относительно оси вращения не может меняться во времени, если система представлена самой себе.

Демонстратор на скамье Жуковского, оттолкнувшись ногою от пола, приводит ее во вращение. Вместе со скамьей вращается и он сам. Во время вращения вращательный импульс системы будет оставаться постоянным. Какие бы внутренние движения ни совершались в системе — внутренние силы не могут изменить вращательный импульс. Если демонстратор разведет руки в стороны, то он увеличит момент инерции системы I, а потому угловая скорость вращения ω должна уменьшиться, чтобы остался неизменным вращательный импульс $I\omega$. Если демонстратор сводит руки к оси вра-

щения, то момент инерции I уменьшается, а угловая скорость увеличивается. Для усиления эффекта демонстратор держит в руках тяжелые гири. При максимальном удалении гирь от оси вращения момент инерции увеличивается в несколько раз. В такое же число раз уменьшается угловая скорость вращения (рис. 60).

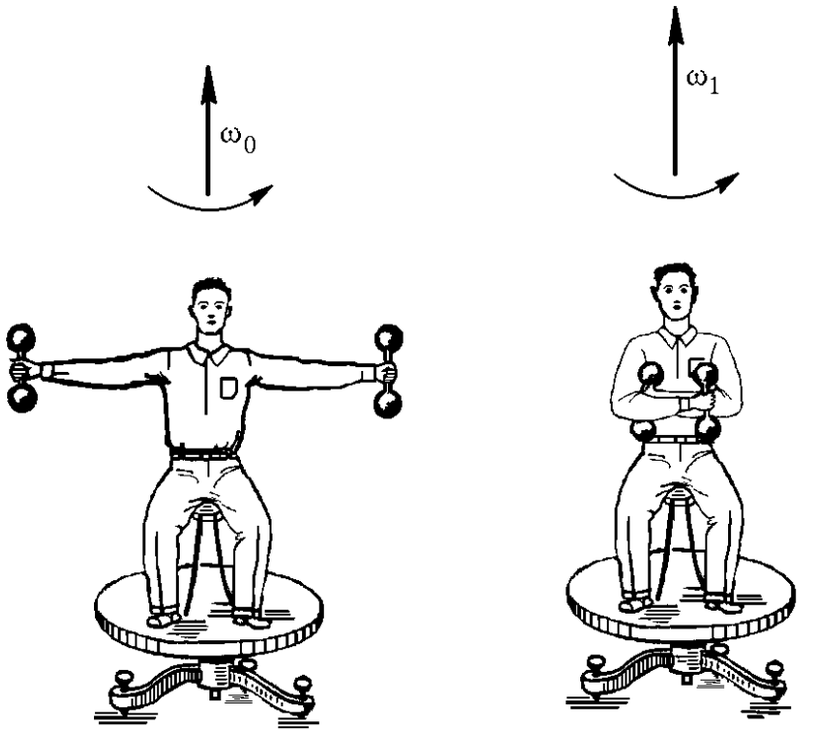


Рис. 60

2. Когда балерина делает пируэт, она вращается на носке, вокруг вертикальной оси. Ноги и руки при этом максимально приближены к оси вращения, и угловая скорость максимальна. Для замедления вращения и остановки балерина разводит руки и отводит ногу в сторону. Наоборот, для сообщения своему телу быстрого вращения балерина отталкивается от пола, получая вращательный импульс, когда момент инерции ее тела максимален. Затем она соответствующим движением уменьшает момент инерции в несколько раз и тем самым увеличивает угловую скорость вращения. Таким образом, она управляет скоростью вращения путем изменения момента инерции своего тела. В сущности, она делает то же самое, что и демонстратор на скамье Жуковского. Тем же самым приемом пользуется гимнаст, выполняющий упражнение на перекладине.

3. Прыгун, чтобы сделать *сальто,* отталкивается от трамплина и тем самым сообщает своему телу вращательный импульс. Этот импульс сохраняется при дальнейшем движении прыгуна в воздухе. Вначале тело прыгуна вытянуто и момент инерции велик. В некоторый момент прыгун свертывается клубком (рис. 61), уменьшая момент инерции в три и большее число раз. Угловая скорость возрастает во столько же раз. С этой угловой скоростью прыгун выполняет один, два и даже три полных оборота. В нужный момент прыгун снова выпрямляет тело и с ма-

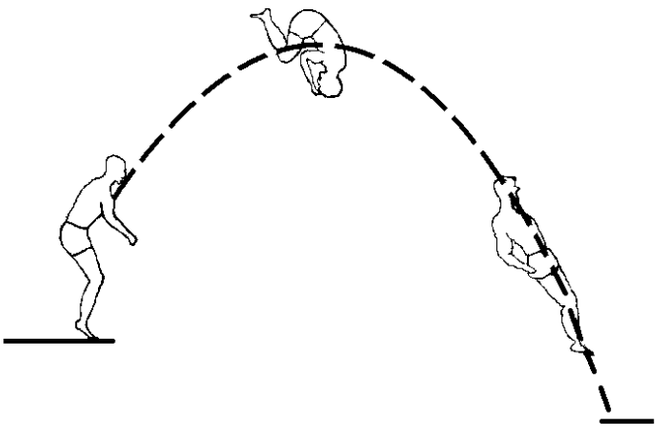


Рис. 61

лой угловой скоростью становится на землю или погружается в воду. Приведя этот пример, мы несколько забежали вперед, так как здесь ось, вокруг которой вращается тело прыгуна, не неподвижна, а движется в пространстве. Однако если движущаяся ось вращения проходит через центр масс прыгуна, то вращение совершается по тем же законам, что и вращение вокруг неподвижной оси (см. § 37).

4. Земля при вращении вокруг собственной оси ведет себя подобно скамье Жуковского. Всякое перемещение масс внутри Земли (выпадение осадков, вулканическая деятельность, горообразование и пр.) меняет момент инерции, а с ним и угловую скорость вращения Земли. Это является причиной нерегулярных колебаний продолжительности суток. Экспериментально обнаружены периодические колебания продолжительности суток с основным периодом в один год и с амплитудой около 0,001 с. Земля подвержена также регулярным внешним воздействиям, прежде всего силам приливного трения, связанным с гравитационным притяжением Луны и Солнца. Благодаря этому средние солнечные сутки увеличиваются примерно на $1{,}640 \cdot 10^{-3}$ с в столетие. Как уже говорилось в § 1, неравномерность вращения Земли можно наблюдать с помощью кварцевых, атомных или молекулярных часов. Ход таких часов управляется колебаниями кристаллической решетки кварца, а также внутриатомными и внутримолекулярными колебаниями при излучении спектральных линий. Указанные колебания обладают значительно большей стабильностью, чем вращение Земли вокруг собственной оси или вокруг Солнца. Это и является причиной, почему в настоящее время эталон времени — секунда — устанавливается именно с помощью таких колебательных процессов, а не с помощью вращения Земли вокруг своей оси или Солнца, как это делалось до недавнего времени (см. § 1).

5. Вернемся к опыту со скамьей Жуковского. При уменьшении момента инерции вращающегося тела его кинетическая энергия увеличивается (при условии, что момент внешних сил равен нулю). Это непосредственно видно из формулы (33.6), так как в рассматриваемом случае вращательный импульс системы $L = I\omega$ не изменяется. Изменение кинетической энергии системы может происходить только за счет работы каких-то сил. Такими силами в нашем примере являются внутренние силы, действующие в системе. Они не могут изменить момент импульса системы. Однако совершаемая ими работа, вообще говоря, отлична от нуля и идет на изменение кинетической энергии вращения системы. Демонстратор на скамье Жуковского должен развить определенную мускульную силу, чтобы удержать вращающиеся гири на их круговых траекториях. Сила, с которой он действует на гирю, есть центростремительная сила $F = m\omega^2 r$, где m — масса гири, а r — расстояние ее от оси вращения. Когда демонстратор приближает гирю к оси вращения, сила F совершает положительную работу. За счет этой работы и происхо-

дит увеличение кинетической энергии системы. При удалении гири работа силы F отрицательна, и кинетическая энергия уменьшается.

Подтвердим эти рассуждения простым расчетом. Чтобы максимально упростить вычисления, схематизируем опыт, заменив реальную систему идеализированной моделью ее. Будем считать, что гири являются материальными точками, а масса рук демонстратора пренебрежимо мала. При такой схематизации момент инерции системы представится выражением $I = I_0 + 2mr^2$, где I_0 — момент инерции системы без гирь, а $2mr^2$ — момент инерции самих гирь (двойка потому, что гирь две). Будем предполагать, что приближение и удаление гирь к оси вращения совершается бесконечно медленно. Тогда в любой момент времени можно пренебречь кинетической энергией радиального движения. Вся работа внутренних сил пойдет на изменение кинетической энергии вращения системы. Вычислим работу A, совершаемую демонстратором, когда он тянет гири к оси вращения, перемещая их с расстояний r_1 до $r_2 < r_1$. Как было показано в § 24, при вычислении работы имеет значение только относительное движение взаимодействующих тел. В нашей задаче это есть движение гирь относительно демонстратора. Каждую гирю демонстратор тянет с силой $m\omega^2 r$. Элементарная работа, совершаемая им, положительна и равна $-2m\omega^2 r dr$ (в нашем случае $dr < 0$). Полная работа A определится интегралом

$$A = -\int_{r_1}^{r_2} 2m\omega^2 r\, dr = -2m \int_{r_1}^{r_2} \frac{(I\omega)^2}{I^2}\, r\, dr = -2m \int_{r_1}^{r_2} \frac{L^2}{I^2}\, r\, dr.$$

Так как момент импульса L во время движения остается постоянным, а $I = I_0 + 2mr^2$, то

$$A = -2mL^2 \int_{r_1}^{r_2} \frac{r\, dr}{(I_0 + 2mr^2)^2} = \frac{L^2}{2} \left\{ \frac{1}{I_0 + 2mr_2^2} - \frac{1}{I_0 + 2mr_1^2} \right\},$$

или

$$A = \frac{L^2}{2I_2} - \frac{L^2}{2I_1} = K_2 - K_1.$$

Та же формула справедлива и при удалении от оси вращения. Она показывает, что *кинетическая энергия вращения изменяется за счет работы мускульной силы демонстратора.*

6. Приведенное объяснение, однако, не отвечает на вопрос, какие силы вызывают изменение угловой скорости вращения системы. Если бы на гирю действовала только центростремительная сила, то она, как сила центральная, не могла бы изменить вращательный импульс гири. Должны были сохраняться в отдельности вращательные импульсы гирь и скамьи Жуковского вместе с демонстратором. Гири и скамья Жуковского вращались бы с различными угловыми скоростями. На самом деле этого нет. При движении гирь по радиусу происходит выравнивание угловых скоростей. Отсюда можно сделать вывод, что во время такого движения помимо центростреми-

тельных сил *на гири действуют силы бокового давления со стороны рук демонстратора. Эти силы не изменяют угловую скорость вращения гирь. Гири, в свою очередь, оказывают боковое давление на руки демонстратора, в результате чего меняется угловая скорость вращения скамьи вместе с демонстратором.* Демонстратор на скамье Жуковского очень хорошо ощущает действие этих сил бокового давления при всяком, в особенности быстром, радиальном перемещении гирь. Дополнительные силы бокового давления перпендикулярны к оси вращения и к относительной скорости гирь. Работы они не производят. Их наличие не может сказаться на результате вычисления работы A, которое было произведено выше. *Силы бокового давления*, однако, *имеют моменты относительно оси вращения и производят перераспределение неизменного момента импульса системы между гирями — с одной стороны — и скамьей Жуковского с демонстратором — с другой.* В результате их действия все эти тела вращаются с общей угловой скоростью. Количественное рассмотрение вопроса будет произведено в § 64.

7. С помощью скамьи Жуковского можно демонстрировать и векторный характер момента импульса. Для этой цели применяется велосипедное колесо с утяжеленным ободом. Если колесо вращается вокруг собственной оси, то вследствие осевой симметрии полный импульс его $\mathbf{p}$ равен нулю. В этом случае, как было показано в § 30, момент импульса $\mathbf{L}$ относительно неподвижной точки не зависит от положения этой точки. С другой стороны, проекция вектора $\mathbf{L}$ на ось вращения колеса равна $I\Omega$, где I — момент инерции колеса, а Ω — его угловая скорость. Проекция вектора $\mathbf{L}$ на любое направление, перпендикулярное к оси колеса, равна нулю ввиду осевой симметрии. Отсюда следует, что вектор момента импульса $\mathbf{L}$ направлен вдоль оси колеса и по модулю равен $I\Omega$.

Демонстратор садится или становится на скамью Жуковского. Ему передают быстро вращающееся колесо с вертикально направленной осью (рис. 62). Полный момент импульса системы направлен вертикально и равен $I\Omega$. Примем вертикальную ось скамьи Жуковского за ось X. Так как момент внешних сил относительно оси X равен нулю, то проекция L_x полного момента импульса системы на эту ось должна сохраняться. В начале опыта весь вращательный импульс сосредоточен в колесе. Затем демонстратор наклоняет ось колеса на угол α. Проекция момента импульса колеса на ось X становится равной $L_x^{\text{кол}} = I\Omega \cos\alpha$, т. е. она уменьшается на $I\Omega\,(1 - \cos\alpha)$. Это уменьшение должно быть скомпенсировано возрастанием соответствующей проекции момента импульса скамьи и демонстратора на величину $L_x^{\text{скам}} = I\Omega\,(1 - \cos\alpha)$. В результате скамья с демонстратором приходит во вращение с угловой скоростью ω, определяемой из уравнения $I_0\omega = I\Omega\,(1 - \cos\alpha)$, где I_0 — момент инерции скамьи. При $\alpha = 90°$ проекция $L_x^{\text{кол}}$ обращается в нуль — она целиком передается скамье и демонстратору. При

$\alpha = 180°$ изменение вращательного импульса колеса становится максимальным $\Delta L_x^{\text{кол}} = 2L_x^{\text{кол}}$, скамья и демонстратор вращаются с максимальной скоростью $\omega_{\text{макс}} = \frac{2I}{I_0}\Omega$. Поворачивая ось, демонстратор придает ей исходной направление — тогда вращение скамьи прекращается. Однако скамья, вообще говоря, не возвращается в исход-

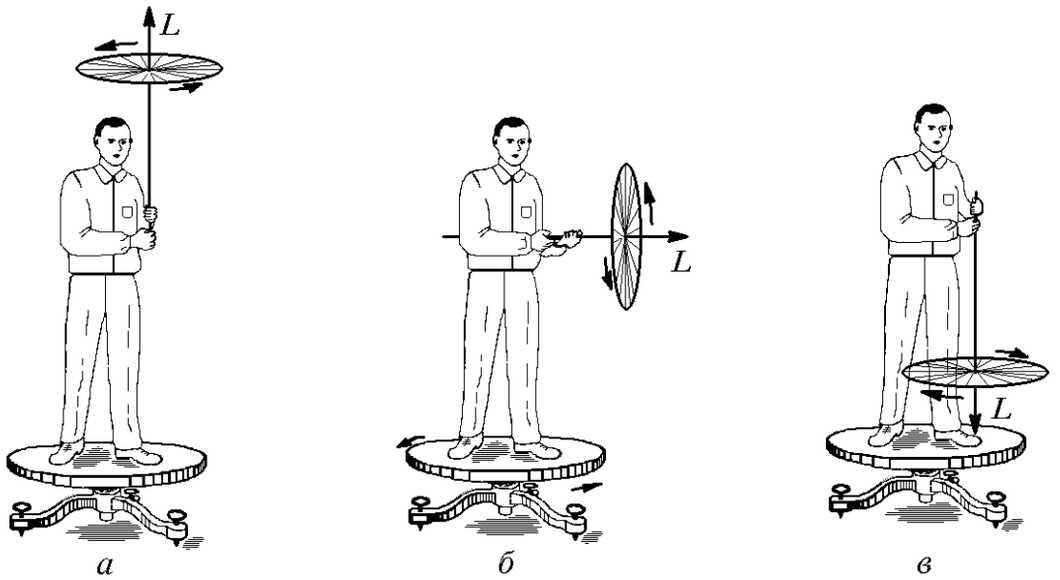


Рис. 62

ное положение, а оказывается повернутой вокруг вертикальной оси на некоторый угол.

Наклоняя ось колеса, демонстратор во время ее движения испытывает значительные силы бокового давления. Колесо как бы стремится вырваться из рук демонстратора. Эти силы направлены горизонтально и притом перпендикулярно как к оси колеса, так и к оси скамьи Жуковского. Их геометрическая сумма равна нулю, но они имеют момент относительно оси X. Последний приводит во вращение скамью Жуковского и демонстратора. Происхождение этих сил будет выяснено в гл. VII.

8. Закончим этот параграф следующим замечанием. Пусть имеется замкнутая система тел (назовем ее лабораторией), которая в начальный момент времени покоилась относительно какой-то неподвижной (инерциальной) системы отсчета S. Можно ли с помощью одних только внутренних движений сместить лабораторию в пространстве и притом так, чтобы все тела в ней вернулись в свои исходные положения? Говоря о смещении лаборатории, мы имеем в виду ее поступательное перемещение без вращения. Отрицательный ответ на этот вопрос дает теорема о движении центра масс. Не так обстоит дело в отношении поворота замкнутой системы тел. *С помощью одних только внутренних движений можно повернуть лабораторию в пространстве на любой угол и притом так, что исходное расположение тел в лаборатории восстановится.* Допустим, например, что лаборатория состоит из замкнутой оболочки A,

в которой помещено всего одно тело B. Пусть тело B начинает вращаться вокруг некоторой оси с угловой скоростью $\dot{\varphi}_B$ (относительно неподвижной системы отсчета). Тогда оболочка A придет во вращение относительно той же оси с угловой скоростью $\dot{\varphi}_A$. По закону сохранения вращательного импульса $I_A\dot{\varphi}_A + I_B\dot{\varphi}_B = 0$, так как в начальный момент вращательный импульс был равен нулю (I_A и I_B — моменты инерции оболочки A и тела B соответственно). Если углы φ_A и φ_B условиться отсчитывать от начальных положений тел A и B, то после интегрирования получится $I_A\varphi_A + I_B\varphi_B = 0$. Угол поворота тела B относительно оболочки A определится разностью $\varphi = \varphi_B - \varphi_A = -(I_A/I_B + 1)\varphi_A$. Если $\varphi = 2\pi n$ (n — целое число), то тело B возвратится в исходное положение относительно оболочки A. При этом угол поворота оболочки φ_A, вообще говоря, не будет равен нулю. Различие в поведении лаборатории при поступательном перемещении и вращении связано со следующим обстоятельством. *При непрерывном поступательном перемещении тела B оно никогда не возвращается в исходное положение относительно тела A.* Различным значениям координаты x соответствуют и различные положения тела. Напротив, *при непрерывном вращении тела B взаимное расположение тел B и A периодически восстанавливается*: значениям угла φ, отличающимся на $2\pi n$, соответствует *одно и то же* относительное расположение тел A и B. Падающая кошка, вращая хвостом и лапами, придает своему телу такое положение, чтобы встать на землю лапами. И это ей удается.

Эти явления можно имитировать на скамье Жуковского. Демонстратор, совершая конические вращения одной или обеими руками, всегда может повернуть скамью Жуковского на произвольный угол. Для усиления эффекта он может взять в руки массивный предмет с большим моментом инерции, например молот.

§ 35. ТЕОРЕМА ГЮЙГЕНСА–ШТЕЙНЕРА

Найдем связь между моментами инерции тела относительно двух различных параллельных осей. Предполагается, что эти оси перпендикулярны к плоскости рисунка и пересекают ее в точках O и A. Ради краткости будем называть сами оси осями также O и A. Разобьем мысленно тело на элементарные массы dm. Радиусы-векторы одной из них, проведенные от осей O и A параллельно плоскости рисунка, обозначим через $\mathbf{r}$ и $\mathbf{r}'$ соответственно. (На рис. 63 изображен такой случай, когда элементарная масса dm лежит в плоскости рисунка.) Тогда $\mathbf{r}' = \mathbf{r} - \mathbf{a}$, где $\mathbf{a}$ означает радиус-вектор $\overrightarrow{OA}$. Следовательно, $r'^2 = r^2 + a^2 - 2(\mathbf{ar})$,

$$\int r'^2\, dm = \int r^2\, dm + a^2 \int dm - 2\left(\mathbf{a}\int \mathbf{r}\, dm\right).$$

Интеграл слева есть момент инерции I_A тела относительно оси A, первый интеграл справа — момент инерции относительно оси O. Последний интеграл можно представить в виде $\int \mathbf{r}\, dm = m\mathbf{R}_C$, где $\mathbf{R}_C$ — радиус-вектор центра масс C тела относительно оси O (точнее, $\mathbf{R}_C$ есть слагающая радиуса-вектора центра масс, параллельная плоскости рисунка). Таким образом,

$$I_A = I_O + ma^2 - 2m(\mathbf{a}\mathbf{R}_C). \qquad (35.1)$$

Допустим, что ось O проходит через центр масс C тела. Тогда $\mathbf{R}_C = 0$, и предыдущая формула упрощается, принимая вид

$$I_A = I_C + ma^2. \qquad (35.2)$$

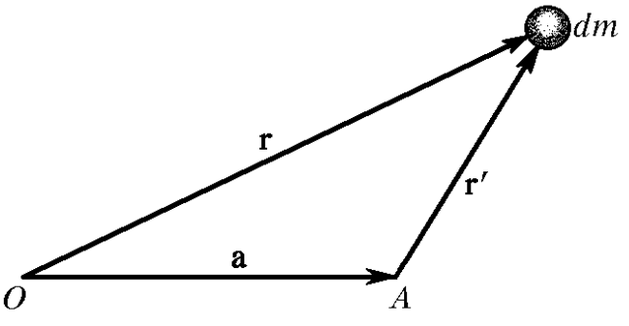


Рис. 63

Это важное геометрическое соотношение называется *теоремой Гюйгенса—Штейнера* (Якоб Штейнер (1796—1863) — швейцарский геометр). *Момент инерции тела относительно какой-либо оси равен моменту инерции его относительно параллельной оси, проходящей через центр масс, сложенному с величиной* ma^2, *где* a *— расстояние между осями.*

§ 36. ВЫЧИСЛЕНИЕ МОМЕНТОВ ИНЕРЦИИ

1. Момент инерции тела относительно какой-либо оси можно найти вычислением или измерить экспериментально*). Если вещество в теле распределено непрерывно, то вычисление момента инерции его сводится к вычислению интеграла

$$I = \int r^2\, dm, \qquad (36.1)$$

в котором r — расстояние от элемента массы dm до оси вращения. Интегрирование должно производиться по всей массе тела. Аналитическое вычисление таких интегралов возможно только в простейших случаях тел правильной геометрической формы. Для тел неправильной формы такие интегралы могут быть найдены численно.

Вычисление моментов инерции во многих случаях можно упростить, используя соображения подобия и симметрии, теорему Гюйгенса—Штейнера, а также некоторые другие общие соотношения, о которых будет сказано ниже.

*) Об одном методе экспериментального определения моментов инерции говорится в § 42.

Рассмотрим два подобных и подобно расположенных относительно оси вращения тела A и B одной и той же плотности. Полные и элементарные массы этих тел относятся как кубы их линейных размеров l. Так как элементарные массы умножаются на квадраты расстояний их до оси вращения, то моменты инерции тел A и B будут относиться как пятые степени тех же размеров. Таким образом, $I \sim l^5$, или

$$I = kml^2. \tag{36.2}$$

Под l следует понимать какой-либо *характерный размер тела* или расстояние какой-либо характерной точки его от оси вращения. Коэффициент пропорциональности k зависит только от формы тела и его расположения относительно оси вращения.

2. Вычисление момента инерции тела относительно оси часто можно упростить, вычислив предварительно *момент инерции* его *относительно точки.* Сам по себе момент инерции тела относительно точки не играет никакой роли в динамике. Он является чисто вспомогательным понятием, служащим для упрощения вычислений. *Моментом инерции тела относительно точки* O называется сумма произведений масс материальных точек, из которых тело состоит, на квадраты их расстояний R до точки O: $\Theta = \sum mR^2$. В случае непрерывного распределения масс эта сумма сводится к интегралу $\Theta = \int R^2\, dm$. Само собой понятно, что момент Θ не следует смешивать с моментом инерции I относительно оси. В случае момента I массы dm умножаются на квадраты расстояний до этой оси, а в случае момента Θ — до неподвижной точки.

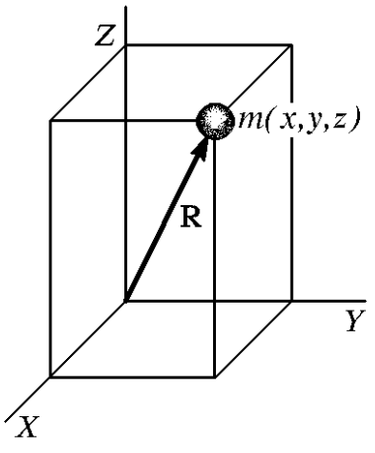


Рис. 64

Рассмотрим сначала одну материальную точку с массой m и с координатами x, y, z относительно прямоугольной системы координат (рис. 64). Квадраты расстояний ее до координатных осей X, Y, Z равны соответственно $y^2 + z^2$, $z^2 + x^2$, $x^2 + y^2$, а моменты инерции относительно тех же осей

$$I_x = m(y^2 + z^2), \quad I_y = m(z^2 + x^2), \quad I_z = m(x^2 + y^2).$$

Сложив эти три равенства, получим

$$I_x + I_y + I_z = 2m(x^2 + y^2 + z^2).$$

Но $x^2 + y^2 + z^2 = R^2$, где R — расстояние точки m от начала координат O. Поэтому

$$I_x + I_y + I_z = 2\Theta. \tag{36.3}$$

Это соотношение справедливо не только для одной материальной точки, но и для произвольного тела, так как тело можно рассматривать как совокупность материальных точек. Таким образом, *сумма моментов инерции тела относительно трех взаимно перпендикулярных осей, пересекающихся в одной точке O, равна удвоенному моменту инерции того же тела относительно этой точки.*

Если повернуть координатные оси X, Y, Z относительно тела, оставляя углы между ними прямыми, то моменты инерции I_x, I_y, I_z, вообще говоря, изменятся. Однако их сумма останется той же самой, так как она равна 2Θ, а величина Θ не зависит от ориентации координатных осей. Таким образом, *сумма моментов инерции I_x, I_y, I_z относительно любых трех взаимно перпендикулярных осей, проходящих через одну точку, зависит только от положения этой же точки и не меняется с изменением ориентации осей.* Более глубокая геометрическая природа этого утверждения раскрывается в тензорной алгебре.

3. Другое следствие можно получить для плоского распределения масс. Допустим, что имеется пластинка произвольной формы с произвольным распределением вещества по ее объему. Если пластинка очень тонкая, то можно считать, что вещество распределено бесконечно тонким слоем по математической плоскости. Примем эту плоскость за координатную плоскость XY. Тогда z-координаты всех материальных точек будут равны нулю, а потому момент инерции Θ пластинки относительно начала координат O представится выражением $\Theta = \sum \Delta m(x^2 + y^2)$, т. е. будет равен моменту инерции пластинки относительно оси Z. Таким образом, в случае плоского распределения масс $I_x + I_y + I_z = 2I_z$, т. е.

$$I_x + I_y = I_z. \tag{36.4}$$

Далее, очевидно, что величина Θ не меньше каждого из моментов инерции I_x, I_y, I_z, например, $\Theta \geqslant I_z$ (знак равенства имеет место только для плоского распределения масс). Вычитая неравенство $2I_z \leqslant 2\Theta$ из равенства (36.3), получим $I_x + I_y - I_z \geqslant 0$, или

$$I_x + I_y \geqslant I_z. \tag{36.4а}$$

Отсюда следует, что из отрезков, длины которых численно равны I_x, I_y, I_z, всегда можно составить треугольник. Для плоского распределения масс (в плоскости XY) формула (36.4а) переходит в формулу (36.4).

После этих предварительных замечаний можно перейти к вычислению моментов инерции конкретных тел.

4. Момент инерции тонкого однородного стержня относительно перпендикулярной оси. Пусть ось проходит через конец стержня A

(рис. 65). Для момента инерции можно написать $I_A = kml^2$, где l — длина стержня. Центр стержня C является его центром масс. По теореме Гюйгенса—Штейнера $I_A = I_C + m(l/2)^2$. Величину I_C можно представить как сумму моментов инерции двух стержней, CA и CB, длина каждого из которых равна $l/2$, масса $m/2$, а следовательно, момент инерции равен $k\frac{m}{2}\left(\frac{l}{2}\right)^2$. Таким образом, $I_C = km(l/2)^2$. Подставляя эти выражения в предыдущую формулу, получим

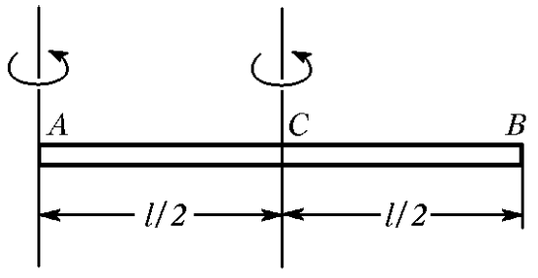


Рис. 65

$$kml^2 = km\left(\frac{l}{2}\right)^2 + m\left(\frac{l}{2}\right)^2,$$

откуда $k = 1/3$. В результате находим

$$I_A = \frac{1}{3}ml^2, \qquad (36.5)$$

$$I_C = \frac{1}{12}ml^2. \qquad (36.6)$$

Существенно, что стержень тонкий. Бесконечно тонкий стержень можно рассматривать как отрезок прямой линии. Он всегда геометрически подобен любой его части. Поэтому коэффициент k будет одним и тем же для всего стержня и для любой его части, например половины. Для стержня конечной толщины подобие между всем стержнем и его частью уже не имеет места. В этом случае числовой коэффициент k имеет разные значения для всего стержня и его половины. По этой причине к стержню конечной толщины формулы (36.5) и (36.6) не применимы. Но ими как приближенными формулами можно пользоваться, когда поперечные размеры стержня очень малы по сравнению с его длиной.

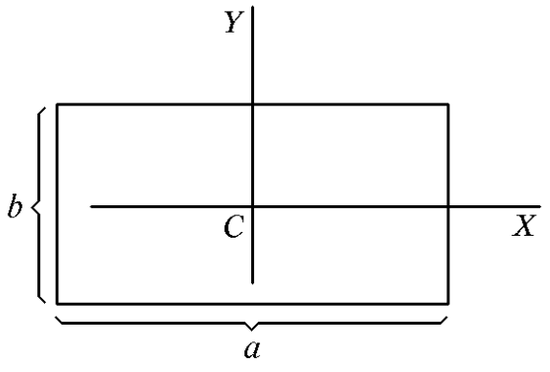


Рис. 66

5. Момент инерции однородных прямоугольных пластинки и параллелепипеда. Пусть координатные оси X и Y проходят через центр пластинки C и параллельны ее сторонам (рис. 66). Представим себе, что все вещество пластинки смещено параллельно оси X и сконцентрировано на оси Y. При таком смещении все расстояния материальных точек до оси X не изменятся. Вместе с ними не изменится и момент инерции I_x относительно оси X. Но в результате смещения пластинка перейдет в бесконечно тонкий стержень длины l, к которому применима формула (36.6). В результате получим

$$I_x = \frac{m}{12}b^2, \quad I_y = \frac{m}{12}a^2. \qquad (36.7)$$

Момент инерции I_z пластинки относительно оси Z, перпендикулярной к ее плоскости, найдется по формуле (36.4), которая дает

$$I_z = \frac{m}{12}(a^2 + b^2). \qquad (36.8)$$

Формула (36.8) годится также для вычисления моментов инерции прямоугольного параллелепипеда относительно его геометрических осей. В этом можно убедиться, если мысленно сжать параллелепипед вдоль одной из геометрических осей в прямоугольную пластинку — при таком сжатии момент инерции относительно этой оси не изменяется. Формула (36.8) дает момент инерции прямоугольного параллелепипеда относительно той его геометрической оси, которая проходит через центр основания с длинами сторон a и b. На рис. 66 эта ось перпендикулярна к плоскости рисунка.

6. Момент инерции бесконечно тонкого круглого кольца (окружности). Момент инерции относительно оси Z (рис. 67), очевидно, равен

$$I_z = mR^2, \qquad (36.9)$$

где R — радиус кольца. Ввиду симметрии $I_x = I_y$. Поэтому из формулы (36.4) находим

$$I_x = I_y = \frac{1}{2} mR^2. \qquad (36.10)$$

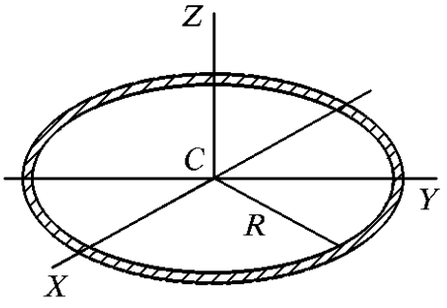


Рис. 67

Формула (36.9), очевидно, дает также момент инерции полого однородного цилиндра с бесконечно тонкими стенками относительно его геометрической оси.

7. Момент инерции бесконечно тонкого диска и сплошного цилиндра. Предполагается, что диск и цилиндр однородны, т. е. вещество распределено в них с постоянной плотностью. Пусть ось Z проходит через центр диска C перпендикулярно к его плоскости (рис. 68). Рассмотрим бесконечно тонкое кольцо с внутренним радиусом r и наружным радиусом $r + dr$. Площадь такого кольца $dS = 2\pi r\, dr$. Его момент инерции найдется по формуле (36.9), он равен $dI_z = r^2 dm$. Момент инерции всего диска определяется интегралом $I_z = \int r^2\, dm$. Ввиду однородности диска $dm = m\frac{dS}{S} = 2m\frac{r\, dr}{R^2}$, где $S = \pi R^2$ — площадь всего диска. Вводя это выражение под знак интеграла, получим

$$I_z = \frac{2m}{R^2}\int_0^R r^3\, dr = \frac{1}{2} mR^2. \qquad (36.11)$$

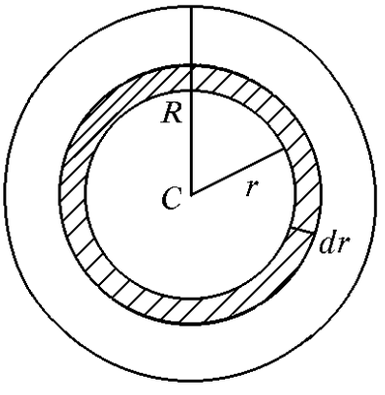


Рис. 68

Момент инерции диска относительно диаметра вдвое меньше, как это непосредственно следует из формулы (36.4) и из соображений симметрии:

$$I_x = I_y = \frac{1}{4} mR^2. \qquad (36.12)$$

Формула (36.11) дает также момент инерции однородного сплошного цилиндра относительно его продольной геометрической оси.

8. Момент инерции однородного сплошного цилиндра относительно поперечной оси. Пусть ось вращения проходит через центр основания цилиндра A перпендикулярно к его продольной геометрической оси (рис. 69). Вырежем мысленно бесконечно короткий цилиндр с массой dm, находящийся от

оси вращения на расстоянии x. Для его момента инерции по теореме Гюйгенса—Штейнера можно написать

$$dI_A = dm \cdot x^2 + \tfrac{1}{4}\, dm \cdot R^2,$$

а для момента инерции всего цилиндра

$$I_A = \int x^2\, dm + \tfrac{1}{4} R^2 \int dm.$$

Первое слагаемое в правой части формально совпадает с выражением для момента инерции однородного бесконечно тонкого стержня, а потому равно $1/3\, ml^2$. Второе слагаемое равно $1/4\, mR^2$. Следовательно,

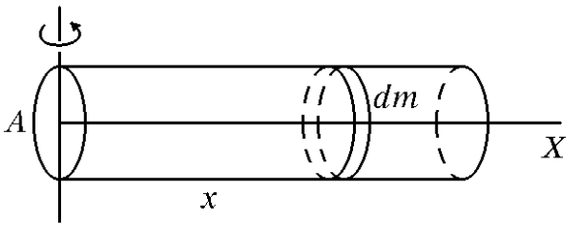


Рис. 69

$$I_A = \tfrac{1}{3}\, ml^2 + \tfrac{1}{4}\, mR^2. \qquad (36.13)$$

Момент инерции I_C относительно поперечной геометрической оси, проходящей через центр масс цилиндра, можно найти по формуле (36.13), если цилиндр разделить на два цилиндра с высотами $l/2$ и массами $m/2$. Получим

$$I_C = \tfrac{1}{12}\, ml^2 + \tfrac{1}{4}\, mR^2. \qquad (36.14)$$

При $R \to 0$ формулы (36.13) и (36.14) переходят в формулы (36.5) и (36.6) для бесконечно тонкого стержня.

9. Момент инерции полого шара с бесконечно тонкими стенками. Сначала найдем момент инерции Θ относительно центра шара. Очевидно, он равен $\Theta = mR^2$. Затем применим формулу (36.3), полагая в ней ввиду симметрии $I_x = I_y = I_z = I$. В результате находим момент инерции полого шара относительно его диаметра

$$I = \tfrac{2}{3}\, mR^2. \qquad (36.15)$$

10. Момент инерции сплошного однородного шара. Сплошной шар можно рассматривать как совокупность бесконечно тонких сферических слоев с массами dm (см. рис. 68). Так как шар по предположению однороден, то $dm = m\,\frac{dV}{V}$, где $dV = 4\pi r^2\, dr$ — объем сферического слоя, а $V = 4/3\, \pi R^3$ — объем всего шара. По формуле (36.15) момент инерции сферического слоя относительно диаметра равен $dI = \frac{2}{3}\, dm\, r^2 = 2m\,\frac{r^4\, dr}{R^3}$. Интегрируя, получаем момент инерции сплошного шара

$$I = \tfrac{2}{5}\, mR^2. \qquad (36.16)$$

11. Момент инерции однородного эллипса. Предполагается, что масса равномерно распределена по площади эллипса. Эллипс можно полу-

чить из круга равномерным сжатием вдоль одного из его диаметров, например вдоль оси Y (рис. 70). При таком сжатии момент инерции относительно оси Y не меняется. Первоначально он был равен $\frac{1}{4} ma^2$ (a — радиус круга, сжатием которого получен эллипс; он равен длине большой полуоси эллипса). Аналогичное рассуждение применимо и для оси X. В результате получим

$$I_x = \frac{1}{4} mb^2, \quad I_y = \frac{1}{4} ma^2. \qquad (36.17)$$

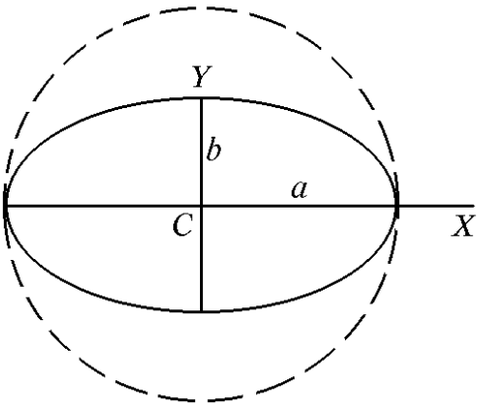


Рис. 70

Момент инерции относительно оси Z, перпендикулярной к плоскости эллипса, найдется по формуле (36.4):

$$I_z = \frac{1}{4} m(a^2 + b^2). \qquad (36.18)$$

Формула (36.18) дает также момент инерции однородного эллиптического цилиндра относительно его продольной геометрической оси.

12. Момент инерции трехосного эллипсоида. Предполагается, что масса равномерно распределена по объему эллипсоида. Координатные оси X, Y, Z направим вдоль главных осей эллипсоида. Длины полуосей эллипсоида обозначим буквами a, b, c. Вычислим момент инерции его относительно главной оси Z. Эллипсоид может быть получен из шара равномерным сжатием или растяжением по трем взаимно перпендикулярным направлениям, например по направлениям осей X, Y, Z. Возьмем однородный шар радиуса a. Его момент инерции $I_{\text{ш}} = \frac{2}{5} ma^2$. Произведем однородное сжатие в направлении оси Z, чтобы шар превратился в бесконечно тонкий круглый диск (конечно, с неравномерным распределением масс). Момент инерции I_z при этом остается неизменным, а моменты инерции I_x и I_y будут равны между собой ввиду симметрии. На основании соотношения (36.4) $I_x = I_y = \frac{1}{2} I_z = \frac{1}{5} ma^2$. Произведем затем равномерное сжатие круглого диска в направлении оси Y, чтобы его размеры в этом направлении сделались равными $2b$. При этом момент инерции I_y останется неизменным, а I_x сделается равным $I_x = \frac{1}{5} mb^2$. Применяя снова соотношение (36.4), для момента инерции I_z полученного эллиптического диска найдем $I_z = I_x + I_y = \frac{1}{5} m(a^2 + b^2)$. Наконец, произведем равномерное растяжение эллиптического диска в направлении оси Z, чтобы он превратился в трехосный эллипсоид с полуосями a, b, c. При этом величина I_z не изменится. Таким образом, момент инерции трехосного эллипсоида относительно оси Z равен

$$I_z = \frac{1}{5} m(a^2 + b^2). \qquad (36.19)$$

Моменты инерции относительно остальных двух главных осей равны соответственно

$$I_x = \frac{1}{5} m(b^2 + c^2), \quad I_y = \frac{1}{5} m(c^2 + a^2).$$

§ 37. УРАВНЕНИЕ МОМЕНТОВ ОТНОСИТЕЛЬНО ДВИЖУЩЕГОСЯ НАЧАЛА И ДВИЖУЩЕЙСЯ ОСИ

1. Уравнение моментов (30.5) справедливо для того случая, когда начало O, относительно которого рассматриваются моменты $\mathbf{L}$ и $\mathbf{M}$, *неподвижно.* Точно так же уравнение (32.2) относится к моментам относительно *неподвижной оси.* В некоторых случаях, однако, целесообразно рассматривать *движущиеся начала* или *движущиеся оси.* Исследует, как меняется в этом случае уравнение моментов. Особый интерес представляют случаи, когда уравнение моментов относительно движущегося начала сохраняет прежний вид (30.5).

2. Рассмотрим сначала материальную точку. Будем понимать под $\mathbf{v}$ и $\mathbf{p} \equiv m\mathbf{v}$ скорость и импульс этой точки относительно неподвижной инерциальной системы отсчета S, а под $\mathbf{r}$ — ее радиус-вектор, проведенный из движущегося начала O. Движение начала O может быть как равномерным, так и неравномерным. Скорость этого движения обозначим через $\mathbf{v}_O$. Момент импульса движущейся точки относительно начала O определим прежним выражением (30.3), т. е. $\mathbf{L} = [\mathbf{rp}]$. Как и раньше, дифференцированием этого выражения найдем

$$\dot{\mathbf{L}} = [\dot{\mathbf{r}}\mathbf{p}] + [\mathbf{r}\dot{\mathbf{p}}].$$

Однако теперь $\dot{\mathbf{r}}$ означает не скорость материальной точки $\mathbf{v}$, а разность между этой скоростью и скоростью движущегося начала $\mathbf{v}_O$. Таким образом,

$$\dot{\mathbf{L}} = [(\mathbf{v} - \mathbf{v}_O)\mathbf{p}] + [\mathbf{r}\dot{\mathbf{p}}],$$

или ввиду уравнения Ньютона $\dot{\mathbf{p}} = \mathbf{F}$ и коллинеарности векторов $\mathbf{v}$ и $\mathbf{p}$ $\dot{\mathbf{L}} = [\mathbf{rF}] - [\mathbf{v}_O\mathbf{p}]$, или, наконец,

$$\dot{\mathbf{L}} = \mathbf{M} - [\mathbf{v}_O\mathbf{p}]. \tag{37.1}$$

Чтобы обобщить это уравнение на случай системы материальных точек, напишем его для i-й материальной точки $\dot{\mathbf{L}}_i = \mathbf{M}_i - [\mathbf{v}_O\mathbf{p}_i]$, а затем просуммируем по всем i. Таким путем снова получим уравнение (37.1). Однако теперь $\mathbf{p}$ будет означать импульс *всей системы* материальных точек, а $\mathbf{M}$ — момент действующих на нее внешних сил.

Импульс $\mathbf{p}$ можно представить в виде $\mathbf{p} = m\mathbf{v}_C$, где $\mathbf{v}_C$ — скорость центра масс системы. Таким образом,

$$\dot{\mathbf{L}} = \mathbf{M} - m[\mathbf{v}_O\mathbf{v}_C]. \tag{37.2}$$

Это и есть *уравнение моментов относительно движущегося начала.*

Если движущееся начало O совпадает с центром масс C системы, то $\mathbf{v}_O = \mathbf{v}_C$, и формула (37.2) *переходит в прежнее уравнение* (30.5). *Уравнение моментов относительно центра масс имеет такой же вид, что и относительно неподвижного начала.* Более того, в этом случае скорости $\mathbf{v}$ материальных точек не обязательно

рассматривать относительно неподвижной системы отсчета S. Их можно брать и относительно самого центра масс C, считая его как бы неподвижным. Если центр масс C движется прямолинейно и равномерно, то это утверждение непосредственно следует из принципа относительности. Но оно справедливо и в случае ускоренного движения центра масс. В самом деле, скорость каждой материальной точки может быть представлена в виде $\mathbf{v} = \mathbf{v}_C + \mathbf{v}_{\text{отн}}$, где $\mathbf{v}_{\text{отн}}$ — скорость точки относительно центра масс C. Поэтому

$$\mathbf{L} = \sum [\mathbf{r} m \mathbf{v}] = \sum [\mathbf{r} m \mathbf{v}_C] + \sum [\mathbf{r} m \mathbf{v}_{\text{отн}}].$$

Предпоследняя сумма в этом равенстве равна нулю. Действительно, так как скорость $\mathbf{v}_C$ одна и та же для всех слагаемых суммы, то ее можно вынести из-под знака суммы, что дает

$$-[\mathbf{v}_C \sum m\mathbf{r}] = -[\mathbf{v}_C \mathbf{r}_C] \sum m,$$

где $\mathbf{r}_C$ — радиус-вектор центра масс. Он равен нулю, так как начало координат O по условию помещено в центр масс. Итак,

$$\mathbf{L} = \sum [\mathbf{r} m \mathbf{v}_{\text{отн}}],$$

что и доказывает наше утверждение.

Второй, более общий, случай, когда уравнение (37.2) переходит в простую формулу (30.5), получается тогда, когда скорости $\mathbf{v}_O$ и $\mathbf{v}_C$ *коллинеарны.* В этом случае векторное произведение $[\mathbf{v}_O \mathbf{v}_C]$ обращается в нуль. Таким образом, *когда скорость движущегося начала O параллельна скорости центра масс C, уравнение моментов принимает простую форму* (30.5). В этом случае, однако, при вычислении момента импульса $\mathbf{L}$ надо брать скорости всех материальных точек обязательно *относительно инерциальной системы отсчета S, а не относительно центра масс.*

3. Аналогичные результаты справедливы и для поступательного движущихся осей, когда при движении оси она все время остается параллельной своему исходному направлению. Нет необходимости формулировать эти результаты отдельно, так как уравнение моментов относительно оси получается из соответствующего уравнения относительно точки путем проецирования на эту ось.

ЗАДАЧИ

1. Определить ускорение тел и натяжения нити на машине Атвуда, предполагая, что $m_2 > m_1$ (рис. 71). Момент инерции блока относительно геометрической оси равен I, радиус блока r. Массу нити считать пренебрежимо малой.

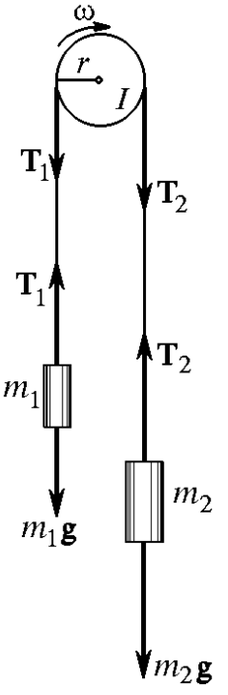


Рис. 71

Р е ш е н и е. Ввиду того, что масса нити пренебрежимо мала, изменения натяжений T_1 и T_2 вдоль нити можно не учитывать. Уравнения движения грузов и блока будут

$$m_1 a = T_1 - m_1 g,$$

$$m_2 a = m_2 g - T_2,$$

$$I \frac{d\omega}{dt} = r(T_2 - T_1).$$

Если нет проскальзывания нити по блоку, то

$$r \frac{d\omega}{dt} = a.$$

Решая эти уравнения, получим

$$a = \frac{m_2 - m_1}{m_1 + m_2 + \dfrac{I}{r^2}} g,$$

после чего находим T_1 и T_2. Если масса блока пренебрежимо мала, то $T_1 = T_2$.

2. К шкиву креста Обербека (рис. 72) прикреплена нить, к которой подвешен груз массой $M = 1$ кг. Груз опускается с высоты $h = 1$ м до нижнего положения, а затем начинает подниматься. В это время происходит «рывок», т. е. увеличение натяжения нити. Найти натяжение нити T при опускании или поднятии груза, а также оценить приближенно натяжение во время рывка $T_{\text{рыв}}$. Радиус шкива $r = 3$ см. На кресте укреплены четыре груза массой $m = 250$ г каждый на расстоянии $R = 30$ см от его оси. Моментом инерции самого креста и шкива пренебречь, по сравнению с моментом инерции грузов. Растяжение нити во время рывка не учитывать.

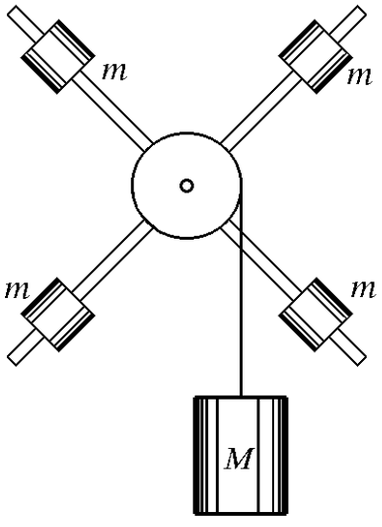


Рис. 72

О т в е т. $T = \dfrac{Mg}{1 + \dfrac{Mr^2}{I}} = \dfrac{Mg}{1 + \dfrac{Mr^2}{2mR^2}} = 0{,}99T_0,$

где I — момент инерции системы, а T_0 — натяжение нити при неподвижном грузе. Среднее натяжение нити во время рывка $T_{\text{рыв}}$ можно оценить следующим образом. Надо вычислить максимальную скорость груза M в нижнем положении. Обозначим ее через v. За время полуоборота шкива $\Delta t = \pi r/v$ импульс груза M меняется на $2Mv$. Это изменение равно импульсу силы, действующей на груз M, за то же время, т. е. $(T_{\text{рыв}} - Mg)\,\Delta t$. Вычисления дают $T_{\text{рыв}} = Mg + \dfrac{Mhr}{\pi m R^2} T \approx 1{,}42T_0$.

3. Монета массой m и радиусом r, вращаясь в горизонтальной плоскости вокруг своей оси с угловой скоростью ω, вертикально падает на горизонтальный диск и прилипает к нему. В результате диск приходит во вращение вокруг своей оси. Возникающий при этом момент сил трения в оси диска постоянен и равен M_0. Через какое время вращение диска прекратится? Сколько оборотов N сделает диск до полной остановки? Момент инерции ди-

ска относительно его геометрической оси I_0. Расстояние между осями диска и монеты равно d.

Ответ. $t = \dfrac{mr^2}{2M_0}\omega$; $N = \dfrac{M_0}{2I}t^2$, где $I = I_0 + m\left(d^2 + \dfrac{r^2}{2}\right)$.

4. Сплошной однородный короткий цилиндр радиусом r, вращающийся вокруг своей геометрической оси со скоростью n об/с, ставят в вертикальное положение на горизонтальную поверхность. Сколько оборотов N сделает цилиндр, прежде чем вращение его полностью прекратится? Коэффициент трения скольжения между основанием цилиндра и поверхностью, на которую он поставлен, не зависит от скорости вращения и равен k.

Ответ. $N = \dfrac{3\pi r n^2}{4kg}$.

5. Тонкий стержень массой m и длиной L (рис. 73) подвешен за один конец и может вращаться без трения вокруг горизонтальной оси. К той же оси подвешен на нити длиной l шарик с такой же массой m. Шарик отклоняется на некоторый угол и отпускается. При какой длине нити шарик после удара о стержень остановится? Удар абсолютно упругий.

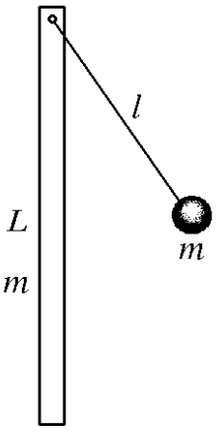


Рис. 73

Ответ. $l = \dfrac{L}{\sqrt{3}}$.

6. Математический маятник массой m и стержень массой M (рис. 74) подвешены к одной и той же точке A, вокруг которой они могут свободно колебаться. Длина нити маятника равна длине стержня. Шарик маятника отклоняют в сторону, так что он приподнимается на высоту H относительно своего нижнего положения. Затем шарик отпускают, и он сталкивается неупруго с палкой. Как будут двигаться шарик и нижний конец палки после удара и на какие высоты они поднимутся?

Решение. Скорость шарика в нижнем положении до удара $v_0 = \sqrt{2gH}$. Так как удар неупругий, то непосредственно после удара шарик и нижний конец стержня в нижнем положении будут иметь одну и ту же скорость v. Она найдется из закона сохранения момента импульса относительно оси A:

$$mlv_0 = mlv + I\omega,$$

где $I = \frac{1}{3}Ml^2$ — момент инерции стержня относительно той же оси. Так как $v = l\omega$, то написанное уравнение дает

$$v = \frac{ml^2}{I + ml^2}v_0 = \frac{3m}{M + 3m}v_0.$$

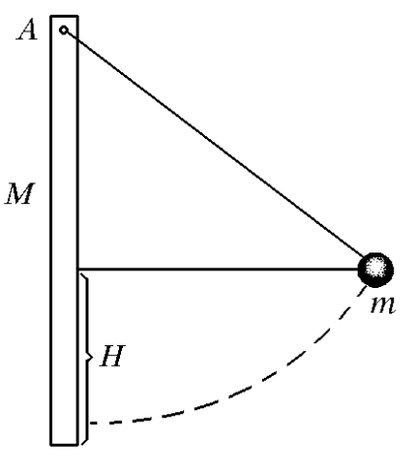


Рис. 74

Теперь надо решить, будут ли шарик и стержень после столкновения двигаться вместе или при дальнейшем движении они разойдутся. С этой целью вычислим скорость шарика v_1 и нижнего конца стержня v_2 при поднятии на какую-либо одну и ту же высоту h_1, если бы при этом они двигались независимо друг от друга. Эти скорости найдутся из уравнения

сохранения энергии

$$v^2 - v_1^2 = 2gh_1, \quad \frac{1}{2}\frac{I}{l^2}(v^2 - v_2^2) = Mg\frac{h_1}{2}.$$

Преобразовав второе уравнение к виду

$$v^2 - v_2^2 = 3gh_1,$$

видим, что $v_1 > v_2$. Поэтому в любом положении шарик будет стремиться обогнать стержень. А так как шарик движется позади стержня, то он все время будет прижиматься к стержню. Отсюда следует, что после удара шарик и стержень будут подниматься как единое тело. Высоту поднятия h легко определить из закона сохранения энергии. Она равна

$$h = \frac{I + ml^2}{(M + 2m)gl^2}v^2 = \frac{6m^2}{(M + 2m)(M + 3m)}H.$$

7. Решить предыдущую задачу в предположении, что до удара был отклонен стержень (нижний конец его был бы поднят на высоту H).

Ответ. После удара шарик поднимается на высоту

$$h_1 = \frac{3}{2}\left(\frac{M}{M + 3m}\right)^2 H,$$

нижний конец стержня — на высоту

$$h_2 = \left(\frac{M}{M + 3m}\right)^2 H = \frac{2}{3}h_1.$$

8. Твердый стержень длиной l и массой M может вращаться вокруг горизонтальной оси A, проходящей через его конец (рис. 75). К той же оси A подвешен математический маятник такой же длины l и массы m. Первоначально стержень занимает горизонтальное положение, а затем отпускается. В нижнем положении происходит идеально упругий удар, в результате которого шарик и стержень деформируются, и часть кинетической энергии переходит в потенциальную энергию деформации. Затем деформация уменьшается, и запасенная потенциальная энергия вновь переходит в кинетическую. Найти значение потенциальной энергии деформации U в момент, когда она максимальна.

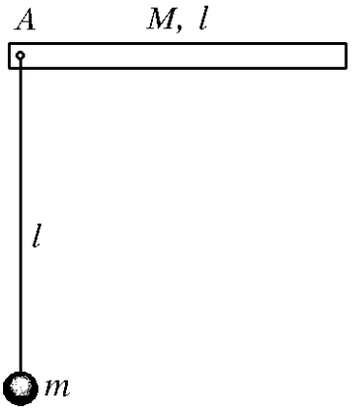


Рис. 75

Ответ. $U = \frac{1}{2}\frac{ml^2}{I + ml^2}Mgl = \frac{3}{2}\frac{Mm}{M + 3m}gl$, где I — момент инерции стержня.

9. Вертикально висящая однородная доска длиной $L = 1{,}5$ м и массой $M = 10$ кг может вращаться вокруг горизонтальной оси, проходящей через ее верхний конец. В нижний конец доски ударяет пуля, летящая горизонтально с начальной скоростью $V_0 = 600$ м/с. Пуля пробивает доску и летит далее со скоростью V. Определить скорость V, если после выстрела доска стала колебаться с угловой амплитудой $\alpha = 0{,}1$ рад. Масса пули $m = 10$ г.

Ответ. $V = V_0 - \frac{M}{m}\sqrt{\frac{2}{3}gL}\sin\frac{\alpha}{2} = 444$ м/с.

10. В общей точке подвеса A (рис. 76) подвешены шарик на нити длиной l и однородный стержень длиной L, отклоненный в сторону на некоторый угол. При возвращении стержня в положение равновесия происходит упругий удар. При каком соотношении между массами стержня M и шарика m шарик и точка удара стержня будут двигаться после удара с равными скоростями в противоположных направлениях? При каком соотношении между массами M и m описанный процесс невозможен?

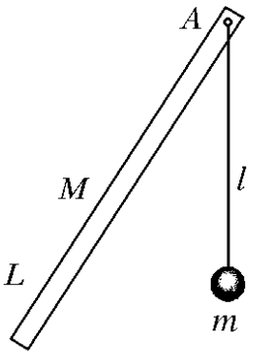


Рис. 76

Ответ. $ML^2 = ml^2$. Так как $L \geqslant l$, то для возможности процесса необходимо $M \leqslant m$. При $M > m$ процесс невозможен.

11. На горизонтальный диск, вращающийся вокруг геометрической оси с угловой скоростью ω_1, падает другой диск, вращающийся вокруг той же оси с угловой скоростью ω_2. Моменты инерции дисков относительно указанной оси равны соответственно I_1 и I_2. Оба диска при ударе сцепляются друг с другом (при помощи острых шипов на их поверхности). На сколько изменится общая кинетическая энергия вращения системы после падения второго диска? Чем объясняется изменение энергии? Геометрические оси обоих дисков являются продолжением одна другой.

Ответ. Кинетическая энергия вращения уменьшается на

$$\Delta K = \frac{1}{2}\frac{I_1 I_2}{I_1 + I_2}(\omega_1 - \omega_2)^2.$$

12. Шкивы двух маховиков соединены ремнем (рис. 77). Радиусы шкивов равны R_1 и R_2. Моменты инерции маховиков относительно их геометрических осей равны I_1 и I_2. Удерживая второй маховик и ремень неподвижными, раскручивают первый маховик до угловой скорости ω_0, вследствие чего между осью первого маховика и ремнем возникает скольжение. Затем ремень и второй маховик отпускают. Пренебрегая всеми силами трения, за исключением сил трения скольжения между ремнем и осями маховиков, найти установившиеся скорости вращения маховиков ω_1 и ω_2, т. е. скорости после прекращения скольжения. Найти также потерю ΔK кинетической энергии на трение скольжения. Массой ремня пренебречь.

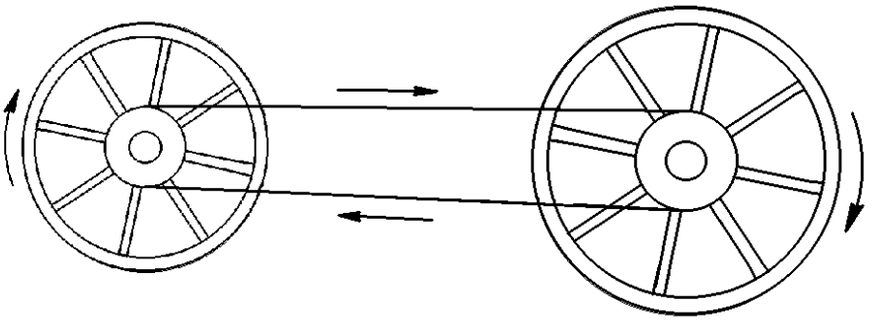


Рис. 77

Решение. Благодаря трению скольжения натяжение ремня сверху T_1 и снизу T_2 будет разным. Применяя к маховикам уравнение (33.4), получим

$$I_1 \frac{d\omega_1}{dt} = (T_1 - T_2)R_1, \quad I_2 \frac{d\omega_2}{dt} = (T_2 - T_1)R_2.$$

Поделим эти уравнения соответственно на R_1 и R_2, сложим и проинтегрируем. Тогда получим

$$\frac{I_1\omega_1}{R_1}+\frac{I_2\omega_2}{R_2}=\text{const}.$$

Входящая сюда постоянная равна $I_1\omega_0/R_1$, так как в начальный момент $\omega_1=\omega_0$, $\omega_2=0$. Когда скольжение прекратится, то $\omega_1R_1=\omega_2R_2$. Решая полученную систему уравнений, найдем угловые скорости ω_1 и ω_2 после прекращения скольжения:

$$\omega_1=\frac{I_1R_2^2}{I_1R_2^2+I_2R_1^2}\,\omega_0,\quad \omega_2=\frac{I_1R_1R_2}{I_1R_2^2+I_2R_1^2}\,\omega_0.$$

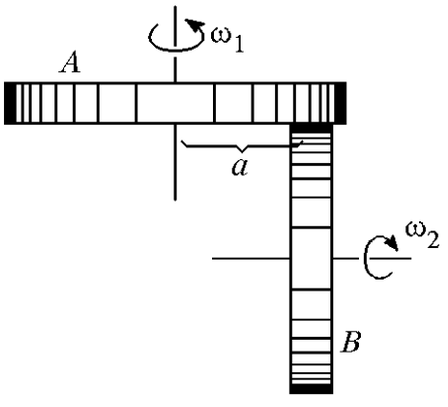


Рис. 78

Потеря кинетической энергии на трение равна

$$\Delta K=\frac{1}{2}\,\frac{I_1I_2R_1^2}{I_1R_2^2+I_2R_1^2}\,\omega_0^2.$$

13. Почему в предыдущей задаче полный момент импульса системы не сохраняется?

14. Однородный диск A массой M_1 и радиусом r_1 (рис. 78) раскручен до угловой скорости ω_0 и приведен в контакт с диском B, ось вращения которого перпендикулярна к оси диска A. Масса диска B равна M_2, а расстояние между точкой соприкосновения и осью диска A равна a. Найти установившиеся угловые скорости дисков ω_1 и ω_2 и потерю энергии в процессе установления. Трением в осях, а также трением качения пренебречь.

Ответ. $\omega_1=\dfrac{M_1r_1^2}{M_1r_1^2+M_2a^2}\,\omega_0$, $\omega_2=\dfrac{M_1r_1^2}{M_1r_1^2+M_2a^2}\,\dfrac{a}{r_2}\,\omega_0=\dfrac{a}{r_2}\,\omega_1$.

Потеря энергии

$$\Delta K=\frac{M_1M_2r_1^2a^2}{4(M_1r_1^2+M_2a^2)}\,\omega_0^2.$$

15. Вертикальный столб высотой l подпиливается у основания и падает на землю, поворачиваясь вокруг нижнего основания. Определить линейную скорость его верхнего конца в момент удара о землю. Какая точка столба будет в этот момент иметь ту же скорость, какую имело бы тело, падая с той же высоты, как и данная точка?

Ответ. $v=\sqrt{3gl}$. Искомая точка находится на расстоянии $x=\frac{2}{3}l$ от основания столба.

16. Изменится ли ответ в предыдущей задаче, если столб первоначально стоял в вертикальном положении на абсолютно гладком льду, а затем начал падать под действием силы тяжести? Чем будет отличаться движение столба в этом случае от движения в предыдущем случае?

Решение. По теореме Кёнига кинетическая энергия столба слагается из кинетической энергии движения его центра масс $\frac{1}{2}\,mv_C^2$ со скоростью v_C и кинетической энергии вращения $\frac{1}{2}\,I\omega^2$ вокруг центра масс с угловой скоростью ω. За время падения центр масс проходит путь $l/2$. При этом совершается работа $mgl/2$, которая идет на приращение кинетической энергии:

$$\frac{1}{2}\,mv_C^2+\frac{1}{2}\,I\omega^2=\frac{1}{2}\,mgl.$$

В нижнем положении, когда столб горизонтален, $v_C = \frac{1}{2} l\omega$. Имея это в виду, а также используя выражение $I = \frac{1}{2} ml^2$, получим $v_C = \frac{1}{2}\sqrt{3gl}$. Скорость верхнего конца столба вдвое больше, т. е. равна $v = \sqrt{3gl}$. Отсюда видно, что результаты будут такими же, что и в предыдущей задаче. Однако характер движения будет другим. В предыдущем случае столб при падении вращался вокруг своего нижнего основания. При этом центр масс столба двигался по дуге окружности. В рассматриваемом случае, поскольку все действующие силы направлены вертикально, центр масс столба при его падении все время будет находиться на одной и той же вертикали.

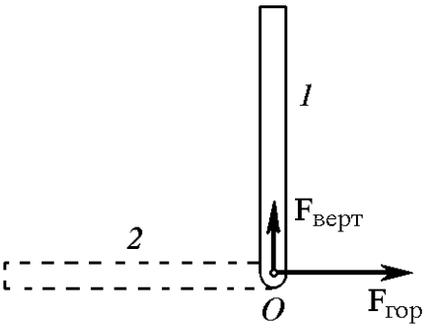


Рис. 79

17. Однородный стержень массой m и длиной l (рис. 79) падает без начальной скорости из положения 1, вращаясь без трения вокруг неподвижной горизонтальной оси O. Найти горизонтальную $F_{гор}$ и вертикальную $F_{верт}$ составляющие силы, с которыми ось O действует на стержень в горизонтальном положении 2.

Р е ш е н и е. Кинетическая энергия стержня в горизонтальном положении $\frac{1}{2} I\omega^2 = \frac{1}{2} mgl$. Центростремительное ускорение центра масс стержня в том же положении $\frac{1}{2}\omega^2 l$. Отсюда по теореме о движении центра масс

$$F_{гор} = m\omega^2 \frac{l}{2} = \frac{ml^2}{2I} mg = \frac{3}{2} mg.$$

Применив к вращению стержня в положении 2 уравнение (33.4), получим

$$I \frac{d\omega}{dt} = mg \frac{l}{2}.$$

Отсюда находим вертикальную составляющую ускорения центра масс в том же положении:

$$a = \frac{l}{2} \frac{d\omega}{dt} = \frac{mgl^2}{4I} = \frac{3}{4} g.$$

Далее,

$$ma = mg - F_{верт}.$$

В результате получится

$$F_{верт} = m(g - a) = \frac{1}{2} mg.$$

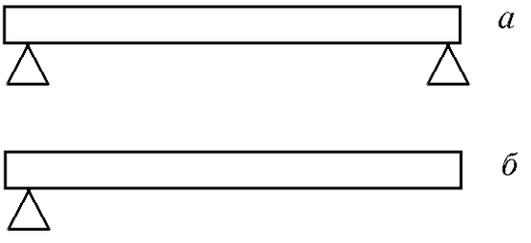


Рис. 80

18. Абсолютно твердая однородная балка весом P лежит своими концами на двух абсолютно твердых опорах (рис. 80 $а$). Одну из них выбивают. Найти начальную силу давления, действующую на оставшуюся опору (рис. 80 $б$).

О т в е т. $F = \frac{1}{4} P$.

Когда балка лежала на двух опорах, на каждую из опор действовала сила $\frac{1}{2}P$. При быстром удалении одной из опор сила, действующая на оставшуюся опору, скачкообразно уменьшается вдвое. Такое скачкообразное изменение связано с идеализацией — балка и опора считаются абсолютно твердыми. Реальные балки и опоры деформируются. При учете этого обстоятельства нагрузка на опору F будет меняться непрерывно.

19. Гимнаст на перекладине выполняет большой оборот из стойки на руках, т. е. вращается, не сгибаясь, вокруг перекладины под действием собственного веса. Оценить приближенно наибольшую нагрузку F на его руки, пренебрегая трением ладоней о перекладину.

Ответ. $F = \left(1 + \dfrac{4a^2m}{I}\right)mg$, где m — масса, I — момент инерции человека относительно перекладины, a — расстояние между осью вращения и центром масс человека. Если при оценке момента инерции моделировать человека однородным стержнем, вращающимся вокруг одного из его концов, то получится $F = 4mg$.

20. Человек на аттракционе «гигантские шаги» движется по замкнутой траектории таким образом, что достигаемая им высота относительно положения равновесия меняется в пределах от $h_{\text{мин}}$ до $h_{\text{макс}}$. Определить максимальную и минимальную скорости человека при таком движении, если длина веревки, на которой он удерживается, равна l.

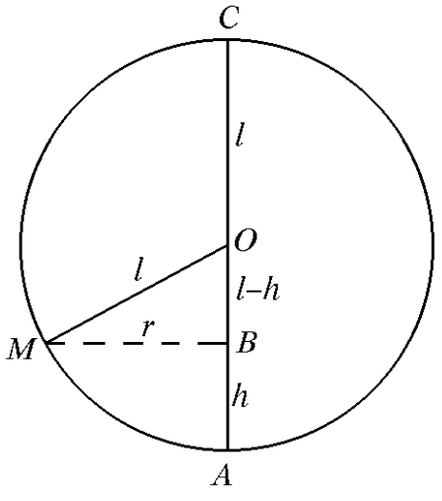


Рис. 81

Решение. На основании закона сохранения энергии

$$v^2 + 2gh = \text{const}. \tag{37.3}$$

Момент силы тяжести относительно точки подвеса не имеет вертикальной составляющей. Момент силы натяжения веревки равен нулю. Поэтому при движении человека вертикальная составляющая его момента импульса остается неизменной. В положениях, где высота h максимальна или минимальна, скорость человека v горизонтальна, а момент импульса равен mvr, где r — расстояние до вертикальной оси, вокруг которой вращается человек. Значит, в этих положениях величина vr одна и та же. В момент, когда высота h максимальна или минимальна, опишем в вертикальной плоскости окружность с центром в точке подвеса O, проходящую через точку нахождения человека M (рис. 81). По известной геометрической теореме $r^2 = AB \cdot BC$, или $r^2 = (2l - h)h$. Поэтому в положениях, где h максимальна и минимальна,

$$(2l - h)hv^2 = \text{const}. \tag{37.4}$$

Запишем соотношения (37.3) и (37.4) для этих положений, имея в виду, что максимуму h соответствует минимум v, и наоборот. Получим

$$v^2_{\text{макс}} + 2gh_{\text{мин}} = v^2_{\text{мин}} + 2gh_{\text{макс}},$$

$$(2l - h_{\text{мин}})h_{\text{мин}}v^2_{\text{макс}} = (2l - h_{\text{макс}})h_{\text{макс}}v^2_{\text{мин}}.$$

Решая эти уравнения, получим

$$v^2_{\text{макс}} = \frac{2gh_{\text{макс}}(2l - h_{\text{макс}})}{2l - (h_{\text{макс}} + h_{\text{мин}})}, \tag{37.5}$$

$$v^2_{\text{мин}} = \frac{2gh_{\text{мин}}(2l - h_{\text{мин}})}{2l - (h_{\text{макс}} + h_{\text{мин}})}. \tag{37.6}$$

При этом учтено, что в реальных условиях $h < l$, так что величина (37.5) действительно максимальна, а (37.6) — действительно минимальна. Если $h_{\text{макс}}$ и $h_{\text{мин}}$ пренебрежимо малы по сравнению с l, то

$$v^2_{\text{макс}} = 2gh_{\text{макс}}, \quad v^2_{\text{мин}} = 2gh_{\text{мин}}. \tag{37.7}$$

21. По внутренней поверхности конической воронки, стоящей вертикально, без трения скользит маленький шарик (рис. 82). В начальный момент шарик находился на высоте h_0, а скорость его v_0 была горизонтальна. Найти v_0, если известно, что при дальнейшем движении шарик поднимается до высоты h, а затем начинает опускаться. Найти также скорость шарика в наивысшем положении v.

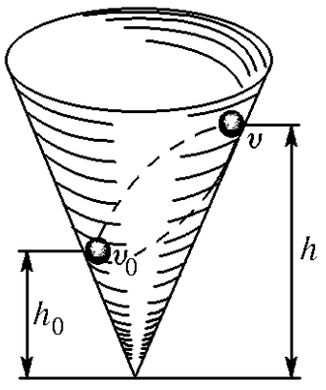


Рис. 82

Ответ. $v_0^2 = \dfrac{2gh^2}{h+h_0}$, $v^2 = \dfrac{2gh_0^2}{h+h_0}$.

22. Тяжелая веревка (линейная плотность ρ) длиной L перекинута через блок с моментом инерции I и радиусом r. В начальный момент блок неподвижен, а больший из свешивающихся концов веревки имеет длину l. Найти угловую скорость вращения блока ω, когда веревка соскользнет с него. Веревка движется по блоку без скольжения, трение в оси блока не учитывать.

Ответ: $\omega^2 = \dfrac{\rho g}{I + L\rho r^2}\,[L^2 + 4r^2 - l^2 - (L - l - \pi r)^2]$.

Указание. Воспользоваться законом сохранения энергии.

23. Метеорит массой $m = 10^5$ т, двигавшийся со скоростью $v = 50$ км/с, ударился о Землю на широте $\vartheta = 60°$. Вся его кинетическая энергия перешла в тепловую (внутреннюю) энергию, а сам он испарился. Какое максимальное влияние мог оказать удар такого метеорита на продолжительность суток?

Ответ: Максимальное изменение продолжительности суток ΔT, вызванное ударом метеорита, определяется формулой

$$\frac{\Delta T}{T} = \pm\frac{mvR\cos\vartheta}{2\pi I}\,T,$$

где $T = 86164$ с — продолжительность суток, $R = 6400$ км — радиус, $M = 6\cdot 10^{21}$ т — масса Земли, I — ее момент инерции. Если считать Землю однородным шаром, то $I = {}^2/_5\,MR^2$ (на самом деле из-за возрастания плотности к центру Земли момент инерции ее несколько меньше и составляет приблизительно $I = {}^1/_3\,MR^2$). В результате получится $\Delta T/T \sim 2\cdot 10^{-5}$, $\Delta T \sim 2\cdot 10^{-10}$ с.

24. Оценить, с какой минимальной скоростью v нужно выпустить на экваторе Земли снаряд массой $m = 1000$ т, чтобы изменить продолжительность земных суток на $\Delta T = 1$ мин?

Ответ. Наивыгоднейшим является выстрел в горизонтальном направлении в плоскости экватора. В этом случае

$$\frac{c-v}{c} \approx \frac{5}{18}\,\frac{m^2c^2T^4}{\pi^2 IM(\Delta T)^2} \approx \frac{25}{36}\,\frac{m^2c^2T^4}{\pi^2M^2R^2(\Delta T)^2} \approx 2\cdot 10^{-22},$$

где c — скорость света в вакууме. Остальные обозначения такие же, как в предыдущей задаче. Относительно приближенных вычислений см. § 22.

25. *Пульсарами* называются небесные объекты, посылающие импульсы радиоизлучения, следующие друг за другом с высокостабильными периодами, которые для известных к настоящему времени пульсаров лежат в пределах примерно от $3\cdot10^{-2}$ до 4 с. Согласно современным представлениям пульсары представляют собой вращающиеся нейтронные звезды, образовавшиеся в результате гравитационного сжатия. Нейтронные звезды подобны гигантским атомным ядрам, построенным из одних только нейтронов. Плотность вещества ρ в нейтронной звезде не однородна, но при грубых оценках ее можно считать одной и той же по всему объему звезды и по порядку величины равной 10^{14} г/см^3. Оценить период вращения T, с каким стало бы вращаться Солнце, если бы оно превратилось в нейтронную звезду. Плотность вещества Солнца возрастает к его центру, а различные слои его вращаются с различными скоростями. При оценке этими обстоятельствами пренебречь и считать, что средняя плотность солнечного вещества $\rho_0 = 1{,}41$ г/см^3, а период вращения Солнца $T_0 = 2{,}2\cdot10^6$ с.

Ответ. $T \approx T_0\left(\frac{\rho_0}{\rho}\right)^{2/3} = 1{,}3\cdot10^{-3}$ с.

26. Гладкий твердый стержень длиной l_0 и массой M равномерно вращается с угловой скоростью ω_0 вокруг неподвижной оси, проходящей через один из концов стержня перпендикулярно к его продольной оси. На стержень надет шарик массой m. Вначале шарик находится на свободном конце стержня и вращается вместе с ним (упор, имеющийся на конце стержня, не позволяет шарику соскользнуть со стержня). В некоторый момент шарику сообщается скорость v, направленная вдоль стержня к оси вращения. Определить наименьшее расстояние l, до которого приблизится шарик к оси вращения, и угловую скорость системы ω в этом положении. В какую сторону будет изогнут стержень, когда шарик движется по направлению к оси вращения? Как изменится изгиб стержня, когда шарик, достигнув наименьшего удаления до оси, начнет двигаться в обратном направлении?

Ответ. $\omega = \omega_0 + \frac{mv^2}{(1/3\,M + m)l_0^2\omega_0}$, $l = l_0\sqrt{\frac{\omega_0}{\omega}}$. При приближении шарика к оси вращения стержень будет изгибаться в сторону, противоположную вращению. При удалении шарика изгиб стержня изменится в обратную сторону.

§ 38. ЗАКОНЫ СОХРАНЕНИЯ И СИММЕТРИЯ ПРОСТРАНСТВА И ВРЕМЕНИ

1. Законы сохранения энергии, импульса и момента импульса связаны с определенными свойствами *симметрии пространства и времени.* Разумеется, их нельзя получить из одних только свойств симметрии какого бы то ни было вида. Законы сохранения относятся к материальным физическим объектам, а последние не являются только пространственно-временными конструкциями. Для вывода законов сохранения в механике нужны какие-то дополнительные исходные положения механического характера. Однако число этих исходных положений можно уменьшить, если использовать свойства симметрии пространства и времени. Так, в этом случае при выводе

законов сохранения импульса и момента импульса не потребуется закон равенства действия и противодействия.

В аналитической механике основные законы обычно выражают через *функцию Лангранжа* (по имени французского математика и механика Жозефа Лангранжа (1736—1813)), и симметрия, о которой идет речь, означает пространственно-временную симметрию самой функции Лангранжа. В общем курсе физики мы не можем идти по этому пути. В основу изложения мы положим *второй закон Ньютона.* Тогда утверждение, приведенное в начале этого параграфа, надо понимать в том смысле, что перечисленные в нем законы сохранения *можно получить из второго закона Ньютона, если к нему присоединить свойства симметрии пространства и времени.* Впрочем, при выводе закона сохранения энергии надо ввести и некоторые более специальные предположения относительно характера действующих сил.

2. Под симметрией пространства и времени понимают *однородность времени, однородность и изотропию пространства.* Однородность времени означает равноправие всех моментов времени. Однородность пространства означает, что в пространстве нет выделенных положений, все точки пространства равноправны. Аналогично изотропия пространства характеризуется отсутствием в нем выделенных направлений, все направления в пространстве эквивалентны.

Более полно:

Однородность времени означает, что если в два любые момента времени все тела замкнутой системы поставить в совершенно одинаковые условия, то начиная с этих моментов все явления в ней будут протекать совершенно одинаково.

Однородность пространства означает, что если замкнутую систему тел перенести из одного места пространства в другое, поставив при этом все тела в ней в те же условия, в каких они находились в прежнем положении, то это не отразится на ходе всех последующих явлений. В том же смысле надо понимать и изотропию пространства, только вместо переноса замкнутой системы надо говорить о ее *повороте* в пространстве на любой угол.

Оговорка относительно замкнутости системы существенна. Положения тела у подножья горы и на ее вершине отнюдь не эквивалентны. На Земле направления вверх и вниз объективно характеризуются разными свойствами: тела на Земле из состояния покоя самопроизвольно всегда падают вниз, но никогда не поднимаются вверх. Одно и то же явление протекает по времени по-разному, если его начать воспроизводить в различные моменты времени, так как при переходе от одного момента времени к другому может измениться обстановка, в которой явление происходит. Но все эти примеры относятся к незамкнутым системам. На поведении замкнутых систем изменения окружающей обстановки не сказываются.

Здесь необходимо сделать такое же замечание, что и в § 15 в связи с принципом относительности Галилея. Нельзя понимать под замкнутой системой тел всю Вселенную. Если поступить так, то

перечисленные свойства симметрии пространства и времени стали бы самоочевидными. Но они стали бы и бессодержательными. Ибо говорить о переносе или повороте системы тел можно только по отношению к каким-то другим телам. Речь идет не о всей Вселенной в целом, а о таких ее частях, которые можно рассматривать как (приближенно) *замкнутые системы.* Отсюда ясно, что свойства пространства и времени, о которых мы говорили, отнюдь не самоочевидны. На них надо смотреть как на *фундаментальные обобщения опытных фактов.*

3. После этих разъяснений обратимся к выводу закона сохранения энергии в механике. Из динамики мы заимствуем следствие второго закона Ньютона, выражающееся формулой

$$A_{12} = K_2 - K_1, \tag{38.1}$$

т. е. работа сил над механической системой равна приращению ее кинетической энергии K (см. § 22). Следующую часть наших рассуждений проведем применительно к одной материальной точке. В случае системы материальных точек все будет обстоять так же, изменится только число аргументов, от которых зависит потенциальная функция U, вводимая ниже. Предположим, что проекции силы F_x, F_y, F_z, действующие на материальную точку, могут быть получены дифференцированием потенциальной функции U:

$$F_x = -\frac{\partial U}{\partial x}, \quad F_y = -\frac{\partial U}{\partial y}, \quad F_z = -\frac{\partial U}{\partial z}.$$

Однако сама потенциальная функция U может зависеть явно не только от координат x, y, z рассматриваемой материальной точки, но и от времени t: $U = U(x, y, z, t)$. Например, это будет так, когда точка находится в силовом поле других тел, которое меняется во времени. Работа, производимая действующими силами над материальной точкой при перемещении ее вдоль некоторой кривой из положения *1* в положение *2*, представляется интегралом

$$A_{12} = -\int \left(\frac{\partial U}{\partial x}\,dx + \frac{\partial U}{\partial y}\,dy + \frac{\partial U}{\partial z}\,dz\right),$$

взятым вдоль той же кривой. Прибавим и вычтем под знаком интеграла член $\frac{\partial U}{\partial t}\,dt$. Тогда, вводя полный дифференциал

$$dU = \frac{\partial U}{\partial x}\,dx + \frac{\partial U}{\partial y}\,dy + \frac{\partial U}{\partial z}\,dz + \frac{\partial U}{\partial t}\,dt,$$

представим предыдущее выражение в виде

$$A_{12} = -\int dU + \int \frac{\partial U}{\partial t}\,dt.$$

В таком виде оно справедливо и для системы материальных точек. Поэтому дальнейшие рассуждения не связаны с предположением,

что система состоит из одной материальной точки. После интегрирования получаем

$$A_{12} = U_1 - U_2 + \int \frac{\partial U}{\partial t}\, dt. \tag{38.2}$$

Комбинация этой формулы с (38.1) приводит к соотношению

$$(K_2 + U_2) - (K_1 + U_1) = \int \frac{\partial U}{\partial t}\, dt. \tag{38.3}$$

До сих пор мы не использовали условие замкнутости системы и свойства однородности времени, поэтому наши рассуждения применимы и для незамкнутых систем. Допустим теперь, что система замкнута. Тогда ввиду однородности времени функция U не может явно зависеть от времени, т. е. $\frac{\partial U}{\partial t} = 0$. В результате получим

$$K_1 + U_1 = K_2 + U_2, \tag{38.4}$$

т. е. уравнение, выражающее механический закон сохранения энергии.

4. Перейдем к доказательству закона сохранения импульса. Допустим, что механическая система замкнута. Все силы $\mathbf{F}_1, \mathbf{F}_2, \ldots$, действующие на материальные точки системы, являются силами внутренними, внешних сил нет. Ввиду однородности пространства энергия системы не изменится, если систему сместить из одного положения пространства в другое. Математически

$$U(\mathbf{r}_1, \mathbf{r}_2, \ldots) = U(\mathbf{r}_1 + \mathbf{R}, \mathbf{r}_2 + \mathbf{R}, \ldots),$$

каковы бы ни были радиусы-векторы материальных точек системы в начальном положении $\mathbf{r}_1, \mathbf{r}_2, \ldots$ и каков бы ни был вектор смещения $\mathbf{R}$. Выберем $\mathbf{R}$ бесконечно малым и направим его вдоль оси X: $\mathbf{R} = \delta x\, \mathbf{i}$. Тогда

$$\left(\frac{\partial U}{\partial x_1} + \frac{\partial U}{\partial x_2} + \ldots\right) \delta x = 0.$$

Ввиду произвольности δx выражение в скобках обращается в нуль. А в силу (29.9) его можно представить в виде $F_{1x} + F_{2x} + \ldots = 0$. Аналогичные соотношения можно написать и для проекций сил на оси Y и Z. Вообще для замкнутой системы $\mathbf{F}_1 + \mathbf{F}_2 + \ldots = 0$. А это есть как раз то условие, при выполнении которого из второго закона Ньютона получается закон сохранения импульса (см. § 12).

5. Закон сохранения момента импульса для замкнутой системы доказывается в точности так же. Используя изотропию пространства, можно доказать, что геометрическая сумма моментов внутренних сил, действующих в системе, равна нулю: $\mathbf{M}_1 + \mathbf{M}_2 + \ldots = 0$ (см. задачу 2 к § 46). Отсюда немедленно следует рассматриваемый закон (см. § 30).

ЗАДАЧИ

1. Пусть $U(\mathbf{r}_1, \mathbf{r}_2)$ означает потенциальную энергию взаимодействия двух материальных точек как функцию радиусов-векторов $\mathbf{r}_1$ и $\mathbf{r}_2$, определяющих их положения в пространстве. Используя однородность пространства, доказать, что U является функцией только разности $\mathbf{r}_2 - \mathbf{r}_1$. Обобщить результат на случай системы n взаимодействующих материальных точек.

Р е ш е н и е. Ввиду однородности пространства потенциальная энергия U не изменится, если обе взаимодействующие точки сместить на один и тот же вектор $\mathbf{a}$. Записанное математически, это условие гласит: $U(\mathbf{r}_1, \mathbf{r}_2) = U(\mathbf{r}_1 + \mathbf{a}, \mathbf{r}_2 + \mathbf{a})$. Это соотношение должно выполняться, каков бы ни был вектор $\mathbf{a}$. Полагая $\mathbf{a} = -\mathbf{r}_1$, получим $U = U(0, \mathbf{r}_2 - \mathbf{r}_1)$, т. е. $U = f(\mathbf{r}_2 - \mathbf{r}_1)$, где f — какая-то функция только разности $\mathbf{r}_2 - \mathbf{r}_1$.

Если система состоит из n взаимодействующих материальных точек, то, рассуждая аналогично, найдем

$$U = f(\mathbf{r}_2 - \mathbf{r}_1, \mathbf{r}_3 - \mathbf{r}_1, \ldots).$$

Разумеется, вместо первой точки можно взять любую из материальных точек системы. Значит, потенциальная энергия U может зависеть только от $n - 1$ векторных аргументов: разностей радиусов-векторов каких-либо $n - 1$ точек системы и радиуса-вектора остальной точки.

2. Какие дополнительные ограничения накладывает на вид функции U изотропия пространства?

О т в е т. Потенциальная энергия U может зависеть только от расстояний каких-либо $n - 1$ материальных точек системы от остальной точки.

3. Используя однородность пространства и галилеевский принцип относительности, показать, что сила взаимодействия материальных точек *1* и *2* не зависит от их координат и скоростей, а может зависеть только *от разностей* этих координат и скоростей.

Р е ш е н и е. В силу однородности пространства и галилеевского принципа относительности ускорение $\mathbf{a}$, а с ним и сила $\mathbf{f} = m\mathbf{a}$ инвариантны относительно переноса начала координат и преобразования Галилея. Возьмем две системы отсчета S и S'. Рассматривая силу $\mathbf{f}$ как функцию координат и скоростей в системе S', напишем $\mathbf{f} = \mathbf{f}(\mathbf{r}_1', \mathbf{r}_2', \mathbf{v}_1', \mathbf{v}_2')$. Систему S' можно выбрать произвольно. Выберем ее так, чтобы в рассматриваемый момент времени материальная точка *1* находилась в начале координат $(\mathbf{r}_1' = 0)$, а ее скорость равнялась нулю $(\mathbf{v}_1' = 0)$. Тогда в этот момент сила $\mathbf{f}$ будет функцией только двух аргументов: $\mathbf{f} = \mathbf{f}(\Delta\mathbf{r}_2', \Delta\mathbf{v}_2')$. Но разности координат и скоростей в обеих системах отсчета одинаковы, а потому $\Delta\mathbf{r}_2' = \mathbf{r}_2' - \mathbf{r}_1' = \mathbf{r}_2 - \mathbf{r}_1$, $\Delta\mathbf{v}_2' = \mathbf{v}_2' - \mathbf{v}_1' = \mathbf{v}_2 - \mathbf{v}_1$. В результате получим

$$\mathbf{f} = \mathbf{f}(\mathbf{r}_2 - \mathbf{r}_1, \mathbf{v}_2 - \mathbf{v}_1).$$

ГЛАВА VI

ГАРМОНИЧЕСКИЕ КОЛЕБАНИЯ

§ 39. КИНЕМАТИКА ГАРМОНИЧЕСКОГО КОЛЕБАТЕЛЬНОГО ДВИЖЕНИЯ

Колебательные явления играют важную роль в самых разнообразных вопросах физики. Подробный разбор их дается в других разделах нашего курса. Здесь же мы ограничимся предварительным рассмотрением простейших *механических колебаний*. Начнем с *колебательного движения* материальной точки. В таком движении точка через равные промежутки времени проходит через одно и то же положение и притом в одном и том же направлении.

Важнейшим среди колебательных движений является так называемое *простое* или *гармоническое колебательное движение*. О нем мы уже говорили в § 11. Характер такого движения лучше всего раскрывается с помощью следующей кинематической модели. Допустим, что геометрическая точка M равномерно вращается по окружности радиуса A с постоянной угловой скоростью ω (рис. 83). Ее проекция N на диаметр, например, на ось X, будет совершать колебательное движение от крайнего положения N_1 до другого крайнего положения N_2 и обратно. Такое колебание точки N и называют простым или гармоническим колебанием. Чтобы его описать, надо найти координату x точки N как функцию времени t. Допустим, что в начальный момент времени $t = 0$ радиус OM образовывал с осью X угол δ. Спустя время t этот угол получит приращение ωt и сделается равным $\omega t + \delta$. Из рис. 83 видно, что

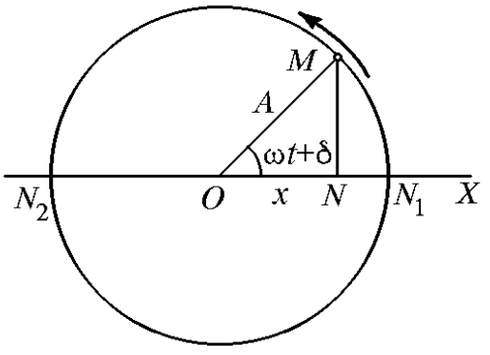


Рис. 83

$$x = A \cos (\omega t + \delta). \qquad (39.1)$$

Эта формула и описывает аналитически гармоническое колебательное движение точки N вдоль диаметра N_1N_2.

Величина A дает максимальное отклонение колеблющейся точки от положения равновесия O. Она называется *амплитудой колебания*. Величина ω называется *циклической частотой*. Величину $\omega t + \delta$ называют *фазой колебания*, а ее значение при $t = 0$, т. е. величину δ, — *начальной фазой*. Если $\delta = 0$, то $x = A \cos \omega t$; если $\delta = -\pi/2$, то $x = A \sin \omega t$ и т. д. Таким образом, при гармоническом

колебании абсцисса x является синусоидальной или косинусоидальной функцией времени t. Для графического изображения гармонического колебательного движения можно откладывать по горизонтальной оси время t, а по вертикальной оси — смещение точки x (см. рис. 22). Тогда получится периодическая кривая — *синусоида.* Форма кривой полностью определяется амплитудой A и циклической частотой ω. Однако ее положение зависит также от начальной фазы δ. По истечении времени

$$T = \frac{2\pi}{\omega} \tag{39.2}$$

фаза получает приращение 2π, а колеблющаяся точка возвращается в свое исходное положение с сохранением начального направления движения. Время T называется *периодом колебания.*

Скорость колеблющейся точки найдется дифференцированием выражения (39.1) по времени. Это дает

$$v = \dot{x} = -\omega A \sin(\omega t + \delta). \tag{39.3}$$

Дифференцируя вторично, получаем ускорение

$$a = \dot{v} = -\omega^2 A \cos(\omega t + \delta), \tag{39.4}$$

или, используя (39.1),

$$a = -\omega^2 x. \tag{39.5}$$

Сила, действующая на материальную точку при гармоническом колебании, равна

$$F = ma = -m\omega^2 x. \tag{39.6}$$

Она пропорциональна отклонению x и имеет противоположное направление. Она всегда направлена к положению равновесия. Такого рода силы часто возникают при малых смещениях материальной точки из положения равновесия.

§ 40. ГАРМОНИЧЕСКИЕ КОЛЕБАНИЯ ГРУЗА НА ПРУЖИНЕ

1. Рассмотрим спиральную пружину, один конец которой закреплен, а к другому подвешено тело массой m (рис. 84). Пусть l_0 — длина недеформированной пружины. Если пружину растянуть или сжать до длины l, то возникает сила F, стремящаяся вернуть тело в положение равновесия. При небольших растяжениях $x = l - l_0$ справедлив *закон Гука* — сила пропорциональна растяжению пружины: $F = -kx$. В этих условиях уравнение движения тела имеет вид

$$m\ddot{x} = -kx. \tag{40.1}$$

Постоянная k называется *коэффициентом упругости* или *жесткостью* пружины. Знак минус означает, что сила F направлена

в сторону, противоположную смещению x, т. е. к положению равновесия.

При выводе уравнения (40.1) предполагалось, что никакие другие силы на тело не действуют. Покажем, что тому же уравнению подчиняется движение тела, подвешенного на пружине в однородном поле тяжести. Обозначим в этом случае буквой X *удлинение пружины,* т. е. разность $X = l - l_0$. Пружина тянет груз вверх с силой kX, сила тяжести — вниз. Уравнение движения имеет вид

$$m\ddot{X} = -kX + mg.$$

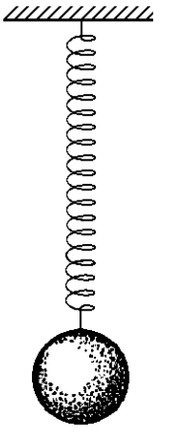


Рис. 84

Пусть X_0 означает удлинение пружины в положении равновесия. Тогда $-kX_0 + mg = 0$. Исключая вес mg, получим $m\ddot{X} = -k(X - X_0)$, тогда уравнение движения примет прежний вид (39.1). Величина x по-прежнему означает смещение груза из положения равновесия. Однако положение равновесия смещается под действием силы тяжести. Кроме того, при наличии тяжести меняется смысл величины $-kx$. Теперь она означает равнодействующую сил натяжения пружины и веса груза. Но все это не затрагивает математическую сторону колебательного процесса. Поэтому можно рассуждать так, как если бы силы тяжести совсем не было. Так мы и поступим.

2. Результирующая сила $F = -kx$ имеет такой же вид, что и сила в выражении (39.6). Если положить $m\omega^2 = k$, то уравнение (40.1) перейдет в

$$\ddot{x} + \omega^2 x = 0. \tag{40.2}$$

Это уравнение совпадает с уравнением (39.5). Функция (39.1) является решением такого уравнения при любых значениях постоянных A и δ. Можно доказать, что это есть *общее решение,* т. е. всякое решение уравнения (40.2) может быть представлено в виде (39.1). Различные решения отличаются друг от друга только значениями постоянных A и δ. (Доказательство приводится в конце параграфа.) Из изложенного следует, что груз на пружине будет совершать гармонические колебания с круговой частотой

$$\omega = \sqrt{\frac{k}{m}} \tag{40.3}$$

и периодом

$$T = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{m}{k}}. \tag{40.4}$$

Период колебаний T не зависит от амплитуды A. Это свойство называется *изохронностью* колебаний. Изохронность, однако, имеет место до тех пор, пока справедлив закон Гука. При больших растяжениях закон Гука нарушается. Тогда и колебания перестают быть

изохронными, т. е. появляется зависимость периода колебаний от амплитуды.

Амплитуда A и начальная фаза δ не могут быть определены из дифференциального уравнения (40.2). Эти постоянные определяются начальными условиями, например начальными значениями смещения x и скорости $\dot{x}$. Дифференциальное уравнение (40.2) справедливо при любых начальных условиях. Оно описывает весь комплекс колебаний, которые может совершать рассматриваемая система. Конкретное колебание выделяется из этого комплекса заданием постоянных A и δ.

3. Потенциальная и кинетическая энергии тела даются выражениями

$$E_{\text{пот}} = \tfrac{1}{2} kx^2, \quad E_{\text{кин}} = \tfrac{1}{2} mv^2 = \tfrac{1}{2} m\dot{x}^2. \tag{40.5}$$

Каждая из них меняется во времени. Однако их сумма E во времени должна оставаться постоянной:

$$E = \tfrac{1}{2} kx^2 + \tfrac{1}{2} m\dot{x}^2 = \text{const}. \tag{40.6}$$

Если воспользоваться выражением (39.1), то из формул (40.5) найдем

$$E_{\text{пот}} = \tfrac{1}{2} kA^2 \cos^2(\omega t + \delta), \quad E_{\text{кин}} = \tfrac{1}{2} m\omega^2 A^2 \sin^2(\omega t + \delta),$$

или в силу соотношения (40.3)

$$E_{\text{кин}} = \tfrac{1}{2} kA^2 \sin^2(\omega t + \delta).$$

Эти формулы можно также записать в виде

$$E_{\text{пот}} = \tfrac{1}{4} kA^2[1 + \cos 2(\omega t + \delta)],$$

$$E_{\text{кин}} = \tfrac{1}{4} kA^2[1 - \cos 2(\omega t + \delta)].$$

Они показывают, что *кинетическая и потенциальные энергии в отдельности не остаются постоянными, а совершают гармонические колебания вокруг общего среднего значения* $\frac{1}{4} kA^2$ *с удвоенной круговой частотой* 2ω. Когда кинетическая энергия проходит через максимум, потенциальная обращается в нуль и обратно. Однако полная энергия $E = E_{\text{кин}} + E_{\text{пот}}$ остается постоянной и связана с амплитудой A соотношением

$$E = \tfrac{1}{2} kA^2. \tag{40.7}$$

Приведенное простое вычисление вместе с тем показывает, что выражение (39.1) является решение дифференциального уравнения (40.6) при условии, что частота ω определяется формулой (40.3), а амплитуда A — формулой (40.7). Таким образом, при заданной

полной энергии E постоянная A не произвольна. Имеется лишь одна произвольная постоянная, определяемая начальными условиями, а именно начальная фаза δ. Для ее определения достаточно знать, например, либо начальное смещение, либо начальную скорость. Наличие в решении только одной произвольной постоянной связано с тем, что уравнение (40.6) — *первого порядка* по времени, в отличие от (40.2), которое является уравнением *второго порядка.* Впрочем, на энергию в уравнении (40.6) можно смотреть как на параметр, который может принимать любые положительные значения, определяющиеся начальными условиями. Тогда уравнение (40.6) становится полностью эквивалентным уравнению (40.2).

4. Все изложенное здесь применимо к гармоническим колебаниям любых механических систем с *одной степенью свободы.* Мгновенное положение механической системы с одной степенью свободы может быть определено с помощью какой-либо *одной* величины q, называемой *обобщенной координатой,* например угла поворота, смещения вдоль некоторой линии и пр. Производная $\dot{q}$ обобщенной координаты по времени называется *обобщенной скоростью.* При рассмотрении колебаний механических систем с одной степенью свободы за исходное удобнее брать не уравнение движения Ньютона, а *уравнение энергии.* Его обычно легче составлять. Кроме того, оно в известном смысле проще уравнения Ньютона, так как является дифференциальным уравнением *первого,* а не второго порядка по времени. Допустим, что механическая система такова, что ее потенциальная и кинетическая энергии выражаются формулами вида

$$E_{\text{пот}} = \tfrac{1}{2}\,\alpha q^2, \quad E_{\text{кин}} = \tfrac{1}{2}\,\beta \dot{q}^2, \tag{40.8}$$

где α и β — положительные постоянные (параметры системы). Тогда закон сохранения энергии приводит к уравнению

$$E = \tfrac{1}{2}\,\alpha q^2 + \tfrac{1}{2}\,\beta \dot{q}^2 = \text{const}. \tag{40.9}$$

Оно отличается от уравнения (40.6) только обозначениями, что при математическом рассмотрении не имеет значения. Из математической тождественности уравнений (40.6) и (40.9) следует, что и общие решения их одинаковы. Поэтому *если уравнение энергии приводится к виду* (40.9), *то*

$$q = q_0 \cos(\omega t + \delta), \tag{40.10}$$

т. е. обобщенная координата q совершает гармоническое колебание с круговой частотой

$$\omega = \sqrt{\alpha/\beta}. \tag{40.11}$$

5. В заключение покажем, как можно найти общее решение дифференциального уравнения (40.2). Из этого уравнения прежде всего вытекает

уравнение энергии (40.6). Поэтому можно сразу исходить из уравнения (40.6). Используя соотношение (40.3), запишем это уравнение в виде

$$\omega^2 x^2 + \dot{x}^2 = \text{const}. \tag{40.12}$$

Левая часть этого соотношения существенно положительна, так как она равна сумме квадратов. Поэтому правую часть можно обозначить через $\omega^2 A^2$, введя тем самым новую постоянную A. Тогда

$$\dot{x}^2 = \omega^2 (A^2 - x^2). \tag{40.13}$$

Так как $\dot{x}^2 \geqslant 0$, то $x \leqslant A$. Поэтому можно положить

$$x = A \cos \Theta, \tag{40.14}$$

где Θ — неизвестная функция времени t. Подставляя это выражение в уравнение (40.13), получим

$$\dot{x}^2 = \omega^2 A^2 (1 - \cos^2 \Theta) = \omega^2 A^2 \sin^2 \Theta,$$

откуда

$$\dot{x} = \pm \omega A \sin \Theta.$$

С другой стороны, дифференцируя выражение (40.14) по времени, находим

$$\dot{x} = -\dot{\Theta} A \sin \Theta.$$

Сравнение полученных выражений для x дает $\dot{\Theta} = \pm \omega$, откуда

$$\Theta = \pm \omega t + \delta,$$

где δ — произвольная постоянная. Таким образом,

$$x = A \cos (\pm \omega t + \delta).$$

Полученные выражения для x: $x_1 = A \cos (\omega t + \delta)$ и $x_2 = A \cos (-\omega t + \delta) = A \cos (\omega t - \delta)$ можно объединить в одно, так как δ — произвольная постоянная. Ее можно во втором выражении переобозначить, заменив на $-\delta$. Итак, в общем случае

$$x = A \cos (\omega t + \delta),$$

что совпадает с выражением (39.1).

§ 41. ФИЗИЧЕСКИЙ МАЯТНИК

1. *Физическим маятником* называется твердое тело, которое может качаться вокруг неподвижной горизонтальной оси. Точка A пересечения ее с вертикальной плоскостью, проходящей через центр масс маятника, называется *точкой подвеса маятника* (рис. 85). Положение тела в каждый момент времени можно характеризовать углом отклонения его из положения равновесия φ. Угол φ играет роль обобщенной координаты q. Кинетическая энергия качающегося физического маятника определяется выражением

$$E_{\text{кин}} = \frac{1}{2} I \dot{\varphi}^2,$$

где I — момент инерции маятника относительно оси A. Потенциальная энергия равна $E_{\text{пот}} = mgh$, где h — высота поднятия центра масс C над его самым нижним положением. Обозначим через a расстояние между центром масс C и точкой подвеса A. Тогда

$$E_{\text{пот}} = mga(1 - \cos\varphi) = 2mga\sin^2(\varphi/2).$$

В случае малых колебаний синус угла $\varphi/2$ можно приближенно заменить самим углом. В этом приближении

$$E_{\text{пот}} = \frac{1}{2}mga\varphi^2.$$

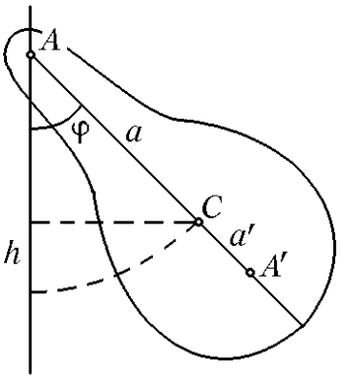


Рис. 85

Таким образом, для малых колебаний потенциальная и кинетическая энергии приводятся к виду (40.8), причем $\alpha = mga$, $\beta = I$. Отсюда следует, что малые колебания физического маятника будут приблизительно гармоническими с циклической частотой

$$\omega = \sqrt{\frac{mga}{I}} \tag{41.1}$$

и периодом

$$T = 2\pi\sqrt{\frac{I}{mga}}. \tag{41.2}$$

Если период колебаний не зависит от амплитуды, то такие колебания называются *изохронными*. Мы видим, что *малые колебания физического маятника изохронны*. Колебания приближенно изохронны, когда угловая амплитуда колебаний не превышает нескольких градусов. При бóльших амплитудах изохронность нарушается. На свойстве изохронности колебаний маятника основано его применение в часах.

Частным случаем физического маятника является *математический маятник*. Так называется маятник, вся масса которого практически сосредоточена в одной точке — в центре масс маятника C. Примером математического маятника может служить шарик, подвешенный на длинной нити. В случае математического маятника $a = l$, $I = ml^2$, где l — длина маятника, и формула (41.2) переходит в

$$T = 2\pi\sqrt{\frac{l}{g}}. \tag{41.3}$$

Сравнивая формулы (41.2) и (41.3), делаем вывод, что *физический маятник колеблется так же, как математический маятник с длиной*

$$l = \frac{I}{ma}, \tag{41.4}$$

которая называется *приведенной длиной* физического маятника. Мы доказали это утверждение только для малых колебаний маятников.

Но оно справедливо и для колебаний с конечными амплитудами, когда колебания не изохронны. Требуется только, чтобы угловые амплитуды физического и математического маятников были одинаковы. Доказательство этого мы предоставляем читателю.

2. Отложим от точки подвеса A вдоль прямой AC отрезок AA', длина которого равна приведенной длине физического маятника l (см. рис. 85). Точка A' называется *центром качания. Центр качания можно определить как математическую точку, в которой надо сосредоточить всю массу физического маятника, чтобы период его колебаний остался без изменений.* По теореме Гюйгенса—Штейнера $I = I_C + ma^2$, где I_C — момент инерции маятника относительно параллельной оси, проходящей через центр масс C. Подставив это выражение в формулу (41.4), придадим ей вид

$$l = a + \frac{I_C}{ma}. \tag{41.5}$$

Отсюда следует, во-первых, что $l > a$, т. е. точка подвеса A и центр качания A' лежат по разные стороны от центра масс C, и, во-вторых, что всем точкам подвеса, одинаково удаленным от центра масс маятника, соответствует одна и та же длина l, а следовательно, один и тот же период колебаний T.

Точка подвеса и центр качания являются *взаимными* или *сопряженными* точками в следующем смысле. *Если маятник подвесить за центр качания A', то его период не изменится и прежняя точка подвеса A сделается новым центром качания.* Это положение называется *теоремой* Гюйгенса. Для ее доказательства обозначим через a' длину отрезка $A'C$ и допустим, что маятник подвешен за точку A'. Тогда его приведенная длина будет

$$l' = a' + \frac{I_C}{ma}.$$

Но $a' = l - a$, или в силу соотношения (41.5) $a' = I_C/(ma)$. Подставив это значение в предыдущую формулу, получим $l' = I_C/(ma) + a$. Таким образом, $l' = l$, т. е. приведенная длина, а с ней и период колебаний физического маятника остались без изменения. Это и доказывает теорему Гюйгенса.

3. Приведем другое доказательство теоремы Гюйгенса, глубже раскрывающее ее содержание. Будем перемещать точку подвеса маятника вдоль одной и той же прямой, проходящей через центр масс C. Посмотрим, как при этом будет меняться его период колебаний. Когда точка подвеса A бесконечно удалена от C, маятник ведет себя как математический. Его период колебаний бесконечно велик. При приближении точки подвеса A к центру масс C период колебаний сначала убывает. Когда точка подвеса совместится с C, маятник при любом отклонении будет в безразличном равновесии. Это значит, что его период колебаний снова становится бесконечно большим. Поэтому по мере приближения точки A к C убывание периода должно смениться возрастанием. Положению точки подвеса, где это происходит, соот-

ветствует минимальный период колебаний. Тогда точка подвеса переходит через точку C на другую сторону прямой AA', период колебаний, перейдя через бесконечность, начинает уменьшаться. При этом двум положениям точки подвеса, находящимся по разные стороны от C на одинаковых расстояниях, соответствуют равные периоды колебаний.

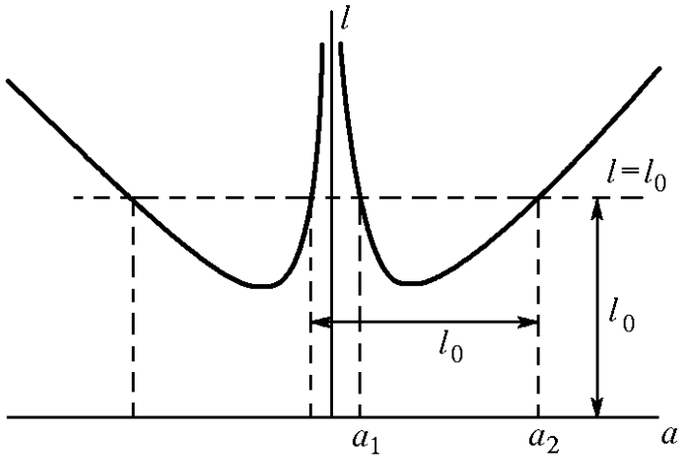


Рис. 86

Вместо периода колебаний можно пользоваться приведенной длиной маятника l, однозначно определяющей его период колебаний. При удалении точки подвеса в бесконечность или при приближении ее к центру масс C приведенная длина l стремится к бесконечности и достигает минимума в каком-то промежуточном положении. Графически это представлено кривой на рис. 86. На оси абсцисс отложена величина a, на оси ординат — приведенная длина l маятника. Кривая состоит из двух ветвей, симметрично расположенных относительно оси ординат. Одна ветвь соответствует случаю, когда точка подвеса расположена по одну, а вторая — по другую сторону от центра масс C. Аналитически кривая изображается уравнением (41.5), которое можно переписать в виде

$$a^2 - la + \frac{I_C}{m} = 0. \tag{41.6}$$

Фиксированному значению приведенной длины l_0 соответствует на рис. 86 горизонтальная прямая $l = l_0$. Точки пересечения ее с кривой определяют положение точек подвеса физического маятника, при которых его приведенная длина равна заданному значению l_0. Вообще говоря, таких точек пересечения четыре. Две из них расположены по одну, две остальные — по другую сторону от центра масс C. Их положение легко найти из квадратного уравнения

$$a^2 - l_0 a + \frac{I_C}{m} = 0. \tag{41.7}$$

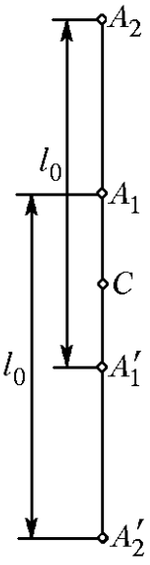


Рис. 87

Если $l_0 > 2\sqrt{I_C/m}$, это уравнение имеет два вещественных положительных корня a_1 и a_2, причем

$$a_1 + a_2 = l_0. \tag{41.8}$$

В этом случае по одну и ту же сторону от центра масс C имеются две точки подвеса A_1 и A_2 (рис. 87), которым соответствует одна и та же приведенная длина l_0. По другую сторону от центра масс C лежит вторая пара симметрично расположенных точек подвеса A_1' и A_2', характеризующаяся той же приведенной длиной l_0. Если $l_0 = 2\sqrt{I_C/m}$, корни уравнения (41.7) совпадают, т. е. обе точки подвеса по каждую сторону от центра масс сливаются в одну. Если $l_0 < 2\sqrt{I_C/m}$, корни

уравнения (41.7) мнимые. Не существует точек подвеса, для которых приведенная длина была бы меньше $2\sqrt{I_C/m}$.

Теорема Гюйгенса теперь становится очевидной. Действительно, из соотношения (41.8) следует, что расстояние между точками A_1 и A_2', а также между точками A_1' и A_2 равно приведенной длине маятника l. Если одну из точек каждой пары принять за точку подвеса, то вторая будет центром качания. Но это и есть теорема Гюйгенса. Наше рассмотрение показывает также, *что точка подвеса и центр качания находятся по разные стороны от центра масс и расположены асимметрично относительно него. Исключение составляет только случай, когда* $l_0 = 2\sqrt{I_C/m}$. *Тогда точки* A_1 *и* A_2 *сливаются в одну точку. Сливаются также и точки* A_1' *и* A_2'. *В этом исключительном случае точка подвеса и центр качания расположены симметрично относительно центра масс.*

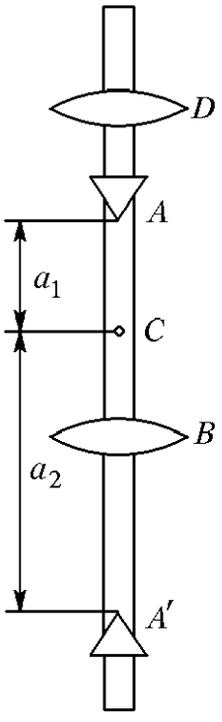


Рис. 88

4. Теорема Гюйгенса используется в *оборотном маятнике* для точных измерений ускорения свободного падения. Существуют разнообразные конструкции оборотного маятника. На рис. 88 схематически изображена одна из них. Маятник состоит из стального стержня, длина которого обычно несколько больше метра. На нем жестко закреплены опорные стальные призмы A и A' и стальная чечевица B, находящаяся между ними. Другая стальная чечевица D находится на одном из концов стержня (не между призмами), она может перемещаться по стержню и закрепляться в нужном положении. Перемещением этой чечевицы достигают совпадения периодов колебаний маятника, когда точками подвеса являются ребра опорных призм A и A'. Эти ребра закреплены *асимметрично* относительно центра масс C. Поэтому при совпадении периодов колебаний расстояние между ними дает приведенную длину физического маятника l. Измерив период колебаний T, можно вычислить g по формуле (41.3).

§ 42. БИФИЛЯРНЫЙ И ТРИФИЛЯРНЫЙ ПОДВЕСЫ

1. Найдем период малых колебаний *бифилярного подвеса*. Так называется устройство, состоящее из двух нитей AB и CD (рис. 89) одинаковой длины, на которых подвешено некоторое тело BD. Если тело повернуть вокруг вертикальной оси OO', то оно начнет совершать крутильные колебания вокруг этой оси. Бифилярный подвес есть система с одной степенью свободы. В качестве координаты, определяющей ее мгновенное положение, удобно взять угол поворота φ тела BD вокруг оси OO', отсчитывая этот угол от положения равновесия.

Кинетическая энергия системы равна $E_{\text{кин}} = \frac{1}{2} I\dot{\varphi}^2$, где I — момент инерции ее относительно оси OO'. Потенциальная энергия равна $E_{\text{пот}} = mgh$, где h — высота поднятия тела BD, отсчитываемая от его нижнего положения. Пусть l означает длину OO' в положении равновесия, $2a$ — рас-

стояние между точками подвеса C и A, $2b$ — расстояние DB. Предполагается, что система симметрична, так что точки O и O' являются серединами отрезков CA и DB. Высота h найдется из условия нерастяжимости нитей AB и CD. Введем прямоугольную систему координат с началом в точке O, ось X направим вдоль прямой OA, ось Z — вниз вдоль прямой OO', ось Y — перпендикулярно к ним. Координаты точки A все время остаются постоянными и равны

$$x_A = a, \quad y_A = 0, \quad z_A = 0.$$

Координаты точки B в положении равновесия равны

$$x_B^{(0)} = b, \quad y_B^{(0)} = 0, \quad z_B^{(0)} = l.$$

При повороте системы на угол φ координаты той же точки становятся равными

$$x_B = b\cos\varphi, \quad y_B = b\sin\varphi, \quad z_B = l - h.$$

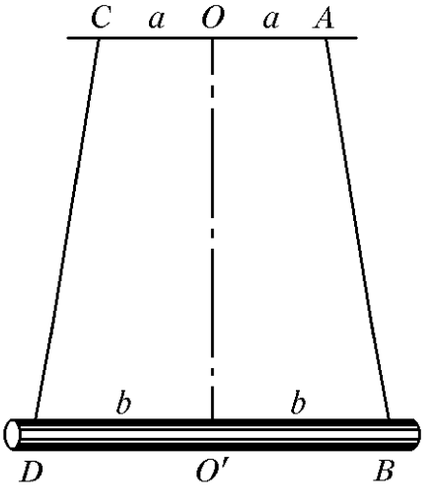


Рис. 89

Условие постоянства длины нити AB можно записать в виде

$$(x_B - x_A)^2 + (y_B - y_A)^2 + (z_B - z_A)^2 = (x_B^0 - x_A^0)^2 + (y_B^0 - y_A^0)^2 + (z_B^0 - z_A^0)^2,$$

или

$$(b\cos\varphi - a)^2 + b^2\sin^2\varphi + (l - h)^2 = (b - a)^2 + l^2.$$

После простых преобразований отсюда находим

$$h = \frac{2ab(1-\cos\varphi)}{2l+h} = \frac{4ab}{2l+h}\sin^2\frac{\varphi}{2}.$$

При малых колебаниях можно положить $\sin(\varphi/2) = \varphi/2$. Кроме того, $h \ll 2l$, и величиной h в знаменателе можно пренебречь. В этом приближении

$$h = \frac{ab}{2l}\varphi^2, \quad E_{\text{пот}} = \frac{mgab}{2l}\varphi^2.$$

Таким образом, потенциальная и кинетическая энергии приводятся к виду (40.9), причем $\alpha = mgab/l$, $\beta = I$. Следовательно, колебания системы будут гармоническими с периодом

$$T = 2\pi\sqrt{\frac{Il}{mgab}}. \tag{42.1}$$

Период колебаний пропорционален корню квадратному из момента инерции и обратно пропорционален корню квадратному из массы системы. Возьмем в качестве тела BD металлический стержень. Выведем его из положения равновесия и заставим совершать крутильные колебания. Они будут сравнительно медленными. Прикрепим затем в точке O' тяжелый груз и снова заставим систему колебаться. Колебания станут значительно более быстрыми. Дело в том, что груз прикреплен на оси вращения, а потому он, значительно увеличивая массу системы, практически не влияет на ее момент инерции. Уменьшение периода колебаний можно объяснить также следующим образом. В положение равновесия система возвращается под действием горизонтальных составляющих сил натяжения нитей. Подвешивая груз, мы сильно увели-

чиваем натяжение нитей, а момент инерции увеличивается незначительно. Это и приводит к тому, что колебания становятся более быстрыми.

2. Формулой (42.1) определяется также период колебаний *трифилярного подвеса* (трифиляра). Он схематически изображен на рис. 90. Точки подвеса A, C и M расположены на окружности радиусом a, точки B, D, N — на окружности радиусом b. Нижний диск может совершать крутильные колебания вокруг вертикальной оси OO'. Вывод формулы (42.1) применим без всяких изменений и к трифилярному подвесу. Это видно уже из того, что при выводе было использовано условие постоянства длины только одной нити AB. Постоянство длины другой нити CD при этом условии выполняется автоматически.

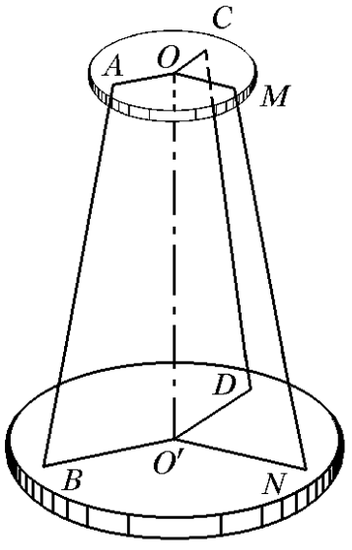


Рис. 90

Трифилярный подвес дает удобный метод измерения моментов инерции тел. Сначала измеряется период колебаний T_0 ненагруженного трифиляра. По этому периоду вычисляется его момент инерции

$$I_0 = \frac{m_0 gab}{4\pi^2 l} T_0^2.$$

Затем на нижний диск трифиляра кладется тело массой m, момент инерции I которого требуется измерить. Пусть T — период крутильных колебаний нагруженного трифиляра. Тогда момент инерции системы относительно оси OO' будет

$$I + I_0 = \frac{(m+m_0)gab}{4\pi^2 l} T^2.$$

Вычитая отсюда предыдущее выражение, находим искомый момент инерции I.

3. Укажем другой метод измерения моментов инерции, который во многих случаях является более предпочтительным. Подвесим тело на стальной проволоке, чтобы оно могло совершать крутильные колебания вокруг вертикальной оси, совпадающей с осью проволоки (рис. 91). При повороте тела на угол φ проволока закручивается, и возникает момент сил M, стремящийся вернуть тело в положение равновесия. Опыт показывает, что момент M в довольно широких пределах пропорционален углу φ: $M = -f\varphi$, где f — постоянная для данной проволоки величина, называемая ее *модулем кручения*. Поэтому

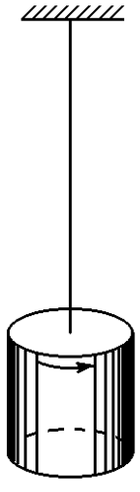


Рис. 91

$$I\ddot{\varphi} = -f\varphi.$$

Это уравнение математически тождественно уравнению (40.1). Значит, тело будет совершать гармонические крутильные колебания с периодом

$$T = 2\pi\sqrt{\frac{I}{f}}. \qquad (42.2)$$

Сняв первое тело, подвесим на той же проволоке другое тело с моментом инерции I'. Тогда период колебаний будет

$$T' = 2\pi\sqrt{\frac{I'}{f}}.$$

Исключая неизвестный модуль кручения f, найдем

$$\frac{I}{I'} = \left(\frac{T}{T'}\right)^2.$$

Если один из моментов инерции, например I, известен, то по этой формуле может быть вычислен момент инерции I' другого тела. Момент инерции I можно вычислить теоретически по геометрическим размерам и массе тела. Для этого надо взять тело правильной геометрической формы, например цилиндр или шар. Формула (42.2) может быть использована также для экспериментального определения модуля кручения проволоки.

ЗАДАЧИ

1. Материальная точка движется в поле тяжести по хорде круга без начальной скорости (рис. 92). Показать, что время ее движения из точки A в нижнее положение B не зависит от положения точки A на окружности. (Этот факт был использован Галилеем для установления законов малых колебаний математического маятника. Для нахождения периода колебаний маятника Галилей заменил малую дугу окружности ADB, по которой движется материальных точка, хордой AB). Вычислить период колебаний маятника в этом приближении и убедиться, что это приближение приводит к правильной зависимости периода колебаний от длины маятника l и ускорения свободного падения g. Сравнить результат с правильной формулой (41.3).

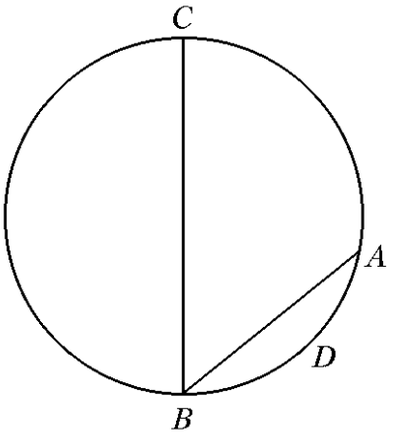


Рис. 92

Ответ. $T = 8\sqrt{\frac{l}{g}}$.

2. Через неподвижный блок с моментом инерции I (рис. 93) и радиусом r перекинута нить, к одному концу которой подвешен груз массой m. Другой конец нити привязан к пружине с закрепленным нижним концом. Вычислить период колебаний груза, если жесткость пружины равна k, а нить не может скользить по поверхности блока.

Ответ. $T = 2\pi\sqrt{\frac{I/r^2+m}{k}}$.

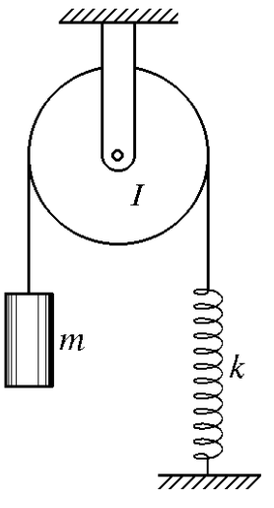


Рис. 93

3. Физический маятник представляет собой однородный стержень длиной l, подвешенный за один из его концов. Определить период колебаний такого маятника.

Ответ. $T = 2\pi\sqrt{\frac{2l}{3g}}$.

4. Тело вращения радиусом a с моментом инерции I (относительно геометрической оси) и массой m катается без скольжения по внутренней поверхности цилиндра радиуса R, совершая малые колебания около положения равновесия (рис. 94). Найти период этих колебаний.

Решение. Рассматривая движение тела как вращение вокруг мгновенной оси (см. § 45) с угловой скоростью ω, напишем для скорости его центра $v = \omega a$. Ту же скорость можно представить в виде $v = (R-a)\dot{\varphi}$.

Приравнивая оба выражения, находим

$$\omega = \frac{R-a}{a}\,\dot{\varphi}.$$

Кинетическая энергия по теореме Кёнига

$$K = \frac{1}{2} I\omega^2 + \frac{1}{2} m(R-a)^2\dot{\varphi}^2 = \frac{1}{2}\left(m + \frac{I}{a^2}\right)(R-a)^2\dot{\varphi}^2.$$

Потенциальная же энергия

$$U = mg(R-a)\ (1-\cos\varphi) \approx \frac{1}{2} mg(R-a)\varphi^2.$$

Применяя общий метод, изложенный в § 40, находим

$$T = 2\pi\sqrt{\left(1 + \frac{I}{ma^2}\right)\frac{R-a}{g}}.$$

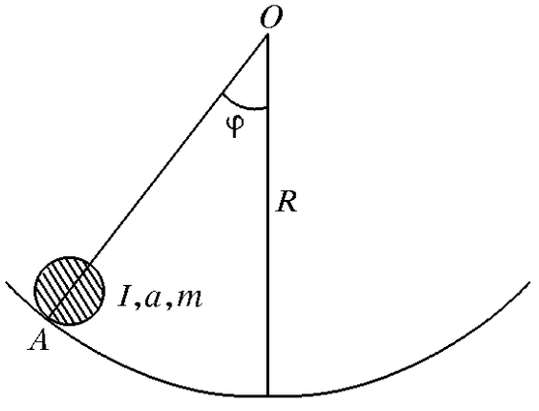


Рис. 94

В частности, для сплошного цилиндра и сплошного шара

$$T = 2\pi\sqrt{\frac{3}{2}\frac{R-a}{g}}, \quad T = 2\pi\sqrt{\frac{7}{5}\frac{R-a}{g}}.$$

5. На горизонтальной плоскости лежит цилиндр с моментом инерции I (относительно продольной геометрической оси), массой m и радиусом r. К оси цилиндра прикреплены две одинаковые горизонтально расположенные спиральные пружины, другие концы которых закреплены в стене (рис. 95, вид сверху). Жесткость каждой пружины равна k, пружины могут работать как на растяжение, так и на сжатие. Найти период малых колебаний цилиндра, которые возникнут, если вывести его из положения равновесия и дать возможность кататься без скольжений по горизонтальной плоскости.

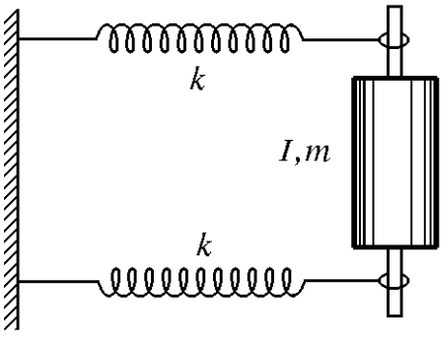


Рис. 95

Ответ. $T = \dfrac{2\pi}{r}\sqrt{\dfrac{I+mr^2}{2k}}$. Для сплошного цилиндра $T = \pi\sqrt{3m/k}$.

6. Однородная квадратная плита подвешена за свои углы к потолку зала на четырех параллельных веревках, длина каждой из которых равна l. Определить период малых крутильных колебаний плиты, которые возникнут, если повернуть ее на малый угол вокруг вертикальной оси.

Ответ. $T = 2\pi\sqrt{\dfrac{l}{3g}}$.

В более общем случае, когда плита не однородна, но центр масс ее совпадает с геометрическим центром плиты,

$$T = 2\pi\sqrt{\frac{2Il}{Mga^2}},$$

где I — момент инерции плиты относительно вертикальной оси, проходящей через ее центр, а a — длина одной из сторон плиты.

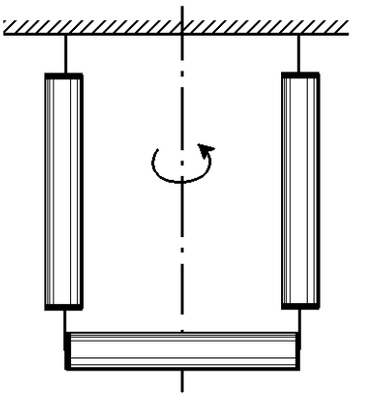


Рис. 96

7. Три однородных стержня длиной l каждый соединены короткими нитями, как указано на рис. 96. Нижний стержень поворачивают на малый угол вокруг вертикальной оси, проходящей через центр системы, и отпускают. Найти период возникших при этом малых колебаний, если массы стержней одинаковы.

Ответ. $T = 2\pi\sqrt{\frac{l}{2g}}$.

8. Шарик массой m подвешен на двух последовательно соединенных пружинках с жесткостями k_1 и k_2 (рис. 97). Определить период его вертикальных колебаний.

Ответ. $T = 2\pi\sqrt{m\left(\frac{1}{k_1}+\frac{1}{k_2}\right)}$.

Указание. Показать, что при растяжениях и сжатиях пружины ведут себя как одна пружина с жесткостью, определяемой соотношением

$$\frac{1}{k} = \frac{1}{k_1} + \frac{1}{k_2}.$$

9. Найти период крутильных колебаний диска, плотно насаженного на составной стержень, состоящий из двух различных последовательно соединенных стержней (рис. 98). Верхний конец A стержня неподвижно закреплен.

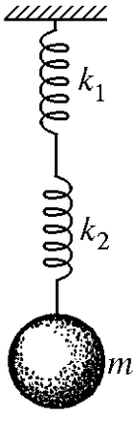


Рис. 97

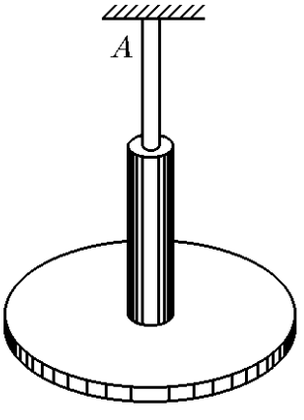


Рис. 98

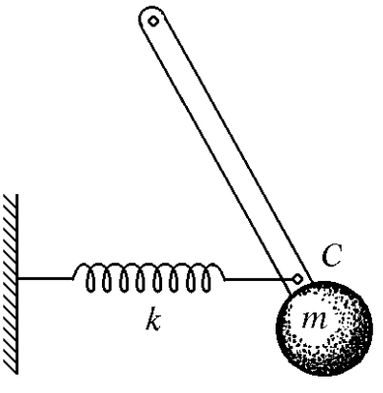


Рис. 99

Если бы диск был насажен только на первый стержень, то период колебаний был бы равен T_1. Если бы он был насажен только на второй стержень, то период колебаний оказался бы равным T_2.

Ответ. $T = \sqrt{T_1^2 + T_2^2}$.

10. Найти период малых колебаний физического маятника массой m, к центру масс C которого прикреплена горизонтальная спиральная пружина с жесткостью k. Другой конец пружины закреплен в неподвижной стенке (рис. 99). Момент инерции маятника относительно точки подвеса равен I, расстояние между точкой подвеса и центром масс маятника равно a. В положении равновесия пружина не деформирована.

Ответ. $T = 2\pi\sqrt{\dfrac{I}{mga + ka^2}}$.

11. Колебательная система состоит из однородного стержня длиной l и массой m, который может вращаться вокруг горизонтальной оси O, проходящей через его конец и перпендикулярной к продольной оси стержня (рис. 100). Другой конец стержня подвешен на пружине с жесткостью k. Расстояние между центром масс стержня и осью вращения $CO = a$. Момент инерции стержня относительно оси O равен I. Найти удлинение пружины x_0 (по сравнению с ее длиной в недеформированном состоянии) в положении равновесия, если в этом положении стержень горизонтален. Определить также период малых колебаний стержня около положения равновесия.

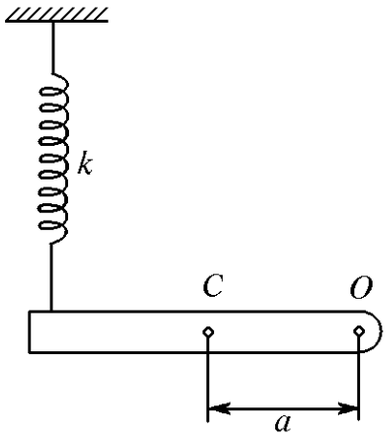


Рис. 100

Ответ. $x_0 = \dfrac{mga}{kl}$, $T = 2\pi\sqrt{\dfrac{I}{kl^2}}$.

12. К концу однородного стержня длиной l и массой m прикреплена короткая упругая пластинка. Пластинку зажимают в тисках один раз так, что стержень оказывается внизу, а другой раз — вверху (рис. 101). Определить отношение периодов малых колебаний стержня в этих случаях. Момент упругих сил пластинки пропорционален углу отклонения стержня от положения равновесия, причем коэффициент пропорциональности равен k.

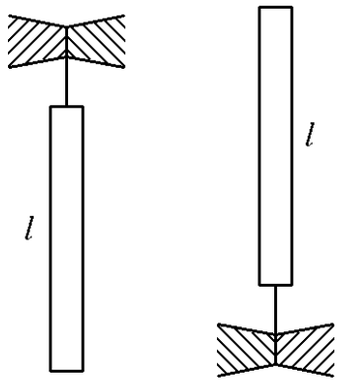


Рис. 101

Ответ. $\dfrac{T_1}{T_2} = \sqrt{\dfrac{2k - mgl}{2k + mgl}}$.

13. Два незакрепленных шарика с массами m_1 и m_2 соединены друг с другом спиральной пружинкой с жесткостью k. Определить период колебаний шариков относительно центра масс системы, которые возникнут при растяжении пружинки.

Ответ. $T = 2\pi\sqrt{\dfrac{m_1 m_2}{(m_1 + m_2)k}}$.

14. Два диска с моментами инерции I_1 и I_2 насажена на общую ось, проходящую через их центры. Осью являются стержень с модулем кручения f. Определить период крутильных колебаний одного диска относительно другого в предположении, что система свободна. Массой стержня пренебречь.

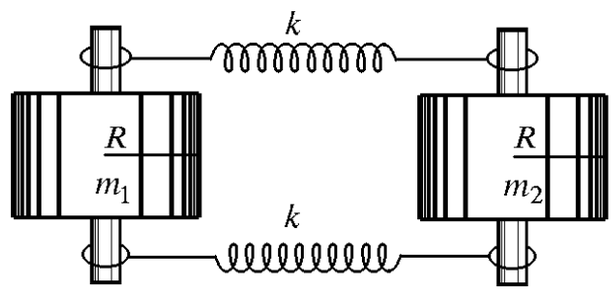


Рис. 102

Ответ. $T = 2\pi\sqrt{\dfrac{I_1 I_2}{f(I_1 + I_2)}}$.

15. Два сплошных однородных цилиндра с одинаковым радиусом R и массами m_1 и m_2 лежат на горизонтальном столе и связаны с помощью двух одинаковых пружин с жесткостью k каждая, как показано на рис. 102 (вид сверху). Определить период малых колебаний, которые возникнут, если растянуть пру-

жины и предоставить систему самой себе, не сообщая ей дополнительной скорости. Цилиндры катаются по столу без проскальзывания. Пружины могут работать как на растяжение, так и на сжатие.

Ответ. $T = \pi \sqrt{\frac{3m_1 m_2}{k(m_1+m_2)}}$.

16. Колебания обычного математического маятника изохронны (точнее, приблизительно изохронны) только тогда, когда их амплитуды малы. Гюйгенс задался целью построить маятник, который совершал бы строго изохронные колебания при любых амплитудах. Он показал, что таковым является *циклоидальный маятник*. Циклоидальный математический маятник представляет собой материальную точку, совершающую колебания, двигаясь под действием силы тяжести по дуге циклоиды. Показать, что колебания циклоидального маятника изохронны, и вывести формулу для его периода.

Решение. Как известно, циклоида представляет собой кривую, описываемую одной из точек окружности, катящейся по неподвижной прямой. Для наших целей надо взять циклоиду, обращенную выпуклостью вниз. В соответствии с этим примем, что окружность расположена ниже горизонтальной прямой, по которой она катится (эта прямая на рис. 103 изображена штриховой линией). За ось X примем параллельную ей прямую, смещенную вниз на диаметр окружности $2a$. Пусть точка A на катящейся окружности, описывающая циклоиду, в исходном положении находится на оси Y в наивысшей точке. Если окружность при качении повернется на угол φ, то ее центр C переместится вправо на расстояние $a\varphi$. При этом точка A сместится относительно центра влево на расстояние $a\sin\varphi$ и вниз на расстояние $a(1-\cos\varphi)$. Поэтому прямоугольные координаты точки A станут

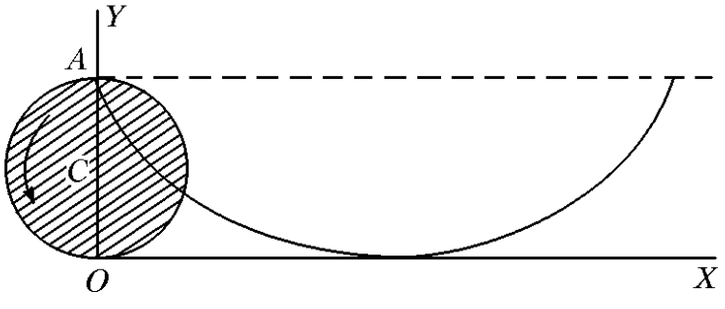


Рис. 103

$$x = a(\varphi - \sin\varphi), \quad y = a(1+\cos\varphi).$$

Это — уравнение циклоиды в параметрической форме. Пусть теперь x и y означают координаты материальной точки, совершающей циклоидальные колебания под действием силы тяжести. Параметр φ становится функцией времени. Потенциальная энергия точки будет $U = mgy$, кинетическая — $K = {}^1/_2 m(\dot{x}^2 + \dot{y}^2)$. Найдя производные $\dot{x}$, $\dot{y}$ и выполнив элементарные преобразования, получим

$$U = 2mga\cos^2\frac{\varphi}{2}, \quad K = 2ma^2\sin^2\frac{\varphi}{2}\cdot\dot{\varphi}^2.$$

Введем обозначение $q = \cos\frac{\varphi}{2}$. Тогда $\dot{q} = -\frac{1}{2}\sin\frac{\varphi}{2}\cdot\dot{\varphi}$. Величина q может быть принята за координату, определяющую положение колеблющейся точки, а ее производная $\dot{q}$ — за соответствующую обобщенную скорость. В этих обозначениях

$$U = 2mgaq^2, \quad K = 8ma^2\dot{q}^2.$$

Потенциальная энергия является квадратичной функцией координаты q, а кинетическая — производной $\dot{q}$ с постоянными коэффициентами. Отсюда

делаем вывод, что при любых амплитудах колебания циклоидального маятника будут изохронными и гармоническими с периодом

$$T = 2\pi \sqrt{\frac{4a}{g}}.$$

17. Маятник подвешен на резинке, растянутой настолько сильно, что ее первоначальной длиной можно пренебречь. Возможны ли горизонтальные гармонические изохронные колебания маятника сколь угодно большой амплитуды? Если возможны, то определить период этих колебаний. Возможны ли круговые движения маятника в вертикальной плоскости? Каково будет движение при любых начальных условиях?

Ответ. И те и другие движения возможны. Их период

$$T = 2\pi \sqrt{\frac{m}{k}},$$

где m — масса маятника, k — жесткость резинки. При произвольных начальных условиях движение маятника будет происходить по эллипсу с периодом обращения T.

18. По штанге, вращающейся в горизонтальной плоскости с постоянной угловой скоростью ω, может скользить без трения груз массой m, удерживаемый на некотором расстоянии от оси вращения пружиной с жесткостью k и начальной длиной r_0. Найти движение груза, которое возникнет, если штангу мгновенно остановить.

Ответ. $r = r_0\left(1 + \dfrac{\omega^2}{\omega_0^2 - \omega^2}\cos \omega_0 t\right)$, где $\omega_0 = \sqrt{\dfrac{k}{m}}$.

При этом должно быть $\omega < \omega_0$. В противном случае равновесие груза на вращающейся штанге в области применимости закона Гука было бы невозможно.

19. На горизонтальной пружине укреплено тело массой $M = 10$ кг, лежащее на гладком столе, по которому оно может скользить без трения (рис. 104). В это тело попадает и застревает в нем пуля массой $m = 10$ г, летящая с горизонтальной скоростью $v = 500$ м/с, направленной вдоль оси пружины. Тело вместе с застрявшей в нем пулей отклоняется от положения равновесия и начинает колебаться относительно него с амплитудой $a = 10$ см. Найти период колебаний тела.

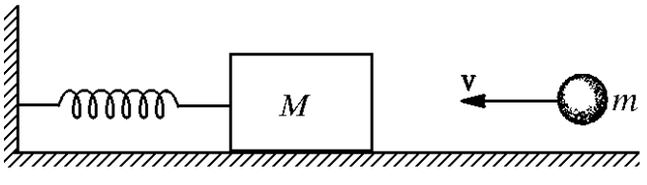


Рис. 104

Ответ. $T = 2\pi \dfrac{M+m}{mv} a \approx 1{,}26$ с.

20. На тонкую стальную спицу надет шарик. Противоположный конец спицы неподвижно закреплен. Показать, что если масса спицы пренебрежимо мала по сравнению с массой шарика, то период малых колебаний, возникающих при отклонении шарика в сторону, пропорционален расстоянию l между шариком и точкой закрепления спицы.

Указание. Рассмотрим вспомогательную однородную спицу, согнутую в кольцо. Если ее разрезать в одном месте и к концам прикрепить шарики A и B, то появятся упругие радикальные силы F, приложенные

к шарикам, стремящиеся распрямить спицу (рис. 105). Силы эти не зависят от места, где произведен разрез. Заметив это, вернемся теперь к нашей задаче. Если шарик сместить в сторону, то спица деформируется. При малых деформациях участок ее между шариком и точкой закрепления спицы можно в первом приближении считать дугой окружности. На основании предыдущего замечания можно утверждать, что при смещении шарика по этому деформированному участку действующая на него сила будет меняться. Пользуясь этим, нетрудно показать, что жесткость k спицы будет обратно пропорциональна квадрату длины l.

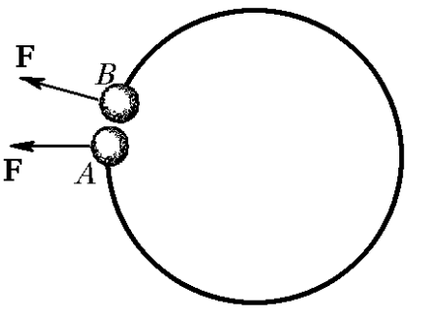


Рис. 105

21. Найти период колебаний физического маятника в зависимости от угловой амплитуды.

Р е ш е н и е. Закон сохранения энергии дает

$$\frac{I}{2}\dot{\varphi}^2 = mga(\cos\varphi - \cos\varphi_0),$$

где φ — угол отклонения маятника из положения равновесия, а φ_0 — максимальное его значение (угловая амплитуда колебаний). Введя приведенную длину маятника (41.4) и выполнив несложные преобразования, получим

$$\frac{d\varphi}{dt} = 2\sqrt{\frac{g}{l}}\sqrt{\sin^2\frac{\varphi_0}{2} - \sin^2\frac{\varphi}{2}}.$$

Разрешив это уравнение относительно dt и интегрируя по φ, найдем период колебаний маятника T как учетверенное время прохождения интервала углов от $\varphi = 0$ до $\varphi = \varphi_0$. При интегрировании удобно ввести новую переменную интегрирования $u = \sin(\varphi/2)/\sin(\varphi_0/2)$. В результате получим

$$T = 4\sqrt{\frac{l}{g}}\int_0^{\pi/2}\frac{du}{\sqrt{1 - k^2\sin^2 u}},$$

где введено обозначение $k \equiv \sin(\varphi_0/2)$. Входящий сюда интеграл не берется в элементарных функциях. Он называется полным *эллиптическим интегралом первого рода*. Его можно представить в виде бесконечного ряда. Так как $|k\sin u| < 1$, то подынтегральное выражение можно разложить в ряд по формуле бинома Ньютона:

$$(1 - k^2\sin^2 u)^{-1/2} = 1 + \frac{1}{2}k^2\sin^2 u + \frac{1\cdot 3}{2\cdot 4}k^4\sin^4 u + \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}k^6\sin^6 u + \ldots$$

Этот ряд равномерно сходится, а потому его можно интегрировать почленно. Сделав это, получим

$$T = 2\pi\sqrt{\frac{l}{g}}\left[1 + \frac{1}{4}\sin^2\frac{\varphi_0}{2} + \left(\frac{1\cdot 3}{2\cdot 4}\right)^2\sin^4\frac{\varphi_0}{2} + \left(\frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\right)^2\sin^6\frac{\varphi_0}{2} + \ldots\right].$$

При малых амплитудах φ_0 эта формула переходит в (41.3).

§ 43. АДИАБАТИЧЕСКИЕ ИНВАРИАНТЫ

1. Энергия, импульс или момент импульса механической системы являются функциями ее координат и скоростей. Если система замкнута, то эти величины сохраняются, т. е. не меняются с течением времени. Если же система не замкнута, а параметры, определяющие ее состояние, изменяются во времени, то указанные величины, вообще говоря, также изменяются. Возьмем, например, математический маятник, нить которого перекинута через гвоздь. Параметрами здесь являются длина нити l и ускорение свободного падения g. Можно тянуть за свободный конец нити, уменьшая или увеличивая l. При этом над маятником совершается внешняя работа, а потому энергия его изменяется. Можно также менять ускорение свободного падения, поднимая или опуская маятник над земной поверхностью. Среди различных изменений внешних параметров играют свободную роль бесконечно медленные изменения, называемые *адиабатическими* *).

При этом параметры, сколь бы медленно они не менялись, могут принимать любые значения, лежащие в допустимых пределах. Для изменения их на конечные величины требуется лишь достаточно длительное время. Изменения параметров системы, даже медленные, влекут за собой и изменения других физических величин. Так, энергия системы, как уже отмечалось, не остается постоянной, поскольку во время изменения параметров над системой производится работа. Но могут встречаться и такие величины, которые остаются постоянными или приблизительно постоянными из-за медленности изменения параметров.

Функции координат, скоростей и параметров системы, остающиеся постоянными при бесконечно медленных изменениях параметров, называются адиабатическими инвариантами. Это определение в дальнейшем будет уточнено, поскольку само понятие «медленности» нуждается в уточнении. Адиабатические инварианты играли большую роль в старой полуклассической теории атома Бора. Но они имеют важное значение и в других разделах физики.

2. Выясним понятие адиабатического инварианта сначала на простейшем, но важном примере гармонического осциллятора, собственная частота которого очень медленно изменяется во времени. Примером может служить математический маятник, медленно изменяющимися параметрами которого являются длина нити l и ускорение свободного падения g (точнее, их комбинация $\omega^2 = g/l$). Другим примером может служить колебание шарика на пружине, жесткость которой k является медленно меняющимся параметром. Все эти системы, называемые *гармоническими осцилляторами*, математически эквивалентны. Для конкретности будем иметь в виду шарик на пружине. Задача о математическом маятнике сводит-

*) В термодинамике термин «адиабатический» применяется в другом смысле. Адиабатическим называют процесс, происходящий без подвода и отвода теплоты.

ся к этому случаю, если ввести обозначение $k = mg/l$. Каким образом производится изменение жесткости или величин, ей эквивалентных, — это не имеет значения, пока задача трактуется как чисто математическая.

Полная энергия гармонического осциллятора равна

$$E = \frac{mv^2}{2} + \frac{kx^2}{2}.$$

Для ее производной по времени можно написать

$$\dot{E} = (mv\dot{v} + kx\dot{x}) + \frac{1}{2}x^2\dot{k}\,.$$

Выражение в скобках обращается в нуль, так как $\dot{x} = v$, действующая сила $F = -kx$ и по закону Ньютона $m\dot{v} = F$. Введя еще потенциальную энергию $U = \frac{1}{2}kx^2$, получим

$$\dot{E} = U(x)\frac{\dot{k}}{k}. \tag{43.1}$$

До сих пор наши вычисления были точными. Используем теперь *медленность изменения* параметра k и его производной $\dot{k}$. Медленность означает, что и при изменяющемся k движение попрежнему будет носить характер колебаний. Только «период» этих колебаний T, а также положение крайних точек, достигаемых осциллятором, будут слегка меняться от колебания к колебанию. Иными словами, за каждое колебание параметр k должен изменяться очень мало. Это обычное представление о медленности. Но в нашей задаче его недостаточно. Надо на изменения k и его производной наложить дополнительное ограничение, потребовав, чтобы за каждый период колебания величина $\dot{k}/k$ оставалась почти постоянной. Точнее, это требование сводится к тому, чтобы на каждом периоде колебания отношение $\dot{k}/k$ могло быть представлено в виде

$$\frac{\dot{k}}{k} = \left(\frac{\dot{k}}{k}\right)_0 (1 + \alpha), \tag{43.2}$$

где $(\dot{k}/k)_0$ — значение этого отношения в какой-либо точке рассматриваемого периода, например в его середине, а α — поправка, стремящаяся к нулю при $\dot{k} \to 0$. Имея это в виду, проинтегрируем выражение (43.1) в пределах от t до $t + T(k)$ для произвольного момента времени t. Получим

$$\mathrm{d}\Delta E = E(t+T) - E(t) = \left(\frac{\dot{k}}{k}\right)_0 \left[\int_t^{t+T} U[x(t')]\,dt' + \beta\right]. \tag{43.3}$$

Здесь β — поправка, обращающаяся в нуль при $\dot{k} \to 0$. (Переменная интегрирования обозначена через t', чтобы не смешивать ее с нижним пределом t.) Входящий сюда интеграл достаточно вычислить в нулевом приближении, т. е. считать при вычислении, что за время T параметр k не меняется. Возникающая вследствие этого ошибка в

выражении для ΔE будет второго или высшего порядка малости по $\dot{k}$. По той же причине можно отбросить поправку β. Наконец, можно опустить индекс нуль в множителе перед интегралом $(\dot{k}/k)_0$. Иными словами, можно написать

$$\Delta E = \frac{\dot{k}}{k} \int_t^{t+T} U[x(t')]\, dt', \tag{43.4}$$

где интеграл (но не множитель перед интегралом) вычисляется в предположении постоянства k. При вычислении интеграла момент времени t можно принять за начало отсчета времени, т. е. положить $t = 0$. Это, конечно, не изменит результата. При $k = \text{const}$ координата x совершает гармонические колебания $x = x_0 \cos(\omega t + \delta)$, а потому

$$U = \frac{1}{2} kx_0^2 \cos^2(\omega t + \delta) = \frac{1}{2} E[1 + \cos(2\omega t + \delta)],$$

так как полная энергия равна $E = \frac{1}{2} kx_0^2$. Используя полученное выражение и выполняя интегрирование, получим

$$\int_0^T U\, dt = \frac{1}{2} E \int_0^T [1 + \cos(2\omega t + \delta)]\, dt = \frac{1}{2} ET.$$

Произведение $T\dot{k}$ с точностью до величин высшего порядка малости по $\dot{k}$ дает приращение Δk параметра k за период T. Таким образом, вместо (43.4) можно написать

$$\Delta E = \frac{\Delta k}{2k} E. \tag{43.5}$$

Приращение Δk за период колебаний может быть сделано сколь угодно малым. Поэтому если энергию E рассматривать как функцию параметра k, то в пределе приближенное соотношение (43.5) перейдет в точное дифференциальное уравнение

$$\frac{dE}{E} - \frac{dk}{2k} = 0.$$

Интегрируя это уравнение, получим

$$\ln \frac{E}{\sqrt{k}} = \text{const},$$

а потому

$$\frac{E}{\sqrt{k}} = \text{const}. \tag{43.6}$$

Используя формулы (40.3) и (40.4), из этого соотношения получим еще два других:

$$ET = \text{const}, \tag{43.7}$$

$$\frac{E}{\omega} = \text{const}. \tag{43.8}$$

Эти соотношения означают, что *величины ET и E/ω для гармонического осциллятора являются адиабатическими инвариантами.* При этом *период колебания T и частота* ω, входящие в эти соотношения, *должны вычисляться так, как если бы при колебаниях параметр k оставался постоянным*, т. е. по формулам (40.3) и (40.4). Например, в случае медленного укорочения нити математического маятника, совершающего малые колебания, его период T медленно уменьшается от колебания к колебанию. Одновременно энергия колебаний возрастает таким образом, что произведение ET остается постоянным.

3. Для правильного понимания доказанной теоремы необходимо точно отдавать себе отчет, что понимается под медленностью изменения параметра осциллятора k. Недостаточно, чтобы изменения параметра k на каждом периоде колебаний были бесконечно малы. Надо, чтобы эти изменения удовлетворяли условию (43.2). Представим себе, например, что вблизи нижнего положения нить математического маятника действием внешних сил немного укорачивается, а вблизи крайних положений удлиняется, принимая исходное значение. Работа внешних сил при укорочении нити вблизи нижнего положения будет больше работы, производимой маятником над внешними полями при удлинении нити вблизи каждого крайнего положения. Причина этого в том, что при колебаниях маятника натяжение нити меняется, достигая максимума в нижнем положении. Поэтому за каждое колебание в систему дважды будет вкладываться энергия. И если число колебаний взять достаточно большим, то и прирост энергии можно сделать также большим, хотя длина маятника, а с ней и период T останутся неизменными. Колебания с периодически меняющимися параметрами называются *параметрическими.* Примером могут служить качели. К параметрическим колебаниям результаты (43.7) и (43.8) не применимы, сколь бы малыми ни были изменения параметра k в пределах каждого периода колебаний. Причина этого в том, что эти изменения не удовлетворяют условию (43.2). Грубо говоря, *условие* (43.2) *сводится к требованию, чтобы изменения параметра k происходили медленно и монотонно.* Так, в приведенном примере адиабатическая инвариантность выражений (43.7) и (43.8) будет иметь место, если длина нити изменяется медленно и монотонно. Если же на такие изменения наложить еще малые изменения колебательного характера, подобные тем, которые имеют место при параметрическом возбуждении колебаний, то к таким случаям теорема об адиабатической инвариантности выражений (43.7) и (43.8) не применима.

4. Полученные результаты можно обобщить на случай негармонических колебаний с одной степенью свободы, т. е. колебаний, совершающихся под действием не квазиупругих сил. В этом случае колеблющаяся величина меняется во времени не синусоидально, а как-то иначе. Период колебаний T определяется не

только параметрами системы, а и их амплитудой. Вместо формулы (43.7) получается

$$\overline{K}T = \text{const}, \tag{43.9}$$

где $\overline{K}$ — кинетическая энергия системы, усредненная по времени за период колебания (черта как раз и означает такое усреднение), т. е.

$$\overline{K}(t) = \frac{1}{T}\int_{t}^{t+T} K(t')\,dt'. \tag{43.10}$$

В случае гармонических колебаний, как нетрудно доказать, средние за период значения кинетической и потенциальной энергий одинаковы, а потому на каждую из них приходится половина полной энергии, т. е. $\overline{K} = \overline{U} = \frac{1}{2} E$. Тогда формула (43.9) переходит в ранее полученную формулу (43.7).

Подставим выражение (43.10) в формулу (43.9) и примем во внимание, что $K = \frac{1}{2} mv^2$, $p = mv$, $dq = v\,dt$, где dq — приращение координаты, определяющей положение математической точки. Тогда получится

$$\oint p\,dq = \text{адиабатический инвариант}, \tag{43.11}$$

причем интегрирование ведется по полному периоду движения материальной точки в предположении, что параметры, характеризующие систему, закреплены. Общее доказательство соотношения (43.11) основано на уравнениях механики в форме Гамильтона. Мы его приводить не будем. Ограничимся только двумя примерами.

5. Первый пример. В цилиндре с гладкими стенками движется вверх и вниз идеально упругий шарик, последовательно отражающийся от основания AB и поршня CD по законам абсолютно упругого удара (рис. 106). Допустим сначала, что поля силы тяжести и прочих силовых полей нет. Заставим поршень CD очень медленно перемещаться со скоростью u. Исследуем, как это скажется на движении шарика. Перейдем в движущуюся систему отсчета, в которой поршень покоится. В этой системе скорость шарика будет $v - u$. После отражения шарика она сохранится по модулю, но изменит знак, т. е. будет равна $-v + u$. В неподвижной системе отсчета та же скорость будет равна $(-v + u) + u = -v + 2u$. Приращение кинетической энергии шарика в результате однократного отражения от движущегося поршня будет поэтому равно

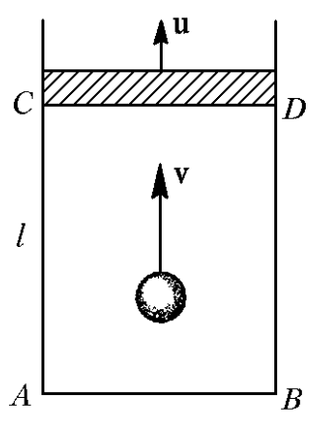


Рис. 106

$$\Delta K = \frac{m}{2}\left[(-v + 2u)^2 - v^2\right] = -2m(uv - u^2).$$

Разделив это соотношение на $K = \frac{1}{2} mv^2$ и пренебрегая квадратом малой скорости u, получим

$$\frac{\Delta K}{K} = -4\frac{u}{v}. \tag{43.12}$$

Посмотрим теперь, как меняется период колебаний шарика T в результате движения поршня. Под периодом T мы понимаем время движения шарика туда и обратно, вычисленное в предположении, что во время такого движения поршень *закреплен*. Если l — расстояние между поршнем и дном цилиндра во время этого движения, то $T = 2l/v$. Спустя время T расстояние l возрастает на uT, а скорость шарика уменьшается на $2u$. Период колебаний в только что указанном смысле изменится и сделается равным

$$T' = \frac{2(l+uT)}{v-2u} = \frac{2(l+uT)\,(v+2u)}{v^2-4u^2},$$

или, пренебрегая квадратом малой скорости u,

$$T' = T + \frac{2uTv+4lu}{v^2} = T + 4T\,\frac{u}{v}.$$

Таким образом, за время T период получает приращение

$$\Delta T = 4T\,\frac{u}{v} = -T\,\frac{\Delta K}{K}.$$

В пределе, когда поршень движется бесконечно медленно, приращения ΔT и ΔK могут рассматриваться как бесконечно малые дифференциалы, и мы получаем уравнение

$$\frac{dT}{T} + \frac{dK}{K} = 0.$$

Интегрируя его, находим

$$TK = \text{const}, \tag{43.13}$$

т. е. величина TK является адиабатическим инвариантом.

6. Второй пример. Учтем теперь наличие силы тяжести. Пусть v_1 — скорость отразившегося шара в верхнем положении (рис. 107). На расстоянии x его скорость v определится соотношением $v^2 = v_1^2 + 2gx$. Интеграл (43.11) в рассматриваемом случае будет

$$J = 2m \int_0^l \sqrt{v_1^2 + 2gx}\; dx,$$

где l — расстояние между поршнем и дном цилиндра. *Интеграл надо вычислить в предположении, что поршень закреплен, т. е. при постоянном l.* Вычисление дает

$$J = \frac{2m}{3g}\,(v_1^2 + 2gl)^{3/2} - v_1^3 = \frac{2m}{3g}\,(v_2^3 - v_1^3),$$

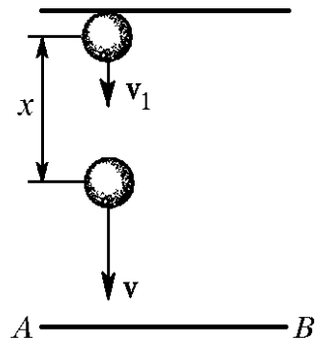


Рис. 107

где v_2 — скорость шара в нижнем положении. Таким образом, надо доказать, что разность $v_2^3 - v_1^3$ является адиабатическим инвариантом. Для этого вычислим значения скоростей v_1 и v_2 спустя период T. Обозначим эти значения через v_1' и v_2' соответственно. Разумеется, вычисление надо по-прежнему провести для неподвижного поршня, но переместившегося в новое положение. За период T поршень переместится вверх на расстояние uT. Шар пройдет это расстояние за время $\Delta T = uT/v_1$ (если пренебречь величинами более высокого порядка малости). При этом под дей-

ствием силы тяжести его скорость уменьшится на $g\,\Delta T = guT/v_1$. Кроме того, при отражении движущегося поршня эта скорость дополнительно уменьшается на $2u$. Поэтому

$$v_1' = v_1 - (g\frac{T}{v_1} + 2)u = v_1 - 2\frac{v_2}{v_1}u,$$

так как

$$v_2 = v_1 + g\frac{T}{2}.$$

На уровне AB скорость шара будет $v_1'' = v_1 - 2u$, а около дна цилиндра

$$v_2'^2 = v_1''^2 + 2gl = (v_1 - 2u)^2 + 2gl = v_2^2 - 4v_1u,$$

если пренебречь квадратом u. Извлекая квадратный корень и снова пренебрегая u^2, получим

$$v_2' = v_2 - 2\frac{v_1}{v_2}u.$$

С той же степенью точности

$$v_2'^3 = v_2^3 - 6v_1v_2u, \quad v_1'^3 = v_1^3 - 6v_1v_2u.$$

Значит, $v_2'^3 - v_1'^3 = v_2^3 - v_1^3$ или $J' - J = 0$, причем это соотношение верно с точностью до членов порядка $u^2 = \dot{l}^2$. Разделив его на время T и отождествив частное $(J' - J)/T$ с производной dJ/dt, получим

$$\frac{dJ}{dt} = A\dot{l}^2,$$

где A от $\dot{l}$ не зависит. Имея в виду, что нас интересуют изменения величины J при *конечных изменениях* l, преобразуем это соотношение, введя вместо дифференциала времени дифференциал длины $dl = \dot{l}\,dt$. Тогда получится

$$\frac{dJ}{dl} = A\dot{l},$$

или в пределе при $\dot{l} \to 0$

$$\frac{dJ}{dl} = 0.$$

Следовательно, $J = \text{const}$, как бы велики ни были изменения параметра l, т. е. *величина J является адиабатическим инвариантом.* Такая адиабатическая инвариантность получилась благодаря тому, что производная dJ/dt оказалась пропорциональной *второй,* а не первой степени $\dot{l}$. Если бы dJ/dt была пропорциональна первой степени производной $\dot{l}$, то адиабатической инвариантности J не получилось бы.

ЗАДАЧИ

1. Шарик математического маятника (или шарик, прикрепленный к пружине с заданной жесткостью) медленно испаряется (система с переменной массой). Будет ли величина ET адиабатическим инвариантом и почему?

Ответ. Нет.

2. Шарику массой m, надетому на тонкую стальную спицу, масса которой пренебрежимо мала по сравнению с m, сообщена продольная скорость в направлении к точке закрепления спицы, а также скорость в перпендикулярном направлении. Предполагая, что за период колебания шарика его смещение вдоль спицы мало по сравнению с ее длиной, и пренебрегая трением, определить характер последующего движения шарика.

Р е ш е н и е. Если $v_\perp$ — поперечная скорость шарика, то величина $v_\perp^2 T$ является адиабатическим инвариантом. Период T пропорционален расстоянию l шарика от точки закрепления спицы (см. задачу 20 к § 42). Поэтому адиабатическим инвариантом будет также $A = v_\perp^2 l$. Кроме того, движение шарика подчиняется закону сохранения энергии, который требует, чтобы полная скорость шарика v сохранялась по абсолютной величине. Если $v_\parallel$ — скорость шарика вдоль спицы, то $v^2 \equiv v_\parallel^2 + v_\perp^2 = \text{const}$. Таким образом, величина $A = (v^2 - v_\parallel^2) l$ есть адиабатический инвариант. На расстоянии l_0, при котором $v^2 l_0 = A$, продольная скорость $v_\parallel$ обратится в нуль. Поэтому шарик не может подойти к точке закрепления ближе расстояния l_0. Достигнув положения $l = l_0$, он должен отразиться.

ГЛАВА VII

МЕХАНИКА ТВЕРДОГО ТЕЛА

§ 44. ТВЕРДОЕ ТЕЛО В МЕХАНИКЕ. УРАВНЕНИЯ ДВИЖЕНИЯ И РАВНОВЕСИЯ ТВЕРДОГО ТЕЛА

1. В двух предыдущих главах уже говорилось о законах движения твердого тела и их применениях к некоторым простейшим движениям. В этой главе будет продолжено изучение избранных вопросов механики твердого тела.

Напомним, что *твердым телом* в механике называют неизменяемую систему материальных точек, т. е. такую идеализированную систему, при любых движениях которой взаимные расстояния между материальными точками системы остаются неизменными. Здесь, как и вообще в классической механике, под материальными точками понимают не атомы или молекулы, а достаточно малые макроскопические части, на которые мысленно можно разделить рассматриваемую механическую систему.

С атомистической точки зрения силы взаимодействия между материальными точками твердого тела являются силами *электрическими.* Но атомистический подход чужд феноменологической механике твердого тела. Последняя рассматривает твердое тело как *сплошную среду*, между различными элементами которой действуют внутренние силы в виде *нормальных* и *касательных напряжений.* Причиной их феноменологическая механика считает *деформации тел.* Если в теле совсем нет деформаций, то не может быть и внутренних напряжений. Однако если деформации, возникающие под действием внешних сил, малы и сами по себе нас не интересуют, то в ряде случаев от них можно отвлечься. Таким путем мы приходим к идеализированной модели тела, совершенно не способного деформироваться, хотя под действием внешних сил в нем и могут возникать внутренние натяжения и давления. Это и есть идеально твердое тело. Допустима или нет такая, как и всякая другая, идеализация — это определяется не только свойствами реальных тел, но и содержанием тех вопросов, на которые надо получить ответ.

2. Твердое тело является механической системой с шестью степенями свободы (см. § 8). Для описания его движения требуется шесть независимых числовых уравнений. Вместо них можно взять два независимых векторных уравнения. Таковыми являются уравнение движения центра масс

$$m\frac{d\mathbf{v}}{dt}=\mathbf{F}_{\text{внеш}} \tag{44.1}$$

и уравнение моментов

$$\frac{d\mathbf{L}}{dt} = \mathbf{M}_{\text{внеш}}. \tag{44.2}$$

Уравнение моментов можно брать относительно произвольного неподвижного начала или относительно центра масс твердого тела. Можно также брать произвольно движущееся начало, если только скорость его в любой момент времени параллельна скорости центра масс (см. § 37). При ограничении свободы движения число независимых уравнений, требующихся для описания движения твердого тела, уменьшается. Оно всегда равно числу степеней свободы.

В уравнения (44.1) и (44.2) входят только внешние силы. Внутренние силы не влияют на движение центра масс и не могут изменить момент импульса тела. Они могут изменять только взаимное расположение и скорости материальных точек тела. Но для абсолютно твердого тела такие изменения невозможны. Таким образом, внутренние силы не влияют на движение твердого тела. Если же сила внешняя, то точку приложения ее можно произвольно перемещать вдоль линии, по которой она действует. Действительно, при таком перемещении не меняются результирующая внешних сил $\mathbf{F}_{\text{внеш}}$ и их момент $\mathbf{M}_{\text{внеш}}$, т. е. уравнения движения (44.1) и (44.2) остаются без изменения. Подобное перемещение недопустимо в случае деформированного тела, так как оно приводит к перераспределению деформаций и изменению внутренних движений тела.

3. Если твердое тело покоится, то уравнения (44.1) и (44.2) переходят в

$$\mathbf{F}_{\text{внеш}} = 0, \quad \mathbf{M}_{\text{внеш}} = 0. \tag{44.3}$$

Это *необходимые условия* равновесия твердого тела. Но они не являются достаточными. При их выполнении центр масс может еще двигаться прямолинейно и равномерно с произвольной скоростью, а само тело может вращаться с сохранением вращательного импульса. Так как при равновесии результирующая внешних сил $\mathbf{F}_{\text{внеш}}$ равна нулю, то момент этих сил $\mathbf{M}_{\text{внеш}}$ в состоянии равновесия не зависит от положения неподвижного начала O, относительно которого он берется [см. формулу (30.7)]. Поэтому при решении любой задачи на равновесие твердого тела начало O можно выбирать произвольно, что можно использовать для упрощения самого решения.

4. Не всегда можно заменять реальные твердые тела идеализированными абсолютно твердыми моделями даже в тех случаях, когда деформации пренебрежимо малы. В качестве примера рассмотрим задачу о равновесии твердой балки. Пусть однородная балка весом P лежит на двух опорах *1* и *2* (рис. 108). Центр масс балки находится посередине между опорами. Найдем силы F_1 и F_2, с которыми балка давит на опоры. Механика твердого тела дает два условия равновесия:

$$F_1 + F_2 = P, \quad F_2 l = P\frac{l}{2}, \tag{44.4}$$

где l — расстояние между опорами. Второе из них означает, что должен обращаться в нуль момент действующих сил относительно опоры *1*. Из этих условий получаем $F_1 = F_2 = P/2$. Это разумный результат. Деформации балки в рассматриваемой задаче не играют существенной роли. Идеализация абсолютно твердого тела допустима.

Рассмотрим теперь балку на трех опорах (рис. 109). Механика твердого тела по-прежнему дает два условия равновесия:

$$F_1 + F_2 + F_3 = P, \quad F_3 x + F_2 l = Pl/2, \tag{44.5}$$

где l — расстояние между опорами *1* и *2*, а x — между опорами *1* и *3*. (Второе уравнение (44.5) получится, если приравнять нулю момент внешних сил относительно опоры *1*.) Двух уравнений недостаточно для определения трех неизвестных сил F_1, F_2, F_3. Одной из этих сил можно придать произвольное значение, тогда из уравнений (44.5) найдутся остальные две. Задача о распределении веса абсолютно твердой балки между тремя опорами, на которых она лежит, оказалась неопределенной. Механические системы, подобные абсолютно твердой балке на трех опорах, называют *статически неопределенными.*

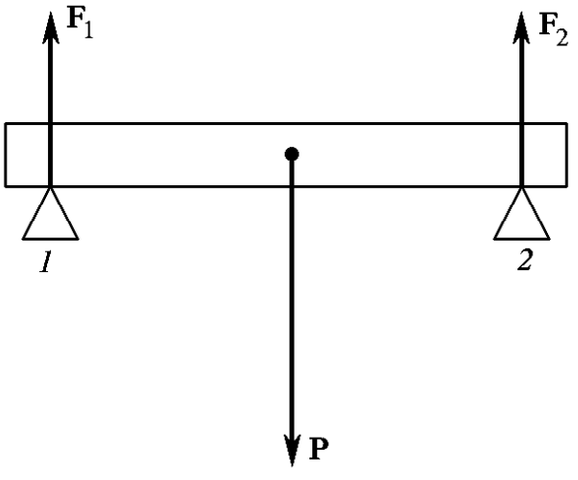


Рис. 108

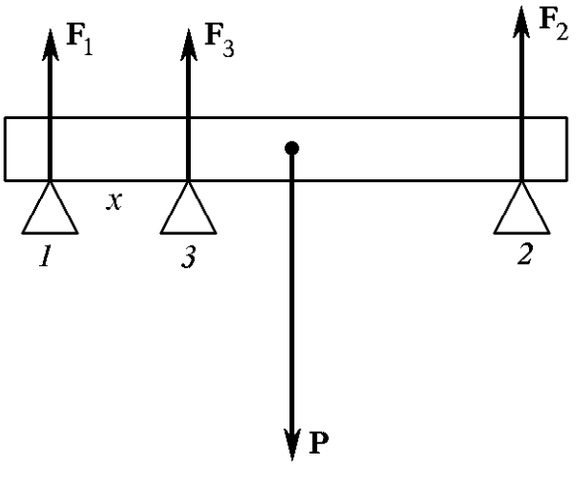


Рис. 109

Аналогичная ситуация встречается также в задаче о равновесии стола, стоящего на горизонтальной плоскости. Механика твердого тела дает в этом случае три независимых уравнения равновесия. Если число ножек стола равно трем, то из этих трех уравнений можно однозначно определить три силы, с которыми ножки давят на плоскость опоры. Но если стол стоит на четырех ножках, то трех уравнений мало для определения четырех сил давления ножек на плоскость опоры. Идеально твердый стол с четырьмя ножками, стоящий на идеально твердой горизонтальной плоскости, является также статически неопределенной системой.

Конечно, вес реальной балки, лежащей на трех опорах, вполне определенным образом распределяется между ними. Точно так же вполне определенным образом распределяется сила давления реального стола между четырьмя ножками, на которых он стоит. Неопределенность, к которой мы пришли, указывает просто на то, что в рассматриваемых задачах балку на трех опорах или стол на четырех ножках *нельзя считать идеально твердыми.* Надо учитывать их деформации, а также деформации опор.

5. Следующее простое рассуждение выясняет, почему деформации могут оказаться существенными. Допустим, что идеально твердая балка лежит на

трех опорах (см. рис. 109). Сместим среднюю опору немного вниз. Так как идеально твердая балка не может деформироваться, то между ней и опорой *3* сразу же возникает зазор — балка перестанет опираться на опору *3* и давить на нее. Сила давления на опору *3* скачком обратится в нуль, а вес балки вполне определенным образом и тоже скачком распределится между оставшимися опорами *1* и *2*. То же произойдет, если у твердого стола бесконечно мало укоротить одну из четырех ножек. Не то будет в случае реальной балки или реального стола. При бесконечно малом опускании опоры *3* балка прогнется и по-прежнему будет на нее опираться. Сила давления на опору *3* не обратится в нуль, а уменьшится бесконечно мало. Аналогично когда стол стоит на четырех ножках, то все ножки деформируются, укорачиваясь под действием силы веса. Если подпилить одну из ножек, укоротив ее бесконечно мало, то деформация ножки уменьшится, ножка удлинится и снова будет касаться плоскости опоры. Таким образом, и в этом случае давление на ножку изменяется бесконечно мало, т. е. непрерывно.

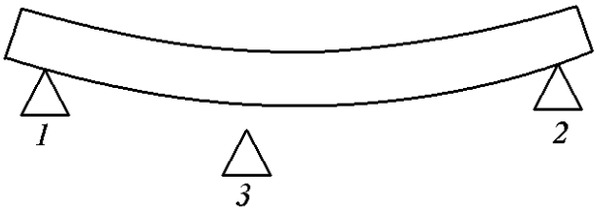


Рис. 110

Допустим теперь, что реальная балка лежит на двух опорах, *1* и *2* (рис. 110). Под действием силы веса балка прогнется. Будем подводить под балку третью опору *3*, непрерывно поднимая ее. Как только опора *3* коснется балки, с балкой не произойдет еще никаких изменений. Но дальнейшее поднятие опоры *3* будет связано с выпрямлением балки, а для этого опора должна действовать на балку с некоторой силой. По мере поднятия опоры *3* эта сила будет непрерывно возрастать, принимая вполне определенное значение при каждом положении опоры *3*. Приведенное рассуждение ясно показывает, почему при решении задачи о распределении веса балки между тремя опорами надо учитывать ее упругие свойства. Эта задача будет решена в гл. X (см. задачу 3 к § 80). Аналогичное рассуждение можно провести и для стола на четырех ножках.

§ 45. МГНОВЕННАЯ ОСЬ ВРАЩЕНИЯ

1. Пусть твердое тело вращается вокруг неподвижной оси. Чтобы получить представление о распределении скоростей в нем, достаточно рассмотреть движение точек тела, лежащих в какой-либо одной плоскости, перпендикулярной к оси вращения. Это значит, что тело можно считать как бы плоским. Соответствующее распределение скоростей показано на схематическом рис. 111. Точка O тела, через которую проходит ось вращения, неподвижна. Все другие точки тела движутся по окружностям с центром в O. Их плоскости пропорциональны радиусам соответствующих окружностей. Модули скоростей могут меняться с течением времени, но ось вращения остается одной и той же.

2. Рассмотрим теперь более общее движение плоского твердого тела. Плоскость вращения совпадает с плоскостью самого тела. Никакой неподвижной оси, вокруг которой происходило бы вращение

тела, не предполагается. Если A и B — две произвольные точки твердого тела (рис. 112), то расстояние между ними остается неизменным, а потому $(\mathbf{r}_B - \mathbf{r}_A)^2 = \text{const}$. Дифференцируя это соотношение по времени, получим $(\mathbf{r}_B - \mathbf{r}_A)(\dot{\mathbf{r}}_B - \dot{\mathbf{r}}_A) = 0$, или

$$\mathbf{r}_{AB}(\mathbf{v}_B - \mathbf{v}_A) = 0, \quad (45.1)$$

где $\mathbf{r}_{AB} \equiv \overrightarrow{AB}$. Допустим, что в рассматриваемый момент времени в теле существует точка, скорость которой в этот момент времени равна нулю. (В § 47 будет показано, что такая точка существует для произвольного плоского движения твердого тела.) Примем ее за точку A. Тогда для рассматриваемого момента времени

$$\mathbf{r}_{AB}\mathbf{v}_B = 0,$$

каково бы ни было положение точки B. Отсюда видно, что скорость $\mathbf{v}_B$ перпендикулярна к $\mathbf{r}_{AB}$, т. е. направлена по касательной к окружности с центром в A. При движении твердого тела всякая прямая в теле остается прямой. Это справедливо и для прямой, соединяющей точки A и B. Поскольку в рассматриваемый момент точка A неподвижна, то модуль скорости $\mathbf{v}_B$ в этот момент пропорционален расстоянию AB от точки B до точки A. На основании всего этого можно сказать, что *мгновенное распределение скоростей в теле в рассматриваемый момент времени будет в точности таким же, как и при вращении вокруг неподвижной оси, проходящей через точку A.* Движение тела в этом случае называют *мгновенным вращением.* Прямая, проходящая через точки тела, скорости которых в рассматриваемый момент времени равны нулю, называется *мгновенной осью вращения.* В нашем примере мгновенная ось проходит через точку A. Словом «мгновенная» хотят подчеркнуть, что это понятие служит для описания распределения скоростей только в какой-то заданный *момент времени.* В отличие от неподвижной оси, сохраняющей свое положение в теле и в пространстве, *мгновенная ось, вообще говоря, перемещается как в теле, так и в пространстве.* Если получить моментальную фотографию распределения скоростей в теле, то по виду этой фотографии нельзя сказать, происходит ли вращение вокруг неподвижной или вокруг мгновенной оси. Чтобы отличить эти два вращения, надо

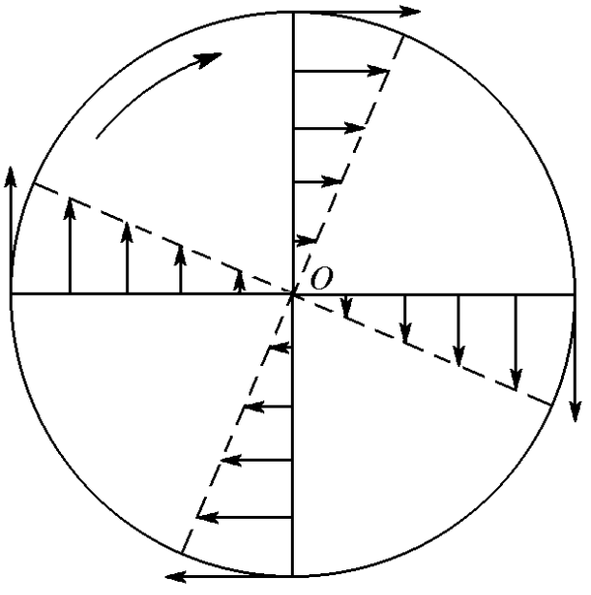


Рис. 111

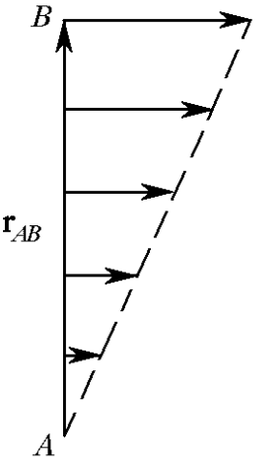


Рис. 112

получить такие фотографии по крайней мере в два различных момента времени.

3. *Мгновенная ось служит для описания мгновенного распределения только скоростей.* Той же осью нельзя пользоваться для описания мгновенного распределения ускорений или высших производных скорости по времени. Распределение ускорений при вращении вокруг мгновенной оси может существенно отличаться от соответствующего распределения ускорений при вращении вокруг неподвижной оси, хотя бы угловые скорости вращения в обоих случаях и совпадали. Дело в том, что для определения ускорений недостаточно знать распределение скоростей только в рассматриваемый момент времени. Надо знать это распределение также в бесконечно близкий момент времени. А в этот момент может оказаться, что движение тела уже перестанет быть вращением вокруг прежней мгновенной оси.

Следующий простой пример хорошо разъясняет суть дела. Рассмотрим качение обруча или диска по плоскости без скольжения (рис. 113). Отсутствие скольжения означает, что точка обруча A, в которой он касается плоскости, в рассматриваемый момент неподвижна. Следовательно, движение обруча можно рассматривать как мгновенное вращение его вокруг мгновенной оси, проходящей через точку касания A. Распределения скоростей при таком движении показано на рис. 113. С течением времени в соприкосновении с плоскостью будут приходить другие точки обруча. При этом точка касания будет перемещаться по плоскости в ту же сторону, куда движется обруч. Это означает, что мгновенная ось перемещается как относительно катящегося обруча, так и относительно плоскости, по которой происходит качение. В этом и состоит смысл утверждения, что мгновенная ось перемещается как в теле, так и в пространстве. Допустим теперь, что качение происходит с постоянной скоростью. Было бы грубой ошибкой вычислять ускорение по формуле $\mathbf{a} = -\omega^2\mathbf{R}$, понимая под $\mathbf{R}$ радиус-вектор, проведенный от мгновенной оси к рассматриваемой точке обруча. Действительно, полная скорость $\mathbf{v}$ любой точки обруча векторно складывается из скорости $\mathbf{v}_C$ поступательного движения центра обруча C и скорости $\mathbf{v}_{вр}$ вращения ее относительно того же центра: $\mathbf{v} = \mathbf{v}_C + \mathbf{v}_{вр}$. Если обруч катится равномерно, то $\frac{d\mathbf{v}_C}{dt} = 0$, и ускорение

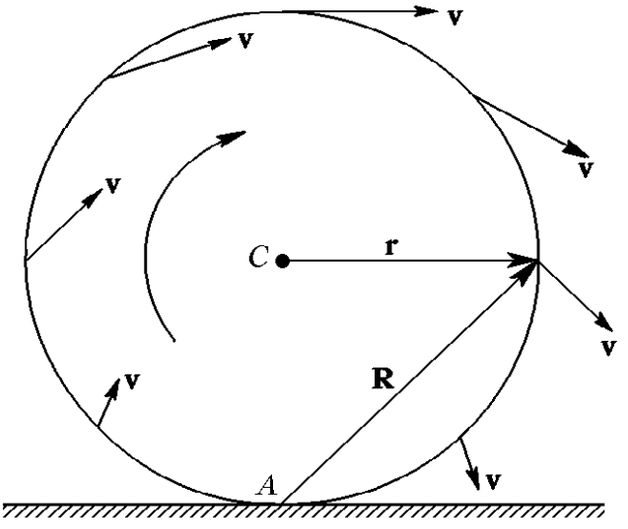


Рис. 113

будет равно $\mathbf{a} = \frac{d\mathbf{v}_{вр}}{dt}$. Поступательное движение не влияет на ускорение $\mathbf{a}$. Оно такое же, как и при вращении вокруг неподвижного центра, т. е. $\mathbf{a} = -\omega^2\mathbf{r}$, где радиус-вектор $\mathbf{r}$ проведен из центра обруча O. Таким образом, при равномерном качении ускорение $\mathbf{a}$ направлено к центру обруча, а не к мгновенной оси.

§ 46. УГЛОВАЯ СКОРОСТЬ КАК ВЕКТОР. СЛОЖЕНИЕ ВРАЩЕНИЙ

1. Пусть твердое тело вращается вокруг неподвижной или мгновенной оси OA с угловой скоростью $\boldsymbol{\omega}$ (рис. 114). Возьмем какую-либо произвольную точку этого тела M, отстоящую от оси вращения на расстояние $\mathbf{r}_\perp$. Линейная и угловая скорости точки M связаны соотношением

$$v = \omega r_\perp. \tag{46.1}$$

Введем аксиальный вектор $\boldsymbol{\omega}$, определяемый векторным произведением

$$\boldsymbol{\omega} = [\mathbf{r}_\perp \mathbf{v}]/\mathbf{r}_\perp^2, \tag{46.2}$$

где $\mathbf{r}_\perp$ — радиус-вектор, проведенный от оси вращения к точке M перпендикулярно к этой оси. Длина вектора $\boldsymbol{\omega}$ в силу соотношения (46.1) численно равна угловой скорости вращения, а направление совпадает с направлением оси вращения. Взаимное расположение векторов $\boldsymbol{\omega}$, $\mathbf{r}_\perp$ и $\mathbf{v}$ мы уясним лучше, если отложим их из общего начала (рис. 115). Эти три вектора взаимно перпендикулярны. Из рисунка видно, что

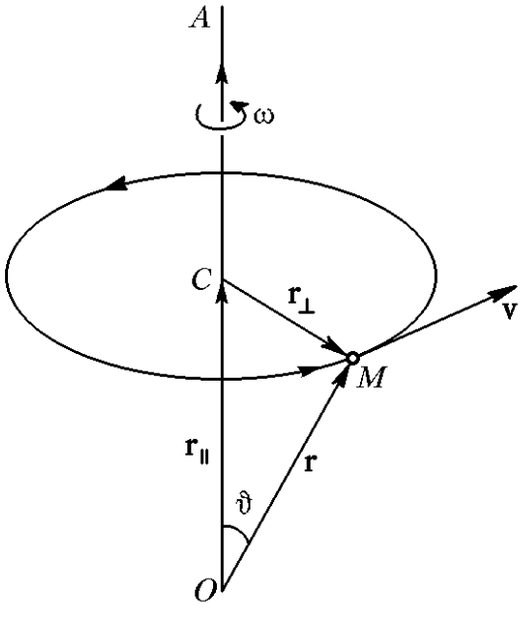


Рис. 114

$$\mathbf{v} = [\boldsymbol{\omega}\mathbf{r}_\perp]. \tag{46.3}$$

Эта формула является обобщением формулы (46.1), поскольку она определяет не только модуль скорости $\mathbf{v}$, но и ее направление. Вектор $\boldsymbol{\omega}$ называется *вектором угловой скорости*, или просто *угловой скоростью вращения*. Таким образом, *угловую скорость можно рассматривать как вектор*. Если расположить буравчик с правой нарезкой параллельно оси вращения и вращать его в ту же сторону, в какую вращается само тело, то направление ввинчивания буравчика укажет направление вектора $\boldsymbol{\omega}$.

Формуле (46.3) можно придать более общий и удобный вид. Возьмем на оси вращения произвольную точку O в качестве начала координат (см. рис. 114). Тогда радиус-вектор $\mathbf{r}$, проведенный из этого начала к точке M, можно представить в виде векторной суммы $\mathbf{r} = \mathbf{r}_\perp + \mathbf{r}_\parallel$, где $\mathbf{r}_\parallel$ — слагающая вектора $\mathbf{r}$ вдоль оси вращения. Так как $[\boldsymbol{\omega}\mathbf{r}_\parallel] = 0$, то вместо формулы (46.3) можно написать более общую формулу

$$\mathbf{v} = [\boldsymbol{\omega}\mathbf{r}]. \qquad (46.4)$$

Из нее получаем $v = \omega r \sin\vartheta$, что совпадает с формулой (46.1), так как $\mathbf{r}\sin\vartheta = \mathbf{r}_\perp$.

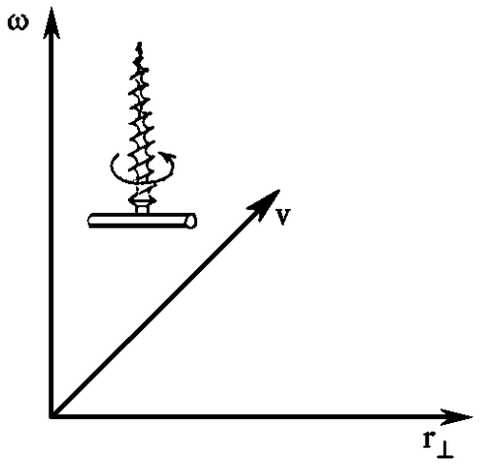


Рис. 115

2. *То что угловая скорость* $\boldsymbol{\omega}$ *есть вектор, не требует специального доказательства, поскольку она определена как векторное произведение двух векторов. Векторный характер* $\boldsymbol{\omega}$ *означает*, разумеется, *только то, что при повороте координатных систем проекции* $\boldsymbol{\omega}$ *на их оси преобразуются так же, как разности координат концов направленного геометрического отрезка.* Над векторами угловых скоростей можно выполнять все математические операции, как над всякими векторами. В частности, можно ввести математическое сложение векторов $\boldsymbol{\omega}_1$ и $\boldsymbol{\omega}_2$ по правилу параллелограмма. Но как будут складываться угловые скорости, если сложение определить с помощью той или иной *физической* операции, — это требует особого исследования. Введем понятие *сложения вращений*, вложив в него следующий смысл. Пусть тело вращается вокруг некоторой оси OA с угловой скоростью $\boldsymbol{\omega}_1$ (рис. 116). Пусть сама ось OA, в свою очередь, вращается с угловой скоростью $\boldsymbol{\omega}_2$, вокруг другой оси OB. Подчеркнем, что в общем случае *речь идет о мгновенных вращениях и притом с нерелятивистскими скоростями.* Первое вращение рассматривается в системе отсчета, в которой (в рассматриваемый момент) ось OA неподвижна. Второе вращение рассматривается в другой системе отсчета — в той, в которой (в тот же момент) неподвижна ось OB. Сложить вращательные движения — значит ответить на вопрос, к какому движению приводит наложение этих двух вращений? При рассмотрении этого вопроса ограничимся случаем, когда оси OA и OB пересекаются между собой.

Вопрос сводится к сложению линейных скоростей в аналогичном физическом смысле (см. § 7; в нерелятивистской механике, как известно, сложение линейных скоростей производится по правилу параллелограмма). Произвольная точка твердого тела M с радиусом-

вектором $\mathbf{r}$ в результате первого вращения (вокруг оси OA) получает линейную скорость $\mathbf{v}_1 = [\boldsymbol{\omega}_1 \mathbf{r}]$, а в результате второго вращения (вокруг оси OB) — линейную скорость $\mathbf{v}_2 = [\boldsymbol{\omega}_2 \mathbf{r}]$. Результирующая линейная скорость будет равна

$$\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 = [(\boldsymbol{\omega}_1 + \boldsymbol{\omega}_2)\mathbf{r}].$$

Если ввести векторную сумму в математическом смысле

$$\boldsymbol{\omega} = \boldsymbol{\omega}_1 + \boldsymbol{\omega}_2, \tag{46.5}$$

то результат запишется в виде

$$\mathbf{v} = [\boldsymbol{\omega}\mathbf{r}]. \tag{46.6}$$

Пусть точка M лежит на оси вектора $\boldsymbol{\omega}$, т. е. на диагонали параллелограмма, построенного на векторах $\boldsymbol{\omega}_1$ и $\boldsymbol{\omega}_2$, или ее продолжении. Тогда $\mathbf{v} = 0$. Все точки указанной оси в рассматриваемый момент времени находятся в покое. Это объясняется тем, что все эти точки в результате первого вращения движутся в одну, в результате второго вращения — в противоположную сторону. Результирующая линейная скорость получается равной нулю. Все прочие точки тела вращаются вокруг оси вектора $\boldsymbol{\omega}$ с угловой скоростью ω. Мгновенную линейную скорость любой точки тела можно вычислить по формуле (46.6). Это значит, что *мгновенное результирующее движение твердого тела есть вращение вокруг мгновенной оси OC*. Эта ось, вообще говоря, непрерывно перемещается как относительно самого твердого тела, так и относительно неподвижной системы отсчета, в которой рассматривается движение.

Итак, мы доказали, что *два вращения с угловыми скоростями* $\boldsymbol{\omega}_1$ *и* $\boldsymbol{\omega}_2$ *складываются* (в рассматриваемом физическом смысле) *в одно вращение вокруг мгновенной оси с угловой скоростью* $\boldsymbol{\omega} = \boldsymbol{\omega}_1 + \boldsymbol{\omega}_2$. *Мгновенная ось в каждый момент времени направлена вдоль диагонали параллелограмма, построенного на векторах* $\boldsymbol{\omega}_1$ *и* $\boldsymbol{\omega}_2$. *Сложение подчиняется правилу параллелограмма.* Физическое сложение в указанном смысле оказалось тождественным с математическим.

3. Поясним изложенное наглядным примером. Пусть по поверхности неподвижного конуса *2* катится без скольжения другой круговой конус *1* (рисунки 117 и 118). Вершины обоих конусов все время находятся в одной и той же точке O. В рассматриваемом движении конус *1* вращается вокруг собственной оси OA с некоторой угловой скоростью $\boldsymbol{\omega}_1$. Сама ось OA описывает коническую поверхность, вращаясь вокруг другой оси OB с угловой скоростью $\boldsymbol{\omega}_2$. Речь идет о сложении этих двух вращений. Так как скольжения нет, то все точки тела, лежащие на прямой OC, по которой конусы

касаются друг друга, неподвижны. Касательная OC является поэтому мгновенной осью вращения конуса *1*. Мгновенная ось вращения перемещается в теле, т. е. в конусе *1*, двигаясь по его поверхности. Но она перемещается также и в пространстве, т. е. по поверхности конуса *2*.

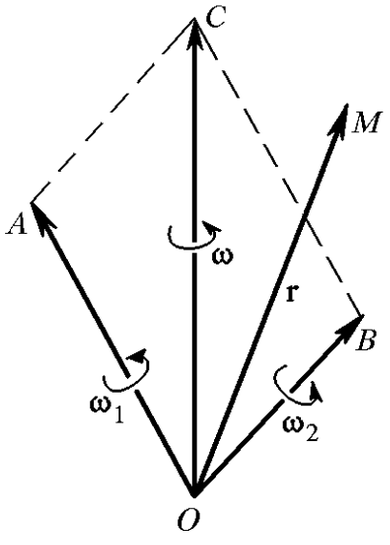


Рис. 116

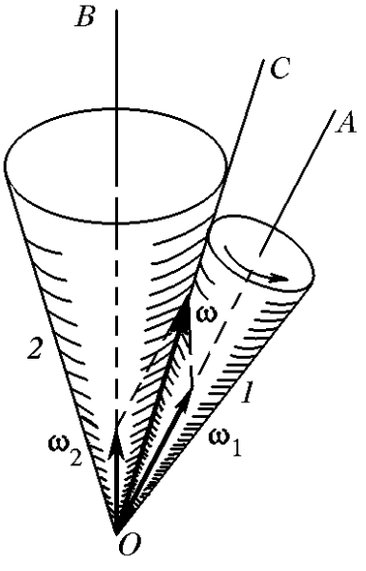


Рис. 117

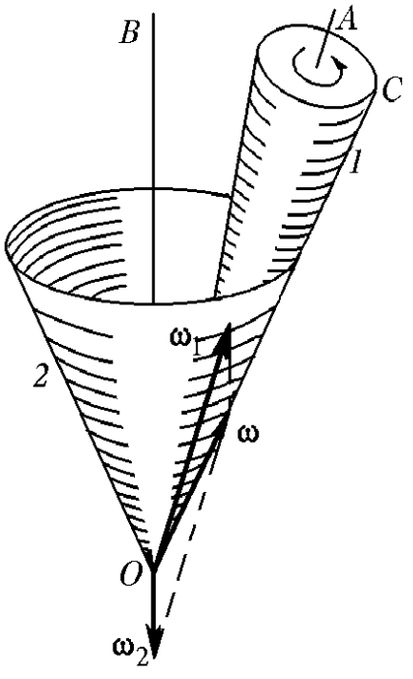
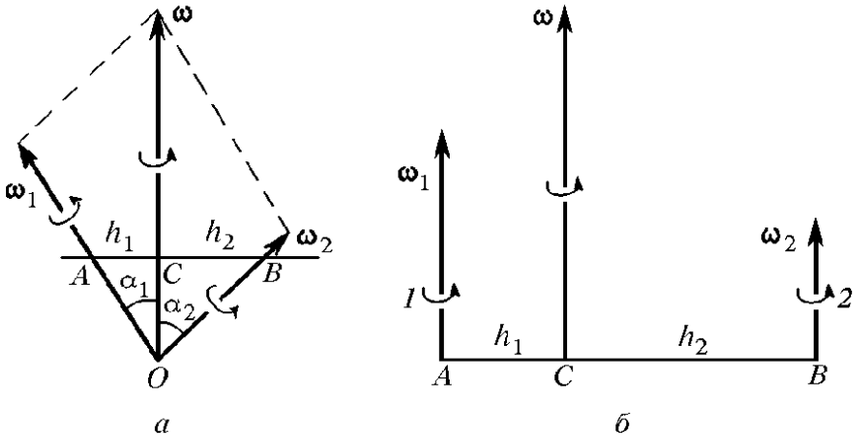


Рис. 118

4. Вращение вокруг параллельных осей можно рассматривать как предельный случай вращений вокруг пересекающихся осей. При сложении таких вращений надо различать два случая: 1) вращения совершаются в одном направлении; 2) вращения совершаются в противоположных направлениях. Рассмотрим первый случай. Построив параллелограмм на векторах $\boldsymbol{\omega}_1$ и $\boldsymbol{\omega}_2$, пересечем его произвольной прямой ACB, перпендикулярной к вектору $\boldsymbol{\omega}$ (рис. 119 *а*). Тогда $h_1 = OC \cdot \operatorname{tg} \alpha_1$, $h_2 = OC \cdot \operatorname{tg} \alpha_2$. Если углы α_1 и α_2 малы, то их тангенсы можно заменить синусами. Сделав это, получим

$$\frac{h_1}{h_2} = \frac{\sin \alpha_1}{\sin \alpha_2} = \frac{\omega_2}{\omega_1}. \qquad (46.7)$$

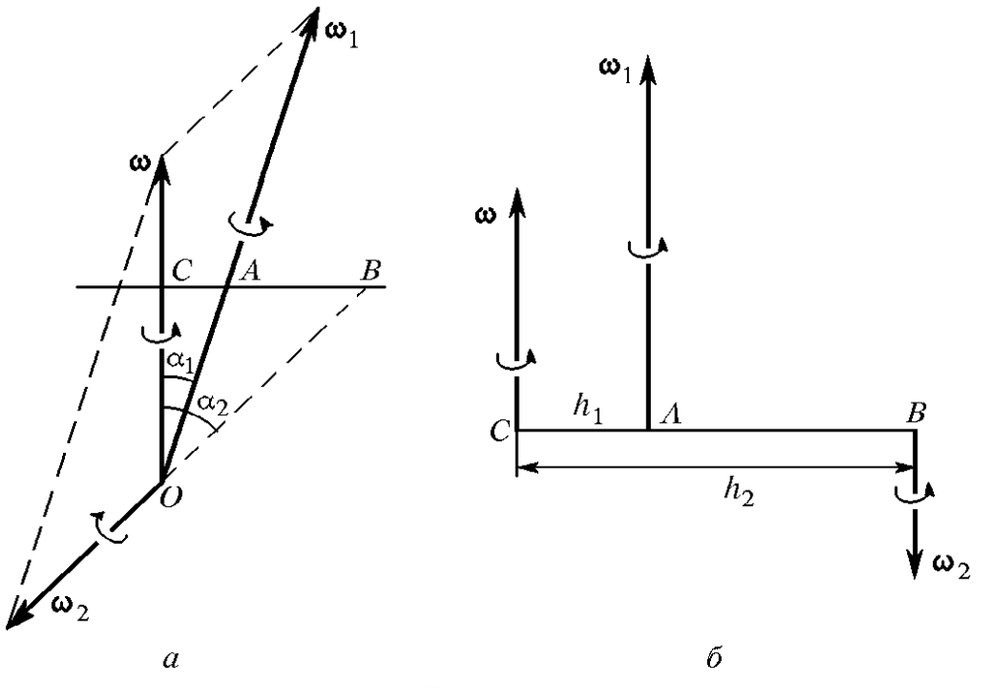


а *б*

Рис. 119

Устремив точку O в бесконечность, получим предельный случай одинаково направленных вращений вокруг параллельных осей (рис. 119 *б*). Такие два вращения складываются в одно вращение вокруг мгновенной оси с угловой скоростью $\omega = \omega_1 + \omega_2$. Мгновенная ось проходит между осями *1* и *2* и делит расстояние между ними обратно пропорционально угловым скоростям ω_1 и ω_2.

Аналогично рассматривается случай, когда векторы $\boldsymbol{\omega}_1$ и $\boldsymbol{\omega}_2$ направлены противоположно. Если $\omega_1 > \omega_2$, то $\omega = \omega_1 - \omega_2$. Мгновенная ось проходит вне отрезка AB со стороны большей угловой ско-



а *б*

Рис. 120

рости (рис. 120). Она делит отрезок AB *внешним образом* на части h_1 и h_2, обратно пропорциональные угловым скоростям ω_1 и ω_2.

5. Рассмотрим, наконец, сложение поступательного и вращательного движений. Если поступательное движение совершается параллельно оси вращения, то при сложении, очевидно, получится *винтовое движение.* Достаточно поэтому ограничиться случаем, когда поступательное движение перпендикулярно к оси вращения. В этом случае все точки тела будут двигаться, параллельно одной и той же плоскости, перпендикулярной к той же оси. Такое движение называется *плоским.* Плоскость, параллельно которой происходит движение, можно принять за плоскость рисунка. Поступательное движение можно рассматривать как вращение вокруг бесконечно удаленной оси. Поэтому разбираемый случай можно свести к сложению двух вращений вокруг параллельных осей, удаляя одну из осей в бесконечность. Ясно, что в результате возникает вращение вокруг какой-то мгновенной оси. Задача сводится к определению положения мгновенной оси и угловой скорости мгновенного вращения. Пусть тело вращается

вокруг оси O с угловой скоростью ω, а сама ось O вращается вокруг параллельной неподвижной оси O_1 с угловой скоростью ω_1 (рис. 121). При сложении возникнет вращение вокруг мгновенной оси A, причем

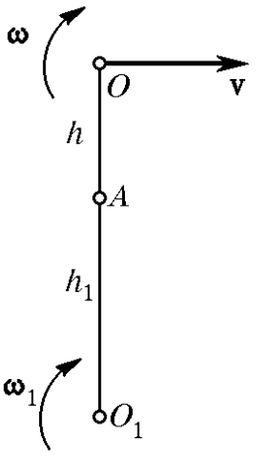


Рис. 121

$$\frac{h}{h_1} = \frac{\omega_1}{\omega}.$$

Вследствие вращения вокруг оси O_1 ось O получает скорость $v = \omega_1(h + h_1)$, перпендикулярную к линии O_1O. Будем удалять O_1 в бесконечность, одновременно уменьшая ω_1 так, чтобы скорость v оставалась неизменной. В пределе вращение оси O вокруг оси O_1 перейдет в поступательное движение со скоростью v. Положение мгновенной оси вращения A определится ее расстоянием до оси O. Это расстояние равно

$$h = \frac{h_1\omega_1}{\omega} = \frac{(h_1+h)\omega_1 - h\omega_1}{\omega} = \frac{v - h\omega_1}{\omega}.$$

Отсюда

$$h\left(1 + \frac{\omega_1}{\omega}\right) = \frac{v}{\omega}.$$

Так как $\omega_1 \to 0$, то в пределе

$$h = \frac{v}{\omega}. \tag{46.8}$$

При этом угловая скорость мгновенного вращения в пределе станет равной ω.

6. Если аксиальный вектор $\boldsymbol{\omega}$ продифференцировать по скалярному аргументу, например по времени t, то в результате получится новый аксиальный вектор $\boldsymbol{\eta} = \frac{d\boldsymbol{\omega}}{dt}$, называемый *угловым ускорением* (см. § 7). Его проекции на координатные оси по определению даются выражениями $\eta_x = \frac{d\omega_x}{dt}$, $\eta_y = \frac{d\omega_y}{dt}$, $\eta_z = \frac{d\omega_z}{dt}$. Аналогично в результате интегрирования $\boldsymbol{\omega}$ по t получается другой аксиальный вектор $\boldsymbol{\varphi} = \int \boldsymbol{\omega}\, dt$ с составляющими $\varphi_x = \int \omega_x\, dt$, $\varphi_y = \int \omega_y\, dt$, $\varphi_z = \int \omega_z\, dt$. Векторный (точнее, псевдовекторный) характер этих величин, как всегда, означает только то, что при повороте (но не инверсии) координатных систем их составляющие преобразуются так же, как разности координат концов направленного геометрического отрезка. Если направление оси вращения не меняется с течением времени, то вектор $\boldsymbol{\varphi}$ направлен параллельно $\boldsymbol{\omega}$, т. е. по оси вращения. Его

длина численно равна углу поворота тела за рассматриваемый промежуток времени. Поэтому $\boldsymbol{\varphi}$ естественно назвать *угловым поворотом тела.* Угловой поворот пропорционален площади сектора OAB, описываемого каким-либо отрезком OA, перпендикулярным к оси вращения, при его переходе из начального положения OA в конечное положение OB (рис. 122). Направление $\boldsymbol{\varphi}$ совпадает с направлением перпендикуляра к плоскости сектора OAB, а его составляющие φ_x, φ_y, φ_z пропорциональны площадям проекций этого сектора на координатные плоскости. Это лишний раз подтверждает векторный характер величины $\boldsymbol{\varphi}$ (см. § 7).

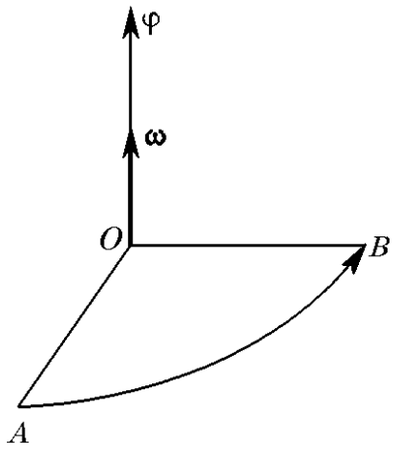


Рис. 122

7. На примере угловых поворотов можно наглядно показать необходимость строгого разграничения между *математическим сложением* векторов (аксиоматически определяемым с помощью правила параллелограмма) и *физическим сложением* их, вводимым с помощью какой-либо физической операции. Введем физическое сложение угловых перемещений в том же смысле, в каком понимается физическое сложение линейных перемещений (см. § 7, п. 6). Пусть материальная точка последовательно совершает вращения вокруг различных осей, проходящих через неподвижную точку O (рис. 123). При таких вращениях она движется вдоль дуг больших кругов по поверхности сферы с центром O. Пусть точка перешла из начального положения A в конечное положение B вдоль дуги большого круга AB. Радиус-вектор точки при этом повернулся на угол φ_1. Затем точка совершила поворот на угол φ_2, перейдя по дуге большого круга BC из положения B в положение C. Каким одним поворотом можно заменить эти два поворота, чтобы перевести точку из того же начального положения A в то же конечное положение C? Ясно, что таким поворотом будет вращение точки по дуге большого круга, проходящей через точки A и C. Обозначим соответствующий угол поворота φ_3. В соответствии со сказанным выше рассматриваемые три поворота можно изобразить векторами $\boldsymbol{\varphi}_1$, $\boldsymbol{\varphi}_2$, $\boldsymbol{\varphi}_3$, перпендикулярными соответственно к плоскостям секторов OAB, OBC и OAC. Поворот $\boldsymbol{\varphi}_3$ можно назвать суммой поворотов $\boldsymbol{\varphi}_1$ и $\boldsymbol{\varphi}_2$ в рассматриваемом физическом смысле. Ясно, что такое сложение не подчиняется правилу параллелограмма. Это

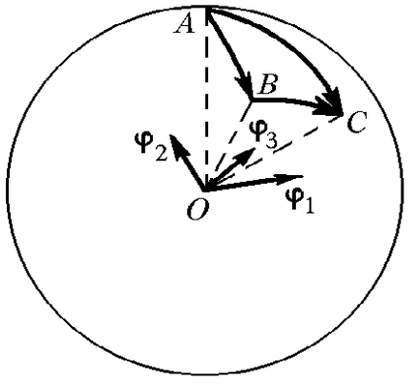


Рис. 123

видно уже из того, что в общем случае вектор $\boldsymbol{\varphi}_3$ не лежит в плоскости векторов $\boldsymbol{\varphi}_1$ и $\boldsymbol{\varphi}_2$.

Особенно очевидным станет это утверждение, если рассмотреть частный случай. За начальное положение материальной точки возьмем полюс A (рис. 124). Затем по дуге меридиана AB совершим первый поворот на угол $\varphi_1 = 90°$, переведя точку в положение B на экваторе. Второй поворот на угол $\varphi_2 = 90°$ совершим по дуге экватора BC. Очевидно, третий поворот φ_3 надо произвести по дуге меридиана AC также на 90°. В рассматриваемом случае все три вектора $\boldsymbol{\varphi}_1$, $\boldsymbol{\varphi}_2$, $\boldsymbol{\varphi}_3$, взаимно перпендикулярны и имеют одну и ту же длину. Ни один из них не может быть геометрической суммой двух других.

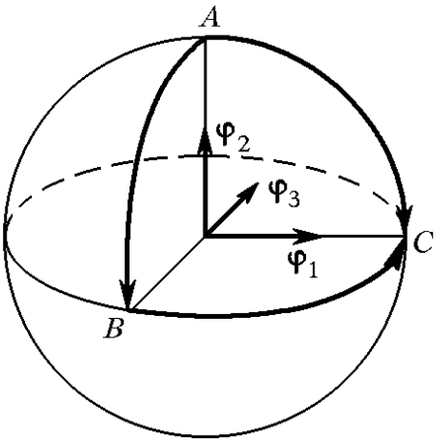


Рис. 124

Если φ_x, φ_y, φ_z означают проекции вектора $\boldsymbol{\varphi}$ на координатные оси, то $\boldsymbol{\varphi} = \varphi_x\mathbf{i} + \varphi_y\mathbf{j} + \varphi_z\mathbf{k}$. Здесь сложение понимается в математическом смысле (по правилу параллелограмма). Однако, как следует из изложенного, слагаемые $\varphi_x\mathbf{i}$, $\varphi_y\mathbf{j}$, $\varphi_z\mathbf{k}$ нельзя рассматривать как последовательно выполняемые повороты вокруг координатных осей, приводящие к единому повороту, представляемому вектором $\boldsymbol{\varphi}$.

8. Допустим, однако, что углы φ_1, φ_2, φ_3 неограниченно стремятся к нулю. Тогда сферический треугольник ABC (см. рис. 123) становится бесконечно малым и может считаться плоским (рис. 125). Дуги больших кругов AB, BC и AC могут рассматриваться как прямолинейные отрезки. Векторы угловых перемещений $\delta\boldsymbol{\varphi}_1$, $\delta\boldsymbol{\varphi}_2$, $\delta\boldsymbol{\varphi}_3$ будут лежать в плоскости треугольника ABC. (Мы пишем $\delta\boldsymbol{\varphi}$ вместо $\boldsymbol{\varphi}$, чтобы подчеркнуть, что речь идет о бесконечно малых углах.) Они, очевидно, перпендикулярны к сторонам AB, BC и AC соответственно, а их длины пропорциональны этим сторонам (рис. 125). Отсюда следует, что бесконечно малый вектор $\delta\boldsymbol{\varphi}_3$ является геометрической суммой векторов $\delta\boldsymbol{\varphi}_1$ и $\delta\boldsymbol{\varphi}_2$. Это значит, что *бесконечно малые угловые перемещения складываются геометрически* (в указанном выше физическом смысле), *т. е. по прави-*

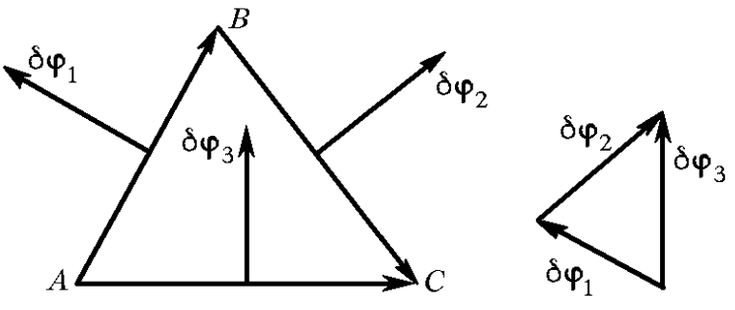


Рис. 125

лу параллелограмма. Иными словами, такое *физическое сложение угловых перемещений в пределе бесконечно малых углов поворота переходит в математическое.*

ЗАДАЧИ

1. Показать, что элементарная работа, совершаемая над системой материальных точек при ее повороте на бесконечно малый угол $\delta\boldsymbol{\varphi}$, выражается скалярным произведением

$$\delta A = (\mathbf{M}\,\delta\boldsymbol{\varphi}),$$

где $\mathbf{M}$ — геометрическая сумма моментов сил, действующих на материальные точки системы, относительно вершины угла поворота.

Решение. $\delta A = \sum (\mathbf{F}_i\,\delta\mathbf{r}_i)$. Здесь суммирование ведется по всем точкам системы. При повороте $\delta\mathbf{r}_i = [\delta\boldsymbol{\varphi}\,\mathbf{r}_i]$, причем угол $\delta\boldsymbol{\varphi}$ один и тот же для всей системы. Подставив это выражение в предыдущую формулу и заметив, что $\mathbf{F}_i[\delta\boldsymbol{\varphi}\mathbf{r}_i] = \delta\boldsymbol{\varphi}[\mathbf{r}_i\mathbf{F}_i] = (\mathbf{M}_i\,\delta\boldsymbol{\varphi})$, получим требуемый результат.

2. Используя изотропию пространства, доказать, что геометрическая сумма моментов внутренних сил, действующих в системе материальных точек, равна нулю (см. § 38).

Решение. Допустим, что система замкнута. Пусть $\mathbf{M}_1$, $\mathbf{M}_2$, ... — моменты внутренних сил, действующие на материальные точки системы, относительно произвольного неподвижного начала O. Повернем всю систему вокруг точки O на произвольный бесконечно малый угол $\delta\boldsymbol{\varphi}$ и притом так, чтобы скорости всех материальных точек повернулись на тот же угол без изменения своего модуля. Ввиду изотропии пространства на такой поворот не требуется затраты работы. Но эта работа представляется скалярным произведением $(\mathbf{M}_1 + \mathbf{M}_2 + \ldots)\,\delta\boldsymbol{\varphi}$. Значит, это скалярное произведение равно нулю, каков бы ни был поворот $\delta\boldsymbol{\varphi}$. Отсюда следует, что для замкнутой системы $\mathbf{M}_1 + \mathbf{M}_2 + \ldots = 0$.

3. Пусть вектор $\mathbf{A}$ неизменной длины вращается вокруг своего начала с угловой скорость $\boldsymbol{\omega}$. Показать, что его производная по времени определяется формулой

$$\dot{\mathbf{A}} = [\boldsymbol{\omega}\,\mathbf{A}]. \tag{46.10}$$

В частности, при вращении координатной системы орты $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ дифференцируются по формулам

$$\frac{d\mathbf{i}}{dt} = [\boldsymbol{\omega}\mathbf{i}],\quad \frac{d\mathbf{j}}{dt} = [\boldsymbol{\omega}\mathbf{j}],\quad \frac{d\mathbf{k}}{dt} = [\boldsymbol{\omega}\mathbf{k}]. \tag{46.11}$$

Решение. Вектор $\mathbf{A}$ неизменной длины можно отождествить с абсолютно твердым тонким стержнем той же длины. Если начало вектора $\mathbf{A}$ неподвижно, то производная $\dot{\mathbf{A}}$ имеет смысл скорости движущегося конца стержня. При такой интерпретации формула (46.10) становится частным случаем формулы (46.4).

4. Движение точки на плоскости можно задать полярными координатами r и φ (рис. 126). Найти выражения для скорости и ускорения точки в этой системе координат.

Р е ш е н и е. Введем единичные векторы **i**, **j**, **k**. Вектор **i** направим вдоль радиуса r. Вектор **j** перпендикулярен к нему и направлен в сторону возрастания угла φ. Вектор **k** (не изображенный на рисунке) перпендикулярен к плоскости рисунка и образует с векторами **i** и **j** правовинтовую систему. При движении точки векторы **i** и **j** вращаются вокруг начала координат с угловой скоростью $\omega = \dot{\varphi}$. Вектор угловой скорости направлен вдоль **k**, так что $\boldsymbol{\omega} = \dot{\varphi}\mathbf{k}$. Применяя формулы (46.11), находим производные векторов **i** и **j**:

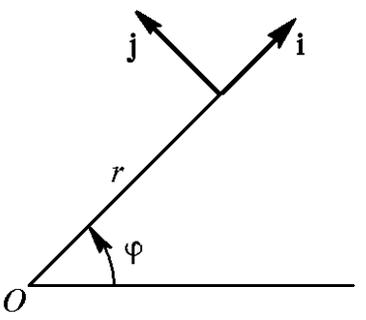


Рис. 126

$$\frac{d\mathbf{i}}{dt} = \dot{\varphi}[\mathbf{ki}] = \dot{\varphi}\mathbf{j}, \quad \frac{d\mathbf{j}}{dt} = \dot{\varphi}[\mathbf{kj}] = -\dot{\varphi}\mathbf{i}. \tag{46.12}$$

Представим радиус-вектор движущейся точки в виде $\mathbf{r} = r\mathbf{i}$. Дифференцируя его один раз, находим скорость:

$$\mathbf{v} = \dot{\mathbf{r}} = \dot{r}\mathbf{i} + r\frac{d\mathbf{i}}{dt} = \dot{r}\mathbf{i} + r\dot{\varphi}\mathbf{j}.$$

Дифференцируя вторично, находим ускорение:

$$\mathbf{a} = \dot{\mathbf{v}} = \ddot{r}\mathbf{i} + r\frac{d\mathbf{i}}{dt} + \dot{r}\dot{\varphi}\mathbf{j} + r\ddot{\varphi}\mathbf{j} + r\dot{\varphi}\frac{d\mathbf{j}}{dt} = (\ddot{r} - \dot{\varphi}^2 r)\mathbf{i} + (2\dot{r}\dot{\varphi} + r\ddot{\varphi})\mathbf{j}.$$

Эти формулы дают разложение скорости и ускорения на радиальные (направленные вдоль радиуса) и азимутальные (направленные по **j**, т. е. в сторону возрастания угла φ) составляющие:

$$v_r = \dot{r}, \quad v_\varphi = r\dot{\varphi}, \tag{46.13}$$

$$a_r = \ddot{r} - \dot{\varphi}^2 r, \quad a_\varphi = 2\dot{r}\dot{\varphi} + r\ddot{\varphi}. \tag{46.14}$$

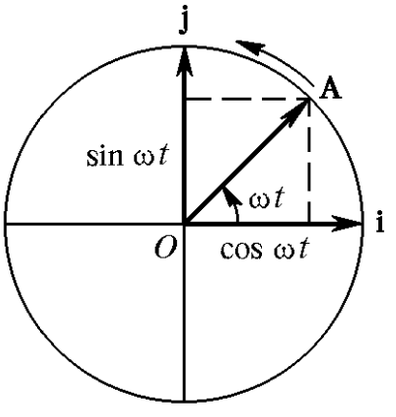


Рис. 127

5. С помощью соотношения (46.10) получить формулы для дифференцирования синуса и косинуса.

Р е ш е н и е. Рассмотрим единичный вектор **A**, равномерно вращающийся вокруг начала координат O (рис. 127). Если координатные оси неподвижны, то

$$\mathbf{A} = \mathbf{i}\cos\omega t + \mathbf{j}\sin\omega t.$$

Производная этого вектора по t равны

$$\dot{\mathbf{A}} = \mathbf{i}\frac{d}{dt}(\cos\omega t) + \mathbf{j}\frac{d}{dt}(\sin\omega t).$$

С другой стороны, ту же производную можно вычислить по формуле (46.10). Так как $\boldsymbol{\omega} = \omega\mathbf{k}$, то эта формула дает

$$\dot{\mathbf{A}} = \omega\,[\mathbf{kA}] = \omega\cos\omega t\,[\mathbf{ki}] + \omega\sin\omega t\,[\mathbf{kj}] = \mathbf{j}\,\omega\cos\omega t - \mathbf{i}\,\omega\sin\omega t.$$

Сравнивая оба результата, получим

$$\frac{d}{dt}(\sin\omega t) = \omega\cos\omega t, \quad \frac{d}{dt}(\cos\omega t) = -\omega\sin\omega t.$$

Можно сказать, что векторная формула (46.10) эквивалентна правилам дифференцирования синуса и косинуса.

§ 47. ТЕОРЕМА ЭЙЛЕРА. ОБЩЕЕ ДВИЖЕНИЕ ТВЕРДОГО ТЕЛА

1. Рассмотрим *плоское движение* твердого тела, т. е. такое движение, когда все точки тела движутся параллельно одной плоскости. Не теряя общности, можно считать само тело плоским, а движение — происходящим в плоскости тела. Положение плоского тела однозначно определяется заданием положений каких-либо двух точек его. Поэтому достаточно ограничиться рассмотрением движения какой-либо одной прямой плоского тела. Пусть выбранная прямая твердого тела перешла из положения AB в положение A_1B_1 (рис. 128). Соединим точку A с точкой A_1, а точку B с точкой B_1. Из середин отрезков AA_1 и BB_1 восстановим перпендикуляры EO и DO, пересекающиеся в точке O. Докажем, что прямую AB можно перевести в положение A_1B_1 путем одного поворота вокруг точки O. Действительно, из построения следует, что точка O равноудалена от точек A и A_1, а также от точек B и B_1. В силу этого прямую AB можно повернуть вокруг точки O так, чтобы точка A совместилась с точкой A_1. Докажем, что при этом точка B также совместится с точкой B_1. Для доказательства допустим, что точка B не совместилась с точкой B_1, а заняла положение B_2. Разумеется, точка B_2 будет находиться на таком же расстоянии от O, что и точка B, а потому $OB_1 = OB_2$. Кроме того, в треугольниках OA_1B_1 и OA_1B_2 сторона OA_1 — общая, а стороны A_1B_1 и A_1B_2 равны, так как тело твердое, а потому расстояние между концами отрезка AB не меняется при его движении. Следовательно, треугольники OA_1B_1 и OA_1B_2 равны. Отсюда делаем вывод, что $\angle OA_1B_1 = \angle OA_1B_2$, так что точка B_2 должна совпадать с точкой B_1.

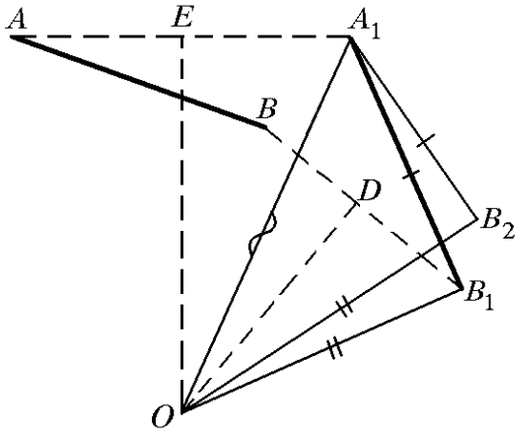


Рис. 128

Таким образом, *при плоском движении твердое тело может быть переведено из любого положения в другое произвольное положение с помощью одного поворота вокруг некоторой оси.* Это положение является частным случаем *теоремы Эйлера* (по имени Леонарда Эйлера (1707—1783)), доказываемой ниже.

Произвольное плоское движение тела можно разбить на ряд следующих друг за другом бесконечно малых перемещений. В результате получится ряд бесконечно близких положений *1*, *2*, *3*, *4*, ..., последовательно проходимых телом. Согласно доказанной теореме

переход тела из положения *1* в положение *2* может быть осуществлен поворотом вокруг некоторой оси O_1; переход из положения *2* в положение *3* — поворотом вокруг другой бесконечно близкой оси O_2 и т. д. Если число промежуточных положений *1*, *2*, *3*, ... стремить к бесконечности, а смещение тела из каждого положения в соседнее — к нулю, то *произвольное плоское движение твердого тела может рассматриваться как вращение вокруг мгновенной оси, движущейся как в теле, так и в пространстве.*

2. Совершенно аналогично формируется *теорема Эйлера.* Согласно теореме Эйлера *твердое тело, имеющее одну неподвижную точку, может быть переведено из произвольного положения в другое произвольное положение путем поворота вокруг некоторой оси, проходящей через эту неподвижную точку.* Доказательство теоремы Эйлера проводится совершенно так же, как и соответствующей теоремы для плоского движения. Если одна из точек твердого тела C неподвижна, то его положение однозначно определяется заданием положений каких-либо двух точек A и B, не лежащих на одной прямой с точкой C. В качестве точек A и B можно взять две точки на поверхности сферы с центром в точке C. Проведем через центр сферы C и точки A и B плоскость. Она пересечет сферу по дуге большого круга AB (см. рис. 128; мы не рисуем отдельно соответствующую сферу и дуги больших кругов, а пользуемся прежним плоским рисунком, мысленно заменяя, где это нужно, прямолинейные отрезки дугами больших кругов; понятно, что центр сферы C на плоском рисунке изобразить нельзя.) Движение дуги AB по поверхности сферы однозначно определяет и движение всего твердого тела. Пусть выбранная дуга перешла из положения AB в положение A_1B_1. Соединим дугами больших кругов точку A с точкой A_1, а точку B с точкой B_1. Через середины этих дуг E и D проведем перпендикулярные к ним дуги больших кругов EO и DO, пересекающиеся в точке O сферы. Точку O соединим с центром сферы C прямой OC. Докажем, что дуга AB может быть переведена в положение A_1B_1 путем поворота вокруг оси CO. Действительно, по построению точки A и A_1, а также точки B и B_1 равноудалены от точки O. Ввиду этого твердое тело можно повернуть вокруг оси CO так, чтобы точка A перешла в положение A_1. Докажем, что при таком повороте точка B также перейдет в положение B_1. Для доказательства допустим, что точка B при повороте перешла не в положение B_1, а в положение B_2. Проведем дуги больших кругов OA_1, A_1B_2 и OB_2. Так как точка B_2 находится на том же расстоянии от O, что и точка B, то $\smile OB_1 = \smile OB_2$. Кроме того, в сферических треугольниках OA_1B_1 и OA_1B_2 дуга OA_1 общая, а дуги

A_1B_1 и A_1B_2 равны, так как тело твердое, и, следовательно, при его движении длина дуги AB не изменяется. Поэтому сферические треугольники OA_1B_1 и OA_1B_2 равны. Значит, $\angle OA_1B_1 = \angle OA_1B_2$, а потому точка B_2 должна совпадать с точкой B_1. Тем самым теорема Эйлера доказана.

Доказанная в начале этого параграфа теорема является частным случаем теоремы Эйлера, так как плоское движение плоского тела может рассматриваться как предельный случай движения по сферической поверхности бесконечно большого радиуса.

Рассуждая так же, как и в случае плоского движения, из теоремы Эйлера можно вывести следующее следствие. *Любое движение твердого тела, имеющего одну неподвижную точку, можно рассматривать как вращение вокруг мгновенной оси, проходящей через эту неподвижную точку. С течением времени мгновенная ось, вообще говоря, непрерывно перемещается как в теле, так и в пространстве.*

3. Рассмотрим теперь самый общий случай движения твердого тела. Выберем в теле произвольную точку O. Всякое движение тела можно разложить на поступательное со скоростью $\mathbf{v}_0$, равной скорости точки O, и вращательное вокруг мгновенной оси, проходящей через эту точку. Обозначая через $\boldsymbol{\omega}$ вектор угловой скорости мгновенного вращения, можем написать для скорости другой произвольной точки A твердого тела

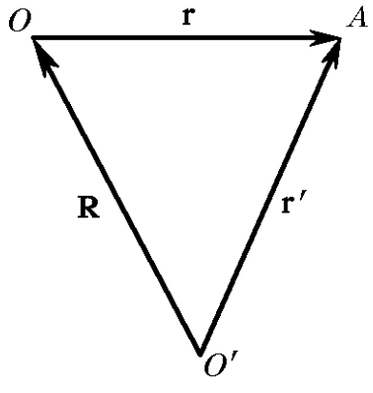


Рис. 129

$$\mathbf{v} = \mathbf{v}_O + [\boldsymbol{\omega}\mathbf{r}], \qquad (47.1)$$

где $\mathbf{r}$ — радиус-вектор, проведенный из точки O в точку A (рис. 129). Скорость поступательного движения $\mathbf{v}_O$, конечно, зависит от выбора точки O. Но *угловая скорость* $\boldsymbol{\omega}$ *не зависит от положения точки* O, *к которой отнесено вращение твердого тела.* Поэтому *можно говорить об угловой скорости вращения твердого тела, не указывая эту точку.* Докажем это.

Выберем другую произвольную точку тела O' и отнесем к ней вращение твердого тела. Соответствующую угловую скорость вращения обозначим через $\boldsymbol{\omega}$. Тогда скорость $\mathbf{v}$ прежней точки A можно представить в другом виде:

$$\mathbf{v} = \mathbf{v}_{O'} + [\boldsymbol{\omega}'r'],$$

где $\mathbf{r}'$ — радиус-вектор, проведенный из O' в A. Так как речь идет о скорости одной и той же точки, то эта величина должна совпадать с (47.1). Это дает

$$\mathbf{v}_O + [\boldsymbol{\omega}\mathbf{r}] = \mathbf{v}_{O'} + [\boldsymbol{\omega}'\mathbf{r}'].$$

для одной материальной точки, но и для всего тела, поскольку последнее можно рассматривать как систему материальных точек, вращающихся вокруг общей оси. Таким образом, формула (47.2) эквивалентна формуле (33.6), полученной ранее иным путем.

§ 48. СКАТЫВАНИЕ ТЕЛ С НАКЛОННОЙ ПЛОСКОСТИ

1. Пусть скатывающееся тело обладает *симметрией вращения* относительно геометрической оси C (рис. 130). Будем предполагать, что при движении не возникает скольжения. Это означает, что скорость тела в точке касания A равна нулю. Отсутствие скольжения обеспечивается действием сил со стороны наклонной плоскости на скатывающееся тело. Эти силы сводятся к силе нормального давления $\mathbf{F}_n$ и к касательной силе трения $\mathbf{F}_\tau$. При отсутствии скольжения сила $\mathbf{F}_\tau$ есть сила трения покоя или сила трения сцепления.

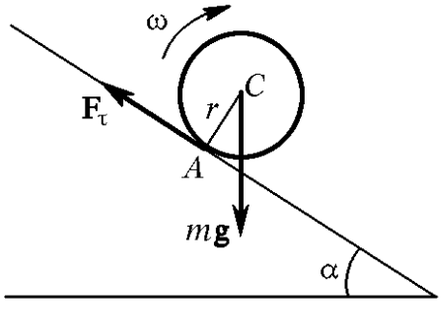


Рис. 130

Модуль силы $\mathbf{F}_\tau$ может принимать любое значение от 0 до kF_n, где k — коэффициент трения (см. § 17). При качении сила устанавливается как раз такой, чтобы не было скольжения. Если касательная сила, требующаяся для этого, превышает kF_n, то чистое качение невозможно — оно будет сопровождаться скольжением.

Решим задачу о скатывании тела тремя различными способами.

Способ 1. Применим уравнение моментов относительно мгновенной оси вращения. При отсутствии скольжения мгновенная ось проходит через точку касания A. Так как мгновенная ось и ось, проходящая через центр масс C, движутся параллельно друг другу, то уравнение моментов имеет обычную простую форму

$$I_A \frac{d\omega}{dt} = M_A, \tag{48.1}$$

где I_A — момент инерции скатывающегося тела относительно мгновенной оси, а M_A — момент внешних сил относительно той же оси. Внешними силами являются сила тяжести $m\mathbf{g}$ и реакция опоры, действующая со стороны наклонной плоскости на скатывающееся тело. Сила реакции опоры выпадает из уравнения моментов, так как она проходит через ось A, и ее момент относительно этой оси равен нулю. Таким образом,

$$I_A \frac{d\omega}{dt} = mgr \sin \alpha.$$

Обозначим через v линейную скорость точки C. Она связана со скоростью v_A точки A тела соотношением $v = v_A + \omega r$. При отсутствии скольжения $v_A = 0$, а потому $v = \omega r$. Для линейного ускорения точки C получаем $a = \frac{dv}{dt} = r\frac{d\omega}{dt}$. Поэтому предыдущее уравнение дает

$$a = \frac{mgr^2}{I_A} \sin \alpha. \tag{48.2}$$

По теореме Гюйгенса—Штейнера $I_A = I_C + mr^2$, где I_C — момент инерции тела относительно оси, проходящей через центр масс C. Следовательно,

$$a = \frac{g \sin \alpha}{1 + I_C/mr^2}. \tag{48.3}$$

Преимущество рассматриваемого способа состоит в том, что в исходное уравнение (48.1) совсем не входит неизвестная реакция опоры.

Способ 2. Применим уравнение моментов относительно оси, проходящей через центр масс C. Оно также имеет простой вид

$$I_C \frac{d\omega}{dt} = M_C,$$

где M_C — момент внешних сил относительно оси C. В это уравнение не войдет сила тяжести, так как она проходит через ось C. Момент создается силой реакции опоры. При этом играет роль только слагаемая F_τ этой силы, параллельная наклонной плоскости, т. е. сила трения сцепления. Ее момент $M_C = rF_\tau$, а потому

$$I_C \frac{d\omega}{dt} = rF_\tau.$$

Это уравнение содержит два неизвестных: угловое ускорение $\frac{d\omega}{dt}$ и силу F_τ. Недостающее уравнение дает теорема о движении центра масс:

$$m\frac{dv}{dt} = mg \sin \alpha - F_\tau. \tag{48.4}$$

Присоединив сюда прежнее соотношение $a = \frac{dv}{dt} = r\frac{d\omega}{dt}$ и разрешив полученные уравнения относительно a, найдем прежний результат (48.3). Кроме того, получаем следующее выражение для силы трения сцепления:

$$F_\tau = \frac{I_C}{I_C + mr^2} mg \sin \alpha. \tag{48.5}$$

Способ 3. Применим закон сохранения энергии. Кинетическая энергия тела равна $K = \frac{1}{2} I_A \omega^2$. Поэтому $\frac{1}{2} I_A \omega^2 = mgh$, где h — высота, с которой опустилось тело при скатывании из состояния покоя. Если оно прошло вдоль наклонной плоскости путь x, то $h = x \sin \alpha$, и, следовательно,

$$\frac{1}{2} I_A \omega^2 = \frac{I_A}{2r^2} v^2 = mgx \sin \alpha.$$

Дифференцируя это соотношение по времени и замечая, что $\frac{dx}{dt} = v$, снова получим формулу (48.2).

2. Так как на скатывающееся тело действует сила трения, то может возникнуть вопрос, почему в рассматриваемой задаче можно применять закон сохранения энергии в его механической форме. Ответ заключается в том, что *при отсутствии скольжения сила трения приложена к тем точкам тела, которые лежат на мгновенной оси вращения.* Мгновенная скорость таких точек равна нулю, а потому приложенная к ним сила трения сцепления *работы не производит* и не влияет на полную кинетическую энергию скатывающегося тела. Роль силы трения сцепления $\mathbf{F}_\tau$ сводится к тому, чтобы привести тело во вращение и обеспечить чистое качение. При наличии силы трения сцепления работа силы тяжести идет на увеличение кинетической энергии не только поступательного, но и вращательного движения тела.

3. Комбинация I_C/m, входящая в формулу (48.3), имеет размерность квадрата длины. Введем для нее обозначение

$$\rho^2 = \frac{I_C}{m}$$

и назовем ρ *радиусом инерции* тела. Формула (48.3) принимает вид

$$a = \frac{g \sin \alpha}{1 + (\rho/r)^2}. \tag{48.6}$$

Величину r можно назвать *радиусом качения* тела. Радиус качения есть расстояние между центром масс скатывающегося тела и мгновенной осью вращения. Для цилиндра или шара радиус качения равен геометрическому радиусу этих тел.

Ускорение скатывающегося тела и приобретенная им скорость поступательного движения зависят от отношения радиуса инерции к радиусу качения. Чем больше это отношение, тем медленнее скатывается тело. Особенно просто этот результат можно уяснить с помощью закона сохранения энергии. Если тело скатывается с высоты h, то вся его потенциальная энергия mgh переходит в кинетическую. Последняя складывается из кинетических энергий поступательного и вращательного движений. Полная кинетическая энергия

тела в нижнем положении равна mgh, т. е. зависит только от высоты h. Чем большая доля кинетической энергии приходится на вращение тела, тем медленнее оно скатывается с наклонной плоскости. Отношение кинетической энергии вращательного движения к кинетической энергии поступательного движения равно

$$\frac{E_{\text{вр}}}{E_{\text{пост}}} = \frac{I_C\omega^2}{mv^2} = \left(\frac{\rho}{r}\right)^2.$$

Максимальное значение для ускорения a получается в случае чистого скольжения при отсутствии сил трения.

Пользуясь выражениями для моментов инерции, полученными в § 36, легко найти соответствующие радиусы инерции, а затем вычислить ускорение a. Таким путем получим следующие результаты:

полый цилиндр (без тормозов): $\rho^2 = r^2$, $a = \frac{1}{2}\, g \sin\alpha$;

сплошной цилиндр: $\rho^2 = \frac{1}{2}\, r^2$, $a = \frac{2}{3}\, g \sin\alpha$;

полый шар: $\rho^2 = \frac{2}{3}\, r^2$, $a = \frac{3}{5}\, g \sin\alpha$;

сплошной шар: $\rho^2 = \frac{2}{5}\, r^2$, $a = \frac{5}{7}\, g \sin\alpha$.

Полые тела скатываются медленнее, чем сплошные тела той же геометрической формы. При одинаковых массах моменты инерции полых тел больше, чем сплошных. Поэтому на долю вращательного движения у полых тел приходится относительно большая кинетическая энергия, чем у сплошных.

Возьмем маховичок, насаженный на ось. Положим ось на наклонные рельсы, чтобы маховичок находился между ними (рис. 131). Радиус качения в этом случае совпадает с радиусом оси маховичка r. Отношение ρ/r здесь велико, и маховичок будет скатываться очень медленно.

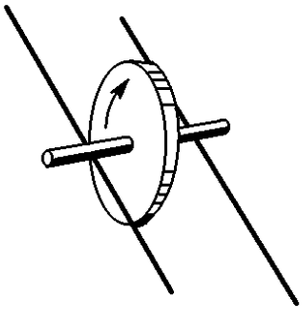


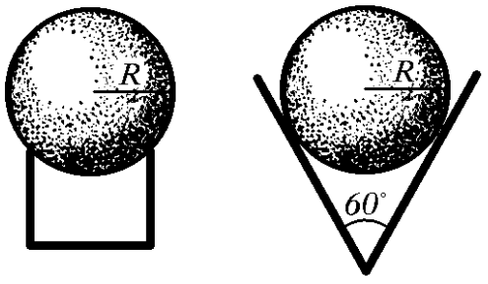
Рис. 131

4. Когда угол наклона α равен нулю, ускорение a обращается в нуль. Вместе с ним обращается в нуль и сила трения сцепления F_τ, как это видно из формулы (48.5). Таким образом, твердое тело, обладающее осевой симметрией, например цилиндр или шар, при отсутствии скольжения катится по твердой горизонтальной плоскости прямолинейно и равномерно, совсем не испытывая силы сопротивления. Этот результат относится к идеализированным моделям тел. Тело и плоскость, по которой оно катится, должны быть идеально твердыми и гладкими. Для реальных тел он не справедлив или справедлив только приближенно. В этом случае тело и плоскость деформируются. На плоскости возникает углубление, тело соприкасается с ней не в одной геометрической точке, а на некотором участке конечной площади. В результате при качении по горизон-

тальной плоскости возникает сила, замедляющая движение. Это есть сила *трения качения*. Она обычно мала по сравнению с силой трения скольжения, и во многих случаях ею можно пренебречь (см. § 17).

ЗАДАЧИ

1. Определить ускорение a центра шарика, скатывающегося без скольжения по наклонному желобу, образующему угол α с горизонтом. Форма поперечного сечения желоба изображена на рис. 132 *а* и *б*.



а *б*

Рис. 132

Ответ. а) $a = \dfrac{R^2-h^2}{\rho^2+(R^2-h^2)}\, g \sin\alpha$, где ρ — радиус инерции шарика, $2h$ — ширина желоба; б) $a = \dfrac{R^2}{4\rho^2+R^2}\, g \sin\alpha$.

2. С какой высоты H должен скатиться шарик с радиусом инерции ρ, для того чтобы он смог без скольжения описать мертвую петлю по желобу радиуса R? Радиусом шарика r по сравнению с R пренебречь.

Ответ. $H = \dfrac{5r^2+\rho^2}{2r^2} R$. Для сплошного шара $H = {}^{27}/_{10}\, R$, для полого $H = {}^{17}/_{6}\, R$.

3. Цилиндр или шар радиусом r катится по плоскости, наклоненной под углом α к горизонту. Определить, при каком значении угла α начнется качение со скольжением, если коэффициент трения скольжения между катящимся телом и плоскостью равен k.

Ответ. $\operatorname{tg}\alpha > \dfrac{r^2+\rho^2}{\rho^2} k$, где ρ — радиус инерции катящегося тела. Для сплошного шара $\operatorname{tg}\alpha > {}^{7}/_{2}\, k$, для полого $\operatorname{tg}\alpha > {}^{5}/_{2}\, k$. Для сплошного цилиндра $\operatorname{tg}\alpha > 3k$, для полого $\operatorname{tg}\alpha > 2k$.

4. Шарик радиусом r скатывается без начальной скорости и без скольжения по поверхности сферы из самого верхнего положения A (рис. 133). Определить точку, в которой он оторвется от сферы и начнет свободно двигаться под действием силы тяжести.

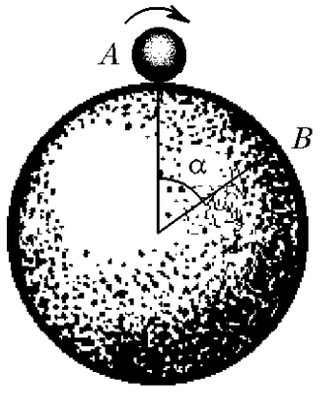


Рис. 133

Ответ. Положение точки B, в которой шарик отрывается от сферы и начинает свободно двигаться под действием силы тяжести, определяется углом α, косинус которого равен

$$\cos\alpha = \frac{2r^2}{3r^2+\rho^2},$$

где ρ — радиус инерции шарика. Результат не зависит от радиуса сферы. Для сплошного шарика $\cos\alpha = {}^{10}/_{17}$ ($\alpha = 54°$), для полого $\cos\alpha = {}^{6}/_{11}$ ($\alpha = 57°$).

5. Цилиндр массой M и радиусом r катится по горизонтальной поверхности стола (рис. 134). Обвитая вокруг цилиндра нить горизонтально проходит через неподвижный блок, к другому концу нити подвешен груз

массой m. Пренебрегая массами блока и нити, найти ускорение центра масс цилиндра.

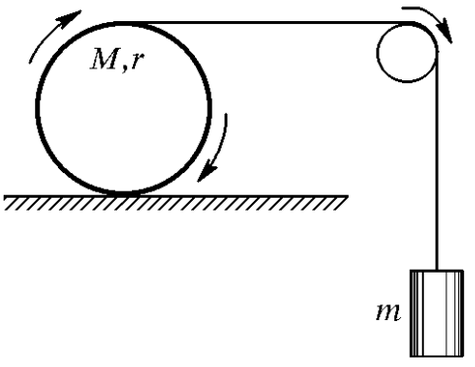


Рис. 134

Ответ. $a = \dfrac{2mr^2}{M(\rho^2+r^2)+4mr^2}\,g$, где ρ — радиус инерции цилиндра. Для сплошного цилиндра $a = \dfrac{4m}{3M+8m}\,g$, для полого $a = \dfrac{m}{M+2m}\,g$.

6. По наклонной плоскости, образующей угол α с горизонтом, скатывается массивный полый цилиндр массой M и радиусом r. По поверхности цилиндра бежит собака таким образом, что она все время занимает наивысшее положение на поверхности цилиндра. Определить, с каким ускорением a скатывается цилиндр, если масса собаки m.

Решение. Метод решения этой задачи поучителен. Для решения проще всего воспользоваться уравнением моментов относительно мгновенной оси вращения A (рис. 135). При этом все движения должны рассматриваться относительно системы отсчета, в которой наклонная плоскость неподвижна. В этой системе собака, все время находится в наивысшей точке цилиндра S, движется параллельно наклонной плоскости и притом с той же скоростью v, с какой движется центр цилиндра. Момент импульса системы L слагается из момента импульса цилиндра $I\omega$ и момента импульса собаки mvh, где $h = r(1+\cos\alpha)$ — длина перпендикуляра, опущенного на наклонную плоскость из точки S. Итак,

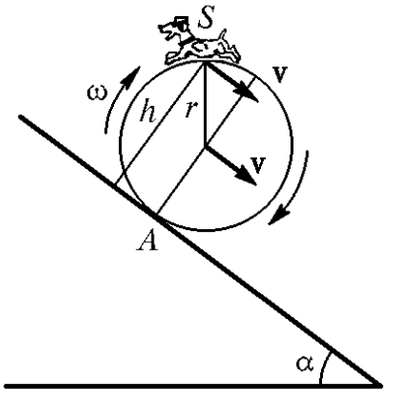


Рис. 135

$$L = I\omega + mrv(1+\cos\alpha),$$

причем под I следует понимать момент инерции цилиндра относительно мгновенной оси, т. е. величину $2Mr^2$. Из-за отсутствия скольжения $v = \omega r$, а потому

$$L = [2M + m(1+\cos\alpha)]rv.$$

Так как центр масс системы и мгновенная ось A движутся параллельно, то производная L по времени должна равняться моменту внешних сил относительно мгновенной оси A, т. е. $(M+m)gr\sin\alpha$. Приравнивая оба выражения, получим

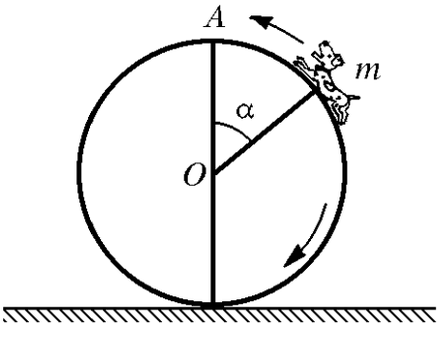


Рис. 136

$$a = \frac{M+m}{2M+m(1+\cos\alpha)}\,g\sin\alpha.$$

7. По поверхности большого полого цилиндра, лежащего на горизонтальной плоскости, начинает бежать собака массой m в направлении к наивысшей точке A и притом так, что она все время находится на одном и том же расстоянии от этой точки (рис. 136). В результате цилиндр начинает ка-

титься по горизонтальной плоскости без скольжения. Масса цилиндра M, а угол AOm равен α. Определить: 1) ускорение оси цилиндра a; 2) силу трения между цилиндром и плоскостью во время качения $F_{тр}$; 3) время t, в течение которого собака способна оставаться на указанном расстоянии от точки A, если максимальная полезная мощность, которую она способна развить, равна $P_{макс}$. Какая при этом будет достигнута максимальная скорость $v_{макс}$ поступательного движения цилиндра? (Полезной мощностью здесь называется мощность, которая затрачивается собакой на увеличение кинетической энергии системы.)

Ответ. $a = \dfrac{mg\sin\alpha}{2M+m(1+\cos\alpha)}$, $F_{тр} = (M+m)a$, $t = \dfrac{P_{макс}}{2M+m}\dfrac{1}{a^2}$, $v_{макс} = \dfrac{P_{макс}}{(2M+m)a}$.

8. Определить ускорение a, с которым цилиндрическая бочка, целиком заполненная жидкостью, скатывается без скольжения с наклонной плоскости, образующей угол α с горизонтом (рис. 137). Трение между жидкостью и стенками бочки считать пренебрежимо малым.

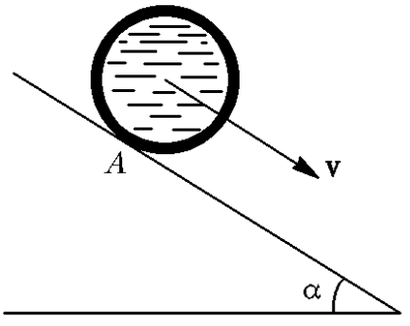


Рис. 137

Решение. При отсутствии трения между жидкостью и стенками бочки вращение бочки не передается жидкости. Жидкость движется поступательно как целое со скоростью v, равной скорости движения центра масс. Момент импульса системы относительно мгновенной оси A равен $L = I_A\omega + mRv$, где R — внешний радиус бочки, I_A — момент инерции ее относительно мгновенной оси A, m — масса жидкости. Из-за отсутствия скольжения $v = \omega R$, так что

$$L = \left(\frac{I_A}{R} + mR\right)v.$$

Центр масс бочки движется параллельно мгновенной оси, а потому

$$\frac{dL}{dt} = \left(\frac{I_A}{R} + mR\right)\frac{dv}{dt} = (M+m)Rg\sin\alpha,$$

где M — масса бочки. Отсюда

$$a = \frac{(M+m)R^2}{I_A+mR^2}g\sin\alpha.$$

В предельном случае, когда бочка не заполнена жидкостью ($m = 0$), получается ранее выведенная формула (48.2). В другом предельном случае, когда толщина стенок бочки пренебрежимо мала по сравнению с радиусом R, $I_A = 2MR^2$,

$$a = \frac{M+m}{2M+m}g\sin\alpha.$$

При этом мы не учитывали моменты инерции днищ бочки, считая их пренебрежимо малыми.

Читателю рекомендуется решить ту же задачу с помощью уравнения моментов относительно центра масс, а также с помощью уравнения сохранения энергии.

9. Диск Максвелла (по имени выдающегося английского физика Джемса Максвелла (1831—1879)) подвешен на очень длинных нитях (рис. 138). Части нитей длиной $l = 50$ см каждая были намотаны на ось диска, после чего диск стал опускаться под действием силы тяжести. Достигнув нижнего положения, диск начал подниматься, сообщив «рывок» нитям. Найти ускорение диска и натяжение нити во время его опускания и поднятия, а также оценить приближенно натяжение нити во время рывка. Масса диска $M = 1$ кг, его радиус $R = 10$ см, радиус оси $r = 0{,}5$ см. Растяжением нити во время рывка пренебречь. (Сравните эту задачу с задачей 2 к § 37.)

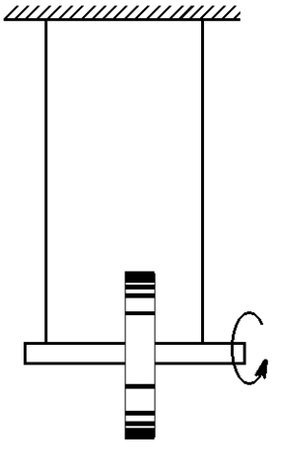


Рис. 138

Ответ. Пока движение совершается без рывка, диск опускается и поднимается с одним и тем же ускорением, направленным вниз:

$$a = \frac{2r^2}{R^2+2r^2}\, g.$$

Натяжение нити при опускании и поднятии диска также одно и то же и равно

$$T_0 = \frac{Mg}{2}\left(1 - \frac{a}{g}\right) \approx 4{,}83 \text{ Н}.$$

Во время рывка нить испытывает дополнительное натяжение ΔT, определяемое приближенным выражением

$$\Delta T \approx \frac{l}{\pi r}\,\frac{2a}{g}\, Mg \approx 3{,}14 \text{ Н}.$$

Полное натяжение нити во время рывка $T = T_0 + \Delta T \approx 8{,}0$ Н.

10. На каком расстоянии l от оси баллистического маятника должно находиться место попадания горизонтально летящего снаряда, чтобы ось маятника при ударе снаряда не испытывала добавочной нагрузки?

Решение. Пусть **F** — горизонтальная сила, с которой ударяющий снаряд действует на маятник (рис. 139). Уравнение моментов относительно точки подвеса O дает

$$I\,\frac{d\omega}{dt} = Fl.$$

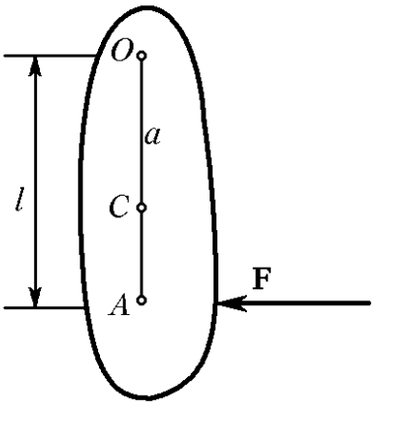


Рис. 139

Так как при ударе ось маятника не испытывает дополнительной нагрузки, то на основании теоремы о движении центра масс можно написать

$$m\,\frac{dv}{dt} = F,$$

где v — скорость центра масс, m — масса маятника. Массой снаряда пренебрегаем. Почленным делением из этого и предыдущего уравнений исключаем силу F и получаем

$$l = \frac{I}{m}\,\frac{d\omega}{dv}.$$

Если a — расстояние между точкой подвеса и центром масс маятника, то $v = \omega a$. В результате находим

$$l = \frac{I}{ma}.$$

Отсюда видно, что l есть приведенная длина физического маятника, а точка A совпадает с центром качения его. Соответствующая ей точка подвеса маятника O называется «*центром удара*». Кузнец точно знает, в каком месте нужно держать рукоятку своего тяжелого молота (именно — в центре удара), чтобы при ударе не ощущать в руке неприятную отдачу.

11. Каким местом шашки следует наносить удар по лозе, чтобы при рубке не ощущалась неприятная отдача? Шашку считать однородной полосой длины l, которую при ударе держать за конец.

Ответ. Расстояние от руки до места удара должно составлять $2l/3$.

12. Твердый цилиндр или шар, положенный на твердую горизонтальную плоскость, катится по ней со скольжением. Показать, что во время качения поступательная и вращательная скорости этого тела связаны соотношением

$$mrv + I\omega = \text{const}, \tag{48.7}$$

где I — момент инерции относительно геометрической оси тела.

Решение. Уравнения движения центра масс и моментов имеют вид

$$m\frac{dv}{dt} = \pm F, \quad I\frac{d\omega}{dt} = \mp M = \mp rF.$$

Верхний знак относится к случаю, когда сила трения F направлена вперед (поступательное движение ускоряется, вращение замедляется), нижний — когда F направлена назад (поступательное движение замедляется, вращение ускоряется). Исключая F и dt, найдем в обоих случаях $mr\,dv = -I\,d\omega$, откуда и следует соотношение (48.7).

13. Согласно уравнению (48.7) качение твердого тела по горизонтальной плоскости не может прекратиться, если нет никаких дополнительных сил, помимо горизонтальной силы трения, действующей в точке касания. В чем причина расхождения этого вывода с опытом?

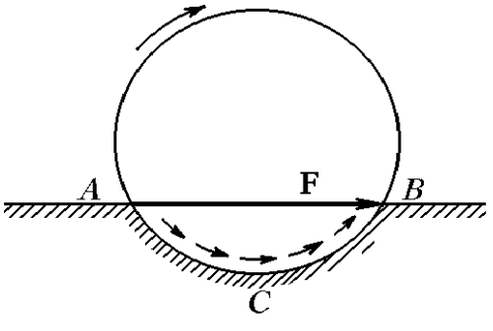


Рис. 140

Решение. Реальные тела деформируемы. На плоскости, по которой катится тело, возникает углубление. Силы трения, действующие на катящееся тело, на рис. 140 изображены маленькими стрелками, их результирующая $\mathbf{F} = \vec{AB}$. Ясно, что момент сил трения M больше момента результирующей, т. е. $M > rF$ (F и M — величины положительные). Из уравнений $m\,dv = \pm F\,dt$, $I\,d\omega = \mp M\,dt$ почленным делением и умножением на r получаем

$$mr\,dv + I\,d\omega\,(rF/M) = 0.$$

С учетом неравенства $rF < M$ отсюда следует, что $mr\,dv + I\,d\omega < 0$, или

$$\frac{d}{dt}(mrv + I\omega) < 0. \tag{48.8}$$

Таким образом, в случае реального качения величина $mrv + I\omega$ убывает со временем и в конце концов обращается в нуль.

14. Сплошному однородному шару радиусом r, лежащему на горизонтальной плоскости, в момент $t = 0$ сообщена скорость v_0 без вращения. Учитывая трение скольжения, но пренебрегая трением качения, найти угловую скорость шара, когда его движение перейдет в чистое качение. Определить потерю кинетической энергии на трение.

Р е ш е н и е. На основании (48.7)

$$mrv_0 = mrv + I\omega = (mr^2 + I)\ \omega,$$

где v — поступательная, а ω — вращательная скорости шара после установления чистого качения. Отсюда и найдется искомая угловая скорость ω. Потеря кинетической энергии равна

$$\Delta K = \frac{1}{2}\frac{I}{I+mr^2}\,mv_0^2 = \frac{1}{7}\,mv_0^2.$$

15. Сплошной однородный шар радиусом r, вращающийся вокруг горизонтального диаметра с угловой скоростью ω_0, ставится на горизонтальную плоскость без сообщения ему поступательного движения. Учитывая трение скольжения, но пренебрегая трением качения, найти линейную скорость v центра шара, когда его движение перейдет в чистое качение. Определить потерю кинетической энергии на трение.

О т в е т. $v = \dfrac{Ir}{I+mr^2}\,\omega_0 = \dfrac{2}{7}\,r\omega_0$, $\Delta K = \dfrac{1}{2}\dfrac{mr^2}{I+mr^2}\,I\omega_0^2 = \dfrac{1}{7}\,mr^2\omega_0^2$.

16. Бильярдный шар катится без скольжения по горизонтальной плоскости со скоростью v и ударяется в покоящийся такой же бильярдный шар, причем линия центров параллельна скорости движения. Определить скорости обоих шаров после того, как их движения перейдут в чистые качения. Какая доля первоначальной кинетической энергии перейдет в тепло? Считать, что при столкновении шаров передачи вращательного движения не происходит. Потерей энергии на трение при чистом качении пренебречь.

О т в е т. Скорость первого шара $v_1 = {}^2/_7\,v$, второго $v_2 = {}^5/_7\,v$. Потеря кинетической энергии на трение составляет 20/49 начального значения кинетической энергии.

17. Два одинаковых бильярдных шара катятся без скольжения навстречу друг другу с одной и той же скоростью v_0 и претерпевают упругий удар. Предполагая, что удар центральный и за время соударения шаров угловые скорости не изменяются, вычислить скорость каждого шара столкновения, когда установится чистое качение.

Р е ш е н и е. При столкновении шары обмениваются поступательными скоростями, тогда как вращательные скорости их сохраняются неизменными. Очевидно, достаточно найти движение одного из шаров. Непосредственно после столкновения начальные скорости рассматриваемого шара будут $v_{\text{нач}} = -v_0$, $\omega_{\text{нач}} = \omega_0 = v_0/r$. Поэтому на основании (48.7) для движения шара после столкновения можно написать $mvr + I\omega = -mv_0r + I\omega_0$. После установления чистого качения $v = \omega r$ и, следовательно,

$$v = \frac{I-mr^2}{I+mr^2}\,v_0 = -\frac{3}{7}\,v_0.$$

18. Бильярдный шар, катящийся без скольжения со скоростью v_0, отражается упруго при нормальном столкновении с неподвижной стенкой. Предполагая, что за время соударения угловая скорость шара не меняется, определить его скорость v после отражения, когда движение перейдет в чистое качение.

Ответ. $v = \frac{I-mr^2}{I+mr^2} v_0 = \frac{3}{7} v_0$.

19. Как надо ударить кием по бильярдному шару, чтобы сила трения шара о сукно бильярдного стола заставляла его двигаться: а) ускоренно; б) замедленно; в) равномерно? Предполагается, что удар наносится горизонтально в вертикальной плоскости, проходящей через центра шара и точку касания его с плоскостью бильярдного стола.

Ответ. Шар будет двигаться равномерно, если точка удара лежит выше его центра на расстоянии 2/5 радиуса. Такие удары называются *нормальными*. Если она лежит еще выше, то движение шара будет ускоренным. Если же точка удара лежит ниже, то шар будут двигаться замедленно. Соответствующие удары называют *высокими* и *низкими*. Решение получено в предположении, что сила трения шара о плоскость стола пренебрежимо мала по сравнению с силой, с которой на шар действует кий во время удара.

20. Как надо ударить кием по бильярдному шару, чтобы при столкновении с другим (неподвижным) шаром: а) оба шара стали двигаться вперед (*удар с накатом*); б) первый шар остановился, а другой двигался вперед; в) второй шар двигался вперед, а первый откатился назад (*удар с оттяжкой*)? Относительно направления и плоскости удара ввести те же предположения, что и в предыдущей задаче.

Ответ. Случай а) реализуется при высоких ударах, случай б) — при нормальных, случай в) — при низких.

21. Виртуозный парижский биллиардист Миньо в 1883 г. в Варшаве ставил на краю бильярда стакан с водой. Бильярдный шар располагался в противоположном конце бильярда. Миньо наносил по шару сильный кратковременный удар кием, наклоненным под большим углом к горизонтальной плоскости, быстро убирая кий после удара. Кий располагался в вертикальной плоскости, проходившей через центр шара и линию, соединявшую его со стаканом. Шар быстро приближался к стакану, но затем внезапно отбрасывался назад, не дойдя 1—2 см до стакана. Объяснить явление.

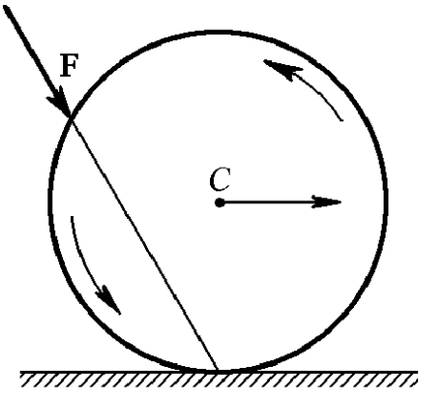


Рис. 140а

Решение. Если удар наклонным кием нанести левее центра шара (рис. 140*а*), то возникает как поступательное движение вправо, так и его вращение со скольжением против часовой стрелки. Из-за скольжения поступательная и вращательная скорости шара будут уменьшаться. Удар кием надо нанести так, чтобы в некоторый момент поступательное движение шара пропало, а вращение продолжилось в том же направлении. Тогда в этот момент шар мгновенно остановится, а затем из-за вращения откатится назад — возникнет частое качение.

22. По бильярдному шару нанесен боковой удар горизонтальным кием, так что вертикальная плоскость не проходит через центр шара. Как отразится такая нецентральность удара на движении шара?

Ответ. Возникнет дополнительное вращение шара вокруг вертикальной оси — верчение. Сила трения верчения и ее момент пренебрежимо малы. Поэтому шар, пока он не ударился о борт бильярда, будет двигаться так же, как и при ударе, нанесенном в вертикальной плоскости, проходящей через центр шара. Верчение скажется при ударе шара о борт бильярда. Так как борт бильярда «мягкий», то время соприкосновения с ним шара конечно. За это время из-за верчения шара на него будет действовать касательная сила трения скольжения, увеличивая или уменьшая касательную скорость поступательного движения шара. Шар отразится от борта бильярда, но угол падения уже не будет равен углу отражения. Он будет больше или меньше, в зависимости от направления верчения.

23. Удар кием по бильярдному шару наносится в какой-то точке в верхней половине шара, причем направление кия может быть каким угодно. Определить в общем случае вид траектории шара в плоскости бильярда.

Решение. Поместим начало координатной системы в центр шара. Координатные оси X и Y направим параллельно плоскости бильярда, а ось Z — вертикально вверх. Пусть F_x, F_y, F_z — слагающие мгновенной силы удара. Сила F_x вызовет вращение вокруг осей Y и Z с угловыми скоростями $\boldsymbol{\omega}_y'$ и $\boldsymbol{\omega}_z'$, сила F_y — вращение с угловыми скоростями $\boldsymbol{\omega}_x'$ и $\boldsymbol{\omega}_z''$, сила F_z — вращения с угловыми скоростями $\boldsymbol{\omega}_x''$ и $\boldsymbol{\omega}_y''$. Вращение с угловыми скоростями $\boldsymbol{\omega}_z'$ и $\boldsymbol{\omega}_z''$ есть верчение вокруг вертикальной оси. Пока шар ударяется о борт бильярда, от верчения можно отвлечься, так как силы трения верчения и их моменты пренебрежимо малы. Надо принять во внимание только угловые скорости $\boldsymbol{\omega}_x = \boldsymbol{\omega}_x' + \boldsymbol{\omega}_x''$ и $\boldsymbol{\omega}_y = \boldsymbol{\omega}_y' + \boldsymbol{\omega}_y''$. Они складываются в угловую скорость $\boldsymbol{\omega} = \boldsymbol{\omega}_x + \boldsymbol{\omega}_y$. Не нарушая общности, оси X и Y можно повернуть, направив ось X вдоль вектора угловой скорости $\boldsymbol{\omega}$. Тогда шар будет вращаться только вокруг оси X в общем случае со скольжением. Помимо вращения силы F_x и F_y вызовут поступательное движение центра шара параллельно осям X и Y, которое передается и движению точки соприкосновения шара с плоскостью бильярда. (Сила F_z на поступательное движение не влияет, так как ее действие компенсируется нормальным давлением.) Благодаря этому изменяется и сила трения скольжения шара по плоскости бильярда. Существенно однако, что результирующая сила трения скольжения постоянна, т. е. не зависит от скорости движения шара. Так как в общем случае эта сила направлена под углом к начальной скорости, сообщенной силой удара, то возникает движение шара по параболе.

24. Вращающийся с угловой скоростью ω_0 сплошной однородный цилиндр радиусом r ставится без начальной поступательной скорости у основании наклонной плоскости, образующей угол α с горизонтальной плоскостью, и начинает вкатываться наверх. Определить время, в течение которого цилиндр достигает наивысшего положения на наклонной плоскости.

Решение. Пусть $\mathbf{F}$ — сила трения, действующая на цилиндр в месте соприкосновения его с наклонной плоскостью (рис. 141). Она заставляет цилиндр подниматься по наклонной плоскости. Сначала, пока не установилось чистое качение, $\mathbf{F}$ является силой трения скольжения. После перехода движения в чистое качение $\mathbf{F}$ переходит в силу трения покоя (сцепления).

Однако независимо от характера движения оно всегда подчиняется уравнению движения центра масс

$$m\frac{dv}{dt} = F - mg\sin\alpha$$

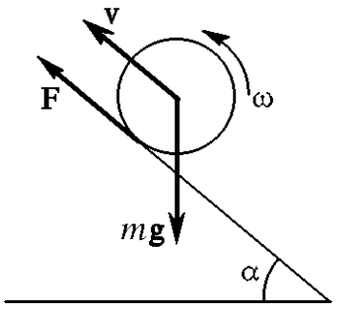


Рис. 141

и уравнению моментов (относительно геометрической оси цилиндра)

$$I\frac{d\omega}{dt} = -Fr.$$

Исключая F, получим

$$mr\frac{dv}{dt} = -I\frac{d\omega}{dt} - mgr\sin\alpha.$$

Интегрирование этого уравнения с учетом начального условия ($\omega = \omega_0$ при $t = 0$) дает

$$mrv = I(\omega_0 - \omega) - mgrt\sin\alpha.$$

Это соотношение справедливо в течение всего времени движения, независимо от того, происходит ли оно со скольжением или является чистым качением. В наивысшей точке должно быть $v = 0$. Отсюда следует, что в той же точке $\omega = 0$. В противном случае цилиндр продолжал бы вкатываться и рассматриваемая точка не была бы наивысшей. Поэтому время подъема t найдется, если в предыдущем уравнении положить $v = \omega = 0$. Это дает

$$t = \frac{I\omega_0}{mgr\sin\alpha} = \frac{r\omega_0}{2g\sin\alpha}.$$

Любопытно, что время поднятия t не зависит от коэффициентов трения между цилиндром и наклонной плоскостью. Результат не изменился бы даже тогда, когда коэффициент трения стал бы переменным. Решение предполагает, однако, что трение достаточно велико, чтобы цилиндр мог вкатываться на наклонную плоскость. При недостаточном трении будет происходить лишь замедление скорости вращения цилиндра. Нетрудно подсчитать, что время замедления определяется прежней формулой.

Напротив, время обратного скатывания цилиндра, а также наибольшая высота поднятия его зависят от коэффициента трения. Такое различие объясняется тем, что скатывание цилиндра все время является чистым качением. Поднятие же его сначала происходит со скольжением, а затем переходит в чистое качение.

25. Считая в предыдущей задаче коэффициент трения скольжения k цилиндра о наклонную плоскость заданным и постоянным, определить: 1) ускорение цилиндра a_1, когда качение происходит со скольжением; 2) время t_1, по истечении которого наступает чистое качение; 3) высоту H_1, которой достигает цилиндр, прежде чем начинается чистое качение; 4) ускорение a_2 при чистом качении; 5) дополнительную высоту H_2, на которую поднимается цилиндр при чистом качении; 6) полную высоту поднятия H; 7) время обратного скатывания цилиндра вниз t. Предполагается, что $k > \operatorname{tg}\alpha$.

Ответ. $a_1 = g(k\cos\alpha - \sin\alpha)$, направлено вверх;

$$t_1 = \frac{I\omega_0 r}{(I+mr^2)a_1 + mr^2 g\sin\alpha} = \frac{\omega_0 r}{(3k\cos\alpha - \sin\alpha)g};$$

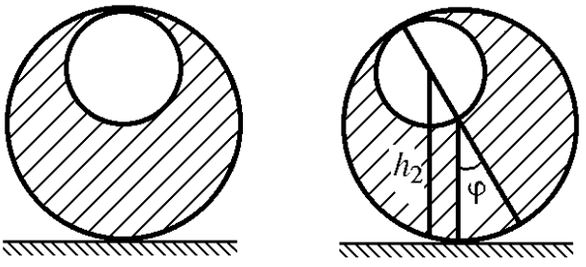
$$H_1 = \frac{1}{2}a_1 t_1^2 \sin\alpha;\ a_2 = \frac{mr^2}{I+mr^2}g\sin\alpha = \frac{2}{3}g\sin\alpha;\ H_2 = \frac{a_1}{a_2}H_1;$$

$$H = H_1 + H_2 = \frac{k\cos\alpha - \sin\alpha}{4g(3k\cos\alpha - \sin\alpha)}\omega_0^2 r^2;$$

$$\overline{t} = \sqrt{\frac{2H}{a_2\sin\alpha}} = \frac{\omega_0 r_0}{2g\sin\alpha}\sqrt{\frac{3(k\cos\alpha - \sin\alpha)}{3k\cos\alpha - \sin\alpha}}.$$

26. Вращающийся с угловой скоростью ω_0 сплошной однородный цилиндр массой m_1 ставится без начальной поступательной скорости на длинную доску массой m_2, лежащую на гладкой горизонтальной плоскости. Начальная скорость доски равна нулю. Пренебрегая силой трения качения, но учитывая трение скольжения между доской и цилиндром, найти угловую скорость вращения цилиндра после того, как его движение перейдет в чистое качение. Доска предполагается настолько длинной, что чистое качение успевает установиться до того, как цилиндр скатится с доски.

Ответ. $\omega = \dfrac{m_1+m_2}{m_1+3m_2}\omega_0$.



а *б*

Рис. 142

27. В сплошном однородном цилиндре радиусом R сделана цилиндрическая полость радиусом $R/2$ с осью, проходящей через середину радиуса цилиндра (рис. 142 *а*). Определить период малых колебаний T, которые возникнут, если положить цилиндр на горизонтальную плоскость и дать ему возможность катиться по ней без скольжения.

Решение. Задача сводится к нахождению выражений для потенциальной и кинетической энергии системы. С этой целью мысленно заполним полость тем же веществом, из которого сделан цилиндр. Образовавшийся таким образом сплошной однородный цилиндр назовем цилиндром *1*, а цилиндр вдвое меньшего радиуса, заполняющий полость, — цилиндром *2*. Массы цилиндров обозначим соответственно через m_1 и m_2. Энергия системы, как потенциальная, так и кинетическая, будет равна разности энергий цилиндров *1* и *2*. При повороте системы из положения равновесия на угол φ (рис. 142 *б*) центр масс цилиндра *1* остается на прежней высоте, его потенциальная энергия U_1 не изменяется. Потенциальная же энергия цилиндра *2* становится равной $U_2 = m_2 g h_2$, где $h_2 = R + \frac{1}{2}R\cos\varphi$ — высота центра масс этого цилиндра над горизонтальной плоскостью, на которой находится система. Полная потенциальная энергия всей системы

$$U = U_1 - U_2 = \text{const} - m_2 gR\left(1 + \frac{1}{2}\cos\varphi\right).$$

Единственное переменное слагаемое, которое оно содержит, есть $-\frac{1}{2}m_2 gR\cos\varphi$. Поэтому при надлежащем выборе аддитивной постоянной

величину U всегда можно представить в виде

$$U = \text{const} + \frac{1}{2} m_2 g R (1 - \cos\varphi) = \text{const} + m_2 g R \sin^2 \frac{\varphi}{2},$$

или для малых углов φ

$$U \approx \text{const} + \frac{1}{4} m_2 g R \varphi^2.$$

Кинетическая энергия системы $K = {}^1/_2 (I_1 - I_2)\dot{\varphi}^2$, где I_1 и I_2 — моменты инерции цилиндров относительно мгновенной оси. При изменении угла φ величины I_1 и I_2 изменяются. Но для малых колебаний этими изменениями можно пренебречь и отнести I_1 и I_2 к тому моменту, когда система находится в положении равновесия. В этом положении с помощью теоремы Гюйгенса—Штейнера нетрудно получить

$$I_1 = {}^3/_2\, m_1 R^2, \quad I_2 = {}^{19}/_8\, m_2 R^2.$$

Приняв еще во внимание, что $m_1 = 4m_2$, найдем

$$K = \frac{29}{16} m_2 R^2 \dot{\varphi}^2.$$

Из полученных выражений для U и K делаем вывод, что малые колебания системы будут гармоническими с периодом

$$T = \pi \sqrt{29 \frac{R}{g}}.$$

28. Большой однородный свинцовый шар массой M лежит на плоской горизонтальной поверхности. Небольшая пуля массой m выпущена из ружья горизонтально со скоростью V в направлении к центру шара. После выстрела пуля застревает внутри шара. Определить линейную скорость шара v после того, как его движение перейдет в чистое качение. При рассмотрении движения шара после удара считать его однородным, пренебрегая массой застрявшей пули. Трением качения пренебречь.

Ответ. $v = \frac{5}{7} \frac{m}{M} V$.

29. Шар массой $M = 1000$ г, лежащий на горизонтальной плоскости, пробивается по диаметру пулей, летящей горизонтально с начальной скоростью $V_0 = 500$ м/с. После удара шар начинает скользить по плоскости. Спустя некоторое время его движение переходит в чистое качение с постоянной скоростью $v = 3$ м/с. Определить скорость пули V после вылета ее из шара, если масса пули $m = 10$ г. Трением качения пренебречь.

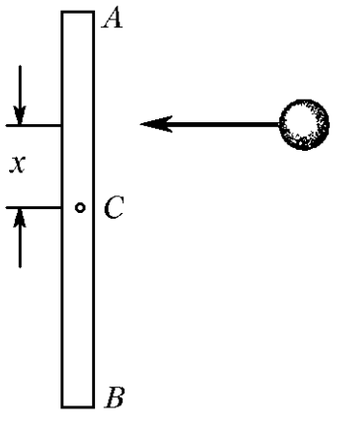


Рис. 143

Ответ. $V = V_0 - \frac{7}{5} \frac{M}{m} v = 80$ м/с.

30. На гладком горизонтальном столе лежит однородный стержень длиной l, который может двигаться по столу без трения (рис. 143). В начальный момент, когда скорость стержня равна нулю, в него ударяется шарик, движущийся перпендикулярно к стержню. На каком расстоянии x от центра стержня C ударился шарик, если непосредственно после удара концы стержня A и B начали двигаться со скоростями v_A и v_B соответственно? (Скорости v_A и v_B считаются положитель-

ными, когда они направлены в ту же сторону, что и скорость шарика до удара, и отрицательными — в противоположном случае.)

Ответ. $x = \frac{l}{6}\frac{v_A - v_B}{v_A + v_B}$. Результат не зависит от характера удара.

31. На идеально гладкой горизонтальной поверхности лежит стержень длиной l и массой M, который может скользить по этой поверхности без трения (см. рис. 143). В одну из точек стержня ударяет шарик массой m, движущийся перпендикулярно к стержню. На каком расстоянии x от середины стержня должен произойти удар, чтобы шарик передал стержню всю свою кинетическую энергию? Удар считать абсолютно упругим. При каком соотношении масс M и m это возможно?

Ответ. $x = \frac{l}{2\sqrt{3}}\sqrt{\frac{M}{m} - 1}$. Для возможности описываемого процесса необходимо $M \geqslant m$. Условие $x \leqslant l/2$ дает еще $M \leqslant 4m$.

32. На гладком горизонтальном столе лежит однородный упругий стержень длиной l и массой M. В конец стержня ударяет упругий шарик массой m, движущийся со скоростью v перпендикулярно к стержню. Найти значение энергии деформации системы в момент, когда она максимальна. Трением между стержнем и столом пренебречь.

Ответ. $U = \frac{M}{M+4m}\frac{mv^2}{2}$. В предельных случаях 1) $M = 0$ и 2) $M = \infty$ получаем 1) $U = 0$; 2) $U = {}^1/_2\, mv^2$.

33. На гладком горизонтальном столе лежит однородный твердый стержень длиной l и массой M, в край которого ударяет твердый шарик массой m, движущийся со скоростью v_0, перпендикулярной к оси стержня. Считая удар идеально упругим и предполагая, что силы трения между поверхностью стола и лежащими на ней телами пренебрежимо малы, вычислить угловую скорость вращения стержня после удара.

Решение. Если F — сила, действующая на шарик во время удара, то

$$m\frac{dv}{dt} = -F, \quad M\frac{dV}{dt} = F, \quad I\frac{d\omega}{dt} = F\frac{l}{2}.$$

Почленным делением исключаем F и получаем

$$\frac{m}{I}\frac{dv}{d\omega} = -\frac{2}{l}, \quad \frac{M}{I}\frac{dV}{d\omega} = \frac{2}{l}.$$

Интегрируя в пределах от начального значения угловой скорости $\omega = 0$ до конечного, найдем

$$v - v_0 = \frac{2}{l} - \frac{I}{m}\omega, \quad V = \frac{2}{l}\frac{I}{M}\omega,$$

причем в этих уравнениях v, V и ω означают величины соответствующих скоростей после удара. Угловая скорость ω найдется из уравнения сохранения энергии. Если в него подставить значения v и V, то для ω получится квадратное уравнение

$$\left[1 + \frac{4I}{l^2}\left(\frac{1}{m} + \frac{1}{M}\right)\right]\omega^2 - 4\frac{v_0}{l}\omega = 0.$$

Один из корней этого уравнения ($\omega = 0$) дает угловую скорость стержня до удара, второй — после удара. По условию задачи надо взять второй корень. С учетом соотношения $I = {}^1/_{12}\, Ml^2$ для него получаем

$$\omega = \frac{12mv_0}{(4m+M)l}.$$

§ 49. ГИРОСКОПЫ. ДВИЖЕНИЕ СВОБОДНОГО ГИРОСКОПА

1. Слово «гироскоп» в буквальном переводе означает прибор для обнаружения вращения. В широком смысле *гироскопом* называется быстро вращающееся твердое тело, ось вращения которого может изменять свое направление в пространстве. Гироскоп, в особенности когда на него действуют внешние силы, может совершать удивительные движения, кажущиеся на первый взгляд неожиданными и непонятными. Они всегда воспринимаются с захватывающим интересом. Быстро вращающийся волчок может служить не только забавной игрушкой, но и прекрасным демонстрационным прибором при изучении законов механики. Все явления, обусловленные быстрым вращением гироскопа, называются *гироскопическими*. Они нашли широкое научно-техническое применение (см. § 51).

Гироскопические эффекты проявляются также у атомов благодаря наличию у них моментов импульса, связанных с внутренними орбитальными движениями или собственными вращениями (спинами) электронов и атомных ядер. Конечно, эти, как и всякие другие, атомные явления должны рассматриваться на основе квантовой механики. Однако есть много общего в гироскопических свойствах атомных и макроскопических систем. Поэтому теория гироскопов может оказаться полезной и при изучении атомной физики.

Наибольшее значение в науке и технике имеют *симметричные гироскопы. Симметричным* называют гироскоп, обладающей симметрией вращения относительно некоторой оси, называемой *геометрической осью* или *осью фигуры гироскопа.* Теория симметричного гироскопа более проста и более важна, чем теория несимметричного гироскопа. В дальнейшем мы ограничимся рассмотрением только симметричных гироскопов. Обычно одна из точек оси фигуры гироскопа бывает закреплена. Закрепленную точку оси фигуры называют *точкой опоры гироскопа.* В более общем смысле *точкой опоры гироскопа* называют такую точку O оси фигуры его, относительно которой рассматривают вращение гироскопа. В общем случае движение гироскопа слагается из движения точки опоры O и вращения вокруг мгновенной оси, проходящей через эту точку. Примером гироскопа с движущейся точкой опоры может служить детская игрушка — *волчок.* Основным в теории является случай, когда точка опоры *неподвижна.* К этому частному случаю можно свести и общий случай, когда точка опоры движется (см. п. 6).

2. Чтобы ось фигуры гироскопа могла свободно поворачиваться в пространстве, гироскоп обычно помещают в так называемом *карданове подвесе* (рис. 144). Маховичок гироскопа закрепляется на его оси фигуры $A'A$, которая может вращается по возможности с малым трением в подшипниках, укрепленных на концах диаметра *внутреннего кольца.* Внутреннее кольцо, в свою очередь, может вращаться вокруг перпендикулярной оси $B'B$, проходящей через

подшипники на концах диаметра *наружного кольца*. Наконец, наружное кольцо может совершать вращение вокруг третьей оси $D'D$, проходящей через неподвижные подшипники подставки. Ось $B'B$ перпендикулярна к оси $A'A$. Все три оси пересекаются в одной точке, называемой *центром карданова подвеса*. Гироскоп в кардановом подвесе имеет *три* степени свободы и может совершать любые повороты вокруг центра подвеса. Во всех вопросах мы будем пренебрегать кинетической энергией и моментами импульсов колец, считая их пренебрежимо малыми по сравнению с кинетической энергией и моментом импульса маховичка гироскопа. Если центр карданова подвеса или точка опоры совпадает с центром масс гироскопа, то гироскоп называется *уравновешенным*.

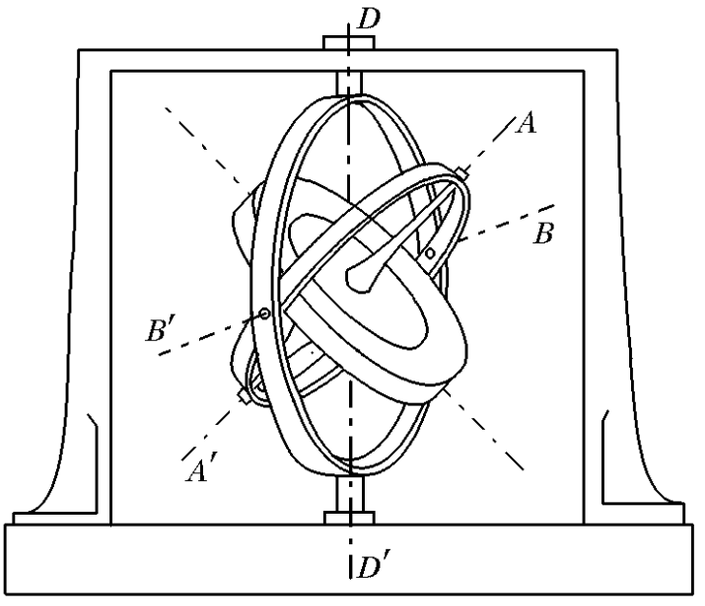


Рис. 144

3. Согласно теореме Эйлера (§ 47) движение гироскопа с неподвижной точкой опоры O можно представить как вращение вокруг мгновенной оси, проходящей через эту точку. Обозначим через $\boldsymbol{\omega}$ вектор мгновенной угловой скорости, с которой вращается гироскоп, $\mathbf{L}$ — момент импульса гироскопа относительно точки O. Найдем связь между векторами $\mathbf{L}$ и $\boldsymbol{\omega}$ для симметричного гироскопа. Если угловая скорость $\boldsymbol{\omega}$ направлена вдоль оси фигуры гироскопа или перпендикулярно к ней, то векторы $\mathbf{L}$ и $\boldsymbol{\omega}$ параллельны между собой. Убедиться в этом проще всего можно следующим образом. Мысленно разобьем все тело гироскопа на пары одинаковых материальных точек, симметрично расположенных относительно оси фигуры гироскопа, как указано на рисунках 145 и 146. Момент импульса такой пары точек относительно точки O будет $d\mathbf{L} = dm[\mathbf{r}_1\mathbf{v}_1] + dm[\mathbf{r}_2\mathbf{v}_2]$, где dm — масса каждой из них. Если гироскоп вращается вокруг оси своей фигуры (рис. 145), то

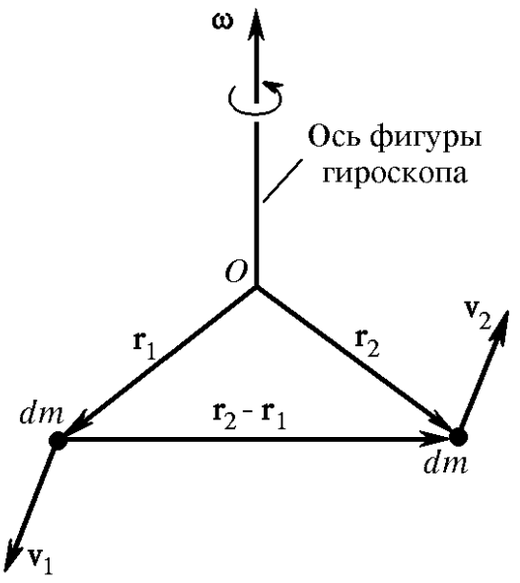


Рис. 145

скорости $\mathbf{v}_1$ и $\mathbf{v}_2$ равны по модулю, но направлены противоположно. В этом случае $d\mathbf{L} = dm[\mathbf{v}_2(\mathbf{r}_2 - \mathbf{r}_1)]$. Векторы $\mathbf{v}_2$ и $(\mathbf{r}_2 - \mathbf{r}_1)$ перпендикулярны к оси вращения. Поэтому вектор $d\mathbf{L}$, а с ним и момент импульса всего гироскопа $\mathbf{L}$ будут направлены вдоль оси вращения. Вектор $\mathbf{L}$ совпадает с моментом импульса относительно оси вращения, а потому $\mathbf{L} = \mathbf{I}_{\parallel}\boldsymbol{\omega}$, где $\mathbf{I}_{\parallel}$ — момент инерции гироскопа относительно оси его фигуры. Если теперь гироскоп вращается вокруг оси, перпендикулярной к оси фигуры (рис. 146), то $\mathbf{v}_1 = \mathbf{v}_2$, а потому $d\mathbf{L} = dm[\mathbf{v}_1(\mathbf{r}_1 + \mathbf{r}_2)]$. Отсюда видно, что $d\mathbf{L}$ и $\mathbf{L}$ опять направлены вдоль оси вращения, причем $\mathbf{L} = I_{\perp}\boldsymbol{\omega}$, где $I_{\perp}$ — момент инерции гироскопа относительно оси, перпендикулярной к его оси фигуры.

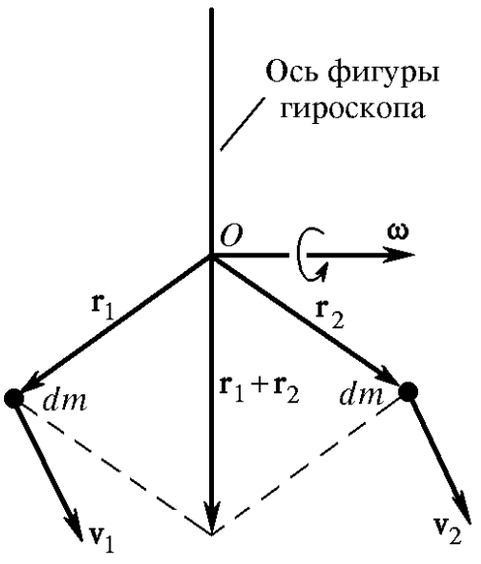


Рис. 146

Допустим теперь, что мгновенная ось направлена под горизонтальным углом к оси фигуры гироскопа. Разложим вектор $\boldsymbol{\omega}$ на две составляющие: направленную вдоль оси фигуры гироскопа $\boldsymbol{\omega}_{\parallel}$ и перпендикулярную к ней $\boldsymbol{\omega}_{\perp}$ (рис. 147). Из общего определения момента импульса (см. § 30) следует, что он выражается линейно через линейные скорости материальных точек, на которые мысленно можно разбить тело гироскопа. В свою очередь эти скорости выражаются линейно через вектор угловой скорости $\boldsymbol{\omega}$, одинаковый для всех точек гироскопа. Отсюда следует, что *вектор* $\mathbf{L}$ *линейно выражается через* $\boldsymbol{\omega}$. Рассматривая его как функцию $\boldsymbol{\omega}$, можно написать $\mathbf{L} = \mathbf{L}(\boldsymbol{\omega}) = \mathbf{L}(\boldsymbol{\omega}_{\parallel} + \boldsymbol{\omega}_{\perp})$, или, в силу указанной линейности, $\mathbf{L} = \mathbf{L}(\boldsymbol{\omega}_{\parallel}) + \mathbf{L}(\boldsymbol{\omega}_{\perp})$. Но функция $\mathbf{L}(\boldsymbol{\omega}_{\parallel})$ была бы равна моменту импульса гироскопа, если бы последний вращался только вокруг его оси фигуры с угловой скоростью $\boldsymbol{\omega}_{\parallel}$. Значит, $\mathbf{L}(\boldsymbol{\omega}_{\parallel}) = I_{\parallel}\boldsymbol{\omega}_{\parallel}$. Аналогично $\mathbf{L}(\boldsymbol{\omega}_{\perp}) = I_{\perp}\boldsymbol{\omega}_{\perp}$. В результате получим

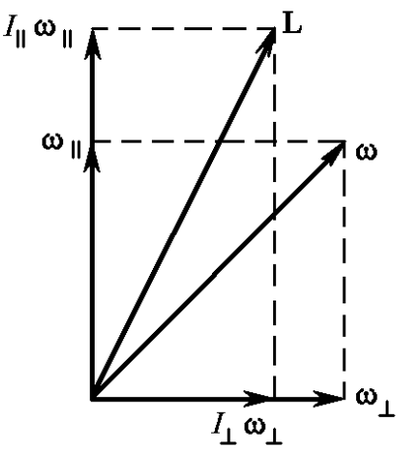


Рис. 147

$$\mathbf{L} = I_{\parallel}\boldsymbol{\omega}_{\parallel} + I_{\perp}\boldsymbol{\omega}_{\perp}. \qquad (49.1)$$

Пользуясь этой формулой, легко найти построением вектор $\mathbf{L}$, если известен вектор $\boldsymbol{\omega}$ (рис. 147). Из построения видно, что векторы $\mathbf{L}$,

$\boldsymbol{\omega}$ и ось фигуры гироскопа лежат в одной плоскости. Однако в общем случае направления векторов $\mathbf{L}$ и $\boldsymbol{\omega}$ не совпадают.

Если воспользоваться формулой (47.2), то из (49.1) можно получить следующие два выражения для кинетической энергии вращающегося гироскопа:

$$K = \frac{1}{2}\left(I_{\parallel}\omega_{\parallel}^2 + I_{\perp}\omega_{\perp}^2\right) = \frac{1}{2}\left(\frac{\mathbf{L}_{\parallel}^2}{I_{\parallel}} + \frac{\mathbf{L}_{\perp}^2}{I_{\perp}}\right). \tag{49.2}$$

Эти выражения показывают, что *кинетическая энергия симметричного гироскопа равна сумме кинетических энергий двух вращений, из которых одно совершается вокруг оси фигуры, а другое — вокруг оси, к ней перпендикулярной.*

На практике гироскоп всегда приводится в быстрое вращение вокруг оси фигуры. По сравнению с этим быстрым вращением вращение, возникающее по тем или иным причинам вокруг перпендикулярной оси, всегда происходит *медленно.* Тогда различие в направлениях векторов $\mathbf{L}$ и $\boldsymbol{\omega}$ становится очень малым. Оба эти направления практически совпадают с направлением оси фигуры гироскопа.

За положительное направление оси фигуры гироскопа принимают направление ее, совпадающее с направлением вектора угловой скорости $\boldsymbol{\omega}$ (точнее, образующее с ним острый угол). Если от точки опоры O отложить отрезок OS единичной длины в положительном направлении оси фигуры гироскопа, то конец этого отрезка S называется *вершиной гироскопа.* Если известно движение вершины гироскопа и угловая скорость вращения его вокруг оси фигуры, то движение гироскопа определено полностью. Поэтому основная задача теории гироскопа сводится к нахождению движения вершины гироскопа и угловой скорости вращения его вокруг оси фигуры.

4. Вся теория гироскопа построена на уравнении моментов

$$\dot{\mathbf{L}} = \mathbf{M}, \tag{49.3}$$

причем моменты $\mathbf{L}$ и $\mathbf{M}$ берутся относительно неподвижной точки опоры гироскопа O. Если момент внешних сил $\mathbf{M}$ равен нулю, то гироскоп называется *свободным.* Для свободного гироскопа $\dot{\mathbf{L}} = 0$, и следовательно,

$$\mathbf{L} \equiv I_{\parallel}\boldsymbol{\omega}_{\parallel} + I_{\perp}\boldsymbol{\omega}_{\perp} = \text{const}. \tag{49.4}$$

Это уравнение выражает сохранение моментов импульса гироскопа. К нему следует присоединить уравнение сохранения энергии

$$K \equiv \frac{1}{2}(\mathbf{L}\boldsymbol{\omega}) = \frac{1}{2}\left(I_{\parallel}\omega_{\parallel}^2 + I_{\perp}\omega_{\perp}^2\right) = \text{const}, \tag{49.5}$$

которое также является следствием уравнения (49.3). Если уравнение (49.4) возвести в квадрат, то получится

$$I_{\parallel}^2\omega_{\parallel}^2 + I_{\perp}^2\omega_{\perp}^2 = \text{const}.$$

Из этого и предыдущего уравнений следует, что *при движении свободного гироскопа длины векторов* $\boldsymbol{\omega}_{\parallel}$ *и* $\boldsymbol{\omega}_{\perp}$ *остаются постоянными.* Вместе с ними *остаются постоянными и обе составляющие момента импульса*: $L_{\parallel} = I_{\parallel}\omega_{\parallel}$ и $L_{\perp} = I_{\perp}\omega_{\perp}$. Следовательно, *остается постоянным угол между векторами* $\mathbf{L}$ *и* $\boldsymbol{\omega}$, как это видно из уравнения (49.5). *Из постоянства* $L_{\parallel}$ *и* $L_{\perp}$ *следует также постоянство угла между вектором* $\mathbf{L}$ *и осью фигуры гироскопа.* В каждый момент времени ось фигуры гироскопа совершает вращение вокруг мгновенной оси с угловой скоростью $\boldsymbol{\omega}$. Векторы $\boldsymbol{\omega}$ и $\mathbf{L}$, как мы видели, лежат в одной плоскости с осью фигуры гироскопа. А так как вектор $\mathbf{L}$ сохраняет неизменным свое направление в пространстве, то мгновенная ось и ось фигуры должны вращаться вокруг этого неизменного направления с одной и той же угловой скоростью. Все это приводит к следующей картине движения свободного гироскопа.

В каждый момент времени движение свободного гироскопа есть вращение вокруг мгновенной оси, проходящей через неподвижную точку опоры. С течением времени мгновенная ось и вектор $\mathbf{L}$ *меняют свое положение в теле, описывая конусы вокруг оси фигуры гироскопа с одной и той же постоянной угловой скоростью* $\boldsymbol{\omega}_1$, *вообще говоря, не равной* $\boldsymbol{\omega}$. *Направление вектора* $\mathbf{L}$ *неизменно в пространстве. Ось фигуры гироскопа и мгновенная ось равномерно вращаются в пространстве вокруг этого направления с той же угловой скоростью* $\boldsymbol{\omega}_1$, *но в противоположном направлении.* Такое движение называется *свободной регулярной прецессией гироскопа.* Слово «регулярная» надо понимать в том смысле, что на конические вращения оси фигуры гироскопа и мгновенной оси не накладываются никакие дрожания.

5. Если гироскоп с достаточно большим моментом инерции привести в быстрое вращение, то он будет обладать большим моментом импульса. Приращение момента импульса, как это следует из уравнения (49.3), определяется интегралом

$$\Delta\mathbf{L} = \int \mathbf{M}\, dt. \tag{49.6}$$

Если внешняя сила действует в течение короткого промежутка времени, то интеграл (49.6), а с ним и приращение момента импульса будут малы. Значит, при кратковременных воздействиях даже очень больших сил движение свободного гироскопа изменяется мало.

Гироскоп как бы сопротивляется всяким попыткам изменить его момент импульса. С этим связана замечательная устойчивость, которую приобретает движение гироскопа после приведения его в быстрое вращение.

Возьмем массивный гироскоп, имеющий конусообразную форму (рис. 148). Вдоль оси его фигуры может ввинчиваться стержень с острым концом, которым гироскоп опирается на подставку. Ввинтим стержень настолько, чтобы точка опоры совпала с центром массы ги-

роскопа. Тогда гироскоп станет уравновешенным. При любом наклоне его оси фигуры он будет находиться в безразличном равновесии. Пока гироскоп не вращается, малейший толчок далеко уводит его из положения равновесия. Приведем теперь гироскоп в быстрое вращение вокруг его оси фигуры. Если палкой нанести сильный удар по стержню гироскопа, то направление стержня в пространстве почти не изменится. Стержень начнет лишь совершать свободную регулярную прецессию, т. е. вращательное движение по поверхности конуса малого угла раствора. Осью конуса будет служить направление момента импульса гироскопа, которое он примет после нанесения удара.

Вот другая демонстрация устойчивости движения быстро вращающегося гироскопа. Гироскоп состоит из массивного металлического маховичка, помещаемого внутри разборного полого металлического шара, состоящего из двух полушарий, которые могут сцепляться между собой. Концы оси маховичка входят в подшипники полого шара, так что маховичок может вращаться вокруг одного из диаметров шара. На ось маховичка наматывается нить, свободный конец которой выходит наружу через отверстие в полом шаре. Дергая за нить, можно привести маховичок в быстрое вращение. Если такой шар попытаться скатить с наклонной плоскости, то он будет упорно сопротивляться этим попыткам. «Послушным» шар будет только тогда, когда ось маховичка горизонтальна и перпендикулярна к направлению скатывания. В этом положении шар может свободно скатываться

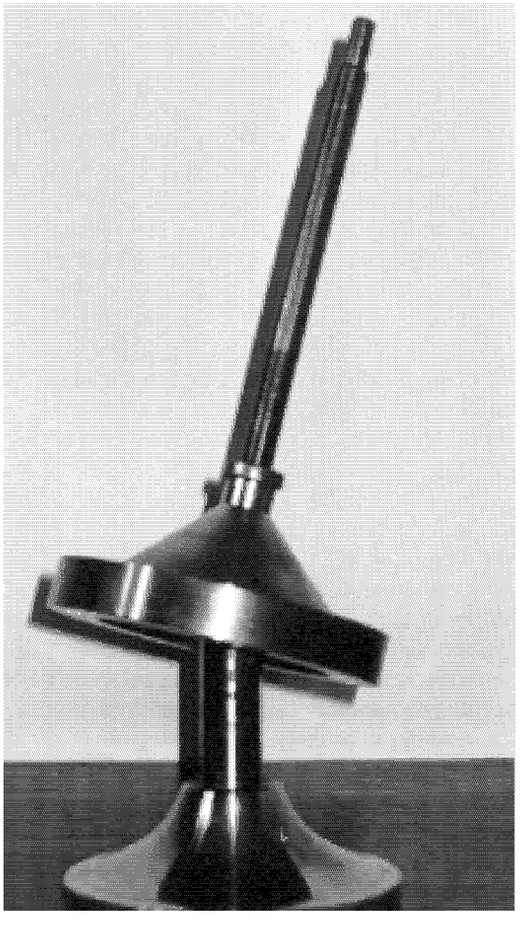


Рис. 148

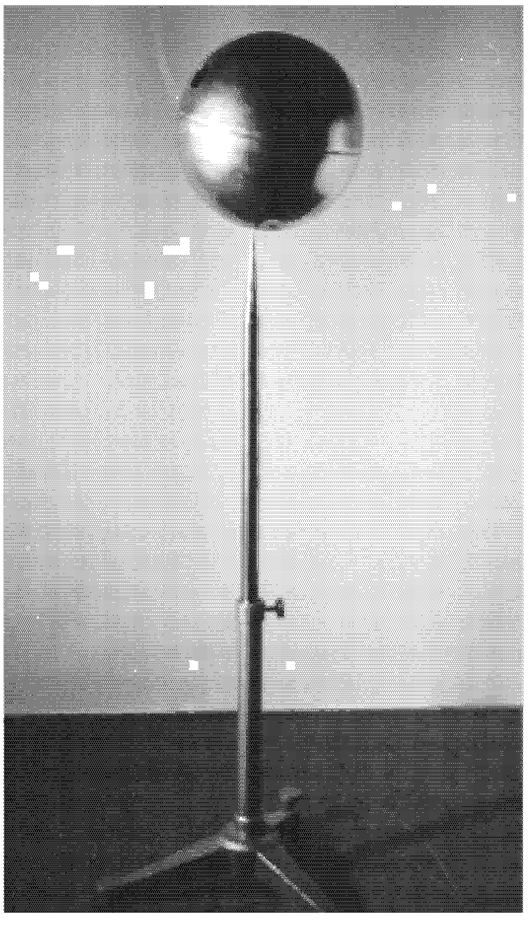


Рис. 149

без изменения направления оси маховичка, т. е. без изменения направления вектора момента импульса $\mathbf{L}$. Во всяком другом положении для «нормального» скатывания ось маховичка, а с ней и вектор $\mathbf{L}$ должны менять направление в пространстве. «Упрямый» гироскоп этого «делать не хочет». Под действием силы тяжести шар гироскопа приобретает медленное вращение и скатывается с наклонной плоскости «боком», стремясь сохранить неизменной ориентацию оси маховичка в пространстве.

Если шар с вращающимся внутри него маховичком поставить на острие иглы даже в наклонном положении, то он не падает (рис. 149), а приобретает медленное вращение вокруг вертикальной оси под действием силы тяжести. Такое вращение называется *вынужденной прецессией*. Вынужденную прецессию мы рассмотрим в следующем параграфе.

6. Посмотрим теперь, как изменится основное уравнение (49.3), если точка опоры гироскопа движется. Ответ можно получить из уравнения (37.2). Скорость каждой точки движущегося гироскопа представим в виде $\mathbf{v} = \mathbf{v}_O + \mathbf{v}_{\text{вр}}$, где $\mathbf{v}_O$ — скорость точки опоры O, а $\mathbf{v}_{\text{вр}} = [\boldsymbol{\omega}\mathbf{r}]$ — скорость, возникающая из-за вращения вокруг этой точки. Тогда момент импульса гироскопа относительно точки опоры представится выражением

$$\mathbf{L} = \int [\mathbf{r}\mathbf{v}_O]\, dm + \mathbf{L}_{\text{вр}},$$

где $\mathbf{L}_{\text{вр}}$ — момент импульса, возникающий только из-за вращения. Если ввести радиус-вектор центра масс $\mathbf{r}_C$, то

$$\mathbf{L} = \mathbf{L}_{\text{вр}} + m[\mathbf{r}_C\mathbf{v}_O]. \tag{49.7}$$

Продифференцировав это выражение и подставив в формулу (37.2), получим

$$\dot{\mathbf{L}}_{\text{вр}} = \mathbf{M} - m[\mathbf{r}_C\dot{\mathbf{v}}_O]. \tag{49.8}$$

Это формула показывает, что от движения точки опоры можно отвлечься. Но тогда к моменту действующих сил $\mathbf{M}$ надо прибавить момент «фиктивной силы», или «силы инерции», $\mathbf{F}_{\text{ин}} = -m\dot{\mathbf{v}}_O$, приложенной к центру масс гироскопа. Этот результат становится совершенно естественным, если отнести движение к системе отсчета, в которой точка опоры гироскопа неподвижна (см. гл. IX).

§ 50. ГИРОСКОП ПОД ДЕЙСТВИЕМ СИЛ. ПРИБЛИЖЕННАЯ ТЕОРИЯ

1. Наиболее интересным видом движения гироскопа является *вынужденная прецессия*. Она возникает под действием внешних сил. Возьмем, например, гироскоп, изображенный на схематическом рис. 150. Он состоит из двух одинаковых маховичков, свободно наса-

женных на общую ось. Гироскоп устроен так, что он может свободно вращаться не только вокруг его оси фигуры OZ, но также вокруг вертикальной и горизонтальной осей OY и OX. Про такой гироскоп говорят, что он имеет *три степени свободы*. Приложим в какой-либо точке A оси фигуры гироскопа постоянную силу $\mathbf{F}$, например подвесим в этой точке небольшой груз P. Когда маховички гироскопа не вращаются, наблюдается привычное явление: под действием веса груза правый маховичок опускается, левый — поднимается.

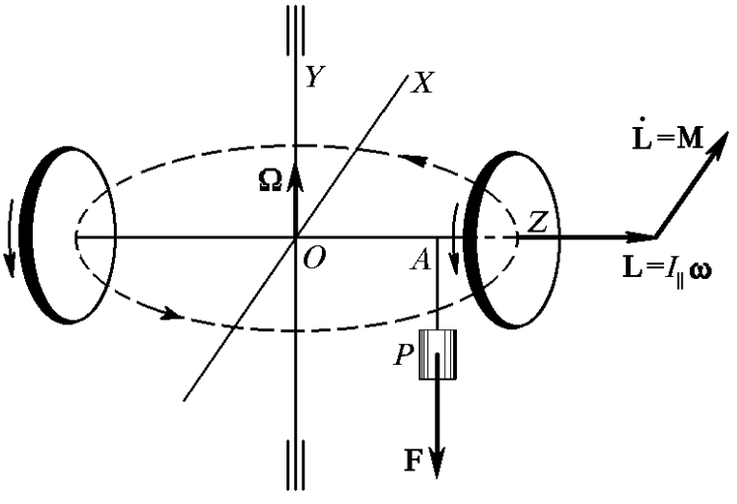


Рис. 150

Однако движение приобретает совсем иной характер, если предварительно маховички были приведены в быстрое вращение в одну и ту же сторону*). В этом случае ось фигуры гироскопа вместе с грузиком P не опускается, а начинает медленно вращаться с постоянной скоростью вокруг вертикальной оси OY. Такое вращение называется *вынужденной прецессией*. Вынужденная прецессия проще всего объясняется *приближенной теорией гироскопа*. Гироскопу всегда стремятся сообщить быстрое вращение вокруг оси его фигуры. Но вследствие различных причин гироскоп, вообще говоря, получает также вращение вокруг перпендикулярной оси. Специфические гироскопические эффекты проявляются тогда, когда это вращение является медленным по сравнению с вращением вокруг оси фигуры гироскопа. В приближенной теории им пренебрегают. В формуле (49.4) отбрасываем второе слагаемое, т. е. полагают

$$\mathbf{L} \approx I_{\parallel}\boldsymbol{\omega}_{\parallel} \approx I_{\parallel}\boldsymbol{\omega}. \qquad (50.1)$$

В этом приближении векторы $\boldsymbol{\omega}$ и $\mathbf{L}$ не отличаются по направлению, оба они направлены вдоль оси фигуры гироскопа. Поэтому о движении оси его фигуры можно судить по изменению направления вектора $\mathbf{L}$, описываемому уравнением (49.3). Если рассматривать $\mathbf{L}$ как радиус-вектор, то производная $\dot{\mathbf{L}}$ геометрически может быть истолкована как скорость движения конца вектора $\mathbf{L}$. Допустим, что точка приложения внешней силы $\mathbf{F}$ лежит на оси фигуры гироскопа. Момент этой силы будет $\mathbf{M} = [\mathbf{aF}]$, где $\mathbf{a}$ — радиус-вектор, проведенный от точки опоры гироскопа к точке приложения силы $\mathbf{F}$. В силу уравнения

*) В демонстрационных опытах маховичок гироскопа приводят в быстрое вращение, прижимая его обод к шкиву электромотора. Существуют гироскопы (например, волчки-компасы), которые сами представляют собой электродвигатели с вращающимся магнитным полем и приводятся во вращение трехфазным током.

(49.3) вектор «скорости» $\dot{\mathbf{L}}$ будет перпендикулярен к оси фигуры гироскопа Z. Такой момент сил может изменить только направление вектора $\mathbf{L}$, а не его длину. Следовательно, если внешняя сила $\mathbf{F}$ постоянна, то вектор $\mathbf{L}$, а с ним и ось фигуры гироскопа должны совершать равномерное вращение вокруг оси OY. Это вращение и есть вынужденная прецессия. Вектор угловой скорости прецессии $\boldsymbol{\Omega}$ в рассматриваемом примере направлен вдоль оси OY.

Если один из маховичков (см. рис. 150) закрутить в одну, а другой — в противоположную сторону с той же угловой скоростью, то прецессии не возникает. В этом случае $\mathbf{L} = 0$, и под действием груза P гироскоп поворачивается вокруг горизонтальной оси OX, как если бы его маховички не вращались.

2. Найдем длину вектора $\boldsymbol{\Omega}$. Вектор $\mathbf{L}$ изменяется только вследствие вращения с угловой скоростью прецессии $\boldsymbol{\Omega}$. Для линейной скорости движения его конца, т. е. производной $\dot{\mathbf{L}}$, можно написать $\dot{\mathbf{L}} = [\boldsymbol{\Omega}\mathbf{L}]$. Поэтому уравнение (49.3) дает

$$[\boldsymbol{\Omega}\mathbf{L}] = \mathbf{M}. \tag{50.2}$$

Из этого уравнения и можно найти угловую скорость прецессии $\boldsymbol{\Omega}$. В нашем примере вектор $\boldsymbol{\Omega}$ перпендикулярен к оси фигуры гироскопа, а потому

$$\Omega = \frac{M}{L} = \frac{M}{I_{\parallel}\,\omega}. \tag{50.3}$$

Легко найти вектор $\boldsymbol{\Omega}$ и в более общем случае, когда ось фигуры гироскопа наклонена к оси, вокруг которой совершается его прецессия. Для этого подставим в уравнение (50.2) выражение $\mathbf{M} = [\mathbf{aF}] = a[\mathbf{sF}]$, где $\mathbf{s}$ — единичный вектор вдоль оси фигуры гироскопа. Так как приближенная теория пренебрегает различием направлений вектора $\mathbf{L}$ и оси фигуры гироскопа, то $\mathbf{L} = L\mathbf{s}$. В результате уравнение (50.2) преобразуется к виду

$$L[\boldsymbol{\Omega}\mathbf{s}] = a[\mathbf{sF}].$$

Отсюда

$$\boldsymbol{\Omega} = -\frac{a}{L}\,\mathbf{F} = -\frac{a}{I_{\parallel}\omega_{\parallel}}\,\mathbf{F}. \tag{50.4}$$

Приведенные рассуждения справедливы при условии $\Omega \ll \omega$, т. е. для быстро вращающегося гироскопа. *Вращение гироскопа считается быстрым, если угловая скорость вращения вокруг его оси фигуры $\omega_{\parallel}$ очень велика по сравнению с угловой скоростью вращения вокруг перпендикулярной оси $\omega_{\perp}$.* В частности, она должна быть очень большой и по сравнению с угловой скоростью прецессии Ω. Для быстро вращающихся гироскопов, применяющихся в технике, угловая скорость прецессии Ω бывает в миллионы раз меньше ω.

3. Для демонстрации вынужденной прецессии совсем не обязательно, чтобы у гироскопа было два маховичка. Можно обойтись и одним маховичком. На рис. 151 изображен небольшой гироскоп с одним маховичком, подвешенный на нити. Вращающийся момент

M создается собственным весом **P** маховичка. Он и вызывает прецессию вокруг вертикальной оси. На схематическом рис. 152 тот же опыт воспроизведен в более крупном масштабе. Маховиком служит массивное велосипедное колесо с наращенной осью, приведенное в быстрое вращение. Колесо подвешивается на длинном проволочном канате за наращенный конец оси. Оси колеса придается приблизительно горизонтальное положение. Колесо прецессирует вокруг вертикальной оси под действием собственного веса. Опыт производит сильное впечатление. Уж очень неожиданным кажется движение колеса, когда оно не опускается под действием собственного веса, а непрерывно «уходит в бок».

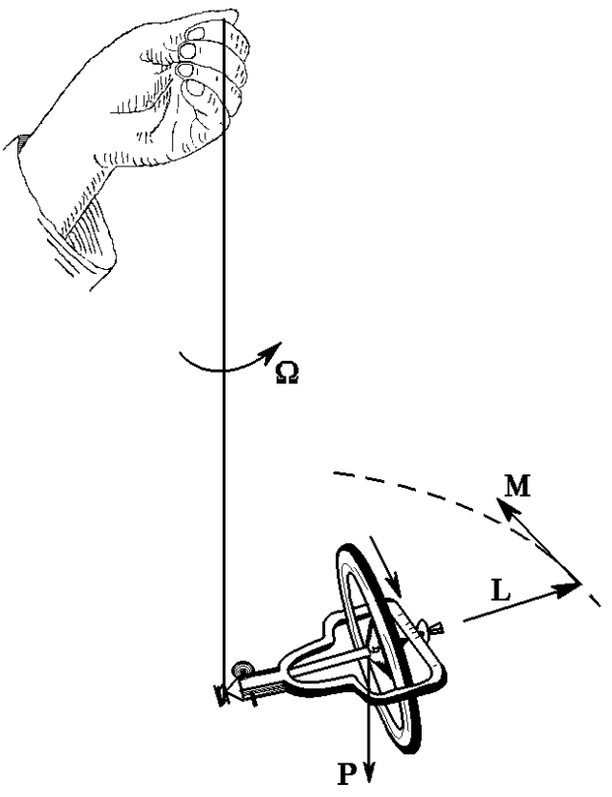


Рис. 151

4. Наконец, для наблюдения прецессии под действием собственного веса гироскопа нить также не обязательна. Можно взять симметричный гироскоп с неподвижной точкой опоры, расположенной на оси его фигуры. Точка опоры может находиться ниже центра масс (как в игрушечном волчке). Но она может находиться и выше центра масс. Тогда гироскоп называется *гироскопическим маятником*. В обоих случаях угловая скорость прецессии **Ω** определяется формулой (50.4), в которой следует положить $\mathbf{F} = m\mathbf{g}$. Для периода прецессии $T = 2\pi/\Omega$ получаем

$$T = 2\pi \frac{I_{\parallel}\omega}{amg}. \qquad (50.5)$$

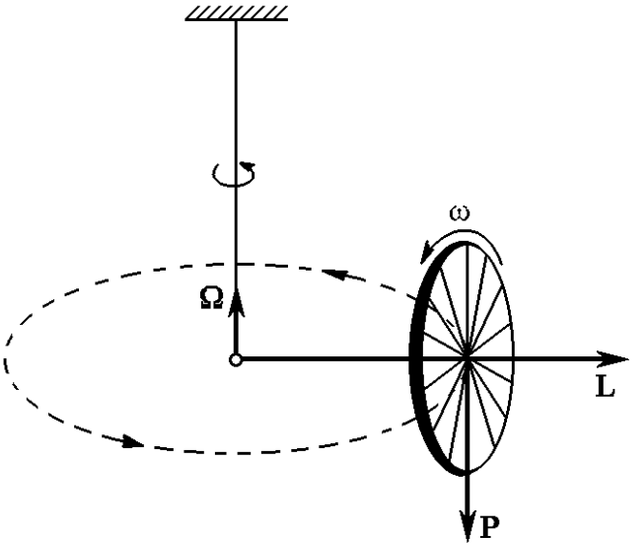


Рис. 152

В случае гироскопического маятника время T называется его *периодом*. Этому периоду можно привести в соответствие *приведенную длину* l гироскопического маятника по формуле

$$T = 2\pi \sqrt{\frac{l}{g}}. \qquad (50.6)$$

Она равна

$$l = \frac{I_{\parallel}^2 \omega^2}{m^2 a^2 g}. \qquad (50.7)$$

При больших скоростях вращения ω и малых a приведенная длина гироскопического маятника может быть сделана очень большой, а его период доведен до нескольких десятков минут. Направление оси фигуры такого гироскопического маятника очень мало подвержено влиянию кратковременных сил и толчков. Гироскопические маятники применяются на самолетах и судах для создания искусственного горизонта (см. § 52) и искусственной вертикали.

Пример. Гироскоп одного из авиагоризонтов характеризуется следующими параметрами: $m = 5 \cdot 10^3$ г, $I_{\parallel} = 8 \cdot 10^4$ г·см^2, $a = 0{,}25$ см. Гироскоп делает 20 000 об/мин и, следовательно, его угловая скорость $\omega = 2094$ рад/с. Подставляя эти данные в формулы (50.6) и (50.7), получим $l = 180$ км, $T = 860$ с $= 14$ мин 20 с. Угловая скорость вынужденной прецессии меньше угловой скорости вращения вокруг оси фигуры гироскопа примерно в $2{,}9 \cdot 10^5$ раз.

5. Поведение гироскопа при вынужденной прецессии на первый взгляд противоречит закону сохранения энергии. Без грузика P ось фигуры гироскопа — неподвижна (см. рис. 150). Как только к гироскопу подвешивается грузик, сразу же возникает прецессионное движение. С этим движением связана дополнительная кинетическая энергия гироскопа. Откуда взялась эта энергия? Единственная сила, которая могла сообщить гироскопу эту кинетическую энергию, есть вес грузика P. Но эта сила направлена вниз, она перпендикулярна к направлению прецессионного движения и поэтому работы не производит. Детальный ответ на этот и аналогичные вопросы дает *точная теория гироскопа*, излагаемая в § 52. Здесь мы ограничимся предварительными, в основном качественными, соображениями.

6. Исследуем сначала, как возбуждается регулярная прецессия. Допустим, что гироскоп приведен во вращение вокруг оси фигуры с угловой скоростью $\omega_{\parallel}$. Пусть ось фигуры совершает дополнительное равномерное вращение с угловой скоростью Ω вокруг неподвижной оси, составляющей произвольный угол с осью фигуры (рис. 153). Выясним, при каких условиях возможно такое движение. К собственному вращению вокруг оси фигуры гироскопа с угловой скоростью $\boldsymbol{\omega}$ при рассматриваемом движении добавляется еще вращение вокруг той же оси с угловой скоростью $\boldsymbol{\Omega}_{\parallel}$. Кроме того, гироскоп совершает дополнительное вращение вокруг перпендикулярной оси с угловой скоростью $\boldsymbol{\omega}_{\perp} = \boldsymbol{\Omega}_{\perp}$. Поэтому его момент импульса будет $\mathbf{L} = I_{\parallel}(\boldsymbol{\omega}_{\parallel} + \boldsymbol{\Omega}_{\parallel}) + I_{\perp}\boldsymbol{\Omega}_{\perp}$. Найдем производную этого вектора по времени. Так как все три вектора $\boldsymbol{\omega}_{\parallel}$, $\boldsymbol{\Omega}_{\parallel}$ и $\boldsymbol{\Omega}_{\perp}$ равномерно вращаются с угловой скоростью $\boldsymbol{\Omega}$, то с той же угловой скоростью будет вращаться и вектор $\mathbf{L}$. Поэтому

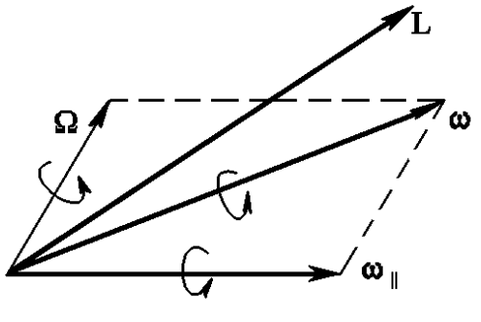


Рис. 153

его производная по времени будет $\dot{\mathbf{L}} = [\boldsymbol{\Omega}\mathbf{L}]$. С другой стороны, уравнение моментов требует, чтобы $\dot{\mathbf{L}} = [\boldsymbol{\Omega}_{\text{прец}}\mathbf{L}]$, где $\boldsymbol{\Omega}_{\text{прец}}$ — угловая скорость регулярной прецессии под действием момента внешних сил $\mathbf{M} = [\mathbf{aF}]$. Она определяется уравнением (50.2). Отсюда следует, что рассматриваемое движение гироскопа возможно, т. е. совместимо с уравнениями механики, если $\boldsymbol{\Omega} = \boldsymbol{\Omega}_{\text{прец}}$. Приведенное рассуждение является вполне точным. Оно показывает, что *вынужденная регулярная прецессия под действием постоянной силы* $\mathbf{F}$ *является одним из возможных точных видов движения, которые может совершать гироскоп.* Однако такое движение может возникнуть только при вполне определенных начальных условиях. В начальный момент надо, очевидно, сообщить вершине гироскопа начальную скорость $\mathbf{v}_0 = [\boldsymbol{\Omega}\mathbf{s}]$, т. е. привести ось фигуры гироскопа в равномерное вращение с угловой скоростью $\boldsymbol{\Omega}$. Энергия, требуемая для этого, сообщается внешним толчком, и никакого противоречия с законом сохранения не возникает. После прекращения действия толчка прецессионное движение гироскопа будет поддерживаться внешней силой $\mathbf{F}$. *Внешняя сила*, таким образом, *не вызывает, а лишь поддерживает регулярную прецессию.*

Поведение гироскопа при вынужденной регулярной прецессии в известном смысле аналогично поведению шарика, привязанного на нити, при равномерном вращении по окружности. Сила натяжения нити тянет шарик к центру окружности, но шарик все время движется перпендикулярно к ней, непрерывно «уходит в бок». Сила натяжения нити не создает, а лишь поддерживает равномерное вращение по окружности. Для создания такого вращения шарику необходимо сообщить дополнительный толчок в боковом направлении. Сила натяжения меняет только направление, но не модуль скорости. Если иметь в виду эту аналогию, то явление «ухода вбок» оси фигуры гироскопа при вынужденной регулярной прецессии представится, быть может, не таким уж странным, каким кажется на первый взгляд.

7. Приведенное исследование, однако, не отвечает на поставленный выше вопрос, касающийся сохранения энергии. Когда, например, мы вешаем грузик (см. рис. 150), то никакого начального толчка гироскопу при этом не сообщается. А между тем гироскоп начинает прецессировать. Ответ заключается в том, что в этом случае возникающая прецессия вовсе *не является регулярной.* На прецессионное движение вершины гироскопа накладываются мелкие вращения и дрожания ее. Они называются *нутациями.* В результате наложения нутаций на прецессионное движение вершина гироскопа описывает траектории *петлеобразного, циклоидального* или *синусообразного* типа, схематически изображенные на рис. 169. Вопрос этот будет разобран в § 52. Вид траектории зависит от начальной скорости, сообщенной оси фигуры гироскопа. Крупномасштабные нутации легко демонстрируются на опыте. Повесим на один из концов оси не очень быстро вращающегося гироскопа грузик (см. рис. 150). Если грузик опустить без начального толчка, то траектория вершины будет циклоидального типа (см. рис. 169 *б*). Если же сообщить гироскопу боковой толчок против направления прецессии, вызываемой грузиком, то на траектории появятся петли (см. рис. 169 *а*). При сообщении бокового толчка в направлении прецессии траектория вершины становится синусообразной (см. рис. 169 *в*). В последнем случае при надлежащей силе толчка траектория может стать круговой, а сама прецессия — регулярной.

Когда мы вешаем грузик, то в первый момент он начинает опускаться под действием силы тяжести. Конец же вектора момента импульса $\mathbf{L}$ приобретает скорость в боковом направлении в соответствии с уравнением $\dot{\mathbf{L}} = \mathbf{M}$. В результате вектор $\mathbf{L}$ перестает быть направленным вдоль оси фигуры гироскопа. Траектория вершины при опускании начинает загибаться вбок. Это и ведет к появлению нутаций. Работа силы тяжести при опускании грузика идет на приращение кинетической энергии прецессионно-нутационного движения. Достигнув нижнего положения, грузик начинает подниматься, кинетическая энергия прецессионно-нутационного движения переходит в потенциальную энергию поднятого грузика. Таким образом, идет непрерывное превращение потенциальной энергии в кинетическую и обратно. Никакого нарушения закона сохранения энергии не получается.

Если нутации малы, то прецессия называется *псевдорегулярной*. Для быстровращающихся технических гироскопов псевдорегулярная прецессия практически не отличается от регулярной. Здесь нутации представляют собой чрезвычайно мелкое и частое дрожание оси фигуры гироскопа, не имеющее никакого значения при изучении основного прецессионного движения (см. пример в § 52, п. 8). Кроме того, мелкомасштабные нутации быстро затухают под действием сил трения, и псевдорегулярная прецессия переходит в регулярную.

8. Существованием нутации объясняется и другое поведение гироскопа, кажущееся парадоксальным. Согласно уравнению (49.3) момент импульса гироскопа $\mathbf{L}$ изменяется только тогда, когда на него действуют внешние силы. Если действие внешних сил прекращается, то мгновенно прекращается изменение вектора $\mathbf{L}$, а с ним и прецессия гироскопа. Ось фигуры гироскопа становится неподвижной. Не противоречит ли закону инерции такая *безынерционность* оси фигуры гироскопа? Действительно, такое противоречие существовало бы, если бы указанная безынерционность относилась к движению самой оси фигуры гироскопа. На самом деле эта безынерционность относится не к оси фигуры, а к вектору $\mathbf{L}$. К выводу о безынерционности движения оси фигуры приводит приближенная теория гироскопа, пренебрегающая нутациями. Мы видим, таким образом, что учет нутаций устраняет противоречия с законом инерции.

9. Согласно изложенному в п. 6 *вынужденная регулярная прецессия должна продолжаться неограниченно долго, если только момент внешних сил* $\mathbf{M}$, *поддерживающих ее, остается постоянным.* Например, под действием груза P гироскоп на рис. 150 должен был бы как угодно долго совершать прецессионное вращение вокруг вертикальной оси Y. При этом груз P все время должен был бы находиться на одной и той же высоте. На самом деле груз P медленно и непрерывно опускается. Это объясняется действием *сил трения* и других *тормозящих сил.* Они создают *тормозящий вращающий момент* $\mathbf{M}_1$, направленный вниз — в отрицательную сторону оси Y. Теперь полный момент внешних сил, действующих на гироскоп, будет $\mathbf{M} + \mathbf{M}_1$. Согласно основному уравнению (49.3) производная $\dot{\mathbf{L}}$, т. е. линейная скорость вращения конца вектора $\mathbf{L}$, направлена вдоль результирующего момента $\mathbf{M} + \mathbf{M}_1$. Она имеет вертикальную составляющую, направленную по $\mathbf{M}_1$. В результате этого конец вектора $\mathbf{L}$, а с

ним и груз P будут опускаться. Можно сказать, что *тормозящий момент* $\mathbf{M}_1$ *вызывает дополнительную прецессию вокруг горизонтальной оси* X, *приводящую к опусканию груза* P.

Правильность такого объяснения легко подтвердить экспериментально. Будем подталкивать прецессирующий гироскоп, действуя против направления прецессионного вращения, вызванного грузом P. Груз начнет опускаться. Если, наоборот, подталкивать гироскоп в направлении прецессионного вращения, то груз начнет подниматься. В первом случае мы создаем вращательный момент $\mathbf{M}_1$, направленный вниз, во втором — вверх. Он вызывает прецессионное вращение вокруг горизонтальной оси X, опуская или поднимая груз. Таким образом, *если пытаться ускорить прецессию, то гироскоп отвечает на эту попытку поднятием груза. Если же тормозить прецессионное движение, то это ведет к опусканию груза.*

10. Интересно переформулировать полученное правило в свете общефизического *принципа Ле Шателье* (по имени французского ученого Ле Шателье (1850—1936)). Согласно этому принципу, установленному его автором в результате рассмотрения отдельных примеров, *на всякое внешнее воздействие система отвечает такими изменениями, которые стремятся ослабить это воздействие.* Когда мы подвешиваем грузик к оси фигуры вращающегося гироскопа, последний отвечает на это воздействие прецессией. Такая прецессия с точки зрения принципа Ле Шателье и должна рассматриваться как реакция гироскопа, ослабляющая внешнее воздействие, т. е. не позволяющая грузику опускаться. Ясно, что если такую реакцию усилить, т. е. вынудить гироскоп прецессировать быстрее, то этот эффект *только усилится.* Иными словами, грузик *должен подниматься.* Наоборот, искусственное замедление прецессионного вращения гироскопа эквивалентно *ослаблению реакции его на воздействие грузика.* В результате такого замедления грузик *должен опускаться.* Изложенная точка зрения оказывается полезной, например, при рассмотрении различного рода стабилизирующих гироскопических приборов, когда надо быстро получить ответ на вопрос, как надо воздействовать на гироскоп, чтобы усилить его стабилизирующее действие.

11. В свете изложенного становится ясным и поведение гироскопа с *двумя степенями свободы.* Для того чтобы свободный гироскоп обладал устойчивостью, а гироскоп под действием внешних сил мог совершать вынужденную прецессию в том виде, как она описана выше, необходимо, чтобы он обладал *тремя степенями свободы*, т. е. мог свободно вращаться вокруг всех трех осей, проходящих через точку опоры O. Одной из таких осей является ось фигуры гироскопа. Вращение вокруг такой оси должно быть возможно во всех случаях. Иначе ни о каких гироскопических эффектах говорить не приходится. Закрепим гироскоп так, чтобы вращение вокруг одной из остальных двух осей сделалось невозможным. Тогда гироскоп становится системой с двумя степенями свободы. Пусть, например,

уравновешенный гироскоп закреплен винтами так, что вращение вокруг вертикальной оси OY происходить не может (см. рис. 150). Если повесить грузик P, то гироскоп начнет вращаться вокруг горизонтальной оси OX как обычное твердое тело, а грузик P — опускается в направлении действия силы веса $\mathbf{F}$. Это радикальное изменение поведения гироскопа так же объясняется обычным правилом прецессии. Сила $\mathbf{F}$ создает момент $\mathbf{M}$, стремящийся вызвать прецессию гироскопа вокруг вертикальной оси OY. Так как гироскоп закреплен, то эта прецессия развиться не может. Она вызовет лишь деформацию кручения вертикальной оси. Со стороны подшипников на эту закрученную ось начнет действовать вращающийся момент, направленный вниз. Этот момент вызовет прецессию гироскопа — его вращение вокруг горизонтальной оси OX. Такому вращению ничто не препятствует. Поэтому указанная прецессия будет действительно происходить. В результате груз P начнет опускаться, что и наблюдается на опыте. Можно сказать, что закрученная ось тормозит прецессию, которая возникла бы под действием груза P, а такое торможение вызывает опускание груза.

Если, не вешая груза, толкнуть один из концов оси фигуры гироскопа вверх или вниз, то он будет продолжать по инерции вращаться в направлении сообщенного толчка вокруг горизонтальной оси OX. Дело опять в том, что во время действия толчка гироскоп стремится прецессировать вокруг вертикальной оси OY. Это приводит к деформации кручения вертикальной оси. Крутящий момент, действуя на гироскоп, будет поднимать или опускать его вершину, т. е. вызовет прецессионное вращение гироскопа вокруг оси OX в направлении сообщенного толчка. Пока происходит такое вращение, вертикальная ось продолжает оставаться деформированной (закрученной).

Таким же путем легко исследовать, как будет вести себя гироскоп, лишенный возможности вращаться вокруг горизонтальной оси OX. Рекомендуем читателю разобраться в этом вопросе.

12. Обычно у волчка центр масс располагается выше точки опоры. Чтобы такой волчок не упал, ему надо сообщить достаточно быстрое вращение вокруг оси фигуры. При медленном вращении вертикальное положение оси фигуры становится неустойчивым, и волчок падает. Если волчок запустить наклонно, то его центр масс начинает подниматься, и волчок принимает вертикальное положение. Причиной этого являются силы трения, действующие в точке опоры.

Посмотрим сначала, как двигался бы волчок, если бы трения не было. Пусть плоскость опоры горизонтальна и абсолютно гладкая. На волчок действуют две силы: сила веса $\mathbf{P}$ и сила нормального давления $\mathbf{F}$ со стороны плоскости опоры (рис. 154 *а*). Момент этих сил направлен горизонтально за плоскость рисунка. Этот момент вызовет прецессию волчка вокруг вертикальной оси. Для рассмотрения этой прецессии воспользуемся уравнением моментов относительно центра масс C. Оно имеет обычный вид (49.3). Пусть прецессия регулярна. Тогда центр масс будет оставаться неподвижным, а потому $\mathbf{F} + \mathbf{P} = 0$ в силу теоремы о движении центра масс. Момент

силы нормального давления **F** заставит вершину волчка уходить за плоскость рисунка. Волчок будет прецессировать в том же направлении, в каком происходит его вращение вокруг оси фигуры. А так как центр масс C остается неподвижным, то точка опоры O будет описывать окружность в направлении, обозначенном стрелками.

Учтем теперь силы трения скольжения, возникающие при движении точки опоры по указанной окружности. Существенно, что результирующая этих

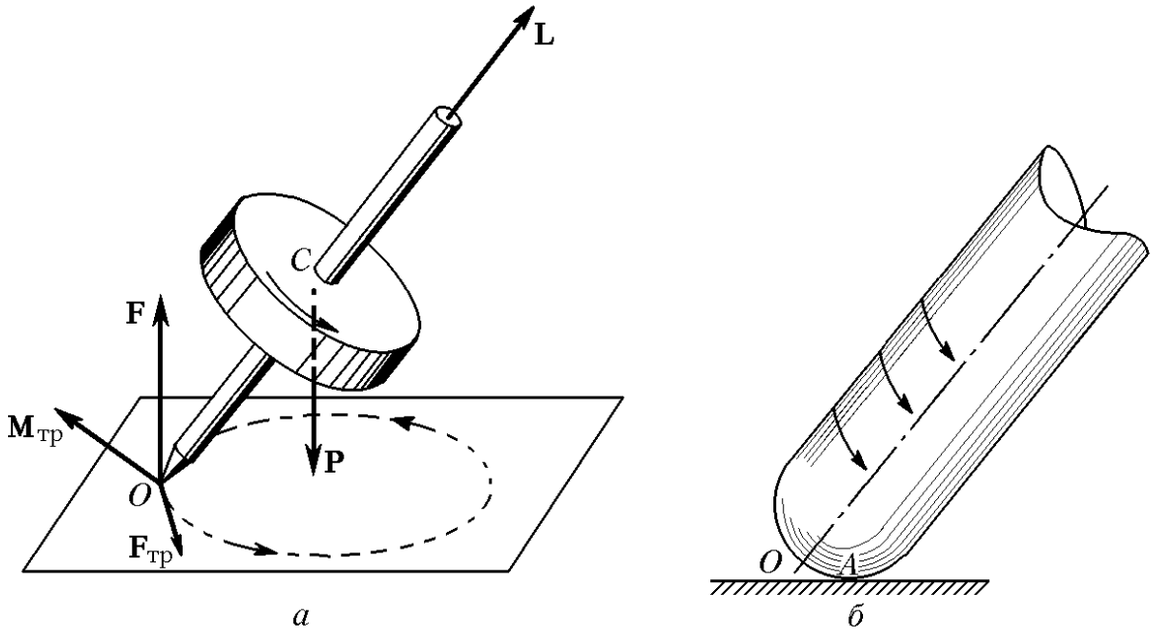


Рис. 154

сил $\mathbf{F}_{тр}$ направлена в ту же сторону, куда движется точка опоры волчка. Это станет ясно, если учесть, что конец ножки волчка, которым он опирается на плоскость опоры, не идеально острый, а закруглен. На рис. 154 *б* такой конец изображен в увеличенном масштабе. Если закрутить волчок вокруг оси фигуры, а затем поставить его на плоскость опоры, то в точке касания A возникнет скольжение. Ясно, что сила трения скольжения, действующая на ножку волчка, будет направлена против ее вращения, т. е. в ту же сторону, что и перемещение при прецессии точки опоры волчка. Заметим, что отмеченная особенность силы трения совершенно не зависит от того, как расположен центр масс относительно точки опоры. Сила трения скольжения $\mathbf{F}_{тр}$ всегда действует в направлении прецессии, вызванной весом волчка и давлением опоры, она стремится ускорить эту прецессию. Согласно общему правилу это должно привести к поднятию центра массы волчка. Если центр масс расположен выше точки опоры, то при таком поднятии ось фигуры волчка будет приближаться к вертикали.

Конечно, можно рассуждать и более детально. Предположим опять, что центр масс расположен выше точки опоры. Момент сил трения $\mathbf{F}_{тр}$ имеет вертикальную составляющую, направленную вверх. Он стремится вызвать прецессию вектора **L**, в результате которой вершина волчка должна подниматься. При неподвижности центра масс C такой подъем вызвал бы опускание точки опоры O. Но такому опусканию препятствует плоскость опоры, и оно происходить не может. Стремление к опусканию точки опоры проявится в возрастании силы нормального давления **F**. Последняя начнет превосходит силу веса **P**. Результирующая этих двух сил становится отличной от нуля.

Она направлена вверх и приводит к поднятию центра масс C, т. е. к выпрямлению оси волчка.

Сила трения $\mathbf{F}_{тр}$, действующая в точке опоры волчка, проявляется и в других явлениях. Она не только замедляет осевое вращение волчка, но и заставляет двигаться его центр масс. Волчок начинает «бегать» по плоскости опоры. Если бы не было других тормозящих сил (например, сопротивления воздуха и пр.), то такое движение непрерывно ускорялось бы. (Это лишний раз подтверждает, что направление силы $\mathbf{F}_{тр}$ совпадает с направлением движения волчка.) При этом скольжение постоянно уменьшается, и движение волчка в конце концов переходит в чистое качение. К этому моменту осевое вращение волчка обычно замедляется настолько сильно, что он теряет устойчивость и падает под действием силы тяжести.

13. Поднятие оси волчка можно объяснить и с помощью уравнения моментов относительно точки опоры O. Надо только принять во внимание, что точка опоры движется ускоренно, и брать уравнение моментов в виде (49.8). Короче говоря, надо учесть действие «сил инерции» (см. § 63 и 64), возникающих из-за ускоренного движения точки опоры. В разбираемом нами вопросе играет роль сила инерции, обусловленная касательным ускорением. Эта «сила» приложена к центру масс C и направлена противоположно касательному ускорению, т. е. на рис. 154 *а* — от читателя. Ее момент относительно точки опоры O имеет вертикальную составляющую, направленную вверх. Этот момент вызывает прецессию, в результате которой вершина волчка поднимается.

14. Случай, когда центр масс лежит ниже точки опоры, рассматривается совершенно аналогично. Силы трения скольжения, возникающие в точке опоры, и в этом случае приводят к поднятию центра масс. Однако здесь оно сопровождается опусканием вершины волчка. Волчок, запущенный наклонно, наклоняется еще больше.

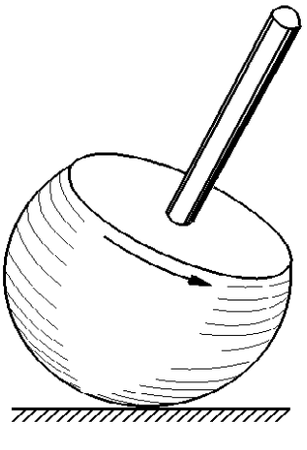


Рис. 155

15. Любопытным примером может служить *китайский волчок*, который имеет форму гриба (рис. 155). Из-за действия сил трения центр масс волчка непрерывно поднимается, а ось фигуры все более и более наклоняется. В конце концов это приводит к *опрокидыванию* волчка. Волчок становится на ножку. Ось фигуры все более и более приближается к вертикальному положению, если только волчок не потеряет устойчивость из-за замедления осевого вращения.

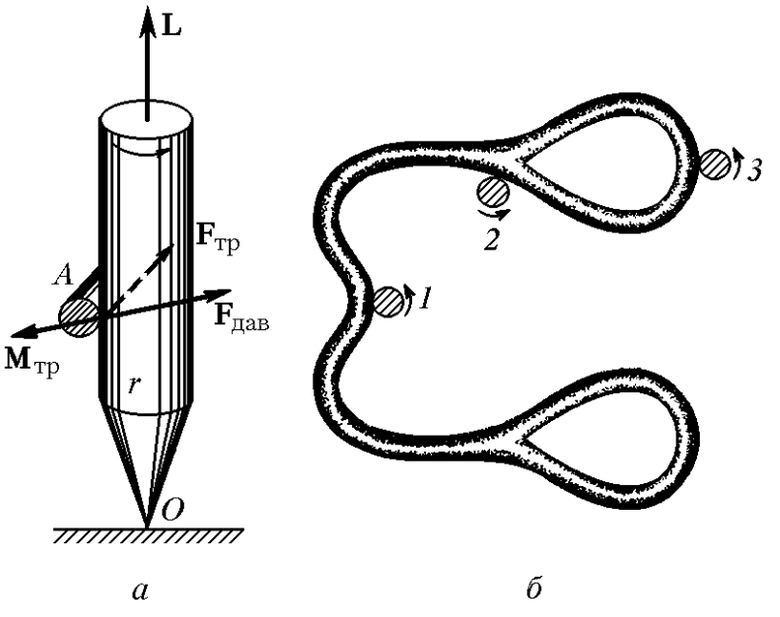
16. Для того чтобы повернуть ось фигуры гироскопа, к нему надо приложить силы, момент которых $\mathbf{M}$ определяется уравнением (50.2), в котором под $\boldsymbol{\Omega}$ следует понимать угловую скорость вынужденного вращения. Такие силы создаются, например, давлением подшипников на ось фигуры гироскопа. Ось фигуры гироскопа действует на подшипники с равными и противоположно направленными силами противодавления. Эти силы противодавления и создаваемые ими вращающие моменты называются *гироскопическими**). Гироскопические силы легко почувствовать, если взять за ось быстро вращающееся велосипедное колесо и попытаться повернуть эту ось. Колесо будет стремиться «вырваться из рук» в перпендикулярном

*) Термин «гироскопические силы» употребляется здесь в ином смысле, чем в § 24.

направлении. При быстром повороте требуется значительное усилие, чтобы удержать ось в руках.

В опыте со скамьей Жуковского (§ 34, п. 7) демонстратор, поворачивающий ось велосипедного колеса, испытывает с его стороны гироскопические силы. Момент этих сил направлен вертикально. Они и приводят во вращение скамью или изменяют ее угловую скорость. Гироскопические силы действуют на подшипники вала турбины корабля или винта самолета, когда при маневрах корабля и самолета направления их движения изменяются быстро. Они вызывают «рыскание по курсу». Для крупных судов это явление незаметно, оно наблюдается у мелких судов и самолетов.

17. Поднесем к оси гироскопа, изображенного на рис. 148, горизонтальный стержень. (Ось фигуры гироскопа изображена на рис. 156*а*. Сечение стержня плоскостью рисунка изображено заштрихованным кружком. Предполагается, что стержень перпендикулярен к плоскости рисунка.) Возникает сила трения скольжения $\mathbf{F}_{тр}$, параллельная стержню и, следовательно, перпендикулярная к плоскости рисунка. Она изображена штриховой стрелкой. Момент этой силы относительно точки опоры $\mathbf{M}_{тр} = [\mathbf{r}\,\mathbf{F}_{тр}]$ лежит в плоскости рисунка и направлен перпендикулярно к стержню и радиусу-вектору $\mathbf{r}$. Он будет стремится вызвать прецессию оси гироскопа в том же направлении. Но эта прецессия возникнуть не может, так как ей препятствует стержень. Она проявится только в том, что ось гироскопа будет прижиматься к стержню. В результате возникает сила давления $\mathbf{F}_{дав}$, действующая со стороны стержня на ось гироскопа. Действие этой силы проявится, во-первых, в увеличении силы $\mathbf{F}_{тр}$. Во-вторых, она вызовет прецессию оси фигуры гироскопа, заставляя эту ось перемещаться вдоль стержня в ту же сторону, куда направлена сила $\mathbf{F}_{тр}$, так как момент силы давления $\mathbf{M}_{дав} = [\mathbf{r}\,\mathbf{F}_{дав}]$ направлен в ту же сторону. Возрастание силы $\mathbf{F}_{тр}$, в свою очередь приведет к возрастанию силы $\mathbf{F}_{дав}$ и к ускорению прецессионного перемещения оси фигуры гироскопа вдоль стержня. Это ускорение прекратится, когда движение оси фигуры со скольжением перейдет в чистое качение. (В описываемом нами опыте для этого может не хватить времени из-за изменения наклона оси фигуры гироскопа при движении.) Начиная с этого момента сила $\mathbf{F}_{тр}$ практически обратится в нуль, сила $\mathbf{F}_{дав}$ станет постоянной, а прецессионное движение оси фигуры гироскопа вдоль стержня — равномерным.



а *б*

Рис. 156

Вместо стержня можно взять какой-либо замкнутый или незамкнутый контур произвольной формы из толстой проволоки. Для демонстрационных

опытов может служить контур, изображенный на рис. 156 *б*. Контур закрепляется на штативе в горизонтальном положении вблизи оси фигуры гироскопа. Если ось фигуры верхним концом привести в соприкосновение с контуром, то она начинает «бегать» по этому контуру, переходя последовательно из положения *1* в положение *2*, *3* и т. д. При этом ось фигуры гироскопа сильно прижимается к контуру. Это очень красивый демонстрационный опыт. Описанное явление называется *периметрическим движением* гироскопа.

ЗАДАЧИ

1. Герой романа Жюля Верна «Вверх дном» предлагал повернуть земную ось, выпустив с Земли тяжелый снаряд. Оценить, с какой минимальной скоростью v нужно выпустить на полюсе Земли снаряд массой $m = 1000$ т, чтобы повернуть в пространстве мгновенную ось вращения Земли на угол $\alpha = 1°$. Масса Земли $M = 6 \cdot 10^{21}$ т. Длина градуса земного меридиана $l = 111$ км. Землю считать однородным твердым шаром (см. задачи 23 и 24 к § 37).

Р е ш е н и е. Максимальный поворот получится, когда скорость снаряда v перпендикулярна к земной оси. Снаряд уносит момент импульса $\mathbf{L} = \frac{m[\mathbf{rv}]}{\sqrt{1-v^2/c^2}}$, перпендикулярный к скорости $\mathbf{v}$. Земля получает такой же момент в обратном направлении. При этом вектор угловой скорости вращения Земли $\boldsymbol{\omega}$ отклоняется вбок на угол $\alpha = L/I\omega$. Подставив сюда $I = {}^2/_5 Mr^2$ и учтя, что разность $c - v$ очень мала, получим

$$\frac{c-v}{c} \approx \frac{25\, m^2 c^2}{8M^2 l^2 \omega^2} \approx 1{,}3 \cdot 10^{-24}.$$

2. Симметричный волчок, ось фигуры которого наклонена под углом α к вертикали (см. рис. 154), совершает регулярную прецессию под действием силы тяжести. Точка опоры волчка O неподвижна. Определить, под каким углом β к вертикали направлена сила, с которой волчок действует на плоскость опоры.

О т в е т. $\operatorname{tg}\beta = \frac{a^3 m^2 g \sin\alpha}{I^2\omega^2}$.

3. Гироскопический маятник, используемый в качестве авиагоризонта характеризуется параметрами, приведенными в п. 4 этого параграфа. Когда самолет двигался равномерно, ось фигуры маятника была вертикальна. Затем в течение времени $\tau = 10$ с самолет двигался с горизонтальным ускорением $\dot{v}_0 = 1$ м/с^2. Определить угол α, на который отклонится от вертикали ось фигуры гироскопического маятника за время ускорения.

О т в е т. $\alpha \approx \frac{ma\dot{v}_0\tau}{I_\parallel\omega} \approx 0{,}43° \approx 25'$.

4. Однородный гладкий сплошной шар, находящийся на горизонтальном столе, быстро вертится вокруг своего вертикального диаметра с угловой скоростью $\boldsymbol{\omega}_0$ (рис. 157). В него ударяется второй, в точности такой же шар. Происходит абсолютно упругий удар без передачи вращения. Ударяемый шар начинает двигаться по столу со скольжением. Коэффициент трения скольжения k считается не зависящим от скорости. Найти угол α между мгновенной осью вращения ударяемого шара и вертикальной линией для

любого момента времени t, когда еще не прекратилось скольжение. Найти также значение этого угла в момент, когда движение переходит в чистое качение. Трением верчения и трением качения пренебречь.

Р е ш е н и е. После удара центр ударяемого шара начнет двигаться с начальной скоростью v_0. По теореме о движении центра масс его скорость в момент времени t будет $v = v_0 - kgt$. Пусть $\boldsymbol{\omega}$ — мгновенное значение вектора угловой скорости. Момент силы трения относительно центра шара будет $kmgr\mathbf{i}$, где $\mathbf{i}$ — единичный вектор, направленный за плоскость рисунка и перпендикулярный к ней.

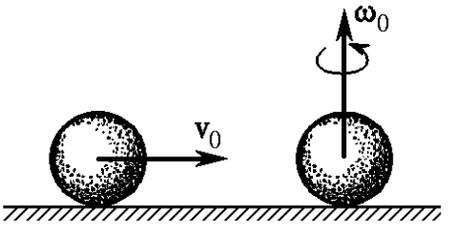


Рис. 157

Из уравнения моментов $I\frac{d\boldsymbol{\omega}}{dt} = kmgr\mathbf{i}$ получаем $\frac{2}{5}r\frac{d\boldsymbol{\omega}}{dt} = kg\mathbf{i}$. Отсюда $\boldsymbol{\omega} = \boldsymbol{\omega}_0 + 5kgt\mathbf{i}/2r$. Мгновенная ось вращения всегда лежит в плоскости, перпендикулярной к плоскости рисунка. Угол α определяется уравнением $\operatorname{tg}\alpha = 5kgt/(2r\omega_0)$. Определим теперь момент времени начала чистого качения. Скорость поступательного движения шара зависит только от горизонтальной составляющей вектора $\boldsymbol{\omega}$. Момент времени начала чистого качения найдется из условий $^5/_2\, kgt = v_0 - kgt$. С этого момента угол α становится и продолжает оставаться постоянным, причем $\operatorname{tg}\alpha = {}^5/_7\, v_0/r\omega_0$. Если $v_0 = \omega_0 r$, то $\operatorname{tg}\alpha = 5/7$, $\alpha = 35°32'$. Вращение шара вокруг фиксированного диаметра неустойчиво. Поэтому найденное решение определяет поворот оси вращения относительно *внешнего пространства*, а не внутри самого шара.

5. Гироскоп, изображенный на рис. 148, совершает установившееся параметрическое движение по круглому металлическому кольцу радиусом R, плоскость которого горизонтальна. Радиус стержня гироскопа r мал по сравнению с R ($r \ll R$). Ось гироскопа наклонена к вертикали под углом α. (В рассматриваемом случае она движется по поверхности кругового конуса с вершиной в точке опоры O.) Найти силу $\mathbf{F}_{\text{дав}}$, с которой стержень гироскопа давит на металлическое кольцо.

Р е ш е н и е. В установившемся режиме периметрическое движение оси фигуры гироскопа есть чистое качение. Если $r \ll R$, то угловая скорость прецессии найдется из условия $\Omega R = \omega r$. Пользуясь этим, легко найти искомую силу:

$$F_{\text{дав}} = \frac{I_{\parallel}\omega^2 r}{R^2}\sin^2\alpha.$$

Пусть $\omega = 100$ об/с $= 628$ рад/с, $I_{\parallel} = 2\cdot 10^4$ г·см^2, $r = 0{,}5$ см, $R = 8$ см, $\alpha = 20°$. Тогда $F_{\text{дав}} \approx 70$ Н.

6. Гироскопические эффекты используются в дисковых мельницах. Массивный цилиндрический каток (бегун), могущий вращаться вокруг своей геометрической оси, приводится во вращение вокруг вертикальной оси (с угловой скоростью Ω) и катится по горизонтальной опорной плите (рис. 158). Такое вращение можно рассматривать как вынужденную прецессию гироскопа, каковым является бегун. При вынужденной прецессии возрастает сила давления на горизонтальную плиту, по которой он катится. Эта сила растирает и измельчает материал, подсыпаемый под каток на плиту. Вычислить полную силу давления катка на опорную плиту.

Ответ. $F_{\text{дав}} = P + I_{\parallel}\Omega^2/r = P + {}^1/_2\, m\Omega^2 r$, где P — вес бегуна, m — его масса, r — радиус. Пусть $r = 50$ см. Тогда при рабочей скорости 1 об/с и, следовательно, $\Omega = 2\pi$ рад/с получаем ${}^1/_2\, m\Omega^2 r \approx mg = P$. Следовательно, $F_{\text{дав}} \approx 2P$. Обратить внимание, что полный момент импульса L не направлен вдоль оси фигуры бегуна, так как имеется еще момент, возникающий из-за вращения вокруг вертикальной оси. Однако последний момент остается неизменным при вращении катка, а потому при решении задачи его можно не принимать во внимание.

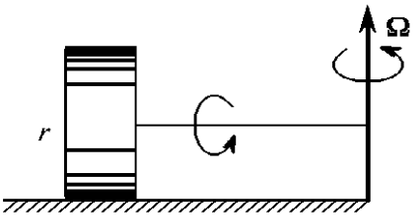


Рис. 158

7. Диск радиуса r, вращающийся вокруг собственной оси с угловой скоростью ω, катится без скольжения в наклонном положении по горизонтальной плоскости, описывая окружность за время T. Определить T и радиус окружности R, если $R \gg r$, а угол между горизонтальной плоскостью и плоскостью диска равен α.

Ответ. $T = \dfrac{3\pi\omega r}{g}\operatorname{tg}\alpha$, $R = \dfrac{3}{2}\dfrac{\omega^2 r^2}{g}\operatorname{tg}\alpha$.

§ 51. ПРИМЕНЕНИЯ ГИРОСКОПОВ

1. Научно-технические применения гироскопов весьма разнообразны. В курсе физики о них можно дать лишь общее представление. Рассмотрим принцип действия некоторых *гироскопических приборов*, совершенно отвлекаясь от деталей конструктивного или технического характера. Будем предполагать, что все приборы и условия, в которых они работают, являются *идеальными*. Так, будем считать, что сил трения и прочих вредных сил нет, что моменты инерции и моменты импульса кардановых колец пренебрежимо малы и т. д. В действительности все эти факторы оказывают существенное, иногда решающее, влияние на поведение реального гироскопа. Однако мы ограничиваем свою задачу выяснением лишь основных идей и принципов, на которых основано действие гироскопических приборов.

2. Начнем с *уравновешенного* (*астатического*) гироскопа с тремя степенями свободы. Пусть он быстро вращается вокруг своей оси фигуры. На направление оси фигуры гироскопа не оказывают влияния сила тяжести, вращение Земли, а также любые ускоренные движения точки опоры. В отсутствие сил, создающих вращающие моменты относительно точки опоры, ось фигуры уравновешенного гироскопа сохраняла бы неизменное направление относительно звезд. Если ось фигуры гироскопа направить на какую-либо звезду, то при перемещении последней по небесному своду она будет поворачиваться относительно Земли, оставаясь все время направленной на ту же звезду. Такой гироскоп позволяет обнаружить суточное вращение Земли, что и было впервые качественно продемонстрировано французским физиком Леоном Фуко (1819—1868). Трудности подобных опытов очень велики. Они связаны с тем, что невозможно полностью освободиться от неизбежного трения в подшипниках карданова подвеса и других вредных сил, создающих вращающие моменты относительно центра подвеса.

3. Свойство уравновешенного гироскопа сохранять неизменным направление оси своей фигуры используется для автоматического управления дви-

жением самодвижущихся мин (торпед), самолетов, судов, ракет и прочих аппаратов. Момент импульса гироскопа $\mathbf{L} = I_{\|}\boldsymbol{\omega}$ должен быть достаточно большим, чтобы уменьшить влияние трения в подшипниках карданова подвеса и прочих вредных сил. Ось фигуры вращающегося гироскопа задает курс движения аппарата. При всяком отклонении аппарата от курса (например, вследствие удара волн или действия порывов ветра) направление оси фигуры гироскопа в пространстве сохраняется. Значит, ось фигуры гироскопа вместе с рамами карданова подвеса поворачивается относительно движущегося аппарата. Поворот рам карданова подвеса с помощью тех или иных приспособлений включает двигатели, приводящие в действие *рули управления*. Последние и возвращают движение аппарата к заданному курсу. В случае торпеды, поскольку ее движение совершается в горизонтальной плоскости (по поверхности моря), достаточно одного гироскопа с осью фигуры, ориентированной в направлении движения. В случае самолета требуется два гироскопа. Один, с вертикальной осью, задает горизонтальную плоскость, в которой должен оставаться самолет. Другой, с горизонтальной осью, ориентированной вдоль оси самолета, задает его курс. Такими «*автопилотами*», освобождающими летчика от необходимости непрерывного управления самолетом, оборудованы почти все современные самолеты, предназначенные для длительных полетов.

4. Важным применением неуравновешенного гироскопа с тремя степенями свободы является создание *искусственных горизонта* и *вертикали*. Это необходимо в навигации в условиях отсутствия видимости линии горизонта. Направление вертикали в каждом месте земного шара можно просто определить с помощью обыкновенного маятника, применяемого в качестве отвеса. Однако такой способ не годится на корабле или самолете ввиду неизбежных ускорений, которые они получают при наборе скорости, поворотах, качке и пр. В этом случае вместо обыкновенного маятника используется *гироскопический маятник* (*гирогоризонт*) с очень большой приведенной длиной (см. § 50, п. 4). При отсутствии ускорения ось гироскопического маятника устанавливается вертикально. Если аппарат движется ускоренно, то появляется прецессия, уводящая ось маятника от вертикального положения. Однако если период прецессии T очень велик, а время ускорения мало по сравнению с T, то за это время прецессия, ввиду ее медленности, не успеет заметно отклонить ось гиромаятника от вертикали (см. задачу 3 к § 50). Эти условия соблюдаются, например, при поворотах движущегося аппарата. Время поворота всегда мало по сравнению с периодом T. Еще меньше чувствителен гиромаятник к качке корабля. Период качки всегда много меньше периода прецессии T, а главное, при качке ускорение за время T многократно и периодически меняет знак. Качка приводит лишь к малозаметным колебаниям оси гироскопического маятника около вертикального положения. Наиболее неблагоприятно на направление оси гиромаятника влияют увеличение и уменьшение скорости, которые могут длиться значительное время и вызывать хотя и не очень большие, но все же заметные отклонения оси фигуры гироскопа.

5. Важнейшим применением гироскопа является *гироскопический компас*, получивший широкое распространение на кораблях. Обычный магнитный компас подвержен действию разнообразных возмущений земного магнитного поля (магнитных бурь). На его показания влияют возмущения магнитного поля, вызываемые большими массами железа на корабле, а также

различные электродинамические воздействия со стороны сложного электротехнического оборудования корабля. В этих условиях использование магнитного компаса на корабле и становится практически невозможным. Гироскопический компас свободен от этих недостатков.

Идея гироскопического компаса впервые была высказана Фуко в 1852 г. Он предложил для этой цели использовать гироскоп с *двумя степенями свободы* в кардановом подвесе. Чтобы лучше уяснить идею гироскопического компаса, а также некоторых других гироскопических приборов, поставим вопрос шире и исследуем, как ведет себя гироскоп, закрепленный на вращающемся основании.

6. Закрепим неподвижно наружное кольцо карданова подвеса (см. рис 144). Гироскоп будет лишен той устойчивости, какая была свойственна ему, когда он обладал тремя степенями свободы. Причина этого, как было выяснено в предыдущем параграфе, заключается в том, что закрепление наружного кольца лишает гироскоп возможности совершать прецессию вокруг горизонтальной оси. Утраченную степень свободы можно, однако, в известной степени восстановить, если закрепить наружное карданово кольцо на основании, которое может свободно вращаться вокруг вертикальной оси. Такое закрепление просто эквивалентно увеличению момента инерции наружного кольца. Подвесим к оси гироскопа грузик. Он вызовет прецессионное вращение вокруг вертикальной оси. Это вращение передастся основанию, на котором закреплено карданово кольцо. Благодаря наличию у основания собственного момента инерции угловая скорость вращения его Ω будет меньше угловой скорости Ω_0, с которой прецессировал бы гироскоп, если бы он не был закреплен. Наличие основания, таким образом, ведет к торможению прецессии, вызванной грузиком. По этой причине грузик будет опускаться (см. § 50, пп. 9 и 10). Если, воздействуя на основание, увеличить угловую скорость его вращения, чтобы она сделалась равной Ω_0, то при прецессии грузик будет оставаться на постоянной высоте. Если же основание вращать со скоростью, большей Ω_0, то грузик начнет подниматься, пока ось фигуры гироскопа не примет вертикальное положение, причем положительный конец ее будет обращен вверх. При вращении со скоростью $\Omega < \Omega_0$ или в противоположном направлении ось фигуры установится также вертикально, но положительный конец ее будет обращен вниз.

Описанное поведение гироскопа объясняется тем, что гироскоп вместе с основанием вынужден вращаться вокруг вертикальной оси с угловой скоростью ω, отличной от скорости Ω_0, с которой он прецессировал бы только под действием веса грузика. Благодаря этому возникает деформация кручения вертикальной оси $D'D$ (см. рис. 144). Деформация кручения создает вращающийся момент $\mathbf{M}$, параллельный той же оси. Под действием этого момента возникает прецессионное вращение вокруг горизонтальной оси $B'B$, в результате которого ось гироскопа устанавливается вертикально в том или ином направлении, в зависимости от того, с какой скоростью и в какую сторону вращается основание. Такая ориентация оси фигуры гироскопа будет наблюдаться и в предельном случае, когда масса грузика равна нулю, т. е. когда грузика нет и $\Omega_0 = 0$. Легко сообразить, что в этом случае *ось фигуры гироскопа устанавливается параллельно оси вращения основания и притом так, что оба вращения совершаются в одинаковых направлениях* (правило Фуко). Про такие оси говорят, что они *одноименно параллельны.*

Все это легко демонстрировать с помощью небольшого гироскопа в карданове подвесе. Для большей наглядности мы заменили на рис. 159 круглые кольца прямоугольными рамками. Раскрутив гироскоп вокруг оси фигуры, возьмемся руками за нижнюю рамку и будем медленно поворачивать ее вокруг вертикальной оси. Ось гироскопа примет вертикальное положение и притом такое, что оба вращения — вращение рамки и вращение гироскопа вокруг собственной оси — будут происходить в одну и ту же сторону. Если начать вращать рамку в противоположном направлении, то произойдет «опрокидывание» гироскопа, т. е. поворот оси его фигуры вокруг горизонтальной оси $A'A$ на 180°. В результате оба вращения будут снова совершаться в одном направлении. И такое опрокидывание гироскопа будет наблюдаться всякий раз, когда мы меняем направление вращения наружной рамки. Во время опрокидывания гироскопа демонстратор испытывает заметное воздействие гироскопических сил, стремящихся повернуть его вокруг горизонтальной оси, перпендикулярно к плоскости наружной рамки.

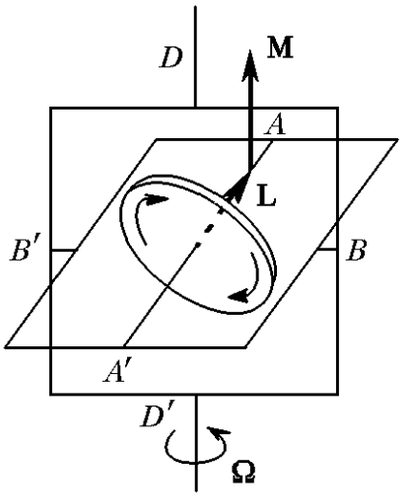


Рис. 159

В другой демонстрации уравновешенный гироскоп с двумя степенями свободы ставят на горизонтальный диск, который может вращаться вокруг вертикальной оси (рис. 160). При вращении диска ось гироскопа становится вертикально. При изменении направления на противоположное гироскоп опрокидывается.

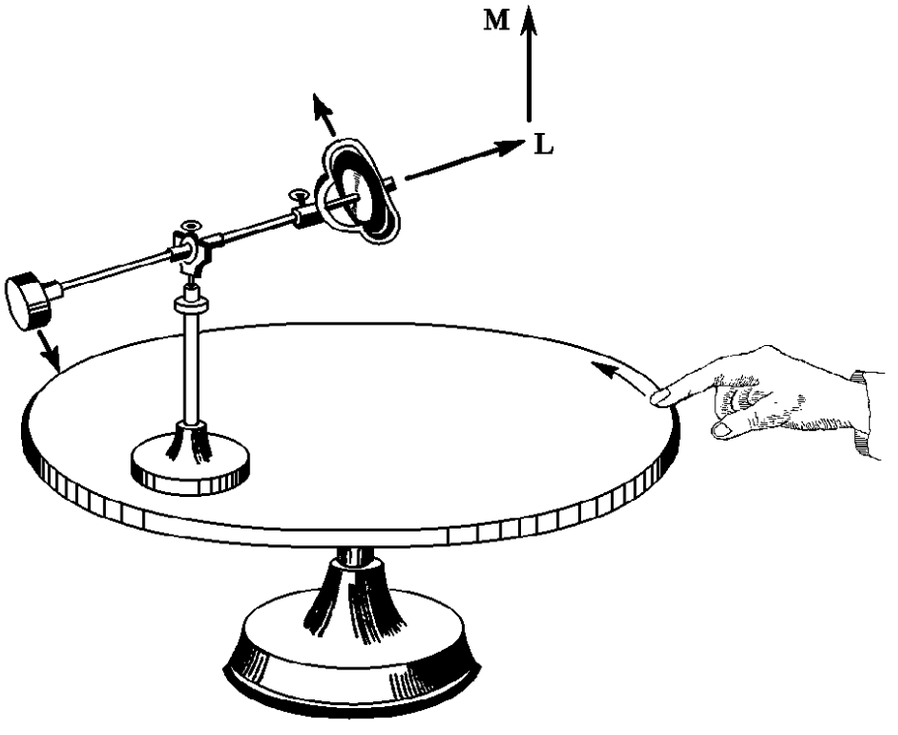


Рис. 160

7. Рассмотрим теперь идею гироскопического компаса, предложенную Фуко. Пусть наружное кольцо в карданове подвесе гироскопа может свободно вращаться вокруг вертикальной оси $D'D$ (см. рис. 159). Внутреннее

кольцо жестко закреплено в наружном под прямым углом. При этих условиях ось фигуры гироскопа $A'A$ вынуждена оставаться в горизонтальной плоскости, совпадающей с плоскостью внутреннего кольца. Она может свободно вращаться в этой плоскости вокруг вертикальной оси $D'D$. Гироскоп ставится на горизонтальную подставку. Последняя, конечно, участвует в суточном вращении Земли. Пусть $\boldsymbol{\Omega}$ — угловая скорость вращения Земли вокруг своей оси. Разложим вектор $\boldsymbol{\Omega}$ на вертикальную $\boldsymbol{\Omega}_в$ и горизонтальную $\boldsymbol{\Omega}_г$ составляющие. Вертикальная составляющая не влияет на поведение гироскопа, так как вокруг вертикальной оси он может вращаться совершенно свободно. Поэтому от наличия $\Omega_в$ можно отвлечься. Горизонтальная составляющая $\Omega_г$ лежит в плоскости меридиана, т. е. направлена вдоль *полуденной линии*. Таким образом, на поведение гироскопа сказывается вращение Земли лишь вокруг полуденной линии рассматриваемого места. Пусть плоскость рис. 161 совпадает с горизонтальной плоскостью. Разложим вектор $\Omega_г$ на составляющую вдоль оси фигуры гироскопа Ω_1 и составляющую Ω_2, к ней перпендикулярную. Первая составляющая на движение гироскопа не влияет, так как вращению вокруг его оси фигуры ничто не препятствует. Остается единственная составляющая $\boldsymbol{\Omega}_2$, изменение которой обусловливается вращением Земли. Гироскоп не может вращаться вокруг оси $\boldsymbol{\Omega}_2$, так как его ось фигуры вынуждена оставаться в горизонтальной плоскости (в плоскости рисунка). Но он может свободно вращаться вокруг своей оси фигуры и вокруг вертикальной оси, т. е. оси, перпендикулярной к плоскости рисунка. Тем самым рассматриваемый случай сведен к случаю, подробно разработанному в п. 6, причем роль вертикали играет направление вектора $\boldsymbol{\Omega}_2$. Поэтому ось фигуры гироскопа должна поворачиваться в плоскости рисунка в направлении к полуденной линии, стремясь стать одноименно параллельной оси вектора $\boldsymbol{\Omega}_2$. Однако при таком вращении длина составляющей $\boldsymbol{\Omega}_2$ уменьшается, а составляющей $\boldsymbol{\Omega}_1$ — увеличивается. Когда ось фигуры гироскопа установится параллельно полуденной линии, $\boldsymbol{\Omega}_2$ обратится в нуль. В этом положении поворот оси фигуры гироскопа, обусловленный вращением Земли, прекратится. При этом, в соответствии с правилом Фуко, собственное вращение гироскопа и вращение Земли вокруг полуденной линии будут происходить в одинаковых направлениях.

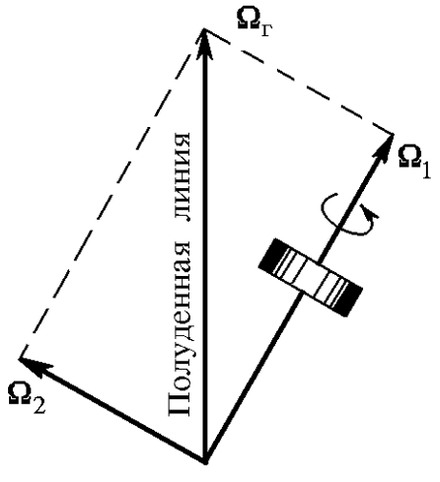


Рис. 161

Идею гирокомпаса Фуко можно уяснить и иначе. Вращение Земли стремится вызвать поворот оси фигуры гироскопа вокруг направления вектора $\boldsymbol{\Omega}_2$. Но такой поворот невозможен, поскольку он выводил бы ось фигуры гироскопа из плоскости рисунка, в которой она вынуждена находиться. Он проявляется лишь в деформациях и в появлении вследствие этого вращающегося момента $\mathbf{M}$, параллельного вектору $\boldsymbol{\Omega}_2$. Этот вращающийся момент передается гироскопу и вызывает прецессию вокруг вертикали, в результате которой ось фигуры гироскопа поворачивается к полуденной линии, стремясь стать одноименно параллельной ей.

8. Фуко указал также, что гироскоп с двумя степенями свободы может быть использован в качестве *инклинометра*, т. е. прибора для определения географической широты места. Закрепим неподвижно наружное кольцо карданова подвеса гироскопа так, чтобы его плоскость совпала с плоскостью географического меридиана. У гироскопа останутся две степени свободы. Он может вращаться вокруг оси своей фигуры и (вместе с внутренним кольцом) вокруг горизонтальной оси $B'B$, перпендикулярной к плоскости меридиана. Пусть плоскость рис. 162 совпадает с плоскостью географического меридиана. Ось фигуры гироскопа может вращаться только в плоскости рисунка вокруг горизонтальной оси $B'B$, перпендикулярной к этой плоскости. Из плоскости рисунка она выходить не может. Разложим угловую скорость осевого вращения Земли $\boldsymbol{\Omega}$ на составляющую $\boldsymbol{\Omega}_1$ вдоль оси фигуры гироскопа и составляющую $\boldsymbol{\Omega}_2$, к ней перпендикулярную. Первая составляющая роли не играет. Существенно только вращение вокруг оси вектора $\boldsymbol{\Omega}_2$. Мы пришли к той же ситуации, что и в предыдущем пункте при разборе гирокомпаса Фуко. Повторив приведенные там рассуждения, видим, что ось фигуры гироскопа будет поворачиваться по направлению к *оси мира* (т. е. к оси собственного вращения Земли). Этот поворот будет сопровождаться уменьшением длины вектора $\boldsymbol{\Omega}_2$. Когда ось фигуры гироскопа станет одноименно параллельной оси мира, вектор $\boldsymbol{\Omega}_2$ обратится в нуль, и дальнейший поворот гироскопа, вызванный осевым вращением Земли, прекратится. Таким образом, ось фигуры гироскопа устанавливается одноименно параллельно с осью мира. Угол между этим направлением и горизонтальной плоскостью и есть географическая широта рассматриваемого места.

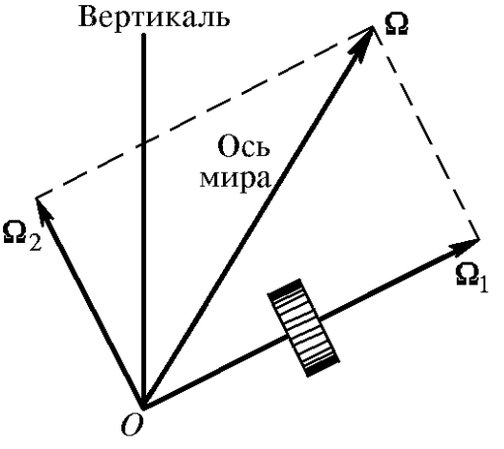


Рис. 162

9. Гирокомпас и гироинклинометр Фуко не получили практического применения. Они лишь теоретически решают поставленные перед нами задачи. Благодаря медленности вращения Земли силы, воздействующие на гироскоп из-за такого вращения, ничтожны и не в состоянии преодолеть (или способны преодолеть с трудом) трение в подшипниках этих приборов. Кроме того, такие приборы в принципе могли бы быть использованы только тогда, когда они установлены на неподвижном (относительно Земли) основании. Они не годятся на самолетах и судах, так как при движении последних развиваются угловые скорости вращения, а также ускорения, во много раз превосходящие соответствующие величины при суточном вращении земного шара.

Задача создания гирокомпаса была поставлена на практическую основу только после того, как стали использовать гироскоп не с двумя, а с *тремя степенями свободы*. Гироскоп должен быть *астатическим*. Но астатический гироскоп с тремя степенями свободы не подвержен влиянию вращения Земли. Эту трудность можно преодолеть, если взять гироскоп с каким-либо приспособлением, которое подвергалось бы воздействию указанного вращения и в свою очередь воздействовало бы на гироскоп. Поясним эту идею на примере одной из старых моделей гирокомпаса, по-

строенного известным американским строителем гироскопических приборов Сперри в 1911 г. и оказавшегося вполне пригодным навигационным прибором. Приспособлением, о котором говорилось выше, здесь является *маятник*, жестко связанный с внутренним кольцом карданова подвеса. Маятником служит тяжелая дуга Q, плоскость которой перпендикулярна к плоскости внутреннего кольца, а значит, параллельна плоскости маховичка гироскопа (рис. 163). Принцип действия прибора чрезвычайно прост. Допустим, что ось наружного кольца $D'D$ установлена вертикально, а плоскость внутреннего кольца вместе с осью гироскопа $A'A$ — горизонтально. В этом положении на гироскоп не действуют никакие моменты сил. Ось фигуры гироскопа $A'A$, если ее направить на какую-либо звезду, начнет двигаться вместе с ней. Допустим, что ось фигуры гироскопа отклонена от полуденной линии, например, к востоку (рис. 164). Звезда, на которую направлена эта ось, будет подниматься. Вместе с ней начнет подниматься и положительный конец оси фигуры гироскопа A. Но тогда начнет поворачиваться вокруг оси $B'B$ и дуга Q. Момент силы тяжести кольца $\mathbf{M}$ относительно точки O будет стремиться опустить точку A и вызовет прецессию гироскопа вокруг вертикальной оси $D'D$, в результате которой ось фигуры гироскопа будет поворачиваться к полуденной линии, стремясь установиться одноименно параллельно ей. То же самое произойдет, если первоначально положительный конец оси фигуры A был отклонен от полуденной линии к западу.

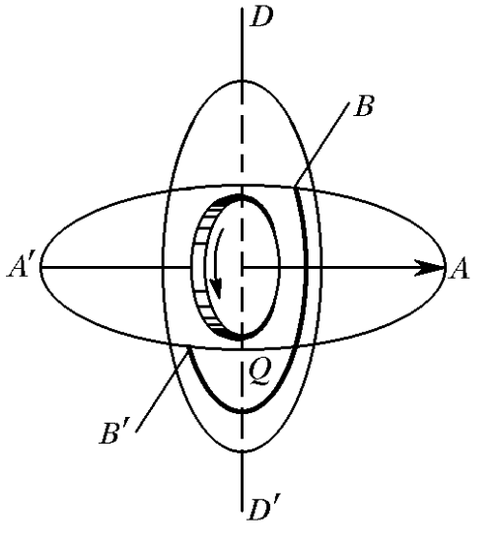


Рис. 163

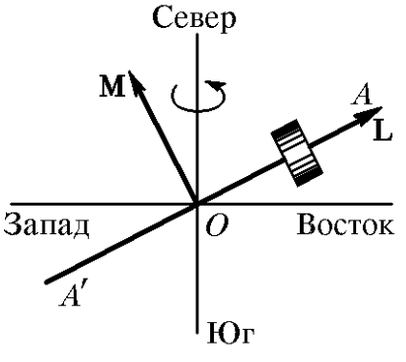
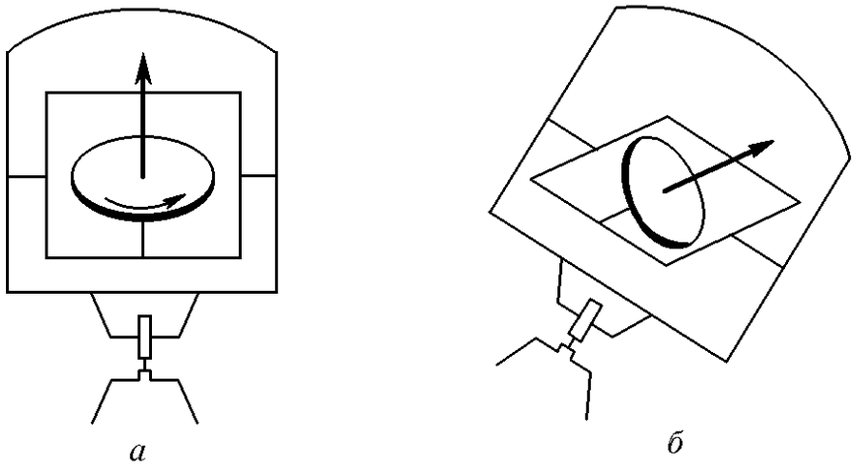


Рис. 164

10. В заключение рассмотрим идею однорельсовой железной дороги. Вагон, катящийся по одному рельсу, неустойчив. Для стабилизации его движения можно применить массивный гироскоп с тремя степенями свободы, установленный внутри вагона, как указано на рис. 165 *а*. Роль наружного



а *б*

Рис. 165

кольца карданова подвеса выполняют стенки вагона. Допустим, что вагон накренился вправо. Сила тяжести еще больше будет стремиться опрокинуть вагон в ту же сторону. Она создает вращающийся момент, направленный за плоскость рисунка параллельно продольной оси вагона. Через подшипники этот момент передастся гироскопу. Гироскоп начнет прецессировать, что вызовет наклон внутренней рамы (рис. 165 *б*). Если каким-либо способом ускорить эту прецессию, то возрастает вращающийся момент сил противодействия со стороны гироскопа (см. § 50, пп. 9 и 10). Центр тяжести вагона начнет подниматься, и вагон вернется в вертикальное положение. Такое вынужденное ускорение прецессионного движения рамы должно выполняться автоматически. В устройстве соответствующего автомата и заключается вся трудность практического осуществления идеи однорельсовой дороги.

§ 52. ОСНОВЫ ТОЧНОЙ ТЕОРИИ СИММЕТРИЧНОГО ГИРОСКОПА

1. Точная теория симметричного гироскопа учитывает различие направлений мгновенной оси вращения, оси фигуры и момента импульса гироскопа относительно его точки опоры. Она справедлива при любых соотношениях между любыми угловыми скоростями $\omega_{\parallel}$ и $\omega_{\perp}$, с которыми гироскоп вращается вокруг своей оси фигуры и перпендикулярной к ней оси. Однако наиболее важные гироскопические эффекты, которым гироскоп обязан своими научными и техническими применениями, проявляются лишь при соблюдении условия $\omega_{\parallel} \gg \omega_{\perp}$.

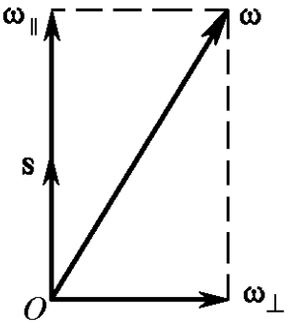


Рис. 166

Отложим от точки опоры O в положительном направлении оси фигуры гироскопа единичный вектор $\mathbf{s}$ (рис. 166). Конечная точка этого вектора называется *вершиной гироскопа*. Производная $\dot{\mathbf{s}}$ имеет смысл линейной скорости движения вершины гироскопа, а потому может быть представлена в виде $\dot{\mathbf{s}} = [\boldsymbol{\omega}\mathbf{s}] = [\boldsymbol{\omega}_{\perp}\mathbf{s}]$. Три вектора $\dot{\mathbf{s}}$, $\boldsymbol{\omega}_{\perp}$ и $\mathbf{s}$ взаимно перпендикулярны и образуют правовинтовую систему, как указано на рис. 167. Из этого рисунка видно, что $\boldsymbol{\omega}_{\perp} = [\mathbf{s}\dot{\mathbf{s}}]$. Поэтому

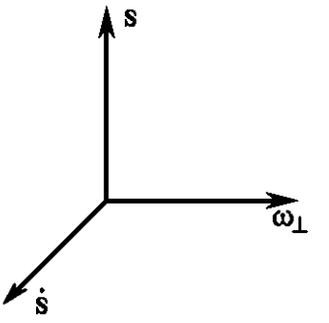


Рис. 167

$$\mathbf{L} = I_{\parallel}\dot{\boldsymbol{\omega}}_{\parallel} + I_{\perp}\boldsymbol{\omega}_{\perp} = I_{\parallel}\boldsymbol{\omega}_{\parallel}\mathbf{s} + I_{\perp}[\mathbf{s}\dot{\mathbf{s}}]. \qquad (52.1)$$

Подставив это выражение в уравнение (49.3), получим

$$I_{\parallel}\boldsymbol{\omega}_{\parallel}\mathbf{s} + I_{\parallel}\boldsymbol{\omega}_{\parallel}\dot{\mathbf{s}} + I_{\perp}[\mathbf{s}\ddot{\mathbf{s}}] = \mathbf{M}. \qquad (52.2)$$

Это основное уравнение точной теории симметричного гироскопа. Его удобно разделить на два уравнения. Первое уравнение получается из (52.2) скалярным умножением на $\dot{\mathbf{s}}$. С учетом соотношения $(\mathbf{s}\dot{\mathbf{s}}) = 0$ такое умножение дает

$$I_{\parallel}\dot{\boldsymbol{\omega}}_{\parallel} = \mathbf{M}_{\parallel}, \qquad (52.3)$$

где $\mathbf{M}_{\parallel} \equiv (\mathbf{Ms})$ — проекция вектора $\mathbf{M}$ на ось фигуры гироскопа. Второе уравнение найдем также из (52.2), но векторным умножением на $\mathbf{s}$. Учитывая при этом тождество $[\mathbf{s}[\mathbf{s}\ddot{\mathbf{s}}]] = -\mathbf{s}^2\ddot{\mathbf{s}} + (\mathbf{s}\ddot{\mathbf{s}})\mathbf{s} = -\ddot{\mathbf{s}} + (\mathbf{s}\ddot{\mathbf{s}})\mathbf{s}$, получим

$$I_{\parallel}\omega_{\parallel}[\mathbf{s}\dot{\mathbf{s}}] - I_{\perp}\ddot{\mathbf{s}} + I_{\perp}(\mathbf{s}\ddot{\mathbf{s}})\ \mathbf{s} = [\mathbf{sM}].$$

Дифференцируя соотношение $(\mathbf{s}\dot{\mathbf{s}}) = 0$, найдем $(\mathbf{s}\ddot{\mathbf{s}}) + \dot{\mathbf{s}}^2 = 0$. С учетом этого преобразуем последнее уравнение к виду

$$I_{\perp}\ddot{\mathbf{s}} = [\mathbf{Ms}] + I_{\parallel}\omega_{\parallel}[\mathbf{s}\dot{\mathbf{s}}] - I_{\perp}\dot{\mathbf{s}}^2\mathbf{s}. \tag{52.4}$$

2. Уравнение (52.3) определяет изменение во времени угловой скорости вращения гироскопа $\omega_{\parallel}$ вокруг оси фигуры. Оно совпадает с соответствующим уравнением вращения твердого тела вокруг закрепленной оси.

Уравнение (52.4) определяет ускорение $\ddot{\mathbf{s}}$, с которым движется вершина гироскопа. Запишем его в виде

$$\mathbf{I}_{\perp}\ddot{\mathbf{s}} = \mathbf{f}, \tag{52.5}$$

где введено обозначение

$$\mathbf{f} = [\mathbf{Ms}] + I_{\parallel}\omega_{\parallel}[\mathbf{s}\dot{\mathbf{s}}] - I_{\perp}\dot{\mathbf{s}}^2\mathbf{s}. \tag{52.6}$$

В этом виде уравнение (52.5) формально совпадает с уравнением Ньютона. Роль массы играет величина $I_{\perp}$, роль силы — вектор $\mathbf{f}$. Вершина гироскопа движется по поверхности неподвижной сферы единичного радиуса $\mathbf{s}^2 = 1$. Ее ускорение слагается из ускорения $(\ddot{\mathbf{s}})_{\perp}$, направленного по касательной к этой сфере, и радиального, или центростремительного, ускорения $(\ddot{\mathbf{s}})_{\parallel} = -\frac{\dot{s}^2}{s}\mathbf{s} = -\dot{\mathbf{s}}^2\mathbf{s}$, т. е. $\ddot{\mathbf{s}} = (\ddot{\mathbf{s}})_{\perp} - \dot{\mathbf{s}}^2\mathbf{s}$. Подставив это выражение в уравнение (52.5), видим, что центростремительное ускорение из него выпадает. Уравнение принимает вид

$$I_{\perp}(\ddot{\mathbf{s}})_{\perp} = [\mathbf{Ms}] + I_{\parallel}\omega_{\parallel}[\mathbf{s}\dot{\mathbf{s}}]. \tag{52.7}$$

Следовательно, уравнение (52.5) или эквивалентное ему уравнение (52.7) определяют не полное ускорение вершины гироскопа $\ddot{\mathbf{s}}$, а только его составляющую $(\ddot{\mathbf{s}})_{\perp}$, касательную к поверхности единичной сферы $\mathbf{s}^2 = 1$. Этого достаточно для нахождения движения вершины по начальным условиям (например, по начальному положению и начальной скорости вершины гироскопа). Действительно, движение вершины гироскопа аналогично движению не свободной, а связанной материальной точки, вынужденной находиться на заданной поверхности.

Воображаемую материальную точку, масса которой равна $I_{\perp}$, помещенную в вершине гироскопа, мы иногда будем называть *изображающей точкой*. На правую часть в уравнении (52.7) можно смотреть как на некоторую «силу», сообщающую ускорение изображающей точке. Первое слагаемое в этой «силе» связано с действием реальных сил, возникающих при взаимодействии гироскопа с окружающими телами. Его мы будем называть *реальной силой*

$$\mathbf{f}_{\text{реал}} = [\mathbf{Ms}].$$

Второе слагаемое $I_{\parallel}\omega_{\parallel}[\mathbf{s}\dot{\mathbf{s}}]$ к взаимодействию тел не имеет отношения. Это есть фиктивная «сила», возникающая при вращении гироскопа вокруг оси фигуры. Она называется *отклоняющей силой*:

$$\mathbf{f}_{\text{откл}} = I_{\parallel}\omega_{\parallel}[\mathbf{s}\dot{\mathbf{s}}]. \tag{52.9}$$

Отклоняющая сила отлична от нуля только тогда, когда изображающая точка движется. Она перпендикулярна как к оси фигуры гироскопа, так и к скорости движения изображающей точки. Эта сила стремится отклонить вершину гироскопа вбок от направления ее движения. Действием отклоняющей силы объясняются все характерные гироскопические эффекты.

Таким образом, основное уравнение движения симметричного гироскопа может быть записано в виде

$$I_{\perp}(\ddot{\mathbf{s}})_{\perp} = \mathbf{f}_{\text{реал}} + \mathbf{f}_{\text{откл}}. \tag{52.10}$$

3. Приближенная теория гироскопа рассматривает такие движения его, при которых ускорением $(\ddot{\mathbf{s}})_{\perp}$ в уравнении (52.10) можно пренебречь. Действительно, в этом случае $\mathbf{f}_{\text{реал}} + \mathbf{f}_{\text{откл}} = 0$ или

$$I_{\parallel}\omega_{\parallel}[\mathbf{s}\dot{\mathbf{s}}] + [\mathbf{M}\mathbf{s}] = 0.$$

Так как оба вектора $\mathbf{s}$ и $\mathbf{M}$ не имеют составляющих вдоль оси фигуры, то отсюда получаем

$$I_{\parallel}\omega_{\parallel}\dot{\mathbf{s}} = \mathbf{M},$$

а это и есть основное уравнение приближенной теории гироскопа.

4. К движению изображающей точки, поскольку оно описывается уравнением (52.10), формально можно применять все теоремы механики точки, например уравнение сохранения энергии. При этом надо только иметь в виду, что отклоняющая сила, как перпендикулярная к скорости $\dot{\mathbf{s}}$, работы не производит. Работа производится только реальной силой $\mathbf{f}_{\text{реал}}$.

5. На основе точного уравнения движения симметричного гироскопа можно, конечно, исследовать движение свободного гироскопа. Поскольку, однако, относящиеся сюда результаты уже были получены в § 49, мы не будем заниматься этим исследованием, а рассмотрим на основе точной теории вынужденную прецессию и нутации симметричного гироскопа.

Допустим, что действующая сила $\mathbf{F}$ постоянна и приложена в одной из точек оси фигуры гироскопа (рис. 168). Радиус-вектор этой точки, проведенной из точки опоры, обозначим через $\mathbf{a}$. Если точка опоры O не совпадает с центром масс гироскопа, то роль силы $\mathbf{F}$ может выполнять вес самого гироскопа. Момент силы $\mathbf{F}$ равен $\mathbf{M} = [\mathbf{a}\mathbf{F}] = [\mathbf{a}\mathbf{F}_{\perp}]$, где $\mathbf{F}_{\perp}$ — слагающая этой силы, перпендикулярна к оси фигуры гироскопа. Следовательно, $f_{\text{реал}} = [\mathbf{M}\mathbf{s}] = [[\mathbf{a}\mathbf{F}_{\perp}]\mathbf{s}] = a\mathbf{F}_{\perp}$, так как векторы $\mathbf{a}$ и $\mathbf{s}$ коллинеарны. Таким образом, уравнение (52.10) примет вид

$$I_{\perp}(\ddot{\mathbf{s}})_{\perp} = a\mathbf{F}_{\perp} + I_{\parallel}\omega_{\parallel}[\mathbf{s}\dot{\mathbf{s}}]. \tag{52.11}$$

Теперь поставим вопрос, можно ли вершине гироскопа сообщить такую начальную скорость, чтобы она совершала регулярную прецессию, т. е. равномерно вращалась вокруг оси, параллельной направлению действующей силы $\mathbf{F}$ и проходящей через точку гироскопа O. Угловую скорость такого вращения обозначим через $\boldsymbol{\Omega}$. Конкретно под силой $\mathbf{F}$ будем понимать вес

самого гироскопа: $\mathbf{F} = m\mathbf{g}$. За положительное направление вектора $\boldsymbol{\Omega}$ примем направление вверх, т. е. направление, противоположное силе $\mathbf{F}$ (см. рис. 168). Ответ на поставленный вопрос легко получить из уравнения (52.11). Для этого спроецируем уравнение (52.11) на направление вектора $\mathbf{F}_\perp$. Вершина гироскопа при регулярной прецессии движется со скоростью $\dot{\mathbf{s}} = [\boldsymbol{\Omega}\mathbf{s}]$ и ускорением $\ddot{\mathbf{s}} = -\Omega^2\mathbf{r}$, где $\mathbf{r}$ — радиус-вектор, проведенный от оси прецессионного вращения к вершине гироскопа ($r = s\sin\alpha = \sin\alpha$, причем α означает угол между осью фигуры гироскопа и вертикальным направлением). Взяв от ускорения $\ddot{\mathbf{s}}$ его составляющую, перпендикулярную к оси фигуры, и выполнив указанное проецирование, получим после сокращения на $\sin\alpha$

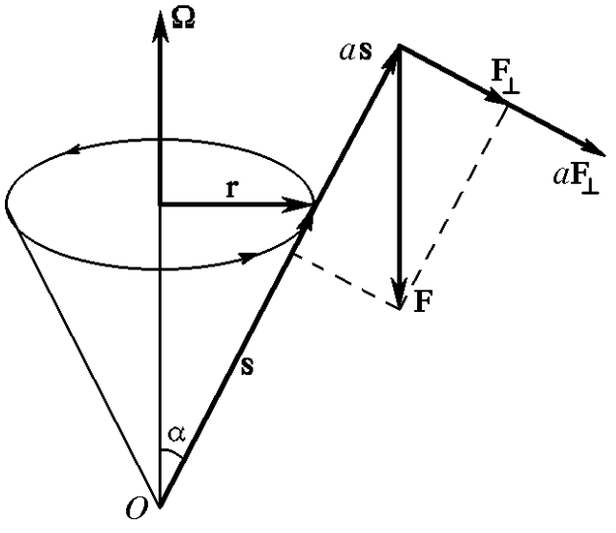


Рис. 168

$$I_\perp\Omega^2\cos\alpha - I_\parallel\omega_\parallel\Omega + aF = 0, \tag{52.12}$$

откуда

$$\Omega = \frac{1}{2I_\perp\cos\alpha}\left(I_\parallel\omega_\parallel \pm \sqrt{I_\parallel^2\omega_\parallel^2 - 4aFI_\perp\cos\alpha}\right).$$

Если центр масс гироскопа лежит выше точки опоры, то угол α — острый (см. рис. 168). В этом случае при недостаточно быстром собственном вращении гироскопа подкоренное выражение в формуле (52.13) может оказаться отрицательным. Тогда рассматриваемая регулярная прецессия становится невозможной, а положение гироскопа — неустойчивым. Вообще, для устойчивости гироскопа необходимо выполнение условия

$$I_\parallel^2\omega_\parallel^2 - 4aFI_\perp\cos\alpha > 0. \tag{52.14}$$

Это условие выполняется всегда, когда центр масс гироскопа лежит ниже точки опоры. Если же центр масс расположен выше точки опоры, то гироскоп должен вращаться достаточно быстро.

Допустим, что условие (52.14) выполнено. Тогда квадратное уравнение (52.12) имеет два вещественных корня. В этом случае регулярная прецессия возможна, и притом не одна, а две. Прецессия, которой соответствует меньший по абсолютной величине корень уравнения (52.12), называется *медленной*. Прецессия, соответствующая другому корню, называется *быстрой*.

6. Допустим, что выполнено условие $I_\parallel^2\omega_\parallel^2 \gg |4aFI_\perp\cos\alpha|$. Тогда для квадратного корня в формуле (52.13) можно написать приближенно

$$I_\parallel\omega_\parallel\left(1 - \frac{4aFI_\perp\cos\alpha}{I_\parallel^2\omega_\parallel^2}\right)^{1/2} \approx I_\parallel\omega_\parallel - \frac{2aFI_\perp\cos\alpha}{I_\parallel\omega_\parallel}.$$

В результате получится

$$\Omega_{\text{медл}} \approx \frac{aF}{I_\parallel\omega_\parallel}, \tag{52.15}$$

$$\Omega_{\text{быстр}} \approx \frac{I_\parallel}{I_\perp\cos\alpha}\,\omega_\parallel. \tag{52.16}$$

Формула (52.15) совпадает с формулой (50.4), к которой приводит приближенная теорема гироскопа. Таким образом, регулярная прецессия, о которой говорится в приближенной теории, есть *медленная прецессия*. Угловая скорость быстрой прецессии, как видно из формулы (52.16), по порядку величины совпадает с $\omega_{\parallel}$. Здесь не выполнено основное условие применимости приближенной теории $\omega_{\perp} \ll \omega_{\parallel}$. Поэтому быструю прецессию нельзя рассматривать в рамках приближенной теории. Регулярная прецессия свободного гироскопа, рассмотренная в § 49, есть частный случай быстрой прецессии, при котором $F = 0$.

7. Для того чтобы у читателя не сложилось впечатления, что быстрая прецессия является каким-то чисто умозрительным явлением, рассмотрим тривиальный пример конического маятника, когда $\omega_{\parallel} = 0$ и ни о каких гироскопических эффектах говорить не приходится. Разумеется, в этом случае центр масс должен лежать ниже точки подвеса. Поэтому угол α целесообразно заменить дополнительным углом $\beta = \pi - \alpha$, который ось маятника образует с вертикалью, направленной вниз. Формула (52.13) переходит в

$$\Omega = \pm\sqrt{\frac{aF}{I\cos\beta}},$$

т. е. в известную формулу круговой частоты конического маятника.

8. Регулярная прецессия, как медленная, так и быстрая, является весьма специальным частным случаем движения вершины гироскопа, реализующимся при вполне определенных начальных условиях. Для исследования общего случая в уравнении (52.7) сделаем замену $\dot{\mathbf{s}} = \mathbf{v}_{\text{п}} + \mathbf{v}_{\text{н}}$. Вектор $\mathbf{v}_{\text{п}}$ определим из условия $[\mathbf{Ms}] + I_{\parallel}\omega_{\parallel}[\mathbf{sv}_{\text{п}}] = 0$. Тогда $I_{\perp}(\ddot{\mathbf{s}})_{\perp} = I_{\parallel}\omega_{\parallel}[\mathbf{sv}_{\text{н}}]$. Величина $\mathbf{v}_{\text{п}}$ есть скорость вершины гироскопа, с которой она двигалась бы, если бы совершала медленную регулярную прецессию. (Вторая слагающая скорости $v_{\text{н}}$ будет описывать нутацию.) Если пренебречь ускорением при такой прецессии, то $\ddot{\mathbf{s}} = \mathbf{v}_{\text{н}}$, а потому

$$I_{\perp}\dot{\mathbf{v}}_{\text{н}} = I_{\parallel}\omega_{\parallel}[\mathbf{sv}_{\text{н}}], \tag{52.17}$$

причем мы опустили у $\mathbf{v}_{\text{н}}$ значок $\perp$, так как слагающая ускорения вдоль оси фигуры гироскопа сейчас не представляет интереса, и от нее можно отвлечься. Если на правую часть уравнения (52.17) смотреть как на аналог силы, то эта сила будет перпендикулярна к скорости $\mathbf{v}_{\text{н}}$, а потому она не может производить работы. Поэтому модуль скорости $\mathbf{v}_{\text{н}}$ меняться не может, и уравнение (52.17) описывает равномерное движение по окружности. Если r — радиус такой окружности, а $\Omega_{\text{н}}$ — угловая скорость вращения, то $v_{\text{н}} = \Omega_{\text{н}}r$, $|\dot{\mathbf{v}}_{\text{н}}| = \boldsymbol{\Omega}_{\text{н}}^2 r$. При этом ввиду перпендикулярности между $\mathbf{s}$ и $\mathbf{v}_{\text{н}}$ из уравнения (52.17) получается

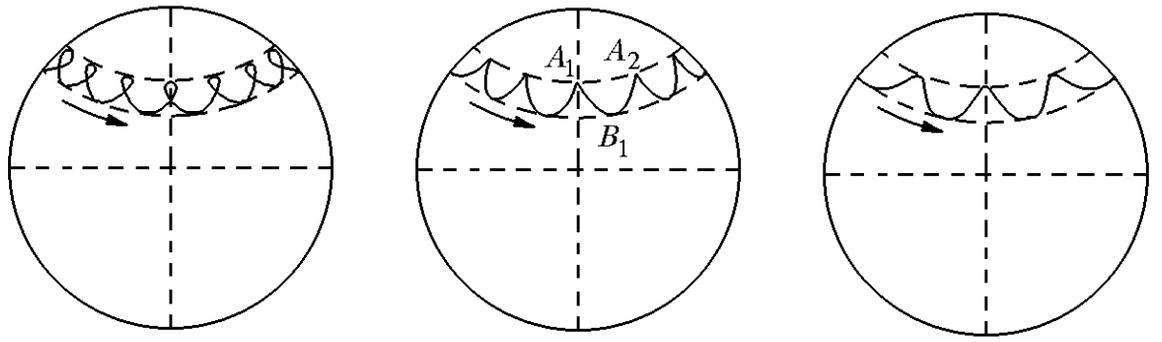
$$I_{\perp}\Omega_{\text{н}}^2 r = I_{\parallel}\omega_{\parallel}\Omega_{\text{н}}r,$$

откуда

$$\Omega_{\text{н}} = \frac{I_{\parallel}}{I_{\perp}}\omega_{\parallel}. \tag{52.18}$$

Таким образом, в общем случае на медленное прецессионное движение вершин гироскопа накладывается равномерное круговое движение с круговой частотой $\Omega_{\text{н}}$, определяемой уравнением (52.18). Радиус кругового движения равен $r = \frac{v_{\text{н}}}{\Omega_{\text{н}}} = \frac{v_{\text{н}}}{\omega_{\parallel}}\frac{I_{\perp}}{I_{\parallel}}$. В результате такого наложения траектория вершины

гироскопа может быть либо циклоидального типа (рис. 169 *б*), либо петлеобразного (рис. 169 *а*), либо она будет напоминать синусоиду (рис. 169 *в*). Какой из этих вариантов осуществляется в каждом конкретном случае, зависит от начальных условий, т. е. от положения вершины гироскопа в начальный момент времени и скорости, которая ей была сообщена в тот же момент. Наложением кругового движения на медленную прецессию и объ-



а *б* *в*

Рис. 169

ясняются *нутации*, о которых говорилось в § 50. Радиус кругового движения r есть не что иное, как амплитуда нутационных колебаний. При $r = 0$ нутаций не будет, и движение вершины перейдет в регулярную прецессию.

Пример. В авиагоризонте, рассмотренном в примере § 50, $I_\perp = {}^3/_5 I_\parallel$. Число нутаций на один прецессионный оборот равно

$$N = \frac{\Omega_н}{\Omega_п} = \frac{I_\parallel}{I_\perp}\frac{\omega_\parallel}{\Omega_п} = 4{,}77 \cdot 10^5.$$

Если начальная скорость вершины гироскопа равна нулю, то $\mathbf{v}_п + \mathbf{v}_н = 0$, а потому $r = v_п/\Omega_н$. Но $v_п = R\Omega_п$, где R — радиус прецессии. Таким образом,

$$\frac{r}{R} = \frac{\Omega_п}{\Omega_н} = \frac{1}{N} = \frac{1}{4{,}77 \cdot 10^5}.$$

Этот пример наглядно показывает, насколько мелким и частым дрожанием являются нутации в быстровращающихся технических гироскопах.

9. В заключение рассмотрим, как можно качественно объяснить характер траектории вершины гироскопа при наличии нутаций. Мы исходим непосредственно из уравнения движения вершины (52.11). Пусть на рис. 169 ось фигуры гироскопа своим положительным концом направлена в сторону читателя. Пусть в начальный момент времени вершина неподвижна и занимает положение A_1 (см. рис. 169 *б*). В этот момент скорость $\dot{\mathbf{s}}$, а потому и отклоняющая сила $I_\parallel \omega_\parallel [\mathbf{s}\dot{\mathbf{s}}]$ равны нулю. Под действием силы тяжести вершина получает скорость, направленную вниз. Но тогда появляется и боковая отклоняющая сила. Она начинает загибать траекторию вершины влево (если, встав на плоскость рисунка, идти в сторону движения). В положении B_1 скорость вершины становится горизонтальной, а отклоняющая сила — вертикальной. По величине отклоняющая сила превосходит силу веса, а вершина гироскопа начинает подниматься. В верхнем положении A_2 скорость вершины обращается в нуль. Это непосредственно следует из уравнения энергии, которому формально подчиняется движение вершины. Затем дви-

жение неограничено повторяется. Получается траектория циклоидального типа. Траектория с петлями рис. 169 *а* получится, если в начальный момент сообщить вершине скорость в направлении против прецессии. Если же начальная скорость сообщена в направлении прецессии, то получится траектория типа рис. 169 *в*. В последнем случае скорость можно подобрать такой, что возникает регулярная прецессия без нутаций.

§ 53. ТЕНЗОР И ЭЛЛИПСОИД ИНЕРЦИИ

1. Вычислим момент инерции I твердого тела относительно производной оси OA (рис. 170). Без ущерба для общности можно принять, что ось проходит через начало координат O. Координаты будем обозначить либо через x, y, z, либо через x_1, x_2, x_3. Таким образом, $x_1 \equiv x$, $x_2 \equiv y$, $x_3 \equiv z$. Разложим радиус-вектор $\mathbf{r}$ элемента массы тела dm на составляющие вдоль оси OA и перпендикулярную к ней: $\mathbf{r} = \mathbf{r}_{\parallel} + \mathbf{r}_{\perp}$. По определению момента инерции

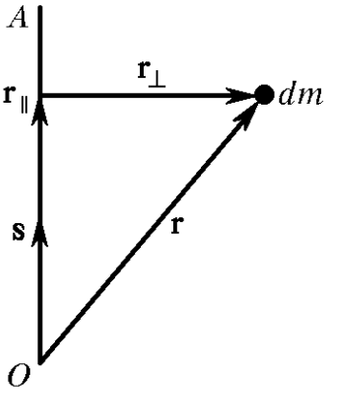


Рис. 170

$$I = \int \mathbf{r}_{\perp}^2 dm = \int (r^2 - r_{\parallel}^2)\, dm.$$

Если $\mathbf{s}$ — единичный вектор вдоль оси OA, то $r_{\parallel} = (\mathbf{rs}) = xs_x + ys_y + zs_z$. Кроме того, $r^2 = x^2 + y^2 + z^2$. Учтя эти соотношения, а также соотношение $s_x^2 + s_y^2 + s_z^2 = 1$, получим

$$I = I_{xx}s_x^2 + I_{yy}s_y^2 + I_{zz}s_z^2 + 2I_{xy}s_xs_y + 2I_{yz}s_ys_z + 2I_{zx}s_zs_x, \tag{53.1}$$

где I_{xx}, I_{yy}, I_{zz}, $I_{xy} \equiv I_{yx}$, $I_{yz} \equiv I_{zy}$, $I_{zx} \equiv I_{xz}$ — постоянные, определяемые выражениями

$$\begin{aligned} I_{xx} &= \int (y^2 + z^2)\, dm, \quad I_{xy} \equiv I_{yx} = -\int xy\, dm, \\ I_{yy} &= \int (z^2 + x^2)\, dm, \quad I_{yz} \equiv I_{zy} = -\int yz\, dm, \\ I_{zz} &= \int (x^2 + y^2)\, dm, \quad I_{zx} \equiv I_{xz} = -\int zx\, dm. \end{aligned} \tag{53.2}$$

Для этих постоянных будем пользоваться обозначениями I_{11}, I_{22}, ..., I_{13}. Величины I_{xx}, I_{yy}, I_{zz}, очевидно, имеют смысл моментов инерции тела относительно координатных осей X, Y, Z соответственно. Совокупность девяти величин

$$\begin{matrix} I_{xx} & I_{xy} & I_{xz} \\ I_{yx} & I_{yy} & I_{yz} \\ I_{zx} & I_{zy} & I_{zz} \end{matrix} \tag{53.3}$$

называются *тензором инерции тела относительно точки O*, а сами эти величины — *компонентами этого тензора**). *Тензор инерции симметри-*

*) Тензором называют упорядоченную совокупность девяти величин, заданную в каждой системе координат, причем при повороте координатных осей эти величины преобразуются как произведения компонентов двух векторов.

чен, т. е. $I_{ij} = I_{ji}$. Поэтому он полностью определяется заданием *шести* компонентов. Формулу (53.1) можно записать в более краткой и симметричной форме:

$$I = \sum_{i,\,j=1}^{3} I_{ij}\, s_i s_j. \tag{53.4}$$

Если известны для какой-либо координатной системы все шесть компонентов тензора инерции, то по формуле (53.1) или (53.4) можно вычислить момент инерции тела относительно произвольной оси, проходящей через начало координат O. Момент инерции относительно другой оси, не проходящей через начало координат, можно вычислить с помощью теоремы Гюйгенса—Штейнера.

2. Формула (53.4) допускает наглядную геометрическую интерпретацию. Через начало координат O будем проводить прямые во всевозможных направлениях и на них откладывать отрезки длиной $\mathbf{r} = 1/\sqrt{I}$. Геометрическим местом концов таких отрезков будет некоторая поверхность. Найдем ее уравнение. Согласно построению радиус-вектор точки, лежащей на этой поверхности, определяется выражением $\mathbf{r} = \mathbf{s}/\sqrt{I}$, а координаты той же точки — выражением $x_i = s_i/\sqrt{I}$. Исключая с помощью этих соотношений величины s_i из (53.4), получим уравнение искомой поверхности

$$\sum_{i,\,j} I_{ij} x_i x_j = 1. \tag{53.5}$$

Эта поверхность второго порядка, очевидно, является эллипсоидом, так как момент инерции I, а с ним и длина радиуса-вектора $\mathbf{r}$ имеют конечные значения, каково бы ни было направление оси $\mathbf{s}$. Она называется *эллипсоидом инерции* тела относительно точки O, являющейся его центром. При перемещении начала координат O относительно тела будет меняться и эллипсоид инерции тела. Если в качестве O взят центр масс тела, то и соответствующий эллипсоид называется *центральным*.

3. Как и всякий тензор, тензор инерции зависит от выбора начала координат и направления координатных осей. При изменении координатной системы меняются и значения компонентов тензора инерции тела. Существенно, однако, что какова бы ни была координатная система, всегда могут быть найдены все шесть компонентов тензора инерции, хотя бы по формулам (53.2). В частности, координатные оси можно направить вдоль главных осей эллипсоида инерции. В этой координатной системе в уравнении (53.5) пропадают члены, содержащие произведения координат, и это уравнение примет вид

$$I_x x^2 + I_y y^2 + I_z z^2 = 1, \tag{53.6}$$

или

$$\sum_{i} I_i x_i^2 = 1. \tag{53.7}$$

Тензор инерции приводится к *диагональному* виду

$$\begin{matrix} I_x & 0 & 0 \\ 0 & I_y & 0 \\ 0 & I_z & 0, \end{matrix} \tag{53.8}$$

причем *диагональные элементы* тензора мы обозначили с помощью одного индекса. Второй индекс в системе главных осей эллипсоида инерции опущен как излишний.

Таким образом, *для всякого твердого тела, где бы ни было выбрано начало координат O, существуют три взаимно перпендикулярные оси, совпадающие с главными осями эллипсоида инерции тела относительно точки O, для которых недиагональные элементы тензора инерции обращаются в нуль.* Эти оси называются также *главными осями тензора инерции.* Они, очевидно, жестко связаны с телом. Точно так же жестко связан с твердым телом и эллипсоид инерции. Если известно положение эллипсоида инерции, то в тот же момент будет известно и положение всего тела. Поэтому задача о вращении твердого тела вокруг неподвижной точки сводится к задаче о вращении его эллипсоида инерции вокруг той же точки. Этим воспользовался французский математик и механик Луи Пуансо (1777—1859) для наглядной геометрической интерпретации вращения твердого тела вокруг неподвижной точки. Она будет рассмотрена в следующем параграфе. Главные оси центрального эллипсоида инерции называют также *главными осями самого тела.*

Направление главных осей тела часто можно определить, пользуясь соображениями симметрии. Так, например, главные оси однородного прямоугольного параллелепипеда параллельны его ребрам. Если тело обладает симметрией вращения вокруг некоторой оси, то его эллипсоид инерции обладает такой же симметрией. К телам такого родя относится, например, цилиндр. В этом случае моменты инерции тела относительно всех осей, перпендикулярных к оси симметрии, одинаковы. Одной из главных осей тела является ось его симметрии. Всякая прямая, к ней перпендикулярная, также будет главной осью тела. Таким образом, существует бесконечное множество троек взаимно перпендикулярных главных осей тела, у которых одна ось, а именно ось симметрии, будет общей. Для шара эллипсоиды инерции относительно всех осей, проходящих через центр шара, одинаковы. В этом случае любая ось будет главной осью тела.

Для динамики вращательного движения твердого тела существенна симметрия не самого тела, а симметрия соответствующего ему эллипсоида инерции. Все тела с одинаковыми эллипсоидами инерции динамически эквивалентны. Чтобы эллипсоид инерции обладал симметрией вращения, не обязательно, чтобы само тело обладало той же симметрией. Возьмем, например, однородный параллелепипед с квадратным основанием. Поместим начало координат O в любой точке геометрической оси параллелепипеда. Тогда нетрудно показать, что эллипсоид инерции будет эллипсоидом вращения, ось симметрии которого совпадает с геометрической осью параллелепипеда. В динамическом отношении движение такого параллелепипеда описывается такими же уравнениями, что и движение однородного цилиндра. Если параллелепипед вырождается в куб, а начало координат помещено в его центре, то эллипсоид инерции вырождается в сферу. В динамическом отношении однородный куб ведет себя так же, как однородный шар.

4. Допустим теперь, что твердое тело равномерно вращается вокруг закрепленной оси, например оси, проходящей через неподвижные подшипники. Со стороны подшипников тело подвергается действию сил. Пусть это единственные внешние силы, действующие на тело. Их равнодействующая $\mathbf{F}$ найдется по теореме о движении центра масс. Она равна

$$\mathbf{F} = -m\omega^2\mathbf{r}_C,$$

где $\mathbf{r}_C$ — радиус-вектор центра масс тела, проведенный от оси вращения перпендикулярно к ней. Момент внешних сил относительно начала координат равен

$$\mathbf{M} = -\int [\mathbf{r}\omega^2\mathbf{r}_\perp]\, dm = \omega^2 \int [\mathbf{r}_\perp \mathbf{r}_\parallel]\, dm.$$

Примем ось вращения за координатную ось X, тогда $\mathbf{r}_\parallel = x\mathbf{i}$, $\mathbf{r}_\perp = y\mathbf{j} + z\mathbf{k}$. Учтя соотношения $[\mathbf{ij}] = \mathbf{k}$, $[\mathbf{ik}] = -\mathbf{j}$, получим

$$\mathbf{M} = \omega^2\mathbf{j} \int zx\, dm - \omega^2\mathbf{k} \int xy\, dm,$$

или

$$\mathbf{M} = \omega^2 (I_{xy}\mathbf{k} - I_{zx}\mathbf{j}).$$

Уберем подшипники и спросим себя, при каких условиях движение тела не изменится, т. е. останется вращением вокруг прежней оси X. Для этого необходимо, чтобы $\mathbf{F} = \mathbf{M} = 0$. Следовательно, ось вращения должна проходить через центр масс и, кроме того, должно быть $I_{zx} = I_{xy} = 0$. Последнее условие означает, что ось вращения должна быть одной из главных осей тела. Найденные условия являются и достаточными. Это следует из того, что при их выполнении удаление подшипников не меняет уравнения движения центра масс и уравнения моментов относительно центра масс. Эти же уравнения (при заданных начальных условиях) однозначно определяют движение твердого тела.

5. Итак, *во всяком твердом теле существуют три взаимно перпендикулярные оси, совпадающие с главными осями центрального эллипсоида инерции тела, вокруг которых тело может вращаться без воздействия внешних сил.* Такие оси называются поэтому *свободными* или *перманентными осями вращения*. Последним термином хотят подчеркнуть, что вращение твердого тела по инерции в отсутствие возмущений может продолжаться сколь угодно долго. Иное дело, будет ли это вращение устойчивым по отношению к малым возмущениям, с которыми в реальных условиях всегда надо считаться. Если при наличии таковых характер движения тела меняется мало, т. е. мгновенная ось вращения хотя и непрерывно изменяет свое положение в теле и пространстве, но все время проходит очень близко от соответствующей свободной оси, то вращение вокруг последней будет *устойчивым*. Если же сколь угодно малое возмущение существенно меняет характер движения тела, т. е. далеко уводит мгновенную ось от исходного направления, вокруг которого первоначально вращалось тело, то это вращение называется *неустойчивым*. В следующем параграфе будет показано, что *вращение вокруг оси с наибольшим или наименьшим моментом инерции является устойчивым, а вращение вокруг оси с промежуточным значением момента инерции — неустойчивым*. Для демонстрации можно взять картонную коробку прямоугольной формы, у которой длины всех ребер различны. Ось с наибольшим моментом инерции будет, очевидно, параллельна наиболее короткому ребру, с наименьшим моментом инерции — наиболее длинному ребру, с промежуточным — ребру промежуточной длины. Коробку подбрасывают вверх, сообщая ей быстрое вращение вокруг одной из этих осей. Во время полета ось вращения сохраняется, если она является осью с наибольшим или наименьшим моментом инерции. Если же первоначальное вращение было сообщено вокруг оси с промежуточным значением момента инер-

ции, то мгновенная ось вращения во время полета коробки непрерывно качается, далеко уходя от первоначального направления в теле. Движение коробки приобретает сложный и запутанный характер.

6. Допустим теперь, что твердое тело вращается вокруг какой-то закрепленной или мгновенной оси OA с постоянной или непостоянной угловой скоростью $\boldsymbol{\omega}$. Найдем его момент импульса $\mathbf{L}$ относительно начала координат O, а также кинетическую энергию $\mathbf{K}$. По определению

$$\mathbf{L} = \int [\mathbf{rv}]\, dm.$$

Подставим сюда $\mathbf{v} = [\boldsymbol{\omega}\mathbf{r}]$ и воспользуемся формулой $[\mathbf{r}[\boldsymbol{\omega}\mathbf{r}]] = \mathbf{r}^2\boldsymbol{\omega} - (\boldsymbol{\omega}\mathbf{r})\mathbf{r}$. Тогда получим

$$\mathbf{L} = \boldsymbol{\omega} \int \mathbf{r}^2\, dm - \int (\boldsymbol{\omega}\mathbf{r})\mathbf{r}\, dm.$$

В проекциях на координатные оси это соотношение записывается так:

$$\begin{aligned} L_x &= I_{xx}\omega_x + I_{xy}\omega_y + I_{xz}\omega_z, \\ L_y &= I_{yx}\omega_x + I_{yy}\omega_y + I_{yz}\omega_z, \\ L_z &= I_{zx}\omega_x + I_{zy}\omega_y + I_{zz}\omega_z, \end{aligned} \tag{53.9}$$

или короче

$$L_i = \sum_{j=1}^{3} I_{ij}\omega_j \quad (i = 1, 2, 3). \tag{53.10}$$

Таким образом, *компоненты вектора момента импульса тела являются линейными однородными функциями компонентов вектора угловой скорости*. В системе главных осей формулы (53.9) упрощаются и принимают вид

$$L_x = I_x\omega_x, \quad L_y = I_y\omega_y, \quad L_z = I_z\omega_z. \tag{53.11}$$

Формулы ясно показывают, что в общем случае направления векторов $\mathbf{L}$ и $\boldsymbol{\omega}$ не совпадают. Кинетическую энергию вращающегося твердого тела легко найти по формуле (47.2). Она равна

$$K = \tfrac{1}{2}(\mathbf{L}\boldsymbol{\omega}) = \tfrac{1}{2}\sum_{ij} I_{ij}\omega_i\omega_j. \tag{53.12}$$

§ 54. ВРАЩЕНИЕ ТВЕРДОГО ТЕЛА ПО ИНЕРЦИИ ВОКРУГ НЕПОДВИЖНОЙ ТОЧКИ

1. Пуансо дал простую и наглядную интерпретацию движения твердого тела по инерции вокруг неподвижной точки опоры O. С твердым телом связывается его эллипсоид инерции с центром в точке опоры O. Движение тела заменяется движением этого эллипсоида (см. § 53, п. 3). В основе интерпретации Пуансо лежат три теоремы, которые мы и докажем. Для краткости будем называть *полюсом* точку пересечения P мгновенной оси с поверхностью эллипсоида инерции.

Теорема 1. *Радиус-вектор, соединяющий точку опоры O с полюсом P, пропорционален мгновенной угловой скорости вращения тела.*

При доказательстве исходим из уравнения энергии $\sum_{i,j} I_{ij}\omega_i\omega_j = 2K = \text{const}$. Возьмем на мгновенной оси точку Q с радиусом-вектором $\mathbf{r} = \boldsymbol{\omega}/\sqrt{2K}$. Тогда из уравнения энергии найдем, что координаты точки Q должны удовлетворять уравнению $\sum_{ij} I_{ij}x_ix_j = 1$. Это значит, что точка Q лежит на поверхности эллипсоида инерции. Так как она лежит на мгновенной оси, то она совпадает с полюсом P. Итак, радиус-вектор полюса P связан с вектором угловой скорости $\boldsymbol{\omega}$ соотношением $\boldsymbol{\omega} = \sqrt{2K}\,\mathbf{r}$. Отсюда и следует доказываемая теорема.

Теорема 2. *Касательная плоскость к эллипсоиду инерции в точке нахождения полюса P перпендикулярна вектору $\mathbf{L}$ момента импульса тела относительно точки опоры O.*

При доказательстве можно воспользоваться уравнением эллипсоида инерции в любой системе координат. Но проще взять уравнение этой поверхности в системе главных осей эллипсоида, т. е. $I_x x^2 + I_y y^2 + I_z z^2 = 1$. Левую часть этого уравнения обозначим через $F(x, y, z)$, т. е. запишем само уравнение в виде $F(x, y, z) = 1$. Как было показано в § 29 (п. 3), вектор

$$\mathbf{N} = \frac{\partial F}{\partial x}\mathbf{i} + \frac{\partial F}{\partial y}\mathbf{j} + \frac{\partial F}{\partial z}\mathbf{k} \equiv \operatorname{grad} F$$

направлен по нормали к поверхности эллипсоида. Иными словами, вектор $\mathbf{N}$ перпендикулярен к касательной плоскости, о которой говорится в теореме. Он равен

$$\mathbf{N} = 2(I_x x\mathbf{i} + I_y y\mathbf{j} + I_z z\mathbf{k}).$$

На основании предыдущей теоремы его можно представить в виде

$$\mathbf{N} = \mu(I_x\omega_x\mathbf{i} + I_y\omega_y\mathbf{j} + I_z\omega_z\mathbf{k}),$$

т. е.

$$\mathbf{N} = \mu\mathbf{L},$$

где μ — некоторый скаляр. Это соотношение доказывает теорему.

Ввиду отсутствия внешних сил относительно точки опоры O вектор $\mathbf{L}$ не меняется по времени. Поэтому не будет менять свое направление и касательная плоскость к эллипсоиду инерции, о которой говорится в теореме.

Теорема 3. *Длина перпендикуляра, опущенного из точки опоры O на плоскость, касательную к эллипсоиду инерции в точке нахождения полюса P, не меняется с течением времени.*

Применим для доказательства уравнение энергии в виде $(\mathbf{L}\,\boldsymbol{\omega}) = 2K = \text{const}$, или $L\omega_L = 2K = \text{const}$, где ω_L — проекция вектора $\boldsymbol{\omega}$ на неизменное направление вектора $\mathbf{L}$. Так как величины L и K постоянны, то отсюда следует, что постоянна и проекция ω_L. Но, как было показано при доказательстве теоремы 1, эта проекция связана с длиной перпендикуляра r_L соотношением $\omega_L = \sqrt{2K}r_L$, что и требовалось доказать.

Из доказанной теоремы следует, что *касательная плоскость к эллипсоиду инерции в точке нахождения полюса неизменна не только по направлению, но и по своему положению в пространстве.* Поэтому эту плоскость часто называют *неизменяемой плоскостью.*

2. Теперь интерпретация Пуансо напрашивается сама собой. Связав с движущимся телом его эллипсоид инерции с центром в точке опоры O, проведем в какой-нибудь момент времени касательную плоскость в точке нахождения полюса в этот момент. Это будет неизменяемая плоскость в соответствии с теоремами 2 и 3. При этом в полюсе P не может быть скольжения между эллипсоидом инерции и неизменяемой плоскостью, так как через эту точку проходит мгновенная ось вращения тела. Если катить без скольжения эллипсоид инерции тела по неизменяемой плоскости с угловой скоростью, пропорциональной радиусу-вектору точки касания (т. е. полюса), то в соответствии с теоремой 1 при таком качении будет воспроизведено (в ускоренном или замедленном темпе) вращение твердого тела, связанного с эллипсоидом инерции.

3. Полюс P одновременно находится и на поверхности эллипсоида инерции, и на неизменяемой плоскости. Допустим для наглядности, что неизменяемая плоскость закрашена, например, покрыта сажей. При качении эллипсоида инерции на его поверхности и на неизменяемой плоскости остаются следы, показывающие, через какие точки проходил полюс. Кривая, которую описывает полюс на поверхности эллипсоида инерции, называется *полодией*. Плоская же кривая, описываемая тем же полюсом на неизменяемой плоскости, называется *герполодией*. Если эллипсоид инерции касается неизменяемой плоскости некоторой точкой, то спустя некоторое время он будет касаться той же плоскости той же точкой, но, вообще говоря, уже в другом месте. Иными словами, полюс на поверхности эллипсоида инерции вернется в свое исходное положение. Это показывает, что *полодия является замкнутой кривой*. Что касается герполодии, то она, вообще говоря, *не замкнута*.

Соединив прямыми точки полодии и точки герполодии с точкой опоры O, получим две конические поверхности. Одна коническая поверхность жестко связана с вращающимся телом. Она называется *конусом полодии*. Другая неподвижна в пространстве и называется *конусом герполодии*. Обе поверхности касаются друг друга вдоль прямой, совпадающей с мгновенной осью вращения. Поэтому между ними нет скольжения. *Движение тела можно рассматривать как качение без скольжения конуса полодии по неподвижному конусу герполодии с угловой скоростью, пропорциональной радиусу-вектору, проведенному из точки опоры к полюсу*. Эта интерпретация, также предложенная Пуансо, только словесно отличается от предыдущей интерпретации.

4. Допустим, что свободное тело вращается вокруг одной из главных осей центрального эллипсоида инерции. Тогда в интерпретации Пуансо эллипсоид инерции будет опираться на неизменную плоскость одной из своих вершин, причем соответствующая главная ось будет перпендикулярна к этой плоскости. Полодия и герполодия выродятся в точки, совпадающие с полюсом P. Отсюда видно, что вращение вокруг главной оси центрального эллипсоида инерции может продолжаться вечно. Это совпадает с доказанным выше утверждением, что главные оси центрального эллипсоида инерции являются также свободными осями вращения.

5. С помощью такой интерпретации Пуансо легко также исследовать вопрос, вращение вокруг каких свободных осей является устойчивым, а вокруг каких — неустойчивым. Вопрос этот сводится к отысканию уравнения конуса полодии относительно координатной системы, связанной с телом. Выберем в качестве таковой систему главных осей. Пусть ось X является осью наибольшего, а ось Z — наименьшего моментов инерции. Таким образом, мы полагаем

$$I_x > I_y > I_z. \tag{54.1}$$

В каждый момент времени движение тела есть вращение вокруг мгновенной оси. При вращении сохраняется кинетическая энергия тела:

$$I_x\omega_x^2 + I_y\omega_y^2 + I_z\omega_z^2 = 2K = \text{const}. \tag{54.2}$$

Кроме того, сохраняется момент импульса тела:

$$\mathbf{L} = I_x\omega_x\mathbf{i} + I_y\omega_y\mathbf{j} + I_z\omega_z\mathbf{k}.$$

Возведя это соотношение в квадрат, получим

$$I_x^2\omega_x^2 + I_y^2\omega_y^2 + I_z^2\omega_z^2 = L^2 = \text{const}. \tag{54.3}$$

Умножив уравнение (54.2) на $h^2 \equiv L^2/(2K)$ и вычтя из уравнения (54.3), получим однородное уравнение

$$I_x(I_x - h^2)\,\omega_x^2 + I_y(I_y - h^2)\,\omega_y^2 + I_z(I_z - h^2)\,\omega_z^2 = 0, \tag{54.4}$$

которому должны удовлетворять компоненты вектора угловой скорости $\boldsymbol{\omega}$. Уравнение мгновенной оси можно записать в виде $\mathbf{r} = \rho\boldsymbol{\omega}$, где ρ — переменный параметр, который может принимать любые значения. Найдя отсюда ω_x, ω_y, ω_z и подставив полученные значения в предыдущее уравнение, видим, что координаты точек мгновенной оси должны удовлетворять однородному уравнению второго порядка

$$I_x(I_x - h^2)\,x^2 + I_y(I_y - h^2)\,y^2 + I_z(I_z - h^2)\,z^2 = 0. \tag{54.5}$$

Это значит, что мгновенная ось вращения лежит на поверхности (54.5), т. е. на поверхности конуса второго порядка. Этот конус и будет конусом полодии, так как по самому определению конус полодии есть линейная поверхность, образованная последовательными положениями в теле мгновенной оси вращения.

6. Вид конуса полодии (54.5) зависит от значения параметра $h^2 \equiv L^2/(2K)$. Очевидно, все коэффициенты уравнения (54.5) не могут иметь одинаковые знаки, так как в этом случае уравнение не может удовлетворяться вещественными значениями x, y, z. Отсюда следует, что $I_x - h^2 > 0$. Действительно, если бы $I_x - h^2$ было меньше нуля, то в силу условия (54.1) величины $I_y - h^2$ и $I_z - h^2$ тем более были бы меньше нуля, т. е. все три коэффициента в уравнении (54.5) были бы отрицательны. А это, как мы показали, невозможно. Заметив это, видим, что могут представляться только два случая:

1) $(I_y - h^2) > 0$, $(I_z - h^2) < 0$;

2) $(I_y - h^2) < 0$, $(I_z - h^2) < 0$.

В первом случае уравнение (54.5) имеет вид $Ax^2 + By^2 - Cz^2 = 0$, где A, B, C — положительные постоянные, причем $A > B > C$. Сечение конуса полодии плоскостью $z = a = \text{const}$ есть эллипс $Ax^2 + By^2 = Ca^2$, а потому конус полодии окружает ось наименьшего момента инерции Z. Напротив, сечения его плоскостями $x = \text{const}$ и $y = \text{const}$ имеют гиперболическую форму.

Во втором случае уравнение конуса полодии имеет вид $Ax^2 - By^2 - Cz^2 = 0$ с положительными постоянными A, B, C. В этом случае в сечении получается эллипс, если оно производится плоскостью $x = \text{const}$. При сечении же плоскостями $y = \text{const}$ и $z = \text{const}$ образуются гиперболы.

Таким образом, в зависимости от значения параметра h *конус полодии окружает либо ось наибольшего, либо ось наименьшего моментов инерции тела. Но он никогда не окружает ось промежуточного момента инерции.*

7. Теперь вопрос об устойчивости вращения относительно свободных осей тела решается тривиально. Если тело вращается по инерции вокруг одной из свободных осей, то при наличии возмущения это вращение будет искажено. После прекращения возмущения мгновенная ось начнет описывать в теле конус полодии. Если вращение происходило вокруг оси с наибольшим или наименьшим моментами инерции, а возмущение было мало, то после прекращения последнего возникает конус полодии малого раствора, окружающий эту ось. Двигаясь по нему, мгновенная ось все время будет проходить вблизи свободной оси, вокруг которой было возбуждено первоначальное вращение тела. Это значит, что вращение вокруг такой оси является устойчивым. Напротив, если тело первоначально вращалось вокруг оси промежуточного момента инерции, то после воздействия малого возмущения возникает конус полодии широкого раствора, окружающий либо ось наибольшего, либо ось наименьшего моментов инерции. Двигаясь по такому конусу, мгновенная ось вращения далеко уйдет от своего исходного направления. Следовательно, вращение вокруг свободной оси с промежуточным моментом инерции является неустойчивым.

8. Если момент инерции относительно каких-либо главных осей, например X и Y, совпадают между собой ($I_x = I_y$), то эллипсоид инерции и конус полодии будут обладать симметрией вращения относительно оси Z. Конус полодии имеет вид $A(x^2 + y^2) - Cz^2 = 0$, где A и C — положительные постоянные. Его сечение плоскостью, перпендикулярной к оси Z, будет круговым. Сечения же плоскостями, параллельными этой оси, будут гиперболическими. Конус полодии, таким образом, окружает ось Z. Вращение вокруг этой оси будет устойчивым, а вращение вокруг перпендикулярной к ней оси — неустойчивым. Действительно, если вращение совершалось, например, вокруг оси X и подверглось возмущению, то после прекращения такового мгновенная ось начнет описывать круговой конус полодии с осью симметрии Z. Если возмущение мало, то это будет конус очень большого раствора. Его образующие будут наклонены к оси симметрии Z под углом, близким к 90°. Двигаясь по такому конусу, мгновенная ось вращения далеко уйдет от своего исходного положения в теле. Однако она все время будет оставаться почти перпендикулярной к оси Z. Всякая прямая, перпендикулярная к оси и проходящая через центр масс тела, может служить перманентной осью вращения.

Когда моменты инерции I_x, I_y и I_z совпадают между собой, то коэффициенты уравнения (54.5) тождественно обращаются в нуль. Это означает, что любая ось, проходящая через центр масс тела, может быть свободной осью вращения.

9. Аналитическое исследование вращения твердого тела вокруг неподвижной точки можно производить также на основе *уравнений Эйлера*. Последние являются следствием уравнения момента импульса твердого тела относительно точки. Они определяют производную по времени угловой скорости вращения твердого тела $\boldsymbol{\omega}$. Вводится вращающаяся система прямоугольных координат, начало которой совпадает с центром масс тела, а координатные оси направ-

ГЛАВА VIII

ТЯГОТЕНИЕ

§ 55. ЗАКОНЫ КЕПЛЕРА И ЗАКОН ВСЕМИРНОГО ТЯГОТЕНИЯ

1. В результате длительной обработки многолетних наблюдений датского астронома Тихо Браге (1546—1601) немецкий ученый Кеплер эмпирически установил три закона движений планет. Эти законы формулируются следующие образом:

1) *каждая планета движется по эллипсу, в одном из фокусов которого находится Солнце*;

2) *радиус-вектор планеты в равные времена описывает равные площади*;

3) *квадраты времен обращений планет относятся как кубы больших осей эллиптических орбит, по которым они движутся вокруг Солнца.*

Первые два закона были опубликованы Кеплером в 1609 г., последний — в 1619 г. *Законы Кеплера* естественным путем привели Ньютона к открытию *закона всемирного тяготения.* Рассмотрим этот вопрос.

Из первого закона Кеплера следует, что траектория планеты — плоская кривая. С учетом этого обстоятельства, как было показано в § 31, из второго закона Кеплера следует, что сила, заставляющая планету двигаться по замкнутым орбитам, направлена к Солнцу. Определим теперь, как эта сила изменяется с изменением расстояния от Солнца и как она зависит от массы планеты. Для упрощения расчетов допустим сначала, что планета движется не по эллипсу, а по кругу, в центре которого находится Солнце. Для планет Солнечной системы такое допущение не является особенно грубым. Эллипсы, по которым на самом деле движутся планеты, весьма мало отличаются от кругов. Ускорение планеты при равномерном движении по круговой орбите радиуса r выражается формулой

$$\mathbf{a}_r = -\omega^2 \mathbf{r} = -\frac{4\pi^2}{T^2}\,\mathbf{r}.$$

Для планет, движущихся по круговым траекториям, третий закон Кеплера записывается в виде

$$T_1^2 : T_2^2 : T_3^2 : \ldots = r_1^3 : r_2^3 : r_3^3 : \ldots,$$

или

$$\frac{r^3}{T^2} = \mathscr{K},$$

где $\mathscr{K}$ — постоянная для всех планет Солнечной системы. Она называется *постоянной Кеплера.* Через параметры эллиптической орбиты постоянная Кеплера выражается формулой

$$\mathscr{K} = \frac{a^3}{T^2}, \tag{55.1}$$

где a — длина большой полуоси орбиты.

Выразив T через $\mathscr{K}$ и r, для ускорения планеты при движении по круговой орбите получим

$$a_r = -\frac{4\pi^2\mathscr{K}}{r^2}. \tag{55.2}$$

Сила, действующая на планету, равна

$$F = -\frac{4\pi^2\mathscr{K}m}{r^2}, \tag{55.3}$$

где m — масса планеты.

Мы доказали, что ускорения двух разных планет, обращающихся вокруг Солнца по круговым орбитам, обратно пропорциональны квадратам расстояний их от Солнца. Но мы еще не доказали, что такая закономерность справедлива и для одной и той же планеты, обращающейся вокруг Солнца по эллиптической орбите. Чтобы доказать это, надо от рассмотрения круговых движений перейти к исследованию движений по эллипсу. Это будет сделано в следующем параграфе. Но можно обойтись и круговыми движениями, если использовать добавочное предположение, что сила взаимодействия между Солнцем и планетой зависит только от мгновенного расстояния между ними, но не зависит от формы траектории, по которой движется планета. Тогда формулы (55.2) и (55.3) можно применять не только к разным планетам, обращающимся по круговым орбитам на разных расстояниях от Солнца, но и к различным положениям одной и той же планеты, движущейся по эллиптической траектории.

2. Коэффициент пропорциональности $4\pi^2\mathscr{K}$, входящий в формулы (55.2) и (55.3), — один и тот же для всех планет, а потому он не может зависеть от массы планеты. Он может, однако, зависеть от параметров, характеризующих Солнце, поскольку последнее является источником сил, заставляющих планеты двигаться по замкнутым орбитам. Но Солнце и планета в их взаимодействии выступают как *равноправные тела.* Различие между ними только *количественное.* Они отличаются друг от друга массами. И если сила взаимодействия F пропорциональна массе планеты m, то она должна быть пропорциональна также и массе Солнца M. Для этой силы можно поэтому написать

$$F = G\frac{Mm}{r^2}, \tag{55.4}$$

где G — новая постоянная, уже не зависящая ни от массы Солнца, ни от массы планеты. Сравнивая эту формулу с (55.3), получаем следующее выражение для постоянной Кеплера:

$$\mathscr{K} \equiv \frac{a^3}{T^2} = \frac{GM}{4\pi^2}. \qquad (55.5)$$

3. Далее, Солнце и планеты отличаются друг от друга и от других тел только *количественно* — значениями масс. Поэтому естественно предположить, что притяжение существует не только между Солнцем и планетой, но и между планетами, а также между любыми другими телами, и что сила притяжения определяется формулой (55.4), в которой под M и m следует понимать массы взаимодействующих тел. Это предположение было введено Ньютоном и подтвердилось на опыте. Он сформулировал закон всемирного тяготения, согласно которому *любые два тела (материальные точки) притягиваются друг к другу с силами, пропорциональными произведению их масс и обратно пропорциональными квадрату расстояния между ними.* Такие силы называются *гравитационными* или *силами всемирного тяготения.* Коэффициент пропорциональности G, входящий в формулу (55.4), один и тот же для всех тел. В этом смысле он является универсальной постоянной. Это одна из важнейших мировых постоянных, называемая *гравитационной постоянной.*

В приведенной формулировке закона всемирного тяготения предполагается, что взаимодействующие тела являются *точечными.* Физически это означает, что размеры тел очень малы по сравнению с расстоянием между ними. Здесь, как и всегда в физике, слова «велик» и «мал» употребляются в относительном смысле — велик или мал *по сравнению с чем-то.* Указанное условие хорошо выполняется для взаимодействий Солнца с планетами, планет между собой и со спутниками. Но если речь идет о гравитационном притяжении двух тел с размерами 10 см, когда расстояние между их центрами масс составляет, например, 20 см, то такие тела не могут рассматриваться как точечные. Чтобы рассчитать их гравитационное взаимодействие, надо мысленно разбить каждое тело на очень малые части, подсчитать по формуле (55.4) силы притяжения между такими частями, а затем и эти силы геометрически сложить (проинтегрировать). В основе этого вычисления лежит *принцип суперпозиции гравитационных полей.* Согласно этому принципу *гравитационное поле, возбуждаемое какой-либо массой, совершенно не зависит от наличия других масс.* Кроме того, *гравитационное поле, создаваемое несколькими телами, равно геометрической сумме гравитационных полей, возбуждаемых этими телами в отдельности.* Принцип этот является обобщением опыта.

Пользуясь принципом суперпозиции легко доказать, что *два однородных шара притягиваются между собой так, как если бы их*

массы были сконцентрированы в их центрах (см. задачи 2, 3, 4 к этому параграфу.

Заметим еще, что каждая планета подвергается гравитационному притяжению не только Солнца, но и других тел Солнечной системы. Однако масса Солнца является преобладающей. Она более чем в 700 раз превосходит общую массу планет и всех остальных тел Солнечной системы. Благодаря этому Солнце является основным телом, управляющим движением планет. Законы Кеплера можно вывести из закона всемирного тяготения Ньютона (см. § 62). При таком выводе предполагается, что единственной силой, действующей на планету, является гравитационное притяжение Солнца. Поэтому *законы Кеплера являются приближенными законами*, не учитывающими гравитационное действие остальных тел Солнечной системы.

4. Во времена Ньютона закон всемирного тяготения был подтвержден только астрономическими наблюдениями над движениями планет и их спутников. Впервые непосредственное экспериментальное доказательство этого закона для земных тел, а также численное определение гравитационной постоянной G были даны английским физиком Г. Кавендишем (1731—1810) в 1798 г. Прибор Кавендиша состоял из легкого горизонтального коромысла (рис. 171*а*), на концах которого укреплялись два одинаковых свинцовых шарика мас-

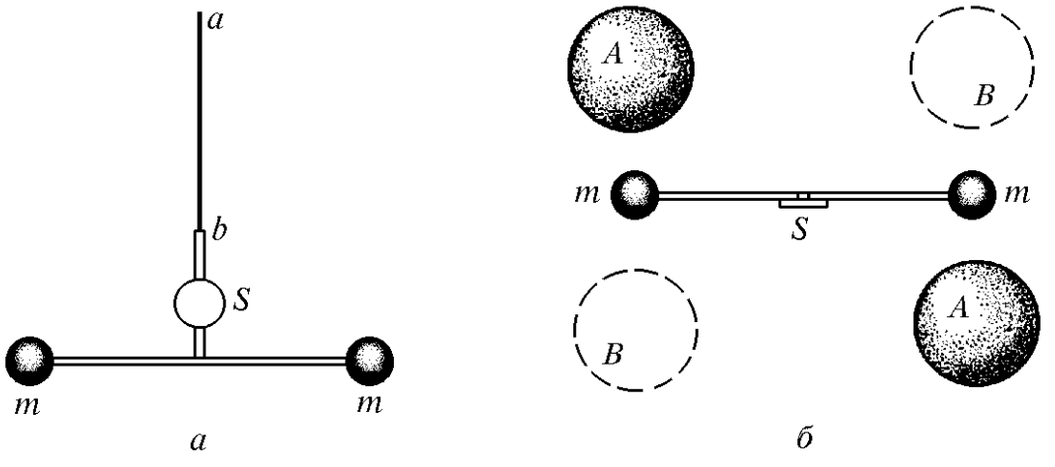


Рис. 171

сой m. Коромысло подвешивалось на тонкой вертикальной нити ab. Вблизи свинцовых шариков массой m помещались два других больших свинцовых шара с массой M каждый, причем $M \gg m$. Шары помещались сначала в положении AA, затем переводились в положение BB (рис. 171*б*). Благодаря гравитационному взаимодействию шариков m с шарами M коромысло поворачивалось из положения равновесия. Угол кручения α измерялся наблюдением луча света, отражавшегося от зеркальца S. Если r — расстояние между центрами малого и большого шаров, а l — длина коромысла, то момент пары гравитационных сил, поворачивающих коромысло, будет $G\frac{Mm}{r^2}l$. В положении равновесия этот вращающий момент должен

быть уравновешен упругим моментом закрученной нити $f\alpha$. Написав условие равновесия для положения свинцовых шаров сначала в AA ($\alpha = \alpha_1$), а затем в BB ($\alpha = \alpha_2$), получим два уравнения

$$f\alpha_1 = G\frac{Mm}{r^2}\,l, \quad f\alpha_2 = -G\frac{Mm}{r^2}\,l.$$

Из них находим

$$f(\alpha_1 - \alpha_2) = 2G\frac{Mm}{r^2}\,l.$$

Модуль кручения f легко найти, наблюдая период свободных колебаний коромысла

$$T = 2\pi\sqrt{\frac{I}{f}} = 2\pi\sqrt{\frac{ml^2}{2f}}.$$

В результате получаем

$$G = \frac{lr^2}{M}\left(\frac{\pi}{T}\right)^2(\alpha_1 - \alpha_2).$$

5. Другой метод определения гравитационной постоянной был предложен немецким физиком-экспериментатором Филиппом Жолли (1809—1880) в 1878 г. На одном из плеч рычажных весов одна под другой подвешены две чашки (рис. 172), между которыми установлено неподвижно тяжелое свинцовое тело массой M правильной геометрической формы.

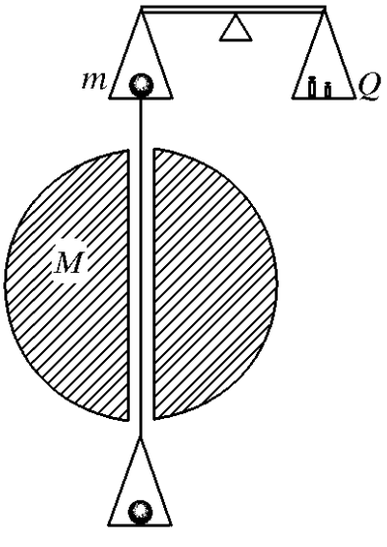


Рис. 172

В этом теле просверлен вертикальный канал, сквозь который свободно проходит проволока, соединяющая обе чашки. Если на верхнюю чашку положить тело массой m, то на него будет действовать вниз сила $Q_1 = mg + F$, где F — сила гравитационного притяжения между массами M и m. Она равна $F = kG\frac{Mm}{r^2}$, где r — расстояние между центрами рассматриваемых масс, а k — числовой коэффициент, зависящий от формы тела M. Для тел правильной геометрической формы его можно вычислить теоретически. Для шара $k = 1$. Если массу m перенести в нижнее положение, то сила F изменит направление. Сила, действующая вниз, станет $Q_2 = mg - F$. Значения Q_1 и Q_2 определяются по весу гирь, которые надо положить на чашку весов, подвешенную к другому плечу коромысла, чтобы весы находились в равновесии. Таким образом,

$$Q_1 - Q_2 = 2F = 2kG\frac{Mm}{r^2}.$$

Из этого соотношения и можно вычислить G.

6. Измерения G современными методами привели к результату

$$G = (6{,}6725 \pm 0{,}0005) \cdot 10^{-8} \text{ дин} \cdot \text{см}^2 \cdot \text{г}^{-2} =$$
$$= (6{,}6726 \pm 0{,}0005) \cdot 10^{-11} \text{ Н} \cdot \text{м}^2 \cdot \text{кг}^{-2}.$$

Гравитационная постоянная, как мы видим, весьма мала. Поэтому и гравитационные взаимодействия между обычными телами, даже считающимися большими с общежитейской точки зрения, ничтожно малы. Нетрудно подсчитать, что два точечных тела с массами по одному килограмму, находящиеся на расстоянии 1 м друг от друга, притягиваются с силой $F = 6{,}67 \cdot 10^{-11}$ Н $= 6{,}67 \cdot 10^{-6}$ дин. Гравитационные силы ничтожны, когда речь идет о взаимодействии *элементарных частиц*. Здесь эти силы, принимая во внимание современное состояние теории элементарных частиц, не учитываются. Но, возможно, их надо будет учитывать при переходе к взаимодействиям на расстояниях порядка 10^{-33} см (см. задачу 1 к этому параграфу). Они являются *основными силами*, управляющими движением и эволюцией небесных тел, массы которых очень велики. В этих случаях наиболее интенсивные — *ядерные* — силы совсем не проявляются, поскольку их радиус действия всего порядка 10^{-13} см. *Электрические силы*, как и силы всемирного тяготения, являются силами *дальнодействующими*. Они убывают также обратно пропорционально квадрату расстояния. Однако на движение астрономических тел электрические силы не оказывают влияния, так как они могут быть и силами притяжения, и силами отталкивания. Все тела в высокой степени *электрически нейтральны*, действие положительных зарядов тела компенсируется равным и противоположно направленным действием отрицательных зарядов. Иное дело — гравитационные силы. Они всегда являются *силами притяжения*. Гравитационные поля тел складываются, а не вычитаются. Это обстоятельство и является причиной того, почему из всех фундаментальных сил гравитационные силы остаются *единственными силами*, управляющими движением массивных астрономических тел.

7. Ньютон ограничился констатацией наличия гравитационных сил и их количественным описанием. Но он воздержался от каких бы то ни было высказываний относительно их физической природы, справедливо считая, что по этому вопросу в его время, кроме фантастических измышлений, ничего сказать было нельзя. После Ньютона было немало попыток дать наглядное физическое объяснение гравитационного притяжения. Никакого научного и даже исторического интереса эти попытки в настоящее время не имеют. Теория тяготения получила дальнейшее развитие в *общей теории относительности* Эйнштейна. Но в ней речь идет не о наглядном физическом объяснении тяготения, а о *новом способе описания его и об обобщении ньютоновского закона тяготения*.

Отказ Ньютона от объяснения тяготения, от сведения его к другим физическим явлениям был воспринят его приверженцами как

общефизическая концепция *непосредственного действия на расстоянии*. Эта концепция не только считает тяготение неотъемлемым свойством материи, но идет значительно дальше. Она считает, что каждому телу присуща способность непосредственно воздействовать на другие тела, находящиеся в других местах пространства, причем это воздействие осуществляется без какого бы то ни было участия промежуточной среды или других физических агентов.

Непосредственное действие на расстоянии отвергается современной наукой. Современная физика считает, что *все взаимодействия осуществляются полями*. Однако она не пытается представить механизм действия поля как-то наглядно. Она наделяет поле лишь способностью к объективному существованию и к передаче взаимодействий. Тело A не непосредственно действует на тело B. Оно создает вокруг себя гравитационное поле. Это поле и воздействует на другое тело B и проявляется в виде силы, действующей на него.

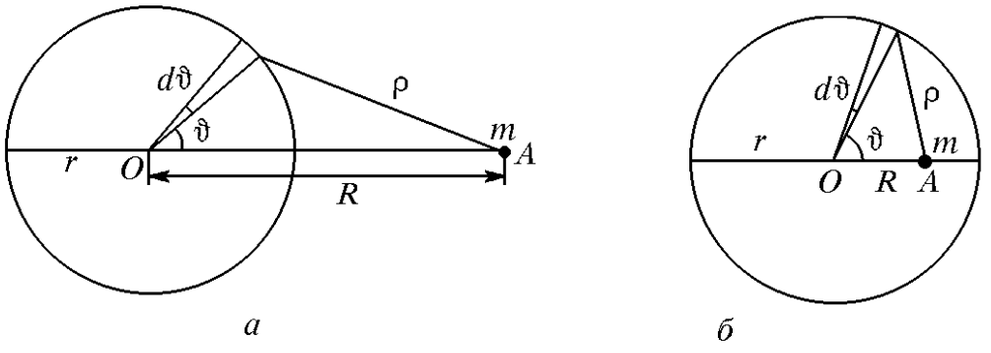
ЗАДАЧИ

1. Найти отношение силы гравитационного притяжения между двумя электронами (и двумя протонами) к силе их электростатического отталкивания.

Ответ. $\frac{F_{гр}}{F_{эл}} = \frac{Gm^2}{e^2}$, где $e = 4{,}8 \cdot 10^{-10}$ ед. СГСЭ — элементарный заряд. Подставляя в формулу массу электрона $m_e = 9{,}11 \cdot 10^{-28}$ г и массу протона $m_p = 1{,}67 \cdot 10^{-24}$ г, получим для электрона $F_{гр}/F_{эл} = 2{,}4 \cdot 10^{-43}$, для протона $F_{гр}/F_{эл} = 8 \cdot 10^{-37}$.

2. Найти потенциальную энергию и силу гравитационного притяжения между однородной полной сферой массой M и материальной точкой массой m.

Решение. Потенциальная энергия гравитационного взаимодействия двух точечных масс определяется формулой (25.6). Соединим центр сферы O с точкой A, в которой помещена точечная масса m (рис. 173 *а* и *б*). Из точки O, как из вершины, опишем два круговых конуса с общей осью OA,



а *б*

Рис. 173

образующие которых наклонены к этой оси под углами ϑ и $\vartheta + d\vartheta$. Они вырежут на поверхности сферы элементарной поясок площадью $dS = 2\pi r^2 \sin\vartheta\, d\vartheta$, где r — радиус сферы. Масса этого пояска $dM = M\frac{dS}{4\pi r^2} = \frac{M}{2}\sin\vartheta\, d\vartheta$. Так как точки пояска равноудалены от точки A, то

потенциальная энергия гравитационного взаимодействия пояска и точечной массы m равна

$$dU = -G\frac{Mm}{2}\sin\vartheta d\vartheta.$$

Перейдем к новой переменной ρ — расстояние между точечной массой m и какой-либо точкой пояска. Эта переменная связана с ϑ соотношением $\rho^2 = R^2 + r^2 - 2Rr\cos\vartheta$, где R — расстояние OA между центром сферы и точечной массой m. При перемещении вдоль поверхности сферы величины R и r остаются постоянными, поэтому

$$\rho d\rho = Rr\sin\vartheta\, d\vartheta,$$

а следовательно,

$$dU = -G\frac{Mm}{2Rr}d\rho,\quad U = -G\frac{Mm}{2Rr}\int\limits_{\rho_{мин}}^{\rho_{макс}} d\rho.$$

Если точка A лежит вне сферы, то максимальное и минимальное значения ρ равны соответственно $\rho_{макс} = R + r$ и $\rho_{мин} = R - r$. В этом случае интегрирование дает

$$U = -G\frac{Mm}{R}. \qquad (55.6)$$

Потенциальная энергия такая же, как если бы вся масса сферы была сосредоточена в одной точке, а именно в центре сферы. То же справедливо и для силы взаимодействия F. Действительно, согласно (29.3), сила F определяется формулой

$$F = -\frac{dU}{dR} = -G\frac{Mm}{R^2}.$$

Можно сказать, что *сфера притягивает материальную точку так, как если бы вся ее масса была сосредоточена в ее центре.* Можно сказать и иначе: *точечная масса притягивает сферу так, как если бы вся масса последней была сосредоточена в ее центре.*

Если же точка A лежит внутри сферической полости (рис. 173 *б*), то $\rho_{макс} = r + R$, $\rho_{мин} = r - R$, и интегрирование дает

$$U = -G\frac{Mm}{r}. \qquad (55.7)$$

На границе полости выражения (55.6) и (55.7) совпадают. Согласно (55.7) *потенциальная энергия материальной точки внутри полости не зависит от R, она постоянна. Сила F, действующая на материальную точку в этом случае, равна нулю*, так как $U = \text{const}$, а потому $F = -\frac{dU}{dR} = 0$.

3. Доказать, что *две однородные полые сферы притягиваются друг другу так, как если бы их массы были сосредоточены в их центрах.*

Доказательство. Если всю массу первого сферического тела сосредоточить в его центре, то создаваемое им гравитационное поле в объеме второго тела не изменится. Вместе с ним не изменится и сила взаимодействия со стороны первого тела на второе. По третьему закону Ньютона остается неизменной и сила, действующая на первое тело. Задача свелась к взаимо-

действию одной сферы с точечной массой. Повторением рассуждения убедимся, что и массу второго тела можно сосредоточить в одной точке, не меняя силы взаимодействия тел.

4. Доказать, что *два однородных шара притягиваются друг к другу так, как если бы масса каждого шара была сосредоточена в его центре.* Доказать также, что если внутри однородного шара имеется сферическая полость, центр которой совпадает с центром шара, то гравитационное поле внутри такой сферы равно нулю. Показать, что эти результаты справедливы и для шаров с концентрически слоистым распределением масс, т. е. таким, когда плотность вещества ρ в каждом шаре зависит только от расстояния до его центра.

5. Рассчитать напряженность гравитационного поля, т. е. силу, действующую на единицу массы, внутри и вне шара радиусом R, заполненного веществом с постоянной объемной плотностью ρ.

Р е ш е н и е. Поле вне шара равно $g = G\frac{M}{r^2}$, где M — масса шара. Для вычисления поля в точке A (рис. 174), лежащей внутри шара на расстоянии r от центра, проведем через эту точку вспомогательную сферу с центром в точке O. Вещество шара, расположенное вне вспомогательной сферы, не влияет на поле внутри нее. В частности, оно не влияет на поле в точке A. Гравитационное поле в точке A создается только веществом, сосредоточенным внутри вспомогательной сферы. Оно равно $G\frac{m}{r^2}$, где m — масса вещества, ограниченного вспомогательной сферой. Таким образом,

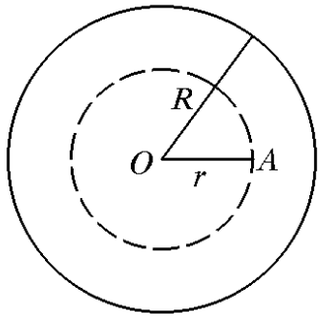


Рис. 174

$$g = \begin{cases} G\dfrac{M}{r^2} = \dfrac{4\pi G}{3}\dfrac{R^3}{r^2}\rho, & \text{если } r \geqslant R, \\ G\dfrac{m}{r^2} = \dfrac{4\pi}{3}G\rho r, & \text{если } r \leqslant R \end{cases}. \qquad (55.8)$$

При $r = R$ оба выражения совпадают.

6. Подсчитать гравитационную энергию U шара радиусом R, равномерно заполненного веществом с объемной плотностью ρ.

Р е ш е н и е. Гравитационная энергия шара есть потенциальная энергия, обусловленная силами тяготения, действующими между материальными точками, на которые можно мысленно разбить шар. Она равна взятой с противоположным знаком работе, которую должны затратить внешние силы, чтобы привести вещество шара в бесконечно разрозненное состояние, когда каждая частица вещества удалена в бесконечность.

Эта работа не зависит от способа, каким шар переводится из начального состояния в конечное. Поэтому при вычислении можно поступить следующим образом. Разобьем мысленно весь шар на бесконечно тонкие концентрические слои и будем последовательно удалять в бесконечность каждый из таких слоев, начиная с самого крайнего. Напряженность поля тяготения в любой точке выделенного слоя, создаваемая веществом, внешним по отношению к этому слою, равна нулю. Поле создается только веществом, которое окружено рассматриваемым слоем. Если m — масса этого вещества, а dm — масса слоя, то работа, затрачиваемая на удаление

слоя в бесконечность, равна $dA = G\frac{m\,dm}{r}$. Но для однородного шара $m = M\left(\frac{r}{R}\right)^3$, где M — масса всего шара. Поэтому $dA = 3G\frac{M^2}{R^6}r^4\,dr$. Учитывая, что $dA = -dU$, и интегрируя, получим

$$U = -3\frac{GM^2}{R^6}\int_0^R r^4\,dr = -\frac{3}{5}\frac{GM^2}{R}. \qquad (55.9)$$

За нуль потенциальной энергии мы приняли энергию шара в бесконечно разрозненном состоянии.

Интересны астрофизические применения формулы (55.9). Физиков давно интересовал вопрос об источниках энергии, излучаемой Солнцем и звездами. В прошлом веке немецкий естествоиспытатель Герман Гельмгольц (1821—1894) и английский физик Уильям Томсон (Кельвин) (1824—1907) выдвинули гипотезу, согласно которой Солнце непрерывно сжимается под действием гравитационных сил. Выделяющееся при этом тепло и идет на излучение Солнца. Максимальная энергия, которая может выделиться в процессе гравитационного сжатия Солнца, соответствует начальному состоянию, в котором вещество Солнца было равномерно распределено по всему бесконечному пространству. Будем считать, что в конечном состоянии плотность солнечного вещества одинакова по всему его объему. В действительности она, конечно, возрастает к центру Солнца. Однако для оценок наше предположение не является очень грубым. Приняв его, можно воспользоваться формулой (55.9). Масса Солнца $M = 2\cdot 10^{33}$ г, радиус $R = 7\cdot 10^{10}$ см. Используя эти данные, получаем для выделившейся энергии

$$E = \frac{3}{5}\frac{GM^2}{R} = 2{,}29\cdot 10^{48}\ \text{эрг}.$$

В настоящее время скорость излучения энергии Солнца составляет $3{,}83\cdot 10^{33}$ эрг/с. Если считать (при грубых оценках это допустимо), что эта скорость была постоянна во времени, то для возраста Солнца получится величина

$$t \approx \frac{2{,}29\cdot 10^{48}}{3{,}83\cdot 10^{33}} = 6\cdot 10^{14}\ \text{с} \approx 1{,}9\cdot 10^{7}\ \text{лет}.$$

Если воспользоваться распределением плотности вещества, соответствующим принятым в настоящее время моделям Солнца, то время t возрастет примерно в $6\cdot 10^7$ лет. Но и эта величина слишком мала. Возраст Земли по геологическим оценкам составляет около $4-4{,}5\cdot 10^9$ лет. Возраст Солнца не меньше. Это показывает, что на настоящей стадии эволюции гравитационное сжатие является слишком слабым источником, чтобы покрыть потери энергии Солнца на излучение. В действительности источником солнечной энергии, равно как и энергии, излучаемой нормальными звездами, являются *ядерные реакции*, идущие в недрах Солнца и звезд. *Конечным итогом основных из этих реакций является превращение водорода в гелий.* Следует, однако, заметить, что гравитационное сжатие становится основным источником энергии на более поздних стадиях эволюции звезд (белые карлики, нейтронные звезды, или пульсары, коллапсары, или «черные дыры»).

7. В сплошном однородном шаре с плотностью вещества ρ сделана сферическая полость, центр которой O_1 смещен относительно центра шара O (рис. 175). Найти гравитационное поле в такой полости.

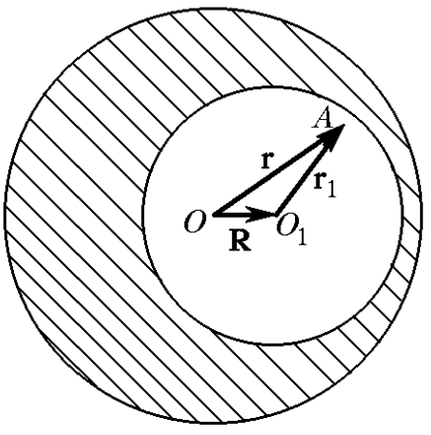


Рис. 175

Р е ш е н и е. Вообразим, что полость заполнена веществом, плотность которого равна плотности шара. Тогда искомое гравитационное поле g представится разностью гравитационных полей двух сплошных шаров с центрами в O и O_1 соответственно. Точка наблюдения A расположена внутри каждого из этих шаров. Поэтому можно воспользоваться формулой (55.8) и написать

$$\mathbf{g} = -\frac{4\pi}{3} G\rho\mathbf{r} - \left(-\frac{4\pi}{3} G\rho\mathbf{r}_1\right) = -\frac{4\pi}{3} G\rho\mathbf{R},$$

где $\mathbf{R}$ — радиус-вектор, проведенный из центра O к центру полости O_1. Поле однородно, т. е. во всех точках полости оно одинаково.

§ 56. УСКОРЕНИЕ ПЛАНЕТ И КОМЕТ ПРИ ДВИЖЕНИИ ПО КОНИЧЕСКИМ СЕЧЕНИЯМ

1. Замена эллиптических орбит круговыми была произведена в предыдущем параграфе исключительно в целях упрощения вычислений. Рассмотрим теперь задачу более строго, не прибегая к такому упрощению. Наши вычисления будут справедливы не только для *планет*, но и для *комет*. Последние, как показывают наблюдения, двигаются по *гиперболам* и *параболам* с фокусом в точке нахождения Солнца, причем это движение подчиняется второму закону Кеплера. Третий закон Кеплера для гиперболических и параболических движений, конечно, теряет смысл. Однако для вычисления ускорения планеты или кометы он не нужен. Действительно, при заданной траектории второй закон Кеплера определяет скорость планеты или кометы на этой траектории. Этого достаточно, чтобы полностью описать движение тела, т. е. указать его положение и скорость в любой момент времени. Зная это, можно вычислить ускорение тела в любой точке траектории. Приведем это элементарное вычисление.

2. Введем полярную систему координат с полюсом в фокусе F_1, где находится Солнце, и полярной осью PA, направленной вдоль большой оси эллипса или гиперболы (рис. 176). Ускорение движущегося тела разложим на радиальную составляющую a_r, направленную вдоль радиуса r, и азимутальную составляющую a_φ, перпендикулярную к радиусу. Они определяются выражениями

$$a_r = \ddot{r} - \dot{\varphi}^2 r, \quad a_\varphi = \frac{1}{r}\frac{d}{dt}(r^2\dot{\varphi}) \tag{56.1}$$

(см. § 46). Величина

$$\sigma = \frac{1}{2} r^2 \dot{\varphi} \tag{56.2}$$

есть секториальная скорость, т. е. площадь, описываемая радиусом-вектором планеты или кометы в единицу времени. По второму закону Кеплера она постоянна, а потому $a_\varphi = \frac{1}{r}\frac{d}{dt}(2\sigma) = 0$. Значит, ускорение рассматриваемого небесного тела не имеет азимутальной составляющей, т. е. направлено к Солнцу. Этот результат был уже получен в § 31 иным путем.

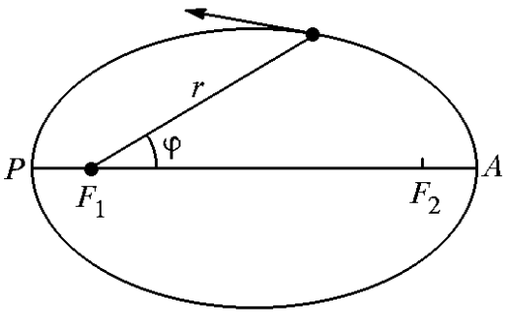


Рис. 176

Чтобы найти радиальное ускорение a_r, надо вычислить производные $\ddot{r}$ и $\dot{\varphi}$. Производная $\dot{\varphi}$ определяется формулой (56.2). Для вычисления производной $\ddot{r}$ воспользуемся уравнением конического сечения в полярной системе координат

$$r(1 - e\cos\varphi) = p, \tag{56.3}$$

где p и e — постоянные величины, из которых первая называется *параметром эллипса*, а вторая — его *эксцентриситетом*. Не нарушая общности, обе эти величины можно считать неотрицательными. Для эллипса $e < 1$, для параболы $e = 1$, для гиперболы $e > 1$. В предельных случаях, когда $e = 0$ и $e = \infty$, получаются круг и прямая линия. Дифференцируя уравнение (56.3) по времени, получим

$$\dot{r}(1 - e\cos\varphi) + er\dot{\varphi}\sin\varphi = 0,$$

или после умножения на r с учетом соотношений (56.2) и (56.3)

$$p\dot{r} + 2e\sigma\sin\varphi = 0.$$

Вторичное дифференцирование дает

$$p\ddot{r} + 2\sigma e\cos\varphi\cdot\dot{\varphi} = 0.$$

Подставляя сюда $\dot{\varphi} = \frac{2\sigma}{r^2}$, $e\cos\varphi = 1 - \frac{p}{r}$, получим

$$\ddot{r} = -\frac{4\sigma^2}{pr^2} + \frac{4\sigma^2}{r^3} = -\frac{4\sigma^2}{pr^2} + \dot{\varphi}^2 r.$$

После этого из первой формулы (56.1) находим

$$a_r = -\frac{4\sigma^2}{pr^2}. \tag{56.4}$$

Таким образом, *из первых двух законов Кеплера вытекает, что ускорение планеты или кометы обратно пропорционально квадрату ее расстояния от Солнца.*

3. Третий закон Кеплера позволяет доказать, что коэффициент пропорциональности $4\sigma^2/p$ — один и тот же для всех планет. Докажем это. Площадь эллипса равна πab, где a и b — длины большой и малой полуосей его. Так как секториальная скорость σ постоянна, то $\sigma = \pi ab/T$, где T — период обращения планеты по ее орбите. Воспользуемся еще формулой аналитической геометрии $p = b^2/a$. Тогда из (56.4) получим

$$a_r = -\frac{4\pi^2 a^3}{T^2}\frac{1}{r^2}. \tag{56.5}$$

(При равномерном вращении по окружности эта формула переходит в известную формулу $a_r = -\frac{4\pi^2 r}{T^2}$.) Вводя постоянную Кеплера (55.1), получим

$$a_r = \frac{4\pi^2 \mathscr{K}}{r^2}. \tag{56.6}$$

Этот результат совпадает с прежней формулой (55.2), но при его выводе здесь были использованы только эмпирические законы Кеплера без привлечения каких бы то ни было дополнительных соображений. Таким образом, формула (55.2) оказалась точной. Этого и следовало ожидать, так как в соответствии с основными положениями механики Ньютона ускорение планеты должно определяться только взаимным расположением Солнца и планеты и не может зависеть от вида траектории и скорости планеты. По той же причине формула (56.6) может служить и для вычисления ускорений комет, хотя третий закон Кеплера для них и не имеет смысла. В этом случае числовое значение постоянной $\mathscr{K}$ будет тем же самым, но она не может быть выражена через параметры орбиты кометы формулами, аналогичными (55.1).

4. Движение по параболе можно рассматривать как предельный случай движения по эллипсу, один из фокусов которого удален в бесконечность. Движение по гиперболе нуждается, однако, в некоторых пояснениях.

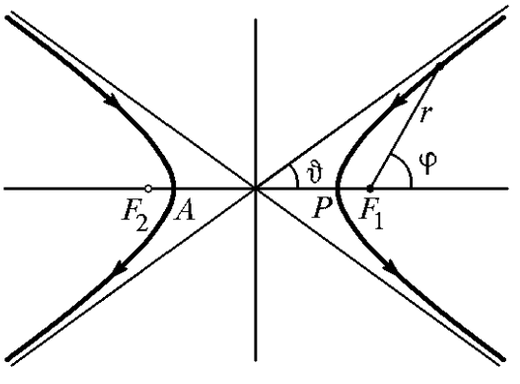


Рис. 177

Гипербола состоит из двух не связанных между собой ветвей. Чтобы обе ветви представлялись единым уравнением (56.3), надо допустить, чтобы расстояние r могло принимать не только положительные, но и *отрицательные* значения. Пусть ϑ — угол, определяемый условием $\vartheta = 1/e$. Он определяет направления *асимптот гиперболы* (рис. 177). Если $|\varphi| > \vartheta$, то r положительно. Этому соответствует правая ветвь гиперболы. Если $|\varphi| < \vartheta$, то r отрицательно. Тогда точку кривой надо искать не в направлении по-

лупрямой, проведенной под углом φ, а в прямо противоположном направлении. Получится левая ветвь гиперболы.

Конечно, движущаяся точка не может перескочить с одной ветви гиперболы на другую. Если на нее действует сила притяжения, то траектория должна быть обращена вогнутостью к силовому центру. Например, если силовой центр (Солнце) находится в фокусе F_1, то возможно движение только по правой ветви гиперболы. Однако, чтобы подметить общие закономерности движений по коническим сечениям, а не только по эллипсам, имеет смысл чисто формально ввести вспомогательную материальную точку, движущуюся по левой ветви гиперболы под действием силы отталкивания, исходящей из того же силового центра F_1. Потенциальная энергия вспомогательной точки представляется выражением $U = + G\frac{Mm}{|r|}$. Она положительна, поскольку силы являются силами отталкивания. Но так как на левой ветви гиперболы величины r отрицательны, то это выражение можно записать в виде $U = - G\frac{Mm}{r}$. Эта формула в точности совпадает с формулой, которой выражается потенциальная энергия действительной точки, движущейся по правой ветви гиперболы. Поэтому если энергия и момент импульса вспомогательной точки относительно фокуса F_1 равны соответствующим величинам для действительной точки, то движения обеих точек будут описываться *одними и теми же уравнениями.* В математических расчетах имеет значение не то, что движется, а то, какими уравнениями движение описывается. Формально математически дело происходит так, как если бы имелась всего одна материальная точка, обладающая способностью «перескакивать» с одной ветви гиперболы на другую. Целесообразность такого искусственного подхода будет проиллюстрирована на одном примере в § 58. Гравитационных сил отталкивания не существует. Но умозрительно их вводить можно. Кроме того, силы отталкивания возникают при электрических взаимодействиях одноименно заряженных частиц. Они, как и силы тяготения, убывают обратно пропорционально квадрату расстояния. Поэтому движение под действием сил отталкивания представляет не только умозрительный, но и физический интерес.

§ 57. УСЛОВИЯ ЭЛЛИПТИЧЕСКОГО, ПАРАБОЛИЧЕСКОГО И ГИПЕРБОЛИЧЕСКОГО ДВИЖЕНИЙ

1. *Когда траектория эллиптическая, движение планеты финитно*, т. е. планета движется в ограниченной области пространства, не уходя в бесконечность. Напротив, в *случае гиперболических и параболических траекторий движение инфинитно* — движение планеты не стеснено определенной областью

пространства, она может удаляться в бесконечность. Таким образом, задача сводится, к нахождению условий финитности и инфинитности движения планеты.

Если E — полная энергия планеты, то

$$\frac{mv^2}{2} - G\frac{Mm}{r} = E = \text{const.} \tag{57.1}$$

Кинетическую энергию Солнца мы не учитываем, считая, что она пренебрежимо мала по сравнению с кинетической энергией планеты. Это справедливо ввиду малости массы планеты по сравнению с массой Солнца. Аналогично если L — момент импульса планеты относительно Солнца, то

$$mr^2\dot{\varphi} = L = \text{const.} \tag{57.2}$$

Исключим из этих уравнений угловую скорость $\dot{\varphi}$. С этой целью разложим полную скорость v на радиальную составляющую v_r и азимутальную составляющую $r\dot{\varphi}$. Тогда

$$\frac{mv^2}{2} = \frac{m}{2}v_r^2 + \frac{m}{2}r^2\dot{\varphi}^2 = \frac{m}{2}v_r^2 + \frac{L^2}{2mr^2},$$

и уравнение (57.1) примет вид

$$\frac{m}{2}v_r^2 - G\frac{Mm}{r} + \frac{L^2}{2mr^2} = E = \text{const.} \tag{57.3}$$

Это уравнение содержит только одну неизвестную — радиальную скорость v_r. Формально оно может рассматриваться как *уравнение энергии для одномерного — радиального — движения точки.* Роль потенциальной энергии играет функция

$$V(r) = -G\frac{Mm}{r} + \frac{L^2}{2mr^2}.$$

2. Задача свелась к нахождению условий финитности и инфинитности одномерного движения с потенциальной энергией $V(r)$. Этот вопрос был исследован в § 25. Наиболее удобен для решения задачи графический метод. На рис. 178 штриховые кривые представляют соответственно графики функций

$$V_1(r) = -G\frac{Mm}{r}$$

и

$$V_2(r) = \frac{L^2}{2mr^2},$$

причем предполагается, что $L \neq 0$. Интересующая нас кривая $V(r)$ найдется сложением ординат этих двух графиков. При $r \to 0$ функция $V_2(r)$ быстрее стремится к бесконечности, чем функция $V_1(r)$. Поэтому при малых r функция $V(r) = V_1(r) + V_2(r)$ положительна

и асимптотически стремится к $+\infty$, когда $r \to 0$. Наоборот, при $r \to \infty$ функция $V_1(r)$ медленнее приближается к нулю, чем $V_2(r)$. Поэтому при больших r функция $V(r)$ отрицательна и асимптотически приближается к нулю, когда $r \to \infty$. График этой функции представлен на рис. 178 сплошной линией. Кривая $V(r)$ имеет вид «потенциальной ямы». Если $L = 0$, то $V(r) \equiv V_1(r)$, минимум на кривой смещается в начало координат и уходит в $-\infty$. Это соответствует случаю, когда планета движется вдоль прямой, проходящей через центр Солнца.

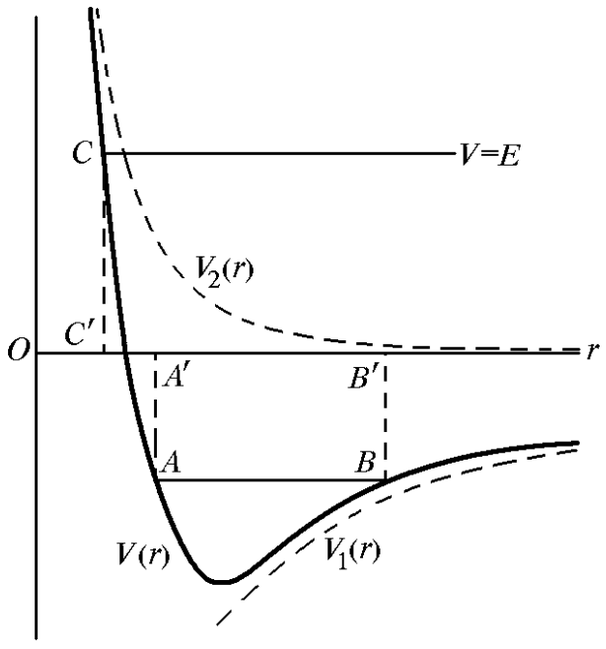


Рис. 178

Так как величина $\frac{1}{2} mv_r^2$ не может быть отрицательной, то из уравнения (57.3) следует, что область, в которой может находиться планета, определяется условием $V(r) \leqslant E$. Проведем горизонтальную прямую $V = E = \text{const}$. Участки кривой $V(r)$, лежащие выше этой прямой, соответствуют точкам пространства, которые не могут быть достигнуты планетой с энергией E. Если $E < 0$, то указанная прямая пересечет кривую $V = V(r)$ в двух точках A и B. Пусть A' и B' — их проекции на горизонтальную ось. Планета может совершать движение только в области между A' и B', она будет «локализована в потенциальной яме» $V = V(r)$. В этом случае движение планеты финитно, и траектория будет *эллиптической*. Если $E > 0$, то прямая пересечет кривую $V(r)$ только в одной точке C, проекцией которой на горизонтальную ось является точка C'. Если планета двигалась справа налево, то в точке C' она переменит направление движения на противоположное и начнет двигаться вправо, монотонно удаляясь в бесконечность. Ее движение инфинитно, а траектория — *гиперболическая*. Наконец, при $E = 0$ движение также инфинитно. Этому промежуточному случаю между эллиптическим и гиперболическим движениями соответствует *движение по параболе*.

Таким образом, *при* $E > 0$ *движение гиперболическое, при* $E < 0$ *— эллиптическое, при* $E = 0$ *— параболическое. В случае сил отталкивания энергия* E *всегда положительна, а потому движение в этом случае всегда гиперболическое* (*в частности, прямолинейно*). Так как при $r \to \infty$ функция $V(r)$ обращается в нуль, то

$$E = \frac{1}{2} mv_\infty^2. \qquad (57.4)$$

Отсюда следует, что *при гиперболическом движении материальная точка приходит в бесконечность с конечной скоростью* v_∞, *при параболическом движении — с нулевой скоростью.* Начальная скорость $v_\text{п}$, которую надо сообщить материальной точке, чтобы она стала двигаться по параболе, называется *параболической скоростью.* Параболическую скорость можно определить из уравнения (57.1), подставив в него $E = 0$. Если r_0 — начальное значение радиуса r, то

$$\frac{mv_\text{п}^2}{2} - G\frac{Mm}{r_0} = 0,$$

откуда

$$v_\text{п} = \sqrt{2G\frac{M}{r_0}}. \tag{57.5}$$

Параболическая скорость связана простым соотношением с «*круговой*» скоростью $v_\text{к}$. Так называется скорость, которой должна обладать планета, чтобы под действием гравитационной силы Солнца двигаться вокруг него по кругу радиуса r_0. Она найдется, если центростремительное ускорение $v_\text{к}^2/r_0$ приравнять гравитационной силе $G\frac{M}{r_0^2}$, действующей на единицу массы. Это дает

$$v_\text{к} = \sqrt{G\frac{M}{r_0}}. \tag{57.6}$$

Таким образом,

$$v_\text{п} = v_\text{к}\sqrt{2}. \tag{57.7}$$

ЗАДАЧИ

1. Допустим, что в результате взрыва тело, двигавшееся по круговой орбите вокруг Солнца, распалось на два осколка одинаковой массы. Один осколок непосредственно после взрыва остановился, другой продолжал движение. По какой траектории будет двигаться второй осколок: эллиптической, гиперболической или параболической?

Ответ. По гиперболической.

2. В условиях предыдущей задачи оба осколка разлетаются в перпендикулярных направлениях с одинаковыми скоростями. По каким орбитам они будут двигаться?

Ответ. Оба осколка будут двигаться по параболам.

§ 58. ВЫЧИСЛЕНИЕ ПАРАМЕТРОВ ОРБИТЫ

1. Длины большой и малой осей эллиптической орбиты планеты можно рассчитать с помощью законов сохранения энергии и момента импульса. В *перигелии* P и в *афелии* A (рис. 179) радиальная

скорость планеты равна нулю. Поэтому, полагая в уравнении (57.3) $v_r = 0$, получим для этих точек

$$r^2 + G\frac{Mm}{E}r - \frac{L^2}{2mE} = 0. \tag{58.1}$$

При $E < 0$ это квадратное уравнение имеет два вещественных положительных корня r_1 и r_2. Один из корней соответствует перигелию P, другой — афелию A. Сумма корней $r_1 + r_2$ дает длину *большой оси* эллипса. Пользуясь для этой длины стандартным обозначением $2a$, получим

$$2a = r_1 + r_2 = -G\frac{Mm}{E} = -G\frac{M}{\varepsilon}, \tag{58.2}$$

где $\varepsilon = E/m$ — полная энергия, приходящаяся на единицу массы планеты. Так как для движения по эллипсу $\varepsilon < 0$, то выражение (58.2) существенно положительно, как это и должно быть.

Круговые траектории являются вырожденными случаями эллиптических. Условие движения по круговой орбите найдется из уравнения (58.2), если в нем положить $r_1 = r_2 = r$. Тогда получится $2E = -GMm/r$ или $2E = U$. Записав это в виде $E = U - E$ и воспользовавшись соотношением $E = K + U$, получим

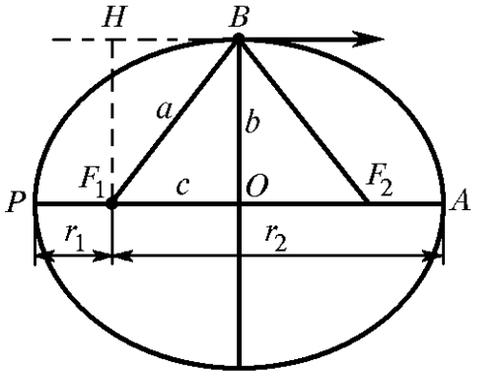


Рис. 179

$$E = -K. \tag{58.3}$$

Таким образом, *при круговом движении сумма полной и кинетической энергий равна нулю.* Нетрудно показать, что это условие снова приводит к формуле (57.6).

Для эллиптического движения формула (58.3) также справедлива, но под K следует понимать *среднее по времени* значение кинетической энергии планеты. Действительно, эллиптическое движение финитно, и к нему можно применить теорему вириала (§ 25, п. 6). Применительно к движению планеты эта теорема дает

$$\overline{K} = -\frac{1}{2}\overline{\mathbf{rF}} = \frac{1}{2}\overline{rF} = \frac{1}{2}GMm\left(\overline{\frac{1}{r}}\right) = -\frac{1}{2}\overline{U}.$$

Вычитая из обеих частей $1/2\,\overline{K}$ и учитывая, что $E = \overline{K} + \overline{U}$, получим

$$\overline{K} = -E.$$

Это и доказывает наше утверждение.

2. Найдем теперь длину *малой полуоси эллипса* b. Для этого помимо энергии надо знать еще момент импульса планеты или ее секториальную скорость $\sigma = \dot{S}$. Большую ось эллипса можно считать известной, поскольку она однозначно определяется энергией планеты. Пусть B — одна из точек, в которых малая ось пересекается с

эллипсом (рис. 179). Так как сумма расстояний любой точки эллипса от его фокусов F_1 и F_2 постоянна и равна $2a$, то $F_1B = a$. Секториальная скорость в точке B равна

$$\sigma = \frac{1}{2} vb, \tag{58.4}$$

так как b есть длина перпендикуляра F_1H, опущенного из фокуса F_1 на направление скорости в этой точке. Скорость v в точке B определится из уравнения энергии. Полагаем в нем $r = a$ и находим

$$\frac{v^2}{2} - G\frac{M}{a} = \varepsilon.$$

Подставив сюда выражение для ε из (58.2), определим v. После этого найдем

$$b = 2\sigma\sqrt{\frac{a}{GM}}. \tag{58.5}$$

3. Распространим теперь полученные результаты на случай *гиперболического движения*. Для этого воспользуемся искусственным приемом, указанным в п. 4 § 56. По правой ветви гиперболы (см. рис. 177) движется комета, по левой — соответствующая ей вспомогательная материальная точка. Эти движения описываются одним и тем же уравнением (57.3). В вершинах гиперболы P и A радиальная скорость v_r равна нулю, и мы снова приходим к квадратному уравнению (58.1). Однако теперь энергия E положительна, так что знаки корней этого уравнения противоположны. Положительный корень r_1 соответствует вершине P, отрицательный r_2 — вершине A. Сумма обоих корней $r_1 + r_2$ отрицательна. По абсолютной величина эта сумма равна расстоянию между вершинами P и A. Используя для этого расстояния стандартное обозначение $AP = 2a$, получим

$$2a = -(r_1 + r_2) = G\frac{Mm}{E} = G\frac{M}{\varepsilon}. \tag{58.6}$$

Эта формула в точности совпала бы с формулой (58.2), если бы условиться считать расстояние между вершинами гиперболы величиной отрицательной.

4. Найдем теперь аналог формулы (58.4) для гиперболического движения. Расстояние между фокусами F_1 и F_2 принято обозначать через $2c$, а под b понимать квадратный корень $b = \sqrt{c^2 - a^2}$. Проведем через фокус F прямую, параллельную одной из асимптот гиперболы (рис. 180). Из фокуса F_1 на прямую F_2M опустим перпендикуляр F_1M. Длину отрезка F_2M можно рассматривать как разность расстояний от фокусов F_1 и F_2 до бесконечно удаленной точки, в которой пересекаются параллельные прямые F_2M и OB. Поэтому в силу известного свойства гиперболы $F_2M = 2a$. На основании теоремы Пифагора делаем вывод, что расстояние F_1M равно $2b$. Секториальную скорость, как величину постоянную, достаточно вычислить для точки, дви-

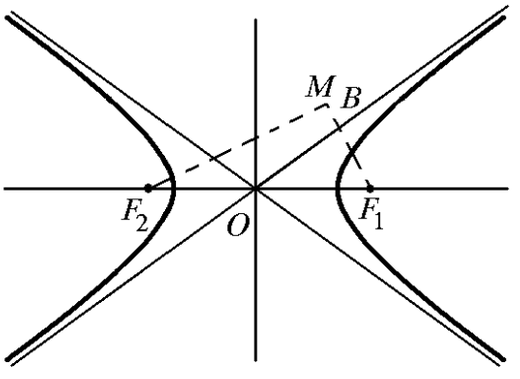


Рис. 180

жущейся в бесконечность. Радиус-вектор такой точки в единицу времени описывает треугольник с основанием v_∞ и высотой $F_1B = b$. Его площадь

$$\sigma = \frac{1}{2} bv_\infty \tag{58.7}$$

и дает секториальную скорость. При этом величина v_∞ определяется формулой (57.4), которую можно записать также в виде

$$\frac{v_\infty^2}{2} = \varepsilon. \tag{58.8}$$

Угол ϑ между асимптотами гиперболы можно вычислить по формуле

$$\operatorname{tg} \frac{\vartheta}{2} = \frac{b}{a} = \frac{bv_\infty^2}{GM}. \tag{58.9}$$

5. Параметр p для эллипса и гиперболы определяется выражением $p = b^2/a$. Подставляя сюда соответствующие значения для b и a, в обоих случаях найдем

$$p = \frac{4\sigma^2}{GM}. \tag{58.10}$$

Той же формулой определяется параметр p и для параболы, поскольку парабола является предельной кривой, в которую переходят эллипс и гипербола. Для параболы параметр p является единственной величиной, определяющей ее форму.

6. Вид траектории планеты, конечно, определяется начальными условиями, т. е. положением и скоростью планеты в некото-

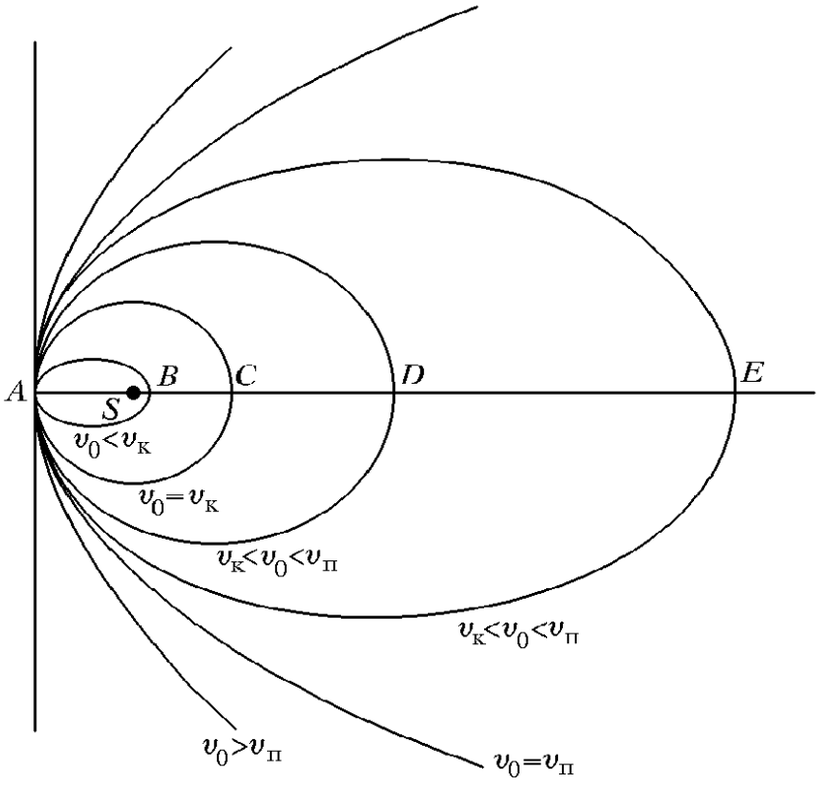


Рис. 181

рый момент времени, который условно можно принять за начальный. Иллюстрируем это следующим примером. Пусть S — Солнце, а A — начальное положение планеты (рис. 181). Расстояние AS обозначим через r_0. Будем сообщать планете в точке A скорость v_0 в направлении, перпендикулярном к AS. Посмотрим, как будет меняться вид траектории при изменении величины v_0. Если полная энергия планеты отрицательна, т. е. v_0 меньше параболической скорости $v_п$, то траекторией планеты будет эллипс. При $v_0 = 0$ эллипс вырождается в прямую, проходящую через Солнце S. Если $v_0 = v_к$, то планета будет двигаться по кругу. В этом случае точки A и C равноудалены от Солнца. Расстояние между ними (большая ось) равно $2r_0$. При уменьшении энергии большая ось эллипса уменьшается. При $v_0 < v_к$ она становится меньше $2r_0$. В этом случае точка A удалена от Солнца S дальше (афелий), чем точка B (перигелий). При $v_0 > v_к$ наоборот, большая ось эллипса больше $2r_0$, т. е. перигелием будет точка A, а афелием — точка D (или E). При $v = v_п = v_к\sqrt{2}$ траекторией будет парабола. При $v > v_п$ она переходит в гиперболу. Все эти результаты представлены в табл. 2:

Таблица 2

Начальная скорость	Траектория планеты
$v_0 = 0$	Прямая, проходящая через Солнце
$v_0 < v_к$	Эллипс с перигелием в точке B и афелием в точке A
$v_0 = v_к$	Окружность с центром в точке нахождения Солнца
$v_к < v_0 < v_п$	Эллипс с перигелием в точке A и афелием в точке D
$v_0 = v_п$	Парабола
$v_0 > v_п$	Гипербола

ЗАДАЧИ

1. Космический корабль движется вокруг Земли по близкой эллиптической орбите. В перигее касательная скорость корабля v увеличивается на $\Delta v = 10$ м/с. Что сделается с перигеем и апогеем при таком изменении скорости?

Решение. Перигей останется на месте, поскольку в нем меняется только касательная скорость. Из формулы (58.2) следует, что $\ln a = -\ln(-\varepsilon) + \text{const}$, и далее, что $\Delta a/a = -\Delta\varepsilon/\varepsilon$. Полная энергия $\varepsilon = \varepsilon_{пот} + \varepsilon_{кин}$. В перигее с точностью до второго порядка потенциальная энергия не меняется. Поэтому $\Delta\varepsilon \approx \Delta\varepsilon_{кин} = \Delta(v^2/2) = v\Delta v$. С учетом (58.2) $\Delta a/a = 2av\Delta v/GM$. Но $GM = gr^2$, где g — ускорение свободного падения в перигее, а r — расстояние до него от центра Земли. С учетом

$v_к^2 = gr$ получаем

$$\frac{\Delta a}{a} = 2av\,\frac{\Delta v}{v_к^2 r}.$$

Приближенно $v = v_к$, $r = a$, так что

$$\frac{\Delta a}{a} \approx 2\,\frac{\Delta v}{v_к} \approx \frac{1}{400}.$$

Полагая $a \approx 6400$ км, $v_к \approx 8$ км/с, найдем $\Delta\,(2a) = 33$ км. Большая ось эллиптической орбиты удлинится примерно на 33 км. Это произойдет за счет удаления апогея, так как перигей остается на месте.

2. Космический корабль движется вокруг Земли по близкой эллиптической орбите. Перигей находится на расстоянии 200 км от поверхности Земли, апогей — на расстоянии 233 км. Как надо изменить касательную скорость в перигее, чтобы корабль начал двигаться по круговой орбите? На сколько будет удалена круговая орбита от поверхности Земли?

Ответ. 200 км; $\Delta v \approx (v_к/2)\,(\Delta a/a) \approx -10$ м/с.

3. В классических опытах Резерфорда исследовалось рассеяние α-частиц на атомных ядрах различных химических элементов. Считая ядро бесконечно тяжелым и полагая, что рассеяние вызывается кулоновскими силами отталкивания, показать, что угол отклонения скорости α-частицы от первоначального направления полета θ связан с прицельным расстоянием b соотношением

$$\operatorname{ctg}\frac{\theta}{2} = \frac{mbv_\infty^2}{2Ze^2}, \tag{58.11}$$

где m — масса α-частицы, v_∞ — ее скорость вдали от ядра, $2e$ — ее заряд, Ze — заряд ядра (e — элементарный заряд, Z — порядковый номер элемента).

Примечание. *Прицельным расстоянием* называется длина перпендикуляра, опущенного из рассеивающего центра (ядра) на исходное направление касательной к траектории, когда рассеиваемая частица находится в бесконечности.

Мы воспользуемся формулой (58.11) в атомной физике при рассмотрении опытов Резерфорда.

§ 59. УЧЕТ ДВИЖЕНИЯ СОЛНЦА

1. При рассмотрении планетных движений мы не учитывали движение Солнца, считая его массу бесконечно большой по сравнению с массой планеты. Для ускорения планеты мы писали

$$m\ddot{\mathbf{r}} = m\mathbf{a}_{абс} = \mathbf{F}, \tag{59.1}$$

где $\mathbf{F} = -G\,\frac{Mm}{r^3}\,\mathbf{r}$ — ньютонова сила гравитационного притяжения, действующая на планету со стороны Солнца. Символом $\mathbf{a}_{абс}$ обозначено ускорение планеты относительно какой-то инерциальной системы отсчета, например системы Коперника. Учтем теперь движение Солнца. Чтобы получить уравнение движения планеты относительно Солнца, надо массу планеты m заменить на приведенную

массу $\mu = \frac{Mm}{M+m}$ (см. § 20). В результате уравнение относительного движения примет вид

$$\mu\ddot{\mathbf{r}} \equiv \mu\mathbf{a}_{\text{отн}} = \mathbf{F}.$$

Подставив выражение для μ, получим

$$m\mathbf{a}_{\text{отн}} = \left(1 + \frac{m}{M}\right)\mathbf{F}. \tag{59.2}$$

Формально дело происходит так, как если бы Солнце оставалось неподвижно, но гравитационная постоянная увеличилась в $(1 + m/M)$ раз. Поэтому *для относительного движения первый и второй законы Кеплера остаются справедливыми.* Зато третий закон должен быть уточнен. Для этого достаточно в формуле (55.5) постоянную G заменить на $G(1 + m/M)$. Это приводит к соотношению

$$\frac{a^3}{T^2(M+m)} = \frac{G}{4\pi^2}. \tag{59.3}$$

Оно показывает, что *отношение* $\frac{a^3}{T^2(M+m)}$ *является универсальной постоянной, т. е. не зависит ни от масс взаимодействующих тел, ни от расстояния между ними.* Таким образом, *третий закон Кеплера для относительного движения не вполне точен.* То обстоятельство, что для планет Солнечной системы он выполняется с большой точностью, связано с тем, что масса планеты очень мала по сравнению с массой Солнца.

Отметим еще соотношение

$$\frac{a_{\text{отн}}}{a_{\text{абс}}} = 1 + \frac{m}{M}, \tag{59.4}$$

которое непосредственно следует из сравнения формул (59.1) и (59.2).

2. На формулах (55.5) и (59.3) основано определение масс планет, имеющих спутников, а также суммы масс двойных звезд. Если масса спутника пренебрежимо мала по сравнению с массой планеты, то для движения спутника справедлив третий закон Кеплера в форме (55.5). Постоянную Кеплера $\mathscr{K}$ можно вычислить, измерив размеры орбиты и время обращения спутника. Зная гравитационную постоянную G, по формуле (55.5) можно вычислить массу планеты M в абсолютных единицах. В астрономии, однако, предпочитают за единицу массы принимать массу Земли. Для определения масс планет в таких единицах не требуется знать числовое значение гравитационной постоянной, известное не очень точно.

В качестве примера найдем отношение массы Солнца $M_{\text{С}}$ к массе Земли $m_{\text{З}}$. Массу Земли будем считать пренебрежимо малой по сравнению с массой Солнца. Точно так же пренебрежем массой Луны по сравнению с массой Земли. Для земной орбиты имеем $a_{\text{З}} = 1{,}496 \cdot 10^8$ км, $T_{\text{З}} = 365{,}26$ суток, для лунной $a_{\text{Л}} = 3{,}844 \cdot 10^5$ км,

$T_{\text{Л}} = 27{,}32$ суток. По формуле (55.5) получаем

$$\frac{M_{\text{С}}}{m_{\text{З}}} = \left(\frac{a_{\text{З}}}{a_{\text{Л}}}\right)^3 \left(\frac{T_{\text{Л}}}{T_{\text{З}}}\right)^2 = 3{,}298 \cdot 10^5.$$

В действительности, как видно из формулы (59.3), таким путем находится отношение $\frac{M_{\text{С}} + m_{\text{З}}}{m_{\text{З}} + m_{\text{Л}}}$. Метод дает отношение масс центральных тел, вокруг которых вращаются спутники, только тогда, когда масса каждого спутника пренебрежимо мала по сравнению с массой соответствующего центрального тела. Это условие идеально соблюдается для искусственных спутников. Например, можно найти отношение масс Луны и Земли, если измерить параметры орбит искусственных спутников, обращающихся вокруг них.

Труднее определить массу планеты, если у нее нет спутников. В этом случае применяются два способа. Во-первых, массу планеты можно вычислить по вызываемому ею возмущению в движении других небесных тел. Примером может служить Меркурий, масса которого была определена по возмущениям орбиты кометы Энке. Во-вторых, массу планеты можно оценить по ее блеску, если сделать правдоподобные предположения относительно ее плотности и альбедо, т. е. величины, характеризующей способность поверхности тела отражать (рассеивать) падающее на нее излучение. Таким путем впервые была оценена масса Плутона. Труднее определять массы звезд. В случае двойной звезды это делается аналогично тому, как и для планеты со спутником, — по периоду обращения компонентов звезды вокруг их центра масс и по расстоянию между ними. Но этот метод определяет не массу каждого компонента звезды, а только сумму этих масс.

ЗАДАЧИ

1. Найти расстояние R между компонентами двойной звезды, если их общая масса $M_1 + M_2$ равна удвоенной массе Солнца M_0, и звезды обращаются по круговым орбитам вокруг их центра масс с периодом $T = 2T_0$, где T_0 — продолжительность земного года. Расстояние от Земли до Солнца $R_0 = 1{,}5 \cdot 10^8$ км.

Ответ. $R = \sqrt[3]{\left(\frac{T}{T_0}\right)^2 \frac{M_1 + M_2}{M_0}} = 2R_0 = 3 \cdot 10^8$ км.

2. Минимальное расстояние между компонентами двойной звезды, обращающимися один относительно другого, равно r_1. Относительная скорость их в этом положении равна v_1. Сумма масс обоих компонентов равна M. Найти расстояние между компонентами r_2 и их относительную скорость v_2 при максимальном удалении относительно друг от друга. При каком минимальном значении относительной скорости v_1 двойная звезда распадается?

Ответ. $r_2 = \left(\frac{2GM}{2GM - r_1 v_1^2} - 1\right) r_1$; $v_2 = \frac{r_1}{r_2} v_1$.

Звезда распадется, если $v_1 \geqslant \sqrt{2GM/r_1}$.

§ 60. ПРИМЕНЕНИЕ ЗАКОНА ВСЕМИРНОГО ТЯГОТЕНИЯ К ПРОБЛЕМЕ ЗЕМНОЙ ТЯЖЕСТИ

По мысли Ньютона, вес тел на Земле является проявлением силы гравитационного притяжения между рассматриваемым телом и Землей*). Для проверки этой идеи Ньютон сравнил ускорение свободного падения тел у поверхности Земли с ускорением Луны на орбите, по которой она движется относительно Земли.

Допустим, что вещество внутри земного шара распределено сферически симметрично, т. е. его плотность зависит только от расстояния до центра Земли. В этом случае, как было показано в § 55, Земля создает во внешнем пространстве такое же гравитационное поле, что и материальная точка той же массы, помещенная в центре Земли. Если верна гипотеза Ньютона, то ускорение свободного падения $g_{\text{абс}}$ на расстоянии r от центра Земли должно определяться формулой

$$g_{\text{абс}} = G\frac{M}{r^2}, \tag{60.1}$$

где M — масса Земли. Той же формулой должно определяться ускорение Луны $a_{\text{Л}}$ на ее орбите:

$$a_{\text{Л}} = G\frac{M}{R^2}, \tag{60.2}$$

где R — радиус лунной орбиты. Таким образом,

$$g_{\text{абс}} = a_{\text{Л}}\left(\frac{R}{r}\right)^2. \tag{60.3}$$

Если $a_{\text{Л}}$ известно, то с помощью этой формулы можно вычислить ускорение свободного падения $g_{\text{абс}}$ на поверхности Земли. Это и было сделано Ньютоном.

Ускорение Луны $a_{\text{Л}}$ можно вычислить, зная R и период обращения Луны по ее орбите T (относительно звезд). Эти величины равны соответственно $R = 3{,}844\cdot 10^5$ км, $T = 27{,}32$ суток. Используя их, находим

$$a_{\text{Л}} = \frac{4\pi^2}{T^2}R = 0{,}2723 \text{ см/с}^2. \tag{60.4}$$

Средний радиус земного шара r, определяемый из условия, чтобы величина $\frac{4}{3}\pi r^3$ равнялась объему Земли, равен $r = 6371$ км. Подставляя эти данные в формулу (60.3), получим $g_{\text{абс}} = 991{,}4$ см/с^2. Эта величина близка к экспериментальным значениям: на полюсе $g_{\text{абс}} = 983{,}2$ см/с^2, на экваторе $g_{\text{абс}} = 981{,}4$ см/с^2. Близкое совпадение может рассматриваться как подтверждение гипотезы Ньютона.

*) Сила веса, о которой идет речь в этом утверждении, строго говоря, равна силе гравитационного притяжения только в том случае, когда взвешивание производится на весах, покоящихся или не имеющих ускорения относительно инерциальной системы отсчета (см. § 66)

Небольшое расхождение обусловлено, главным образом, тем, что мы не учли движение самой Земли. Формула (60.4) дает ускорение Луны относительно Земли $(a_{\text{Л}})_{\text{отн}}$, тогда как в формулу (60.3) должно входить ускорениеЛуны относительно инерциальной системы отсчета $(a_{\text{Л}})_{\text{абс}}$. Согласно формуле (59.4) эти ускорения связаны между собой соотношением

$$(a_{\text{Л}})_{\text{отн}} = \left(1 + \frac{m}{M}\right)(a_{\text{Л}})_{\text{абс}},$$

где m — масса Луны. Следовательно, вычисленное выше значение $g_{\text{абс}}$ надо уменьшить в $(1 + m/M)$ раз. Отношение массы Луны к массе Земли составляет $m/M = 1/81$. Введя эту поправку, получим $g_{\text{абс}} = 979{,}3$ см/с^2, что значительно лучше согласуется с опытом. Оставшееся небольшое расхождение можно объяснить отступлениями формы Земли от шаровой.

Заметим, что с помощью формулы (6.1) можно вычислить массу Земли. Для этого надо знать числовое значение гравитационной постоянной G.

ЗАДАЧИ

1. Показать, что если высота над земной поверхностью мала по сравнению с радиусом Земли R, то зависимость ускорения свободного падения на Земле от высоты определяется приближенной формулой

$$g \approx g_0\left(1 - 2\frac{h}{R}\right) \approx g_0(1 - 0{,}00314h),$$

где g_0 — значение g на земной поверхности. Предполагается, что высота h измеряется в километрах.

2. Для вычисления средней плотности Земли δ Эйри (1801—1892) предложил и осуществил следующий метод. Измеряются ускорения свободного падения g_0 на поверхности Земли и g в шахте глубиной h. Принимается, что плотность Земли в поверхностном слое толщиной h однородна и равна $\delta_0 = 2{,}5$ г/см^3. (Это предположение плохо соответствует действительности.) В опытах Эйри было $g - g_0 = 0{,}000052\, g_0$, $R/h = 16\,000$ (R — радиус Земли). Пользуясь этими данными, вычислить среднюю плотность Земли. (Обратите внимание, что g вблизи поверхности Земли возрастает с глубиной! Чем это объясняется?)

Ответ. $\delta \approx \dfrac{3\delta_0}{2 - \dfrac{g - g_0}{g_0}\dfrac{R}{h}} \approx 6{,}5$ г/см^3.

3. Допустим, что в земном шаре вдоль оси вращения просверлен канал от полюса к полюсу. Как будет двигаться материальная точка, помещенная в такой канал без начальной скорости? Плотность вещества земного шара ρ считать однородной.

Ответ. Точка будет совершать гармонические колебания с круговой частотой, определяемой соотношением $\omega^2 = 4/3\ \pi\rho G = g/R$, где R — радиус земного шара, g — ускорение свободного падения на поверхности Земли. Период

этих колебаний $T = 2\pi\sqrt{R/g} \approx 84$ мин. Интересно отметить, что *период колебаний зависит только от плотности шара, но не зависит от его размеров.*

Определить начальную скорость метеоритов v_∞, если максимальное прицельное расстояние, при котором они еще падают на Землю, равно l ($l > R$, где R — радиус земного шара). Получить числовой ответ при $l = 2R$. (См. примечание к задаче § 58.)

Ответ. $v_\infty = R\sqrt{\frac{2gR}{l^2 - R^2}}$. При $l = 2R$ $v_\infty = \sqrt{\frac{2}{3}gR} \approx 6{,}5$ км/с.

5. Вычислить массу Земли, используя параметры орбиты советского искусственного спутника «Космос-380». Период обращения спутника (относительно звезд) $T = 102{,}2$ мин, расстояние до поверхности Земли в перигее 210 км, а апогее 1548 км. Землю считать шаром с радиусом 6371 км.

Ответ. $M = \frac{4\pi^2}{G}\frac{a^2}{T^2} \approx 6 \cdot 10^{27}$ г, где a — половина длины большой оси эллиптической орбиты спутника.

§ 61. КОСМИЧЕСКИЕ СКОРОСТИ

1. Теория финитных и инфинитных движений планет, изложенная в § 57, полностью применима к движению искусственных спутников Земли и космических кораблей (разумеется, с выключенными двигателями). Сопротивление воздуха мы не будем учитывать, предполагая, что движение происходит в достаточно разреженной атмосфере. Кроме того, при движении вблизи Земли мы будем пренебрегать силами гравитационного притяжения Солнца, Луны и планет. Массу Земли будем обозначать буквой M, массу искусственного спутника — буквой m.

Полная энергия спутника или космического корабля в поле земного тяготения равна

$$E = \frac{mv^2}{2} - G\frac{Mm}{r},$$

или в силу соотношения (60.1)

$$E = \frac{mv^2}{2} - mrg_{\text{абс}}. \tag{61.1}$$

(В дальнейшем будем писать просто g вместо $g_{\text{абс}}$.) Если энергия E отрицательна, то движение финитно и будет происходить по эллиптической траектории. При круговом движении

$$v_{\text{к}} = \sqrt{G\frac{M}{r}} = \sqrt{gr}. \tag{61.2}$$

Если r — радиус земного шара, то получаемая по этой формуле величина называется *первой космической скоростью.* Она приблизительно равна 8 км/с.

Минимальное значение E, при котором движение становится инфинитным, равно нулю. В этом случае получается движение по

параболе со скоростью

$$v_{\text{п}} = \sqrt{2gr} = v_{\text{к}}\sqrt{2} \approx 11{,}2 \text{ км/с}, \qquad (61.3)$$

называемой *параболической* или *второй космической скоростью*. Это есть минимальная скорость, которую необходимо сообщить телу, чтобы оно никогда не вернулось на Землю (при условии, что тело не подвергается гравитационному действию со стороны других небесных тел).

Если, наконец, полная энергия E положительна, т. е. начальная скорость тела превосходит вторую космическую скорость, то его движение станет гиперболическим.

2. Совершенно аналогичные вычисления можно провести и для движений в гравитационном поле Солнца. Среднее расстояние до Солнца составляет 150 000 000 км. Скорость Земли при круговом движении на таком расстоянии 29,8 км/с. Для того чтобы при запуске с такого расстояния тело навсегда покинуло пределы Солнечной системы, надо сообщить ему скорость относительно Солнца не меньше $29{,}8\sqrt{2} \approx 42{,}1$ км/с. Находясь на Земле, тело движется вместе с ней вокруг Солнца со скоростью 29,8 км/с. Если бы тело не подвергалось действию земного притяжения, то ему достаточно было бы сообщить относительно Земли дополнительную скорость $42{,}1 - 29{,}8 = 12{,}3$ км/с в направлении ее движения, чтобы относительно Солнца оно стало двигаться с параболической скоростью и навсегда покинуло пределы Солнечной системы. В действительности же для этого требуется бóльшая скорость, так как тело дополнительно должно преодолеть действие земного притяжения. Скорость относительно Земли, которую необходимо сообщить телу, чтобы оно навсегда покинуло пределы Солнечной системы, называется *третьей космической скоростью*. Значение третьей космической скорости зависит от того, в каком направлении корабль выходит из зоны действия земного тяготения. Она минимальна, если это направление совпадает с направлением орбитального движения Земли вокруг Солнца, и максимальна, когда эти направления противоположны.

Точное вычисление третьей космической скорости довольно кропотливо, так как при этом надо учесть гравитационное взаимодействие трех тел: Солнца, Земли и космического корабля. Однако такое вычисление не представляет большого труда, если пренебречь влиянием поля тяготения Солнца на движение космического корабля в течение всего времени, которое он затрачивает для выхода из зоны действия земного тяготения*). Будем обозначать малыми буквами (v, $v_{\text{к}}$, $v_{\text{п}}$) скорости корабля относи-

*) Более подробное рассмотрение показывает (см. § 65), что в действительности при таком расчете мы пренебрегаем не полем тяготения Солнца, а лишь его *неоднородностью* в той области пространства, где преобладающим является поле тяжести Земли. Однородная составляющая поля тяготения Солнца компенсируется силами инерции, возникающими из-за свободного падения Земли на Солнце. Поэтому ошибка, которую мы делаем при вычислении третьей космической скорости, ничтожна.

тельно Земли. Все скорости относительно Солнца будем обозначать большими буквами (V, $V_к$, $V_п$). Пока корабль движется в поле земного тяготения, его движение удобнее относить к системе отсчета, в которой Земля неподвижна. Считая массу Земли M бесконечно большой по сравнению с массой корабля m, запишем уравнение энергии в виде

$$\frac{mv^2}{2} - G\frac{Mm}{r} = \frac{mv_\infty^2}{2},$$

где v_∞ — скорость корабля в тот момент, когда он практически выходит из зоны действия земного тяготения. Вводя круговую скорость $v_к^2 = GM/r$, получаем $v_\infty^2 = v^2 - 2v_к^2$. После того как корабль выйдет из зоны действия земного тяготения, будем относить его движение к системе отсчета, в которой неподвижно Солнце. В момент выхода из зоны тяготения скорость корабля $\mathbf{V}$ в этой системе равна векторной сумме скорости $\mathbf{v}_\infty$ и скорости кругового движения Земли $\mathbf{V}_к$. Если корабль выходит из зоны земного тяготения под углом ϑ, то такой же угол будет между скоростями $\mathbf{v}_\infty$ и $\mathbf{V}$. Значит,

$$V^2 = V_к^2 + v_\infty^2 + 2V_к v_\infty \cos\vartheta.$$

Третья космическая скорость v_3 найдется из условия $V = V_п \equiv \sqrt{2}V_к$. Подставляя это значение для V в предыдущее соотношение, получим квадратное уравнение для v_∞, из которого найдем

$$v_\infty = (\sqrt{1 + \cos^2\vartheta} - \cos\vartheta)V_к.$$

Положительный знак перед квадратным корнем выбран потому, что величина v_∞ по своему смыслу существенно положительна. После этого получим

$$v_3^2 = (\sqrt{1 + \cos^2\vartheta} - \cos\vartheta)^2 V_к^2 + 2v_к^2 \tag{61.4}$$

Минимальное значение третьей космической скорости получится при $\vartheta = 0$ (запуск в направлении орбитального движения Земли), а максимальное — при $\vartheta = \pi$ (запуск в направлении против орбитального движения Земли). Для этих значений формула (61.4) дает

$$\begin{aligned} v_3^{\text{мин}} &\approx \sqrt{0{,}171\ V_к^2 + 2v_к^2} \approx 16{,}7 \text{ км/с}, \\ v_3^{\text{макс}} &\approx \sqrt{5{,}828\ V_к^2 + 2v_к^2} \approx 72{,}7 \text{ км/с}. \end{aligned} \tag{61.5}$$

Вычислим теперь приближенно *четвертую космическую скорость* v_4. Так называется минимальная скорость, которую надо сообщить ракете, чтобы она могла упасть в заданную точку Солнца. Такая скорость зависит от положения этой точки на поверхности Солнца. На старте ракета движется вокруг Солнца вместе с Землей со скоростью $V_к$. Чтобы ракета упала на Солнце, ее движение надо затормозить. Как и ранее находим, что при выходе из зоны земного притяжения скорость ракеты будет $\mathbf{V} = \mathbf{V}_к + \mathbf{v}_\infty$ (относительно Солнца). Наименьшая энергия, которую нужно затратить для замедления, получится тогда, когда скорости $\mathbf{V}_к$ и $\mathbf{v}_\infty$ направлены противоположно. В этом случае $V = V_к - v_\infty$ (все скорости положительны), а энергия,

приходящаяся на единицу массы ракеты, равна

$$\varepsilon = \frac{1}{2}(V_к - v_\infty)^2 - G\frac{M}{R} = -\frac{1}{2}(V_к^2 + 2V_к v_\infty - v_\infty^2),$$

где $R = CA$ — расстояние ракеты до центра Солнца при ее максимальном удалении (рис. 181a). Если $\varepsilon < 0$, то траекторией ракеты будет эллипс с большой осью

$$2a = -G\frac{M}{\varepsilon} = \frac{2RV_к^2}{V_к^2 + V_к v_\infty - v_\infty^2}.$$

Один из фокусов эллипса находится в центре Солнца. Обозначим через $x = CP$ расстояние от центра Солнца до ближайшей вершины этого эллипса. Расстояние x однозначно определяет форму эллипса, а с ней и линию на поверхности Солнца, на которой будет лежать точка падения. Большая ось эллипса $2a = R + x$. Подставив это значение в предыдущее уравнение, придем к квадратному уравнению для v_∞. Меньший корень этого уравнения равен

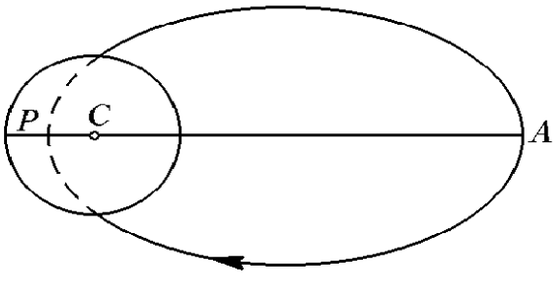


Рис. 181а

$$v_\infty = V_к\left(1 - \sqrt{\frac{2x}{R+x}}\right).$$

Четвертая космическая скорость v_4 ракеты определится из соотношения $v_4^2 = v_\infty^2 + 2v_к^2$, или

$$v_4^2 = V_к^2\left(1 - \sqrt{\frac{2x}{R+x}}\right)^2 + 2v_к^2.$$

Она зависит от параметра x, определяющего место падения. При $x = 0$ (прямолинейное движение по направлению к центру Солнца) скорость v_4 максимальна и равна

$$v_4^{\text{макс}} = (V_к^2 + 2v_к^2)^{1/2} \approx 31{,}8 \text{ км/с}.$$

Ракета упадет в передней точке Солнца. При $x = r$ (r — радиус Солнца) ракета упадет в задней точке Солнца, двигаясь по касательной к его поверхности. В этом случае скорость минимальна и равна

$$v_4^{\text{мин}} \approx \left[V_к^2\left(1 - \sqrt{\frac{2r}{R+r}}\right)^2 + 2v_к^2\right]^{1/2} \approx [V_к^2(1 - \sqrt{2\alpha})^2 + 2v_к^2]^{1/2} \approx 29{,}2 \text{ км/с},$$

где $\alpha = 4{,}65 \cdot 10^{-3}$ рад — средний угловой радиус Солнца.

ЗАДАЧИ

1. Искусственный спутник Земли вращается по круговой орбите радиусом R с периодом T_1. В некоторый момент на очень короткое время был включен реактивный двигатель, увеличивший скорость спутника в α раз, и спутник стал вращаться по эллиптической орбите. Двигатель сообщал ускорение спутнику все время в направлении движения. Определить макси-

мальное расстояние спутника от центра Земли, которого он достигнет после выключения двигателя. Найти также период T_2 обращения спутника по новой (эллиптической) орбите.

Р е ш е н и е. Обозначим через $E_к$ полную энергию спутника при движении по круговой орбите. Согласно (58.3) $E_к = -K$, $U = -2K$. После того как отработал двигатель, скорость спутника возросла в α раз, а кинетическая энергия K — в α^2 раз. Потенциальная энергия не изменилась, так как за время работы двигателя спутник переместился пренебрежимо мало. Таким образом, полная энергия спутника на эллиптической орбите будет

$$E_{эл} = \alpha^2 K + U = (\alpha^2 - 2)K = (2 - \alpha^2)E_к.$$

Большие оси эллиптических орбит обратно пропорциональны полным энергиям (см. формулу (58.2)). Поэтому

$$\frac{a}{R} = \frac{1}{2-\alpha^2}, \quad a = \frac{R}{2-\alpha^2}.$$

Орбита будет эллиптической, если $\alpha^2 \leqslant 2$. Максимальное расстояние спутника от центра Земли (в апогее)

$$R_{макс} = 2a - R = \frac{\alpha^2 R}{2-\alpha^2}.$$

Период обращения T_2 найдется из третьего закона Кеплера и равен

$$T_2 = \frac{T_1}{(2-\alpha^2)^{3/2}}.$$

2. Найти такой радиус R круговой орбиты спутника Земли, движущегося в направлении ее вращения в плоскости земного экватора, чтобы он все время оставался неподвижным относительно Земли. (Такой спутник называется *стационарным*.)

О т в е т. $R = (g/\omega^2 R_0)^{1/3} R_0 \approx 6{,}60\ R_0$. Здесь R_0 — экваториальный радиус Земли, $\omega^2 R_0$ — центростремительное ускорение на экваторе, обусловленное осевым вращением Земли, g — ускорение свободного падения. На экваторе $\omega^2 R_0/g = 1/288$.

3. Силы приливного трения, вызываемые лунными приливами, замедляют осевое вращение Земли. Этот процесс будет продолжаться до тех пор, пока не сделаются равными угловые скорости осевого вращения Земли и орбитального движения Луны вокруг Земли. Определить общую угловую скорость ω обоих вращений, продолжительность земных суток T и радиус лунной орбиты a после того, как это произойдет. В настоящее время угловая скорость осевого вращения Земли равна $\omega_З = 7{,}29 \cdot 10^{-5}$ рад/с, момент импульса Земли относительно своей оси $L_З = 5{,}91 \cdot 10^{40}$ г·см^2/с, момент инерции Земли относительно той же оси $I_З = 8{,}11 \cdot 10^{44}$ г·см^2, радиус лунной орбиты $a_0 = 3{,}84 \cdot 10^{10}$ см, период обращения Луны вокруг Земли (относительно звезд) $T_Л = 27{,}3$ сут, масса

Луны $m = 7{,}35 \cdot 10^{25}$ г. Для упрощения расчета считать, что земная ось перпендикулярна к плоскости лунной орбиты.

Решение. Используя приведенные данные, находим: момент инерции Луны относительно оси вращения Земли $I_Л = ma_0^2 = 1{,}08 \cdot 10^{47}$ г·см^2 (моментом инерции Луны относительно ее собственной оси пренебрегаем), угловая скорость орбитального вращения Луны вокруг Земли $\omega_Л = 2{,}67 \cdot 10^{-6}$ рад/с, момент импульса Луны относительно Земли $L_Л = I_Л\omega_Л = 28{,}9 \cdot 10^{40}$ г·см^2/с, полный момент импульса системы Земля—Луна $L = L_З + L_Л = 34{,}8 \cdot 10^{40}$ г·см^2/с. По закону сохранения момента импульса $(I_З + ma^2)\omega = L$, или, пренебрегая $I_З$, $ma^2\omega = L$. По третьему закону Кеплера $a^3\omega^2 = a_0^3\omega_Л^2$. Из этих двух уравнений можно найти неизвестные a и ω. В указанном приближении

$$a = \frac{L^2}{m^2 a_0^4 \omega_Л^2} a_0 = \left(\frac{L}{L_Л}\right)^2 a_0 = 1{,}45, \quad a_0 = 5{,}58 \cdot 10^{10} \text{ см},$$

$$\frac{\omega}{\omega_Л} = \left(\frac{a_0}{a}\right)^{3/2} = 0{,}573, \quad T = \frac{27{,}3}{0{,}573} = 47{,}7 \text{ сут}.$$

4. Космический корабль подходит к Луне по параболической траектории, почти касающейся поверхности Луны. Чтобы перейти на стелящуюся круговую орбиту, в момент наибольшего сближения включают тормозной двигатель, выбрасывающий газы со скоростью $v = 4$ км/с относительно корабля в направлении его движения. Какую часть общей массы системы будет составлять горючее, использованное для торможения корабля? Средний радиус Луны $R = 1738$ км, ускорение свободного падения на поверхности Луны $g = 162$ см/с^2.

Ответ. $(m_0 - m)/m_0 = (\sqrt{2} - 1)\sqrt{gR}/v \approx 0{,}17$.

5. Искусственный спутник движется вокруг Земли в разреженной атмосфере по круговой (или почти круговой) орбите. Как влияет сопротивление среды на скорость движения спутника и его момент импульса относительно центра Земли?

Решение. Согласно (58.3) при круговом движении $E = -K$. Трение уменьшает полную энергию E. Поэтому кинетическая энергия K возрастает (спутник приближается к Земле).

6. Космический корабль без начальной скорости свободно падает на Землю из удаленной точки. В каком месте следует повернуть направление скорости корабля на 90° (без изменений ее модуля), чтобы он стал двигаться вокруг Земли по круговой траектории?

Ответ. Посередине между центром Земли и начальным положением корабля.

7. Космический корабль движется вокруг Земли по эллиптической орбите. В какой точке орбиты следует изменить направление скорости корабля (без изменения ее модуля), чтобы корабль стал двигаться по круговой орбите?

Решение. Так как энергия корабля зависит только от длины $2a$ большой оси его орбиты, то переход на круговую орбиты произойдет на расстоянии a, т. е. в точке пересечения эллипса с его малой осью. Направление скорости корабля надо повернуть на такой угол, чтобы оно оказалось перпендикулярным к линии, соединяющей корабль с центром Земли.

8. Космический корабль движется вокруг Земли по эллиптической орбите. В точке пересечения эллипса с его малой осью включается двигатель. Как надо изменить скорость корабля в этой точке, чтобы он перешел на параболическую орбиту?

Ответ. Увеличить в $\sqrt{2}$ раз.

9. Какую перегрузку испытывает при старте космонавт в космическом корабле на самом начальном участке полета, когда корабль вместе с ракетой-носителем поднимается вертикально вверх с постоянным ускорением и за время $\tau = 4$ с набирает скорость $v = \alpha v_к$, где $v_к$ — первая космическая скорость, а $\alpha = 0{,}03$? (Перегрузкой называется отношение $n = (P - P_0)/P_0$, где P_0 — вес космонавта на Земле, а P — «вес», который показали бы пружинные весы при взвешивании космонавта в корабле.)

Решение. Примем за положительное направление направление вверх. «Вес» космонавта в корабле будет

$$P = P_0 + m\frac{dv}{dt}.$$

Считая на начальном участке величину P постоянной, находим скорость корабля через время τ:

$$v \equiv \alpha v_к = \frac{P - P_0}{m}\tau.$$

Отсюда

$$\frac{P - P_0}{P_0} = \frac{\alpha v_к}{g\tau} = \frac{\alpha}{\tau}\sqrt{\frac{R}{g}} = 6.$$

§ 62. ВЫВОД ЗАКОНОВ ДВИЖЕНИЯ ПЛАНЕТ ИЗ ЗАКОНА ВСЕМИРНОГО ТЯГОТЕНИЯ НЬЮТОНА

В предыдущих параграфах три закона Кеплера были приняты за исходные. Пользуясь ими, мы пришли к закону всемирного тяготения Ньютона. Теперь поступим наоборот. Примем, что на планету со стороны Солнца действует сила тяготения, подчиняющаяся закону Ньютона. Найдем движение планеты под действием такой силы. Массу Солнца будет считать бесконечно большой по сравнению с массой планеты. К такому случаю сводится и общий случай, когда это условие не выполняется (см. § 59). Возьмем полярную системы координат (r, φ), полюс которой поместим в центре Солнца. Скорость планеты v можно разложить на радиальную скорость $v_r = \dot{r}$ и перпендикулярную к ней азимутальную скорость $v_\varphi = r\dot{\varphi}$. Очевидно, $v^2 = \dot{r}^2 + r^2\dot{\varphi}^2$. Законы сохранения энергии и момента импульса планеты запишем в виде

$$\frac{1}{2}(\dot{r}^2 + r^2\dot{\varphi}^2) - G\frac{M}{r} = \varepsilon, \tag{62.1}$$

$$\frac{1}{2}r^2\dot{\varphi}^2 = \sigma, \tag{62.2}$$

где M — масса Солнца, ε — полная энергия планеты, приходящаяся на единицу ее массы, σ — секториальная скорость, остающаяся постоянной

во время движения. Для нахождения уравнения траектории планеты исключим время. Считая r функцией φ, имеем $\dot{r} = \frac{dr}{d\varphi}\dot{\varphi}$. Подставляя это значение в уравнение (62.1) и исключая $\dot{\varphi}$ с помощью уравнения (62.2), получим

$$\left(\frac{1}{r^2}\frac{dr}{d\varphi}\right)^2 + \frac{1}{r^2} = \frac{1}{2\sigma^2}\left(\varepsilon + G\frac{M}{r}\right). \tag{62.3}$$

Введем новую переменную $\rho = -1/r + 1/p$, p — постоянная, значение которой будет установлено ниже. Тогда уравнение (62.3) примет вид

$$\left(\frac{d\rho}{d\varphi}\right)^2 + \left(\rho - \frac{1}{p}\right)^2 = \frac{\varepsilon}{2\sigma^2} + G\frac{M}{2\sigma^2}\left(-\rho + \frac{1}{p}\right).$$

Подберем постоянную p так, чтобы в этом уравнении исчезали члены, содержащие первые степени ρ. Для этого надо положить

$$p = \frac{4\sigma^2}{GM}. \tag{62.4}$$

При таком выборе постоянной p получим

$$\left(\frac{d\rho}{d\varphi}\right)^2 = \frac{\varepsilon}{2\sigma^2} + \frac{1}{p^2} - \rho^2.$$

Поскольку слева стоит неотрицательная величина, постоянная величина $1/p^2 + \varepsilon/2\sigma^2$ также неотрицательна, ее можно обозначить через A^2:

$$A^2 = \frac{1}{p^2} + \frac{\varepsilon}{2\sigma^2}. \tag{62.5}$$

В результате получим

$$\left(\frac{d\rho}{d\varphi}\right)^2 = A^2 - \rho^2. \tag{62.6}$$

Очевидно, $A^2 \geqslant \rho^2$, а потому можно положить $\rho/A = \cos\Theta$, где Θ — новая неизвестная. Тогда

$$A^2 - \rho^2 = A^2\sin^2\Theta, \quad \frac{d\rho}{d\varphi} = -A\sin\Theta\frac{d\Theta}{d\varphi}.$$

Подставляя в (62.6) и сокращая на $A\sin\Theta$, получим $\frac{d\Theta}{d\varphi} = \pm 1$, откуда $\Theta = \pm\varphi + \varphi_0$. Следовательно, $\rho = A\cos(\pm\varphi + \varphi_0) = A\cos(\varphi \pm \varphi_0)$. В последнем выражении двойной знак перед φ_0 сохранять не имеет смысла, поскольку φ_0 есть постоянная интегрирования. Возвращаясь к прежним обозначениям, получим

$$\frac{1}{r} = \frac{1}{p}[1 - e\cos(\varphi + \varphi_0)], \tag{62.7}$$

где

$$e = pA = \sqrt{1 + \frac{\varepsilon p^2}{2\sigma^2}} = \sqrt{1 + \frac{8\varepsilon\sigma^2}{G^2M^2}}. \tag{62.8}$$

Без ограничения общности можно положить $\varphi_0 = 0$. Это означает просто, что отсчет углов φ ведется от такого положения радиуса-вектора планеты, когда его длина равна $\frac{p}{(1-e)}$. При таком отсчете уравнение (62.7) принимает вид

$$r = \frac{p}{(1-e\cos\varphi)}. \tag{62.9}$$

Это — уравнение конического сечения с эксцентриситетом e и параметром p. Если $\varepsilon < 0$, то $e < 1$ (эллипс); если $\varepsilon = 0$, то $e = 1$ (парабола); если $\varepsilon > 0$, то $e > 1$ (гипербола). Мы пришли к результатам, полученным в § 57 иным путем. Нетрудно теперь вычислить остальные параметры орбиты и в случае эллиптического движения получить третий закон Кеплера. Однако все эти вычисления уже были проделаны ранее, и в новых вычислениях нет необходимости.

ГЛАВА IX

ДВИЖЕНИЕ ОТНОСИТЕЛЬНО НЕИНЕРЦИАЛЬНЫХ СИСТЕМ ОТСЧЕТА

§ 63. СИЛЫ ИНЕРЦИИ ПРИ УСКОРЕННОМ ПОСТУПАТЕЛЬНОМ ДВИЖЕНИИ СИСТЕМЫ ОТСЧЕТА

1. До сих пор мы относили движение к какой-либо одной из бесчисленного множества *инерциальных систем отсчета.* В такой системе отсчета основным уравнением движения материальной точки является уравнение, выражающее *второй закон Ньютона.* Запишем здесь это уравнение в виде

$$m\mathbf{a}_{\text{абс}} = \mathbf{F}, \tag{63.1}$$

снабдив ускорение $\mathbf{a}$ индексом «абс», смысл которого выяснится в дальнейшем. Поставим теперь задачу найти уравнения движения в *неинерциальных системах отсчета*, т. е. таких системах, которые движутся ускоренно относительно инерциальных систем. Задача сводится к установлению *законов преобразования сил и ускорений* при переходе от инерциальной системы к любой неинерциальной системе отсчета. Дорелятивистская физика считала этот вопрос чисто кинематическим и решала его на основе следующих двух допущений: 1) время абсолютно, т. е. промежутки времени между любыми двумя событиями одинаковы во всех системах отсчета; 2) пространство абсолютно, т. е. расстояния между любыми двумя точками (материальными телами) также одинаковы во всех системах отсчета. Таким образом, в дорелятивистской физике считалось, что *расстояния и промежутки времени инварианты по отношению к переходу от одной системы отсчета к любой другой, произвольно движущейся системе отсчета.* Оба допущения казались настолько самоочевидными, что даже явно не формулировались. И только глубокий анализ проблемы пространства и времени в теории относительности выявил *постулативный характер* этих допущений. При этом оказалось, что *оба допущения приближенно верны лишь для медленных движений.* При быстрых движениях они становятся неверными. Ограничимся сейчас нерелятивистским рассмотрением, т. е. *будем предполагать, что все скорости, в том числе и относительные скорости самих систем отсчета, малы по сравнению со скоростью света в вакууме.*

2. Условимся называть *неподвижной* какую-либо произвольно выбранную инерциальную систему отсчета, а движение относительно

нее — *абсолютным*. В формуле (63.1) речь идет об ускорении при абсолютном движении именно в таком смысле. Не следует вкладывать в понятия «неподвижная система отсчета» и «абсолютное движение» что-либо большее по сравнению с тем, что содержится в приведенном определении. Оба понятия чисто условны и не противоречат утверждению, что всякое движение *относительно*. Тело, покоящееся в движущейся системе отсчета, увлекается последней в ее движении относительно неподвижной системы отсчета. Такое движение тела называется *переносным*. Абсолютное движение тела складывается из его относительного и переносного движений.

Цель настоящей главы — изучить *относительное движение*. Для этого прежде всего следует установить уравнения относительного движения. *Под уравнениями движения мы понимаем соотношения, которыми определяются ускорения всех материальных точек механической системы в той системе отсчета, относительно которой рассматривается движение*. Когда система отсчета движется относительно неподвижной системы отсчета прямолинейно и равномерно, она сама является инерциальной системой отсчета. В этом случае уравнения относительного движения совпадают с уравнениями *абсолютного движения*, т. е. даются законами Ньютона. Поэтому достаточно ограничиться рассмотрением только тех случаев, когда рассматриваемая система отсчета движется относительно неподвижной системы отсчета с ускорением.

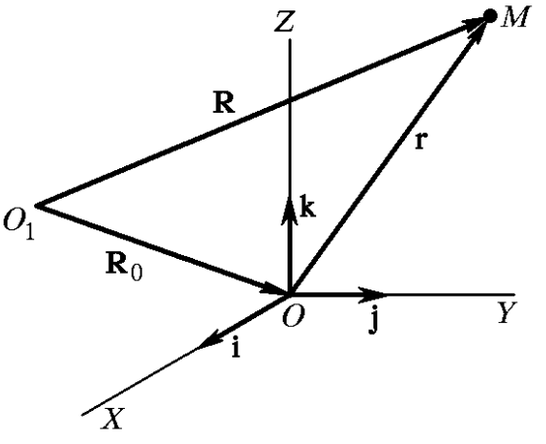


Рис. 182

3. Возьмем две системы отсчета: неподвижную систему S_1 с началом координат в точке O_1 и движущуюся систему S с началом координат в точке O (рис. 182). Обозначим через $\mathbf{R}_0$ радиус-вектор $\overrightarrow{O_1O}$, проведенный из неподвижного начала O_1 к движущемуся началу O. Пусть M — какая-либо материальная точка. Ее положение в неподвижной системе отсчета определяется радиусом-вектором $\mathbf{R}$, а в движущейся — радиусом-вектором $\mathbf{r} = \overrightarrow{OM}$. Векторы $\mathbf{R}$, $\mathbf{R}_0$, $\mathbf{r}$ в каждый момент времени связаны соотношением

$$\mathbf{R} = \mathbf{R}_0 + \mathbf{r}. \tag{63.2}$$

Дважды дифференцируя это соотношение по времени, получим

$$\dot{\mathbf{R}} = \dot{\mathbf{R}}_0 + \dot{\mathbf{r}}, \tag{63.3}$$

$$\ddot{\mathbf{R}} = \ddot{\mathbf{R}}_0 + \ddot{\mathbf{r}}. \tag{63.4}$$

Чтобы лучше выявить идейную сторону вопроса, рассмотрим сначала частный случай, когда система S движется относительно неподвижной системы S_1 *поступательно*. Вектор $\dot{\mathbf{R}}$, очевидно, всегда дает абсолютную скорость $\mathbf{v}_{\text{абс}}$, а вектор $\ddot{\mathbf{R}}$ — абсолютное ускорение $\mathbf{a}_{\text{абс}}$ движущейся точки M. Вектор $\mathbf{v}_0 \equiv \dot{\mathbf{R}}_0$ есть абсолютная скорость, а $\mathbf{a}_0 \equiv \ddot{\mathbf{R}}_0$ — абсолютное ускорение начала координат O системы S. При поступательном движении эти величины совпадают соответственно со скоростью и ускорением любой точки системы S. Таким образом, $\mathbf{v}_0$ и $\mathbf{a}_0$ должны быть интерпретированы как *переносные скорость* и *ускорение*. Точно так же при поступательном движении $\dot{\mathbf{r}}$ и $\ddot{\mathbf{r}}$ дают соответственно *относительную скорость* и *относительное ускорение*, т. е. значения этих величин в движущейся системе отсчета S. Итак, при поступательном движении

$$\mathbf{v}_{\text{абс}} = \mathbf{v}_{\text{отн}} + \mathbf{v}_{\text{пер}}, \tag{63.5}$$

$$\mathbf{a}_{\text{абс}} = \mathbf{a}_{\text{отн}} + \mathbf{a}_{\text{пер}}, \tag{63.6}$$

причем $\mathbf{a}_{\text{пер}} = \mathbf{a}_0$, $\mathbf{v}_{\text{пер}} = \mathbf{v}_0$.

4. Подставим теперь выражение (63.6) в уравнение (63.1) и перенесем член, содержащий $\mathbf{a}_{\text{пер}}$, в правую часть. Получим

$$m\mathbf{a}_{\text{отн}} = \mathbf{F} - m\mathbf{a}_0. \tag{63.7}$$

Это и есть уравнение относительного движения материальной точки. На правую часть этого уравнения формально можно смотреть как на некоторую «силу», действующую на материальную точку в движущейся системе отсчета. Таким образом, в каждом системе отсчета сила определяется как вектор, равный произведению массы материальной точки на ее ускорение в этой системе отсчета. Не обязательно, чтобы «сила» в таком смысле была результатом взаимодействия тел. Однако необходимо располагать каким-то независимым способом, позволяющим выразить «силу» через координаты и скорости движущейся точки. Только при этом условии мы в состоянии написать уравнение движения типа (63.7), а к этому в конце концов сводится реальное содержание законов механики.

«Сила» $\mathbf{F} - m\mathbf{a}_0$ слагается из двух существенно различных составляющих. Первая составляющая $\mathbf{F}$ есть «настоящая сила» в том смысле, что она является результатом взаимодействия тел. Она зависит только от разностей координат и разностей скоростей взаимодействующих материальных точек. В нерелятивистской кинематике все эти разности не меняются при переходе от одной системы отсчета к другой, произвольно поступательно движущейся системе. Поэтому не меняется и сила $\mathbf{F}$. Она *инвариантна* относительно такого перехода.

Совсем иной характер имеет составляющая — $m\mathbf{a}_0$. Это составляющая возникает не из-за взаимодействия тел, а из-за *ускоренного*

движения системы отсчета. Она называется *силой инерции*, точнее *поступательной силой инерции*, поскольку сейчас мы ограничиваемся лишь поступательными движениями систем отсчета. При переходе к другой ускоренной системе отсчета меняются и силы инерции. Они *не инвариантны* относительно такого перехода. Этим силы инерции отличаются от «настоящих сил», возникающих при взаимодействии тел. Второе отличие состоит в том, что *силы инерции не подчиняются закону равенства действия и противодействия.* Если на какое-либо тело действует сила инерции, то не существует противодействующей силы, приложенной к другому телу. Движение тел под действием сил инерции аналогично, таким образом, движению во внешних силовых полях. *Силы инерции всегда являются внешними по отношению к любой движущейся системе материальных тел.*

5. Реальны или фиктивны силы инерции? Ответ на этот вопрос зависит от смысла, который вкладывается в слова «реальный» и «фиктивный». Если придерживаться ньютоновской механики, согласно которой все силы должны быть результатом взаимодействия тел, то на силы инерции надо смотреть как на фиктивные силы, исчезающие в инерциальных системах отсчета. Однако такая точка зрения не обязательна. Все взаимодействия осуществляются посредством силовых полей и передаются с конечными скоростями. И на силы инерции можно смотреть как на действия, которым подвергаются тела со стороны каких-то реальных силовых полей. Правда, эти поля определенным образом преобразуются при переходе от рассматриваемой системы отсчета к другой системе, движущейся относительно нее ускоренно. Но это не является основанием считать эти силы фиктивными. Ведь электрические и магнитные силы также преобразуются при переходе к другой системе отсчета (даже от инерциальной к инерциальной). И тем не менее никто не сомневается в реальном существовании электромагнитных полей.

Независимо от того, какую из этих точек зрения мы примем, существует много явлений, которые могут быть интерпретированы как проявление сил инерции. Когда поезд набирает скорость, пассажир в вагоне испытывает действие силы, направленной против движения поезда. Если пассажир сидит по ходу поезда, то эта сила прижимает его к спинке сиденья. Это и есть сила инерции. При торможении поезда сила инерции меняет направление и стремится отделить тело пассажира от стенки сиденья. Если в ускоренно движущемся вагоне висит маятник, то сила инерции стремится отклонить его в сторону, противоположную ускорению. В состоянии равновесия сила инерции уравновешивается силами тяжести и натяжением нити подвеса. Особо заметно проявляются силы инерции при внезапном быстром торможении поезда. Силы инерции вызывают перегрузки, действующие на летчика или космонавта при больших ускорениях самолета или при запуске и торможении космического корабля.

Конечно, все эти явления можно понять, не пользуясь представлением о силах инерции, а рассматривая движение относительно инерциальной системы отсчета. Так, в примере с маятником маятник движется ускоренно относительно инерциальной системы отсчета.

Маятник должен отклониться назад, чтобы возникла сила натяжения с горизонтальной составляющей, направленной вперед. Эта составляющая и сообщает маятнику ускорение. Однако во многих случаях бывает проще рассматривать явления непосредственно в движущейся системе отсчета, не переходя к инерциальной. Кроме того, иногда затруднительно разделить полную силу, действующую в неинерциальной системе отсчета, на «реальную силу», возникающую из-за взаимодействия тел, и «фиктивную» силу инерции, связанную с ускорением движения системы отсчета.

§ 64. СИЛЫ ИНЕРЦИИ ПРИ УСКОРЕННОМ ПРОИЗВОЛЬНОМ ДВИЖЕНИИ СИСТЕМЫ ОТСЧЕТА

1. Допустим теперь, что система отсчета S (см. рис. 182) движется относительно неподвижной системы S_1 *совершенно произвольно.* Это движение можно разложить на два: *поступательное движение* со скоростью $\mathbf{v}_0$, равной скорости движения начала координат O, и *вращательное движение* вокруг мгновенной оси, проходящей через это начало. Угловую скорость этого вращения обозначим через $\boldsymbol{\omega}$. Она может меняться как по модулю, так и по направлению. Пусть $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ — единичные векторы (орты) координатных осей системы координат S, которую мы будем предполагать прямоугольной. Длины этих векторов, поскольку они единичные, остаются неизменными. Но их направления с течением времени могут измениться. Это переменные векторы. Каждый из них вращается с угловой скоростью $\boldsymbol{\omega}$. Их производные по времени определяются формулами (46.11). Выпишем эти формулы еще раз:

$$\frac{d\mathbf{i}}{dt} = [\boldsymbol{\omega}\mathbf{i}], \quad \frac{d\mathbf{j}}{dt} = [\boldsymbol{\omega}\mathbf{j}], \quad \frac{d\mathbf{k}}{dt} = [\boldsymbol{\omega}\mathbf{k}]. \tag{64.1}$$

Ход рассуждений остается в точности таким же, как и в предыдущем параграфе. Усложняются только вычисления. Формулы (63.2)—(63.4), разумеется, остаются без изменения. Остается неизменной и интерпретация слагаемых $\dot{\mathbf{R}}_0$ и $\ddot{\mathbf{R}}_0$. Первое есть абсолютная скорость $\mathbf{v}_0$, а второе — абсолютное ускорение $\mathbf{a}_0$ начала координат O. Меняются только слагаемые $\dot{\mathbf{r}}$ и $\ddot{\mathbf{r}}$, которые мы и должны найти.

2. Пусть x, y, z — координаты движущейся точки M в движущейся системе S. Тогда

$$\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}. \tag{64.2}$$

Дифференцируя это выражение, получим

$$\dot{\mathbf{r}} = (\dot{x}\mathbf{i} + \dot{y}\mathbf{j} + \dot{z}\mathbf{k}) + \left(x\frac{d\mathbf{i}}{dt} + y\frac{d\mathbf{j}}{dt} + z\frac{d\mathbf{k}}{dt}\right).$$

В первой скобке дифференцируются только координаты x, y, z, как если бы единичные векторы $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$, а с ними и система отсчета S были неподвижными. Такую операцию должен был бы выполнить наблюдатель, покоящийся в системе S, если бы он поставил перед собой задачу найти скорость точки M в этой системе, т. е. по нашей терминологии относительную скорость $\mathbf{v}_{\text{отн}}$. Таким образом,

$$\mathbf{v}_{\text{отн}} = \dot{x}\mathbf{i} + \dot{y}\mathbf{j} + \dot{z}\mathbf{k}. \tag{64.3}$$

Используя далее формулы (64.1) получим

$$x\frac{d\mathbf{i}}{dt} + y\frac{d\mathbf{j}}{dt} + z\frac{d\mathbf{k}}{dt} = x[\boldsymbol{\omega}\mathbf{i}] + y[\boldsymbol{\omega}\mathbf{j}] + z[\boldsymbol{\omega}\mathbf{k}] =$$

$$= [\boldsymbol{\omega}(x\mathbf{i} + y\mathbf{j} + z\mathbf{k})] = [\boldsymbol{\omega}\mathbf{r}].$$

Таким образом,

$$\dot{\mathbf{r}} = \mathbf{v}_{\text{отн}} + [\boldsymbol{\omega}\mathbf{r}]. \tag{64.4}$$

Окончательно для абсолютной скорости можно написать

$$\mathbf{v}_{\text{абс}} = \mathbf{v}_{\text{отн}} + \mathbf{v}_{\text{пер}}, \tag{64.5}$$

т. е. выражение, совпадающее с (63.5). Однако теперь переносная скорость дается выражением

$$\mathbf{v}_{\text{пер}} = \mathbf{v}_0 + [\boldsymbol{\omega}\mathbf{r}]. \tag{64.6}$$

Эта величина есть абсолютная скорость, которую имела бы точка M, если бы она покоилась в движущейся системе отсчета S. Поэтому-то она и называется переносной скоростью. Переносная скорость слагается из двух частей: скорости $\mathbf{v}_0$, с которой движется начало координат O, и скорости $[\boldsymbol{\omega}\mathbf{r}]$, возникающей из-за вращения системы S вокруг этого начала.

3. Несколько сложнее обстоит дело с абсолютным ускорением. Для вычисления абсолютного ускорения продифференцируем выражение (64.5) по времени. С учетом соотношения (64.6) находим

$$\mathbf{a}_{\text{абс}} \equiv \dot{\mathbf{v}}_{\text{абс}} = \dot{\mathbf{v}}_{\text{отн}} + \dot{\mathbf{v}}_0 + [\boldsymbol{\omega}\dot{\mathbf{r}}] + [\dot{\boldsymbol{\omega}}\mathbf{r}].$$

Производная $\dot{\mathbf{v}}_{\text{отн}}$ найдется дифференцированием выражения (64.3). При этом, разумеется, надо дифференцировать не только компоненты относительной скорости $\dot{x}$, $\dot{y}$, $\dot{z}$, но и координатные орты $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$. Это делается в точности так же, как и дифференцирование выражения (64.2). Поэтому по аналогии с формулой (64.4) можно написать

$$\dot{\mathbf{v}}_{\text{отн}} = \mathbf{a}_{\text{отн}} + [\boldsymbol{\omega}\mathbf{v}_{\text{отн}}], \tag{64.7}$$

где

$$\mathbf{a}_{\text{отн}} = \ddot{x}\mathbf{i} + \ddot{y}\mathbf{j} + \ddot{z}\mathbf{k}. \tag{64.8}$$

Последнее выражение дает относительное ускорение. Для его нахождения надо дважды дифференцировать координаты x, y, z, считая координатные орты $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ неподвижными. Именно так поступал бы наблюдатель, изучающий движение относительно системы отсчета S и не подозревающий о ее движении. Поэтому-то величина (64.8) и называется относительным ускорением.

Слагаемое $[\boldsymbol{\omega}\dot{\mathbf{r}}]$ преобразуем, подставив в него выражение (64.6) для $\dot{\mathbf{r}}$

$$[\boldsymbol{\omega}\dot{\mathbf{r}}] = [\boldsymbol{\omega}\mathbf{v}_{\text{отн}}] + [\boldsymbol{\omega}[\boldsymbol{\omega}\mathbf{r}]].$$

Окончательно для абсолютного ускорения найдем

$$\mathbf{a}_{\text{абс}} = \mathbf{a}_{\text{отн}} + 2[\boldsymbol{\omega}\mathbf{v}_{\text{отн}}] + \dot{\mathbf{v}}_0 + [\boldsymbol{\omega}[\boldsymbol{\omega}\mathbf{r}]] + [\dot{\boldsymbol{\omega}}\mathbf{r}]. \tag{64.9}$$

Этому результату можно придать вид

$$\mathbf{a}_{\text{абс}} = \mathbf{a}_{\text{отн}} + \mathbf{a}_{\text{кор}} + \mathbf{a}_{\text{пер}}, \tag{64.10}$$

где

$$\mathbf{a}_{\text{кор}} = 2[\boldsymbol{\omega}\mathbf{v}_{\text{отн}}], \tag{64.11}$$

$$\mathbf{a}_{\text{пер}} = \dot{\mathbf{v}}_0 + [\boldsymbol{\omega}[\boldsymbol{\omega}\mathbf{r}]] + [\dot{\boldsymbol{\omega}}\mathbf{r}]. \tag{64.12}$$

Вектор $\mathbf{a}_{\text{пер}}$ зависит только от движения системы отсчета S относительно неподвижной системы S_1. Только такое ускорение испытывала бы точка, если бы покоилась в системе S. Поэтому вектор $\mathbf{a}_{\text{пер}}$ называется *переносным ускорением*. Наконец, слагаемое $\mathbf{a}_{\text{кор}} = 2[\boldsymbol{\omega}\mathbf{v}_{\text{отн}}]$ зависит как от относительного, так и от переносного движений. Оно называется *кориолисовым ускорением*, по имени Кориолиса, который впервые ввел это понятие в механику. Равенство (64.10) вместе с выражениями для отдельных слагаемых, стоящих в его правой части, выражает так называемую *теорему Кориолиса*. Согласно этой теореме *абсолютное ускорение является векторной суммой относительного, кориолисова и переносного ускорений*.

Исследуем структуру переносного ускорения. Для этого воспользуемся формулой (64.12). Слагаемое $\dot{\mathbf{v}}_0$ есть переносное ускорение, вызванное поступательным движением системы S, тождественным с движением начала координат O. Остальные два слагаемых вызываются вращением системы S. Из них $[\dot{\boldsymbol{\omega}}\mathbf{r}]$ есть часть переносного ускорения, вызванная неравномерностью вращения. При равномерном вращении ($\boldsymbol{\omega} = \text{const}$) это слагаемое пропадает. Другое слагаемое $[\boldsymbol{\omega}[\boldsymbol{\omega}\mathbf{r}]]$, обозначаемое в дальнейшем $\mathbf{a}_{\text{ц}}$, есть

центростремительное ускорение, направленное к мгновенной оси вращения. Действительно, представим радиус-вектор $\mathbf{r}$ в виде $\mathbf{r} = \mathbf{r}_\perp + \mathbf{r}_\parallel$, где $\mathbf{r}_\parallel$ и $\mathbf{r}_\perp$ — компоненты этого радиуса-вектора, направленные вдоль оси вращения и перпендикулярно к ней соответственно. Так как $[\boldsymbol{\omega}\mathbf{r}_\parallel] = 0$, то

$$\mathbf{a}_\text{ц} \equiv [\boldsymbol{\omega}[\boldsymbol{\omega}\mathbf{r}]] = [\boldsymbol{\omega}[\boldsymbol{\omega}\mathbf{r}_\perp]].$$

Раскрыв по известной формуле двойное векторное произведение и приняв во внимание, что $(\boldsymbol{\omega}\mathbf{r}_\perp) = 0$, получим

$$\mathbf{a}_\text{ц} = -\boldsymbol{\omega}^2\mathbf{r}_\perp. \tag{64.13}$$

Эта формула и доказывает наше утверждение.

4. Можно было бы теперь перейти к написанию уравнения относительного движения материальной точки. Однако мы хотим еще раз на частном примере получить теорему Кориолиса. Таким путем мы лучше выясним происхождение кориолисова ускорения и других членов, из которых складывается абсолютное ускорение.

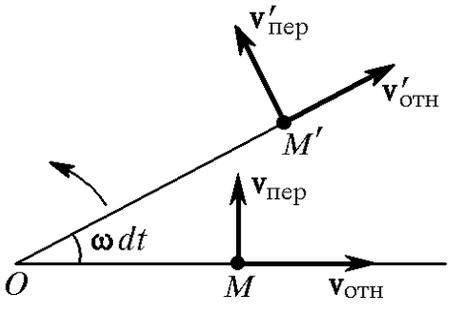


Рис. 183

Пусть шарик M (рис. 183) движется вдоль жесткого стрежня, вращающегося вокруг неподвижной оси с угловой скоростью $\boldsymbol{\omega}$, перпендикулярной к плоскости рисунка. Его абсолютная скорость $\mathbf{v}_\text{абс}$ складывается из двух взаимно перпендикулярных скоростей: скорости вдоль стержня и скорости, к нему перпендикулярной. Первая есть относительная скорость в системе отсчета, в которой стержень покоится. Вторая возникает из-за вращения стержня и потому является переносной скоростью. Таким образом, $\mathbf{v}_\text{абс} = \mathbf{v}_\text{отн} + \mathbf{v}_\text{пер}$, а потому $\mathbf{a}_\text{абс} = \dot{\mathbf{v}}_\text{отн} + \dot{\mathbf{v}}_\text{пер}$. Пусть за время dt стержень повернулся на угол $d\varphi = \omega\,dt$. За то же время шарик перешел из положения M в положение M'. Найдем приращение, которое претерпевает за то же время вектор $\mathbf{v}_\text{отн}$. Если бы не было вращения стержня, то это приращение возникало бы только из-за неравномерности движения вдоль стержня и было бы равно $\mathbf{a}_\text{отн}\,dt$. Но из-за поворота вектор $\mathbf{v}_\text{отн}$ получает дополнительное приращение $[d\varphi\mathbf{v}_\text{отн}]$. Полное приращение вектора $\mathbf{v}_\text{отн}$ будет

$$d\mathbf{v}_\text{отн} = \mathbf{a}_\text{отн}\,dt + [\boldsymbol{\omega}\mathbf{v}_\text{отн}]dt.$$

Теперь найдем приращение вектора $\mathbf{v}_\text{пер} = [\boldsymbol{\omega}\mathbf{r}]$. Очевидно, $d\mathbf{v}_\text{пер} = [d\boldsymbol{\omega}\mathbf{r}] + [\boldsymbol{\omega}d\mathbf{r}]$. Первое слагаемое возникает из-за неравномерности вращения и равно $[\dot{\boldsymbol{\omega}}\mathbf{r}]\,dt$. Второе связано с перемещением точки M в (абсолютном) пространстве и дается выражением $[\boldsymbol{\omega}\mathbf{v}_\text{абс}]\,dt = [\boldsymbol{\omega}\mathbf{v}_\text{отн}]\,dt + [\boldsymbol{\omega}\mathbf{v}_\text{пер}]\,dt$. Таким образом,

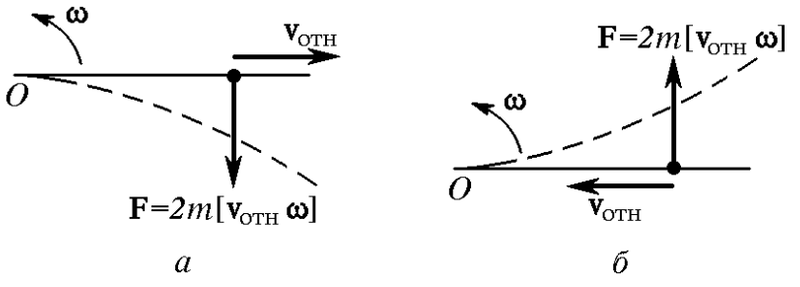
$$d\mathbf{v}_\text{пер} = [\dot{\boldsymbol{\omega}}\mathbf{r}]\,dt + [\boldsymbol{\omega}\mathbf{v}_\text{отн}]\,dt + [\boldsymbol{\omega}\mathbf{v}_\text{пер}]\,dt.$$

Сложив приращения обоих векторов, $\mathbf{v}_{\text{отн}}$ и $\mathbf{v}_{\text{пер}}$, найдем окончательно

$$\mathbf{a}_{\text{абс}} = \mathbf{a}_{\text{отн}} + 2[\boldsymbol{\omega}\mathbf{v}_{\text{отн}}] + [\boldsymbol{\omega}[\boldsymbol{\omega}\mathbf{r}]] + [\dot{\boldsymbol{\omega}}\mathbf{r}].$$

Как ясно из вывода, в рассматриваемом случае кориолисово ускорение слагается из двух равных членов. Первый возникает из-за вращения вектора $\mathbf{v}_{\text{отн}}$ вместе со стержнем. Второй появляется из-за приращения переносной скорости $\mathbf{v}_{\text{пер}}$, которое получается вследствие приближения шарика к оси вращения или удаления от нее. Очевидно, вывод применим и в том случае, когда направление оси вращения меняется с течением времени.

Кориолисово ускорение $2[\boldsymbol{\omega}\mathbf{v}_{\text{отн}}]$ направлено перпендикулярно к вращающемуся стержню. Для того чтобы сообщить такое ускорение телу M, стержень должен оказывать на него боковое давление. Сила бокового давления



а *б*

Рис. 184

равна $2m[\boldsymbol{\omega}\mathbf{v}_{\text{отн}}]$, где m — масса тела M. В свою очередь, тело M действует на стержень с равной и противоположно направленной силой $\mathbf{F} = 2m[\mathbf{v}_{\text{отн}}\boldsymbol{\omega}]$. Если тело удаляется от оси вращения (рис. 184 *а*), то сила $\mathbf{F}$ направлена противоположно вращению и замедляет его. При этом стержень изгибается таким образом, что он выпуклой стороной обращен в сторону вращения, как показано штриховой линией. Напротив, если тело приближается к оси вращения (рис. 184 *б*), то сила $\mathbf{F}$ направлена в сторону вращения стержня. В этом случае угловая скорость вращения стержня увеличивается, а сам стержень изгибается так, что в сторону вращения обращена его вогнутая сторона. В опыте со скамьей Жуковского, описанном в § 34 (см. рис. 60), возникают такие же силы бокового давления, с которыми гири действуют на демонстратора, когда он приближает или удаляет их от оси вращения. Эти силы и изменяют угловую скорость вращения скамьи Жуковского вместе с демонстратором, сидящим на ней. Вообще, действием таких сил объясняются все явления, связанные с изменением угловой скорости вращения изолированного тела при изменении его момента инерции.

5. Обратимся теперь к написанию уравнений относительного движения. Поступим в точности так же, как в предыдущем параграфе. В уравнение (63.1) подставим выражение (64.10) и все члены перенесем в правую часть, за исключением члена, содержащего относительное ускорение. Таким путем получим

$$m\mathbf{a}_{\text{отн}} = \mathbf{F} - m\mathbf{a}_{\text{кор}} - m\mathbf{a}_{\text{пер}}, \tag{64.14}$$

или более подробно

$$m\mathbf{a}_{\text{отн}} = \mathbf{F} + 2m[\mathbf{v}_{\text{отн}}\boldsymbol{\omega}] - m\dot{\mathbf{v}}_0 + m\boldsymbol{\omega}^2\mathbf{r}_\perp - m[\dot{\boldsymbol{\omega}}\mathbf{r}]. \tag{64.15}$$

К «настоящей» силе $\mathbf{F}$ добавились две силы инерции: так называемая *кориолисова сила*

$$\mathbf{F}_{\text{кор}} = -m\mathbf{a}_{\text{кор}} = 2m[\mathbf{v}_{\text{отн}}\boldsymbol{\omega}] \tag{64.16}$$

и *переносная сила инерции*

$$\mathbf{F}_{\text{пер}} = -m\mathbf{a}_{\text{пер}} = -m\dot{\mathbf{v}}_0 + m\boldsymbol{\omega}^2\mathbf{r}_\perp - m[\dot{\boldsymbol{\omega}}\mathbf{r}]. \tag{64.17}$$

Разумеется, к этим силам инерции относятся все общие замечания, которые были высказаны в предыдущем параграфе применительно к силам инерции, возникающим при ускоренном поступательном движении системы отсчета.

6. Переносная сила инерции в общем случае состоит из трех слагаемых. С первым слагаемым $-m\dot{\mathbf{v}}_0$ мы уже познакомились в предыдущем параграфе. Это есть *поступательная сила инерции*, возникающая из-за ускоренного движения начала координат O. Последнее слагаемое $-m[\dot{\boldsymbol{\omega}}\mathbf{r}]$ обусловлено *неравномерностью вращения* системы отсчета. Оно не получило специального названия. Второе слагаемое

$$\mathbf{F}_{\text{ц}} = m\boldsymbol{\omega}^2\mathbf{r}_\perp \tag{64.18}$$

называется *центробежной силой инерции* или просто *центробежной силой*. Действию центробежной силы подвергается, например, пассажир в движущемся автобусе на поворотах. Перегрузки, испытываемые летчиком при выполнении фигур высшего пилотажа на больших скоростях, также в основном вызываются центробежными силами. Если на центробежной машине подвесить несколько шариков на нитях и привести машину в быстрое вращение, то центробежные силы отклонят шарики от оси вращения. Угол отклонения тем больше, чем дальше шарик отстоит от оси. Центробежные силы используются в центробежных сушилках для отжима белья и в сепараторах для отделения сливок от молока.

7. Центробежные силы, как и всякие силы инерции, существуют лишь в ускоренно движущихся (вращающихся) системах отсчета и исчезают при переходе к инерциальным системам. Забыв это, можно прийти к парадоксам, которые часто ставят в тупик школьников. Вот один из самых распространенных парадоксов такого типа. Пусть тело движется по окружности. На него действуют две силы: центростремительная $\mathbf{F}_1$, направленная к центру окружности, и центробежная $\mathbf{F}_2$, направленная в противоположную сторону. Эти силы равны по модулю и уравновешивают друг друга: $\mathbf{F}_1 + \mathbf{F}_2 = 0$. По закону инерции тело должно двигаться прямолинейно и равномерно. Противоречие возникло потому, что движение стали относить к неподвижной (инерциальной) системе отсчета. А в этой системе никаких центробежных сил не существует. Есть только одна центростремительная сила $\mathbf{F}_1$, которая и сообщает телу ускорение. Это может быть, например, натяжение шнура, к которому привязано тело. Вводить центробежную силу

нии она не совершает работы. Кориолисова сила, таким образом, является силой *гироскопической* (см. § 24, п. 6).

ЗАДАЧИ

1. В чем ошибочность следующего рассуждения: пусть A и B — две неподвижные материальные точки, расстояние между которыми равно r. Состояние покоя точки B можно рассматривать как результат сложения двух вращений с одинаковыми, но противоположно направленными постоянными угловыми скоростями: $+\boldsymbol{\omega}$ и $-\boldsymbol{\omega}$. При первом вращении возникает центростремительное ускорение $\mathbf{a}_1 = \omega^2\mathbf{r}$, при втором — центростремительное ускорение $\mathbf{a}_2 = (-\omega)^2\mathbf{r} = \omega^2\mathbf{r} = \mathbf{a}_1$. Результирующее ускорение точки B равно $\mathbf{a} = \mathbf{a}_1 + \mathbf{a}_2 = 2\omega^2\mathbf{r}$. Следовательно, точка A действует на точку B с силой притяжения $\mathbf{F} = 2m\omega^2\mathbf{r}$, где m — масса точки B. Поскольку ω — величина произвольная, получается абсурдный результат, что точки A и B притягиваются друг к другу с произвольной, наперед заданной силой.

Р е ш е н и е. Не учтено кориолисово ускорение. Введем систему отсчета S, равномерно вращающуюся вокруг точки A с угловой скоростью $+\boldsymbol{\omega}$. Пусть точка B вращается относительно этой системы с угловой скоростью $-\boldsymbol{\omega}$. Обозначая вектор $\overrightarrow{AB}$ через $\mathbf{r}$, имеем для скоростей и ускорений точки B:

$$\mathbf{v}_{\text{отн}} = -[\boldsymbol{\omega}\mathbf{r}],\ \mathbf{a}_{\text{отн}} = \mathbf{a}_{\text{пер}} = -\omega^2\mathbf{r},\ \mathbf{a}_{\text{кор}} = 2[\boldsymbol{\omega}\mathbf{v}_{\text{отн}}] = 2\omega^2\mathbf{r}.$$

Следовательно, $\mathbf{a}_{\text{абс}} = \mathbf{a}_{\text{отн}} + \mathbf{a}_{\text{кор}} + \mathbf{a}_{\text{пер}} = 0$.

2. Стрелок и мишень находятся в диаметрально противоположных точках карусели радиусом $R = 5$ м, равномерно вращающейся вокруг вертикальной оси. Период вращения карусели $T = 10$ с, скорость пули $v = 300$ м/с. Пренебрегая максимальной линейной скоростью вращающейся карусели ωR по сравнению со скоростью пули, определить приближенно, под каким углом α к диаметру карусели должен целиться стрелок, чтобы поразить мишень. Задачу рассмотреть как с точки зрения вращающейся, так и с точки зрения неподвижной системы, и сравнить результаты.

О т в е т. $\alpha = 4\pi R/vT = 0{,}0209$ рад $= 1{,}2°$.

3. Тонкий стержень длиной l вращается вокруг одного из концов, описывая круговой конус (физический конический маятник). Найти период движения T в зависимости от угла при вершине конуса 2φ.

У к а з а н и е. В системе отсчета, вращающейся вместе со стержнем вокруг вертикальной оси, стержень покоится. Задача сводится к нахождению условия равновесия подвешенного стержня в этой системе под действием силы тяжести и центробежной силы.

О т в е т. $T = 2\pi\sqrt{2l\cos\varphi/3g}$

4. Физический маятник, состоящий из шарика, насаженного на конец тонкого жесткого стержня, может свободно колебаться вокруг горизонтальной оси A, проходящей через верхний конец стержня. Ось A неподвижно закреплена на геометрической оси горизонтального диска, равномерно вращающегося вокруг этой (вертикальной) геометрической оси с угловой скоростью ω. Таким образом, плоскость колебаний маятника вращается вместе с диском с той же угловой скоростью ω. Найти период малых колебаний

маятника, если масса стержня пренебрежимо мала по сравнению с массой шарика. При каком условии нижнее вертикальное положение стержня станет неустойчивым положением равновесия?

Ответ. $T = 2\pi\sqrt{l/(g - l\omega^2)}$, если $l\omega^2 < g$. При $l\omega^2 > g$ положение равновесия неустойчиво.

5. Представим, себе, что в земном шаре просверлен канал по диаметру в плоскости экватора. Вычислить силу F, с которой будет давить на стенку канала тело, падающее по нему с поверхности Земли, в тот момент, когда оно достигнет центра Земли. Считать, что трения нет, а плотность Земли однородна.

Ответ. $F = (4\pi/T)\sqrt{R/g}T \approx 0{,}12P$, где P — вес тела на поверхности Земли, T — продолжительность звездных суток, R — радиус Земли.

6. Какую центральную силу надо прибавить к силе притяжения Солнца для того, чтобы орбита планеты, не меняя своего вида, вращалась вокруг Солнца? (Задача Ньютона.)

Решение. Обозначим через $\mathbf{F}_1$ силу ньютонова притяжения планеты к Солнцу, $\mathbf{F}_2$ — дополнительную центральную силу, о которой говорится в условии задачи, $\boldsymbol{\omega}$ — угловую скорость вращения орбиты. Вектор $\boldsymbol{\omega}$ перпендикулярен к плоскости орбиты. Полный момент импульса планеты относительно Солнца $\mathbf{L}$ слагается из момента импульса относительно движения $\mathbf{L}_1 = m[r\mathbf{v}_{\text{отн}}]$ и момента импульса дополнительного вращения $\mathbf{L}_2 = mr^2\boldsymbol{\omega}$. Момент $\mathbf{L}$, очевидно сохраняется, так как полная действующая сила $\mathbf{F}_1 + \mathbf{F}_2$ является центральной. Момент $\mathbf{L}_1$ тоже сохраняется. Действительно, таким моментом обладала бы планета, если бы вращения орбиты не было, и все ее движение происходило под действием только одной центральной силы $\mathbf{F}_1$. Поэтому должен сохраняться и момент $\mathbf{L}_2$, а планета должна вращаться с угловой скоростью

$$\boldsymbol{\omega} = \frac{\mathbf{L}_2}{mr^2} = \frac{\text{const}}{r^2}. \tag{64.19}$$

Вращение орбиты неравномерное за исключением случая, когда орбита круговая. В системе отсчета, вращающейся вместе с орбитой с угловой скоростью $\boldsymbol{\omega}$, уравнение движения планеты, с одной стороны, имеет вид

$$m\mathbf{a}_{\text{отн}} = \mathbf{F}_1 + \mathbf{F}_2 + m\omega^2\mathbf{r} - m[\dot{\boldsymbol{\omega}}r] + 2m[\mathbf{v}_{\text{отн}}\boldsymbol{\omega}].$$

С другой стороны, по условию, в этой системе планета должна двигаться по обычному кеплеровскому эллипсу, а потому $m\mathbf{a}_{\text{отн}} = \mathbf{F}_1$. Это дает

$$\mathbf{F}_1 = -m\omega^2\mathbf{r} + m[\dot{\boldsymbol{\omega}}r] - 2m[\mathbf{v}_{\text{отн}}\boldsymbol{\omega}]. \tag{64.20}$$

Дифференцируя (64.19) по времени и принимая во внимание, что $\mathbf{L}_2 = \text{const}$, получим

$$\dot{\boldsymbol{\omega}} = -2\frac{\dot{r}}{r}\frac{\mathbf{L}_2}{mr^2} = -2\frac{\dot{r}}{r}\boldsymbol{\omega}.$$

Скорость $\mathbf{v}_{\text{отн}}$ можно разложить на две составляющие: вдоль радиуса $\frac{\dot{r}}{r}\mathbf{r}$ и перпендикулярную к нему. Последняя возникает из-за вращения планеты

по кеплерову эллипсу с угловой скоростью $\boldsymbol{\omega}_{\text{отн}} = \frac{\mathbf{L}_1}{mr^2}$. Таким образом,

$$\mathbf{v}_{\text{отн}} = \frac{\dot{r}}{r}\mathbf{r} + [\boldsymbol{\omega}_{\text{отн}} r].$$

Подставив это в (64.20), после простых преобразований получим

$$\mathbf{F}_2 = -m\{\omega^2 + 2(\boldsymbol{\omega}\boldsymbol{\omega}_{\text{отн}})\}\,\mathbf{r}, \tag{64.21}$$

или

$$\mathbf{F}_2 = -[L_2^2 + 2(\mathbf{L}_1\mathbf{L}_2)]\frac{\mathbf{r}}{mr^4} = \frac{\text{const}}{r^4}\,\mathbf{r}. \tag{64.22}$$

Отсюда видно, что дополнительная сила $\mathbf{F}_2$ должна меняться обратно пропорционально кубу расстояния планеты от Солнца.

7. Применить теорему Кориолиса для решения *обратной задачи* о движении симметричного гироскопа. Прямая задача механики состоит в том, чтобы по заданным силам определить движение механической системы. Обратная задача сводится к определению сил по заданному движению системы. Пусть гироскоп совершает вынужденную регулярную прецессию. Какие на него должны действовать силы, чтобы эта прецессия имела место?

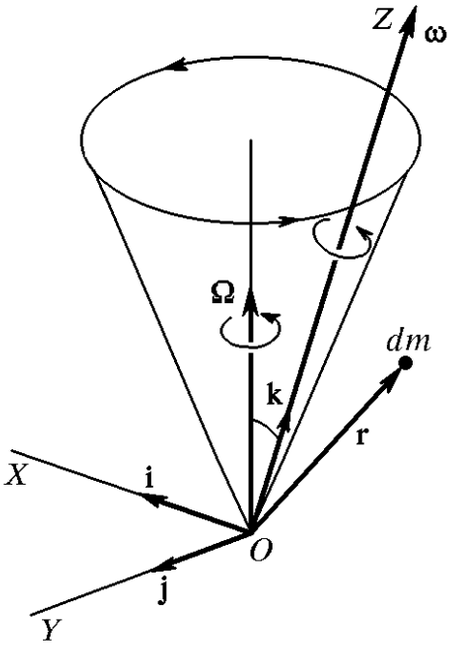


Рис. 185

Р е ш е н и е. Пусть гироскоп равномерно вращается вокруг своей оси фигуры с угловой скоростью $\boldsymbol{\omega}$, а ось фигуры вращается также равномерно с угловой скоростью $\boldsymbol{\Omega}$ (вынужденная прецессия). Перейдем к системе отсчета, вращающейся с угловой скоростью $\boldsymbol{\Omega}$. В этой системе ось фигуры гироскопа неподвижна, так что

$$\mathbf{v}_{\text{отн}} = [\boldsymbol{\omega}\mathbf{r}],\ \mathbf{v}_{\text{пер}} = [\boldsymbol{\Omega}\mathbf{r}],\ \mathbf{a}_{\text{пер}} = [\boldsymbol{\Omega}[\boldsymbol{\Omega}\mathbf{r}]]. \tag{64.23}$$

Мысленно выделим из тела гироскопа элемент массы dm с радиусом-вектором $\mathbf{r}$ (рис. 185). Обозначим через $d\mathbf{f}$ действующую на него (реальную) силу. При использовании формулы (64.9) надо помнить, что угловая скорость вращения системы отсчета обозначена через $\boldsymbol{\Omega}$ (а не через $\boldsymbol{\omega}$, как в этой формуле). Применив к выделенному элементу массы второй закон Ньютона и использовав формулу (64.9) и выражения (64.23), можем написать

$$d\mathbf{f} = dm[\boldsymbol{\omega}[\boldsymbol{\omega}\mathbf{r}]] + 2dm[\boldsymbol{\Omega}[\boldsymbol{\omega}\mathbf{r}]] + dm[\boldsymbol{\Omega}[\boldsymbol{\Omega}\mathbf{r}]]. \tag{64.24}$$

Поскольку гироскоп предполагается идеально твердым телом, его уравнения движения полностью определяются геометрической суммой $\mathbf{f}$ внешних сил и их моментов относительно точки опоры O. Для нахождения $\mathbf{f}$ проинтегрируем выражение (64.24). Векторы $\boldsymbol{\omega}$ и $\boldsymbol{\Omega}$, как постоянные, можно при этом вынести из-под знака интеграла. Кроме того, учтем, что $\int \mathbf{r}dm = m\mathbf{r}_C$, где $\mathbf{r}_C$ — радиус-вектор центра масс гироскопа C. В результате получим

$$\mathbf{f} = m\{[\boldsymbol{\omega}[\boldsymbol{\omega}\mathbf{r}_C]] + 2[\boldsymbol{\Omega}[\boldsymbol{\omega}\mathbf{r}_C]] + [\boldsymbol{\Omega}[\boldsymbol{\Omega}\mathbf{r}_C]]\}.$$

Первые два слагаемых в правой части этого соотношения равны нулю, так как центр масс C лежит на оси фигуры гироскопа, а потому векторы $\mathbf{r}_C$ и $\boldsymbol{\omega}$ коллинеарны. Поэтому окончательно

$$\mathbf{f} = m[\boldsymbol{\Omega}[\boldsymbol{\Omega}\mathbf{r}_C]]. \tag{64.25}$$

Этот результат, конечно, можно было бы написать сразу на основании теоремы о движении центра масс, поскольку ускорение последнего происходит только из-за прецессионного вращения и равно $[\boldsymbol{\Omega}[\boldsymbol{\Omega}\mathbf{r}_C]]$ (центростремительное ускорение). Сила $\mathbf{f}$ возникает автоматически как реакция точки опоры на прецессирующий гироскоп.

Перейдем теперь к вычислению момента сил $\mathbf{M}$. Для этого радиус-вектор $\mathbf{r}$ надо векторно умножить на выражение (64.24) и проинтегрировать по m. Но и без вычислений ясно, что при таком интегрировании слагаемое $dm[\boldsymbol{\omega}t[\boldsymbol{\omega}\mathbf{r}]]$ не внесет никакого вклада в момент $\mathbf{M}$. Действительно, член, возникающий от интегрирования этого слагаемого, не зависит от того, содержатся в сумме (64.24) другие слагаемые или не содержатся. Но рассматриваемое слагаемое соответствует равномерному вращению вокруг оси фигуры гироскопа, которая, как известно, является одной из свободных осей вращения, а для равномерного вращения вокруг свободной оси никаких внешних сил не требуется. По той же причине не повлияло бы на момент $\mathbf{M}$ и последнее слагаемое суммы (64.24), если бы прецессия происходила вокруг оси, перпендикулярной к оси фигуры гироскопа, а его центр масс C совпадал с точкой опоры O. Во всех остальных случаях это не так. Однако при вычислении момента $\mathbf{M}$ мы пренебрежем последним слагаемым в (64.24) и вот по какой причине. Мы будем предполагать прецессионное вращение Ω очень медленным по сравнению с собственным вращением ω и пренебрегать квадратами малой величины Ω. А последнее слагаемое в (64.24) как раз квадратично по Ω. Таким образом, момент $\mathbf{M}$ происходит лишь от второго слагаемого в сумме (64.24). Разложив соответствующее двойное векторное произведение, умножив его векторно на $\mathbf{r}$ и проинтегрировав, получим

$$\mathbf{M} = 2\int (\boldsymbol{\Omega}\mathbf{r})\ [\mathbf{r}\boldsymbol{\omega}]\ dm.$$

Для вычисления интеграла введем прямоугольную систему координат. Ось Z направим вдоль оси фигуры гироскопа, а ось X расположим в плоскости, в которой лежат векторы $\boldsymbol{\omega}$ и $\boldsymbol{\Omega}$ (см. рис. 185). В этой системе координат

$$\mathbf{M} = -2\mathbf{j}\omega\Omega_x\int x^2dm - 2\mathbf{j}\omega\Omega_z\int zxdm + 2\mathbf{i}\omega\Omega_x\int xydm + 2\mathbf{i}\omega\Omega_z\int yzdm.$$

Все входящие сюда интегралы, за исключение первого, обращаются в нуль из-за осесимметричного распределения масс. Первый же член может быть записан в виде

$$\mathbf{M} = -\mathbf{j}\omega\Omega_x\int (x^2+y^2)dm = -\mathbf{j}I_\perp\omega\Omega\sin\vartheta,$$

где ϑ — угол между векторами $\boldsymbol{\omega}$ и $\boldsymbol{\Omega}$, а $I_\perp$ — соответствующий момент инерции гироскопа. В векторной форме

$$\mathbf{M} = I_\perp[\boldsymbol{\Omega}\boldsymbol{\omega}]. \tag{64.26}$$

Векторное произведение $[\boldsymbol{\Omega}\boldsymbol{\omega}]$ есть вектор скорости, с которой при регулярной прецессии движется конец вектора $\boldsymbol{\omega}$, неизменно связанный с осью фигуры гироскопа. Таким образом, вершина гироскопа перемещается не в направлении приложенной силы, а в перпендикулярном к ней направлении — в направлении момента $\mathbf{M}$. Это то, что кажется более всего удивительным в движении гироскопа. Если перейти к системе отсчета, вращающейся с угловой скоростью прецессии $\boldsymbol{\Omega}$, то можно сказать, что в этой системе момент внешних сил должен уравновешивать момент сил инерции Кориолиса.

§ 65. УРАВНЕНИЕ ОТНОСИТЕЛЬНОГО ДВИЖЕНИЯ МАТЕРИАЛЬНОЙ ТОЧКИ В ГРАВИТАЦИОННОМ ПОЛЕ ЗЕМЛИ С УЧЕТОМ ЕЕ ВРАЩЕНИЯ

1. Применим уравнение относительного движения (64.15) к движению тел относительно Земли. Движущуюся систему отсчета S свяжем с вращающейся Землей. Речь идет о вращении Земли относительно инерциальной системы отсчета, например системы Коперника. Начало координат O поместим в центре Земли. Таким образом, под $\mathbf{v}_0$ следует понимать скорость, а под $\dot{\mathbf{v}}_0$ — ускорение центра Земли. Земля вращается практически равномерно, а потому последний член в уравнении (64.15) выпадает. Далее, так как речь будет идти только об относительном движении, условимся опускать в уравнении (64.15) индекс «отн», т. е. будем полагать $\mathbf{v} \equiv \mathbf{v}_{\text{отн}}$, $\mathbf{a} \equiv \mathbf{a}_{\text{отн}}$. Внешнюю силу представим в виде суммы трех сил $\mathbf{F}_З + \mathbf{F}_0 + \mathbf{F}$, где $\mathbf{F}_З$ — сила гравитационного притяжения Земли, $\mathbf{F}_0$ — равнодействующая сил гравитационного притяжения Солнца, Луны, планет, звезд и прочих небесных тел, $\mathbf{F}$ — геометрическая сумма всех остальных сил, действующих на материальную точку. Сила $\mathbf{F}$ слагается, например, из силы сопротивления воздуха, силы трения, силы натяжения нити и пр. В этих обозначениях уравнение (64.15) примет вид

$$m\mathbf{a} = (\mathbf{F}_З + m\omega^2\mathbf{r}_\perp) + 2m[\mathbf{v}\boldsymbol{\omega}] + \mathbf{F} + (\mathbf{F}_0 - m\dot{\mathbf{v}}_0). \qquad (65.1)$$

2. Используем далее фундаментальный физический закон, согласно которому *все тела в одном и том же поле тяготения падают и с одинаковым ускорением*. Этот закон называется *обобщенным законом Галилея*, так как Галилей бы первым, кто установил его справедливость для тел, свободно падающих в поле тяжести Земли. Из этого закона следует, что *сила, действующая на тело в гравитационном поле, зависит не от состава тела, а только от его массы. Она строго пропорциональна массе тела.* В этом отношении силы тяготения ведут себя так же, как и силы инерции. Последние, очевидно, также строго пропорциональны массам тел.

3. Основной вклад в силу $\mathbf{F}_0$ вносят гравитационные поля Солнца и Луны. Эти поля, в особенности гравитационное поле Луны, неоднородны. Они убывают обратно пропорционально квадратам

расстояний от Солнца и Луны. Однако размеры Земли очень малы по сравнению с этими расстояниями. При рассмотрении движений вблизи земной поверхности изменениями гравитационных полей Солнца, Луны и всех прочих внешних гравитационных полей на расстояниях порядка диаметра земного шара можно в первом приближении пренебречь, т. е. считать внешнее гравитационное поле в окрестности Земли *однородным*. Однородное гравитационное поле сообщает одно и то же ускорение всем телам, независимо от того, в каких точках поля эти тела находятся. Значит, в принятом приближении внешнее гравитационное поле сообщает рассматриваемой материальной точке такое же ускорение, что и центру Земли, т. е. $\dot{\mathbf{v}}_0$. Поэтому

$$\mathbf{F}_0 - m\dot{\mathbf{v}}_0 = 0.$$

Таким образом, силы *гравитационного притяжения Солнца, и всех остальных небесных тел выпадают из уравнений относительного движения* (65.1). *Они полностью компенсируются поступательными силами инерции, возникающими из-за ускорения, сообщаемого Земле этими полями.* Этот замечательный результат, как мы видим, является следствием обобщенного закона Галилея.

4. Сила $\mathbf{F}_З$ гравитационного притяжения Земли, а с ней и векторная сумма $\mathbf{F}_З + m\boldsymbol{\omega}^2\mathbf{r}_\perp$, вследствие того же закона Галилея, пропорциональны массе материальной точки m. Эта сумма не зависит от относительного движения точки и характеризует только гравитационное поле Земли и ее вращение. Целесообразно рассматривать эту сумму как единую величину. Для нее мы введем обозначение

$$\mathbf{F}_З + m\omega^2\mathbf{r}_\perp = m\mathbf{g}. \tag{65.2}$$

Тогда уравнение относительного движения примет вид

$$m\mathbf{a} = m\mathbf{g} + 2m[\mathbf{v}\boldsymbol{\omega}] + \mathbf{F}. \tag{65.3}$$

Величина $\mathbf{g}$ одна и та же для всех тел — она может меняться только при переходе из одной точки пространства в другую.

Для установления физического смысла вектора $\mathbf{g}$ допустим, что внешних сил нет ($\mathbf{F} = 0$), а скорость $\mathbf{v}$ материальной точки равна нулю. Тогда из формулы (65.3) следует $\mathbf{a} = \mathbf{g}$. Таким образом, *вектор* $\mathbf{g}$ *есть ускорение свободно падающего тела относительно Земли при условии, что его скорость в рассматриваемый момент равна нулю.* Оговорка относительно скорости тела необходима, так как при наличии скорости $\mathbf{v}$ появляется дополнительное ускорение из-за кориолисовой силы. Мы видим, что ускорение свободного падения состоит из двух слагаемых

$$\mathbf{g} = \mathbf{g}_{\text{абс}} + \omega^2\mathbf{r}_\perp. \tag{65.4}$$

Первое из них, $\mathbf{g}_{\text{абс}} = \mathbf{F}_З / m$, есть ускорение, вызванное силой гравитационного притяжения Земли. Такое ускорение мы получили бы, если бы измеряли ускорение свободного падения относительно неподвижной системы отсчета при условии, что, помимо земного гравитационного поля, никаких других полей нет. Второе слагаемое $\omega^2 \mathbf{r}_\perp$ есть ускорение, сообщаемое центробежной силой инерции и связанное с вращением Земли.

§ 66. ВЕС И ВЗВЕШИВАНИЕ ТЕЛ

1. *Весом тела называется приложенная к нему сила* $\mathbf{P}$, *равная силе, с которой это тело действует не подставку, на которой оно лежит, или тянет за подвес, к которому оно подвешено. При этом предполагается, что тело, подставка и подвес покоятся в той системе отсчета, в которой производится взвешивание. Когда говорят о весе тела, обычно предполагают, что тело, подставка и подвес покоятся относительно Земли.* Допустим ради определенности, что тело лежит на подставке. Оно действует на подставку с силой $\mathbf{P}$, подставка действует на тело с противоположно направленной силой $\mathbf{F}$. По смыслу $\mathbf{P}$ и $\mathbf{F}$ суть силы взаимодействия подставки и тела. Они удовлетворяют третьему закону Ньютона: $\mathbf{F} = -\mathbf{P}$. Предполагая, что тело на подставке покоится, подставим в формулу (65.3) $\mathbf{v} = 0$, $\mathbf{a} = 0$, $\mathbf{F} = -\mathbf{P}$. Тогда для $\mathbf{P}$ найдем

$$\mathbf{P} = m\mathbf{g}. \tag{66.1}$$

Учитывая (65.4), видим, что $\mathbf{P}$ состоит из двух слагаемых:

$$\mathbf{P} = m\mathbf{g}_{\text{абс}} + m\omega^2 \mathbf{r}_\perp = \mathbf{F}_З + m\omega^2 \mathbf{r}_\perp. \tag{66.2}$$

Значит, *вес есть геометрическая сумма силы гравитационного притяжения Земли* $\mathbf{F}_З$ *и центробежной силы инерции* $m\omega^2 \mathbf{r}_\perp$.

Если тело подвешено на нити, то рассуждения остаются теми же самыми. В этом случае направление нити определяет направление силы $\mathbf{P}$, а следовательно, и ускорение свободного падения $\mathbf{g}$. Оно называется *направлением отвеса* или *отвесным направлением.*

2. Вектор $\mathbf{g}_{\text{абс}}$ характеризует гравитационное поле Земли. В каждой точке пространства он определяется только размерами и формой Земли, а также распределением вещества в ней. Если бы Земля была правильным шаром, а вещество внутри нее было распределено сферически-симметрично, то вектор $\mathbf{g}_{\text{абс}}$ был бы направлен точно к центру Земли. Направление отвеса определяется вектором $\mathbf{g}$, т. е. диагональю параллелограмма, построенного на векторах $\mathbf{g}_{\text{абс}}$ и $\omega^2 \mathbf{r}_\perp$ (рис. 186). Таким образом, если бы даже Земля была строго сферически-симметрична, то направление к ее центру не совпало бы с направлением отвеса. Различие между этими двумя направле-

ниями для сферически-симметричной Земли обусловлено центробежной силой. Реальная Земля сплюснута вдоль оси вращения, и это является второй причиной различия указанных двух направлений. Впрочем, ввиду медленности вращения Земли и малости ее сплюснутости, оба направления отличаются друг от друга весьма мало. Для сферически-симметричной Земли угол α между ними определяется формулой

$$\sin\alpha = \frac{\omega^2 r_\perp}{g}\sin\vartheta = \frac{1}{2}\frac{\omega^2 r}{g}\sin 2\vartheta, \qquad (66.3)$$

где ϑ — географическая широта рассматриваемого места (рис. 186). На полюсе и на экваторе угол α обращается в нуль. Для реальной (несферической) Земли формула (66.3) хотя и приближенна, но достаточно точна. Проецируя векторы $\mathbf{g}_{\text{абс}}$ и $\omega^2\mathbf{r}_\perp$ на направление вектора $\mathbf{g}$ и полагая $\cos\alpha \approx 1$, легко получить приближенную формулу

$$g = g_{\text{абс}} - \omega^2 r_\perp \cos\vartheta = g_{\text{абс}} - \omega^2 r\cos^2\vartheta. \qquad (66.4)$$

Ошибка этого расчета порядка α^2.

Величина g может быть найдена путем взвешивания или из опытов по свободному падению тел. Более точно ее можно найти, измеряя период колебаний обратного маятника (см. § 41). Опыты показали, что g зависит от географической широты. На полюсе $g = 983{,}2$ см/с^2, на экваторе $g = 978{,}0$ см/с^2. Зная g, можно по формуле (66.4) вычислить и $g_{\text{абс}}$. На полюсе $g_{\text{абс}} = g = 983{,}2$ см/с^2. На экваторе

$$g_{\text{абс}} = g + \omega^2 r = 978{,}0 + \frac{4\pi^2}{86\,164^2}\cdot 6{,}378\cdot 10^8 = 981{,}4 \text{ см/с}^2.$$

Если бы Земля была правильным шаром со сферически-симметричным распределением вещества в нем, то величина $g_{\text{абс}}$ должна была бы быть одной и той же на полюсе и на экваторе. В действительности на экваторе $g_{\text{абс}}$ меньше, чем на полюсе. Это объясняется сплюснутостью Земли, обусловленной действием центробежных сил. Точки экватора отстоят от центра Земли дальше, чем полюсы. Поэтому они притягиваются к центру Земли слабее, чем такие же точки на полюсе. Разумеется, изменение ускорения свободного падения g на земной поверхности нельзя обнаружить с помощью рычажных весов. Но это можно сделать с помощью пружинных весов.

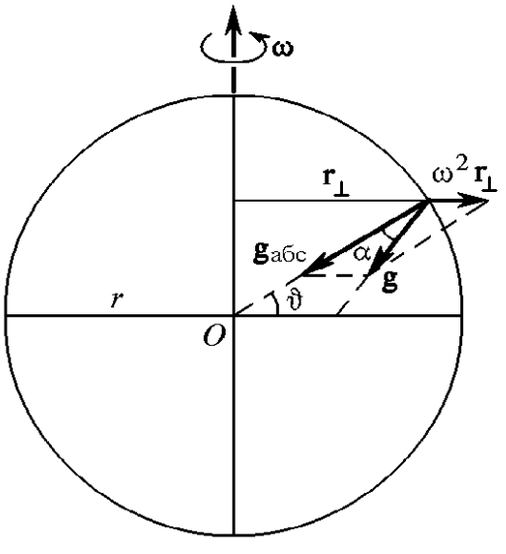


Рис. 186

3. Допустим теперь, что пружинные весы установлены на искусственном спутнике или космическом корабле. Что покажут эти весы, когда взвешиваемое тело покоится относительно корабля? Наши прежние рассуждения не изменятся, только в них Землю надо заменить космическим кораблем. В частности, движущуюся систему отсчета S мы связываем теперь с кораблем. Землю же надо рассматривать как внешнее тело, которое наряду с Солнцем, Луной и прочими небесными телами создает внутри внешнее гравитационное поле. В силу малых размеров корабля это поле внутри корабля можно считать однородным. Оно полностью компенсируется поступательными силами инерции, возникающими в системе отсчета S из-за ускорения, сообщаемого ей этим гравитационным полем. Поэтому если двигатели на корабле выключены и он свободно падает в гравитационном поле, то относительное движение внутри корабля описывается прежним уравнением (65.3). Только ввиду ничтожности гравитационного поля, создаваемого самим кораблем, член $m\mathbf{g}$ теперь обусловлен исключительно вращением корабля и равен $m\omega^2\mathbf{r}_\perp$ (центробежная сила). Если корабль не вращается, а взвешиваемое тело относительно него покоится, то из уравнения (65.3) получаем $\mathbf{F} = 0$. Величина $\mathbf{F}$ есть сила, с которой на тело действует растянутая пружина весов. Мы видим, что пружина не растянута, т. е. вес, показываемый весами, равен нулю. Весы не реагируют на внешние гравитационные поля, последние полностью компенсируются поступательными силами инерции. Такое состояние «*невесомости*» свойственно всем телам внутри космического корабля. Состояние «невесомости» проявляется в том, что в телах полностью отсутствуют внутренние упругие напряжения, которые в обычных условиях возникают под действием силы тяжести.

Если корабль вращается, то появляется центробежная сила, не компенсируемая внешними гравитационными полями. Эта сила создает на корабле «*искусственную тяжесть*».

Наконец, если включены двигатели, сообщающие кораблю дополнительное поступательное ускорение $\mathbf{w}$, то в правой части уравнения (65.3) добавляется член $-m\mathbf{w}$. Весы покажут вес $\mathbf{P} = -m\mathbf{w}$. Все тела внутри космического корабля снова становятся «весомыми». Этим «весом» и обусловлены перегрузки, которые испытывают космонавты на старте или торможении космических кораблей.

ЗАДАЧИ

1. Тело на экваторе взвешивается на пружинных весах в полдень, когда гравитационные силы Земли и Солнца тянут его в противоположные стороны. Одновременно такое же тело взвешивается в полночь в диаметрально противоположной точке земного шара, когда обе эти силы направлены в одну сторону. Вес какого тела будет больше?

Р е ш е н и е. Если пренебречь неоднородностью гравитационного поля Солнца в окрестности Земли, то в обоих случаях получится один и тот же вес. Учтем теперь неоднородность гравитационного поля Солнца, пренебрегая влиянием Луны. Веса тел в диаметрально противоположных точках земного шара *1* (день) и *2* (ночь) будут соответственно

$$P_1 = F_З - F_С(R - r) - m\omega^2 r + mw,$$

$$P_2 = F_З + F_С(R + r) - m\omega^2 r + mw$$

(рис. 187). Здесь $F_З$ и $F_С$ — силы гравитационного притяжения Земли и Солнца соответственно, R — расстояние между их центрами, r — радиус Земли, w — ускорение центра Земли под действием гравитационного притяжения Солнца. Очевидно, $mw = F_С(R)$. Вычитая, находим

$$P_2 - P_1 = [F_С(R + r) - F_С(R)] + [F_С(R - r) - F_С(R)].$$

Разлагая обе разности в квадратных скобках по формуле Тейлора и ограничиваясь квадратичными членами по r, получим

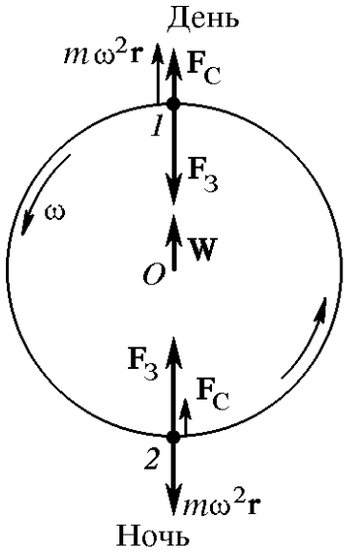


Рис. 187

$$P_2 - P_1 = r^2 \frac{d^2 F_С}{dR^2}.$$

Преобразуем это выражение, используя соотношения

$$F_С = G\frac{Mm}{R^2} = \frac{4\pi^2 R}{T^2} m, \quad P = mg$$

(M — масса Солнца, T — период обращения Земли вокруг Солнца, P — вес тела). После несложных преобразований найдем

$$\frac{P_2 - P_1}{P} = \frac{24\pi^2}{gT^2}\frac{r^2}{R} = \frac{12\pi^2 r^2}{sR}.$$

Здесь $s = 1/2\, gT^2$ означает расстояние, которое проходила бы Земля в течение года, если бы она двигалась равноускоренно с ускорением g. Вычисляя это расстояние, получим $s \approx 5 \cdot 10^{12}$ км и далее

$$\frac{P_2 - P_1}{P} \approx 6{,}5 \cdot 10^{-12}.$$

2. Найти разность между весами одинаковых тел в диаметрально противоположных точках земного шара, обусловленную неоднородностью гравитационного поля Луны. Считать, что центры Луны, Земли и обе рассматриваемые точки *1* и *2* лежат на одной прямой (см. предыдущую задачу).

О т в е т. $\frac{P_2 - P_1}{P} = \frac{M_Л}{M_З}\frac{24\pi^2 r^2}{RgT^2} \approx 8 \cdot 10^{-10}$, где $M_З$ и $M_Л$ — масса Земли и Луны, R расстояние между их центрами, T — период обращения Луны вокруг Земли, r — радиус Земли.

3. Пароход движется на восток вдоль параллели с географической широтой $\vartheta = 60°$. Скорость парохода $v = 10$ м/с. Определить вес тела P на пароходе, если взвешивание происходит на пружинных весах. Вес того же тела, неподвижного относительно Земли, в той же точке земной поверхности равен P_0.

Ответ.

$$P = P_0\left[1 - \frac{2\omega v\cos\vartheta + v^2/R}{g}\right] \approx P_0\left(1 - 2\frac{\omega v}{g}\cos\vartheta\right) \approx P_0(1 - 7{,}5\cdot 10^{-5})$$

(R — радиус Земли).

4. Самолет летит с постоянной скоростью, описывая окружность на постоянной высоте. Какое направление будет укзывать нить отвеса подвешенного в салоне самолета? Найти период малых колебаний математического маятника внутри самолета, если длина маятника равна l , корпус самолета наклонен к направлению горизонта под углом α.

Ответ. Нить отвеса установится перпендикулярно к полу салона самолета. $T = 2\pi\sqrt{l\cos\alpha/g}$.

5. Самолет летит на постоянной высоте по окружности радиусом $R = 25$ км с постоянной скоростью $v = 250$ м/с. В кабине самолета установлены пружинные и маятниковые часы. Какое время полета t' покажут маятниковые весы, если это время, измеренное пружинными часами, равно $t = 1$ ч? Часы считать идеальными. Силу Кориолиса, ввиду ее малости, не учитывать.

Ответ. $t' = t(1 + v^4/(4R^2g^2)) = 1$ ч 56 с.

§ 67. ОТКЛОНЕНИЕ ПАДАЮЩИХ ТЕЛ ОТ НАПРАВЛЕНИЯ ПОДВЕСА

1. Пусть тело свободно падает в поле тяжести Земли. В этом случае $\mathbf{F} = 0$, и уравнение (65.3) переходит в

$$\mathbf{a} = \mathbf{g} + 2[\mathbf{v}\boldsymbol{\omega}]. \qquad (67.1)$$

Это уравнение описывает *свободное падение тел с учетом вращения Земли.* Влияние вращения Земли сводится к действию центробежной и кориолисовой сил. Центробежная сила учитывается автоматически, так как она включена в вес тела $m\mathbf{g}$ как его составная часть. Наличие этой силы не меняет вид уравнения. Только направление к центру Земли заменяется направлением отвеса. В остальном центробежная сила не приводит к качественно новым явлениям. Более существенно влияет на характер движения кориолисова сила. *При падении тел без начальной скорости кориолисова сила проявляется в отклонении свободно падающих тел к востоку и экватору от направления отвеса.* Теория этих явлений сводится к решению дифференциального уравнения (67.1). Если вектор $\mathbf{g}$ постоянен, то векторное уравнение (67.1) эквивалентно системе трех линейных дифференциальных уравнений второго порядка с постоянными коэффициентами. Точное решение такой системы получить нетрудно с помощью общеизвестных методов, излагаемых в теории дифференциальных уравнений. Однако мы по этому пути не пойдем. Он громоздок, а главное, получение точных решений вряд ли оправдано, когда в уравнении (67.1) пренебрегается зависимостью $\mathbf{g}$ от координат. Последнее допустимо лишь тогда, когда движение рассматривается в сравнительно небольшой области пространства, во всех точках кото-

рой вектор $\mathbf{g}$ практически один и тот же. А в этих случаях прекрасно работает приближенный *метод последовательных приближений*, дающий вполне достаточную точность. Вычисления по этому методу просты и лучше выявляют сущность явления. Им мы и воспользуемся.

2. В уравнении (67.1) член $2[\mathbf{v}\boldsymbol{\omega}]$ мал по сравнению с $\mathbf{g}$. Его можно рассматривать как малую поправку и в *нулевом приближении* отбросить. Тогда получатся законы свободного падения без учета вращения Земли:

$$\mathbf{a} = \mathbf{g}, \quad \mathbf{v} = \mathbf{v}_0 + \mathbf{g}t, \tag{67.2}$$

где $\mathbf{v}_0$ — начальная скорость тела.

Пользуясь нулевым приближением, можно учесть и влияние кориолисовой силы. С этой целью в уравнение (67.1) мы подставим значение $\mathbf{v}$ из нулевого приближения и таким путем получим ускорение $\mathbf{a}$ в *первом приближении*:

$$\mathbf{a} = \mathbf{g} + 2[\mathbf{v}_0\boldsymbol{\omega}] + 2t[\mathbf{g}\boldsymbol{\omega}]. \tag{67.3}$$

Интегрирование этого уравнение дает скорость $\mathbf{v}$ в том же приближении:

$$\mathbf{v} = \mathbf{v}_0 + gt + 2t[\mathbf{v}_0\boldsymbol{\omega}] + t^2[\mathbf{g}\boldsymbol{\omega}]. \tag{67.4}$$

С помощью этого выражения снова уточняем выражение для кориолисовой силы. Именно, подставляя его в уравнение (67.1), получаем выражение для ускорения $\mathbf{a}$ во *втором приближении*:

$$\mathbf{a} = \mathbf{g} + 2[\mathbf{v}_0\boldsymbol{\omega}] + 2t[\mathbf{g}\boldsymbol{\omega}] + 4t[[\mathbf{v}_0\boldsymbol{\omega}]\boldsymbol{\omega}] + 2t^2[[\mathbf{g}\boldsymbol{\omega}]\boldsymbol{\omega}], \tag{67.5}$$

а после интегрирования по t — для скорости $\mathbf{v}$ в том же приближении:

$$\mathbf{v} = \mathbf{v}_0 + \mathbf{g}t + 2t[\mathbf{v}_0\boldsymbol{\omega}] + t^2[\mathbf{g}\boldsymbol{\omega}] + 2t^2[[\mathbf{v}_0\boldsymbol{\omega}]\boldsymbol{\omega}] + \tfrac{2}{3}t^3[[\mathbf{g}\boldsymbol{\omega}]\boldsymbol{\omega}] \tag{67.6}$$

Описанный процесс последовательных приближений можно было бы продолжить неограниченно. Оборвем его на втором приближении. Интегрируя (67.6) по t, находим радиус-вектор материальной точки в любой момент времени во втором приближении:

$$\mathbf{r} = \mathbf{r}_0 + \mathbf{v}_0 t + \tfrac{1}{2}\mathbf{g}t^2 + t^2[\mathbf{v}_0\boldsymbol{\omega}] + \\ + \tfrac{1}{3}t^3[\mathbf{g}\boldsymbol{\omega}] + \tfrac{2}{3}t^3[[\mathbf{v}_0\boldsymbol{\omega}]\boldsymbol{\omega}] + \tfrac{1}{6}t^4[[\mathbf{g}\boldsymbol{\omega}]\boldsymbol{\omega}]. \tag{67.7}$$

В частности, если тело падает без начальной скорости, то для его смещения из начального положения $\mathbf{s} = \mathbf{r} - \mathbf{r}_0$ получим

$$\mathbf{s} = \tfrac{1}{2}\mathbf{g}t^2 + \tfrac{1}{3}t^3[\mathbf{g}\boldsymbol{\omega}] + \tfrac{1}{6}t^4[[\mathbf{g}\boldsymbol{\omega}]\boldsymbol{\omega}]. \tag{67.8}$$

3. Чтобы проанализировать полученный результат, введем прямоугольную систему координат, начало которой поместим в точку A, из которой начинает падать рассматриваемое тело (рис. 188). Ось X направим по параллели на восток, ось Y — по меридиану к экватору, ось Z — по направлению отвеса вниз, т. е. вдоль век-

тора $\mathbf{g}$. Спроецируем затем выражение (67.8) на координатные оси. Векторное произведение $[\mathbf{g}\boldsymbol{\omega}]$ направлено на восток, двойное векторное произведение $[[\mathbf{g}\boldsymbol{\omega}]\boldsymbol{\omega}]$ есть вектор, направленный от оси вращения Земли и перпендикулярный к ней. Поэтому, переходя к проекциям, получим

$$z = \frac{1}{2} gt^2 - \frac{1}{6} \omega^2 t^4 g \cos^2 \vartheta, \tag{67.9}$$

$$x \equiv s_{\text{вост}} = \frac{1}{3} \omega t^3 g \cos \vartheta, \tag{67.10}$$

$$y \equiv s_{\text{экв}} = \frac{1}{12} \omega^2 t^4 g \sin 2\vartheta, \tag{67.11}$$

где ϑ — угол географической широты рассматриваемого места. Второе слагаемое в формуле (67.9) есть только малая поправка к нулевому приближению и не меняет качественно характер явления. Это слагаемое можно отбросить и находить время падения по формуле нулевого приближения

$$t = \sqrt{\frac{2z}{g}}. \tag{67.12}$$

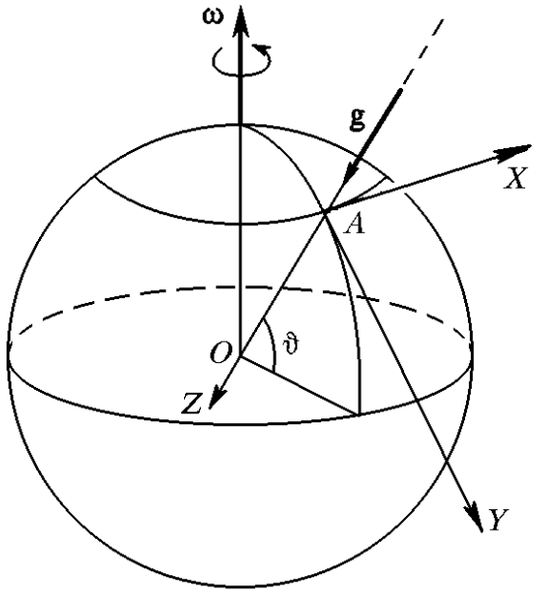


Рис. 188

Иное дело, когда речь идет о формулах (67.10) и (67.11). Здесь в нулевом приближении $x = y = 0$. Вращение Земли сказывается в появлении двух новых эффектов: отклонении свободно падающих тел к востоку и к экватору от направления отвеса (а не от направления к центру Земли, как это иногда ошибочно утверждают). Выражение для восточного отклонения можно записать в виде

$$s_{\text{вост}} = \frac{2}{3} \omega t h \cos \vartheta = \frac{4}{3} \pi \frac{t}{T} h \cos \vartheta, \tag{67.13}$$

где h — высота падения, а $T = 2\pi/\omega$ — период суточного вращения Земли.

Отклонение $s_{\text{вост}}$ очень мало, так как в формулу (67.13) входит малый множитель t/T. Так, при $h = 100$ м $t = 4{,}5$ с, и для широты Москвы ($\vartheta = 56°$) получаем $s_{\text{вост}} = 1{,}2$ см. При падении с высоты $h = 500$ м получилось бы $s_{\text{вост}} = 13{,}8$ см. Несмотря на малость эффекта, его с уверенностью удалось наблюдать в опытах с падением тел в глубоких шахтах уже в середине XIX века.

Экваториальное отклонение связано с восточным соотношением:

$$s_{\text{экв}} = \frac{\omega t \sin \vartheta}{2} s_{\text{вост}}. \tag{67.14}$$

Из-за наличия малого множителя $\omega t = 2\pi t/T$ отклонение к экватору очень мало и по этой причине недоступно наблюдению.

ЗАДАЧИ

1. Из ружья произведен выстрел строго вверх (т. е. параллельно линии отвеса). Начальная скорость пули $v_0 = 100$ м/с, географическая широта места $\vartheta = 60°$. Учитывая осевое вращение Земли, определить приближенно, насколько восточнее или западнее от места выстрела упадет пуля. Сопротивление воздуха не принимать во внимание.

Ответ. Пуля отклонится к западу на расстояние

$$x_{\text{зап}} = \frac{4}{3}\frac{v_0^3\omega}{g}\cos\vartheta \approx 51 \text{ см}.$$

Результат может показаться неожиданным. При движении вверх кориолисова сила отклоняет брошенное тело к западу от направления отвеса, при движении вниз она отклоняет его к востоку. На первый взгляд кажется, что отклонение к западу должно компенсироваться последующим отклонением к востоку. На самом деле это не так. Когда тело движется вверх, его боковая начальная скорость равна нулю. В наивысшую точку тело переходит, однако, с западной составляющей скорости, которую оно приобретает под действием кориолисовой силы. Поэтому обратное падение тела начинается с начальной скоростью, направленной на запад. При этом тело не только смещается к востоку под действием изменившей направление кориолисовой силы, но и продолжает по инерции двигаться на запад. В результате отклонение к западу оказывается больше, чем отклонение к востоку.

2. Под каким углом α к вертикали надо произвести выстрел вверх, чтобы пуля упала обратно в точку, из которой был произведен выстрел? Использовать данные предыдущей задачи.

Ответ. Ствол ружья надо наклонить к востоку под углом

$$\alpha = \frac{2}{3}\frac{v_0\omega}{g}\cos\vartheta \approx 2{,}45\cdot 10^{-4} \text{ рад} \approx 0{,}85' \approx 51''.$$

3. Из орудия, установленного в точке земной поверхности с географической широтой $\vartheta = 30°$, производится выстрел в направлении на восток. Начальная скорость снаряда $v_0 = 500$ м/с, угол вылета снаряда (т. е. угол наклона касательной в начальной точке траектории к плоскости горизонта) $\alpha = 60°$. Пренебрегая сопротивлением воздуха и учитывая вращение Земли, определить приближенно отклонение y точки падения снаряда от плоскости стрельбы. Какое это будет отклонение: к югу или к северу? (Плоскостью стрельбы называется плоскость, проходящая через направление касательной в начальной точке траектории и направление отвеса в той же точке.)

Ответ. $y = \frac{1}{g^2}\cdot 4\omega v_0^3 \sin\vartheta \cos\alpha \sin^2\alpha \approx 71$ м. К югу.

§ 68. МАЯТНИК ФУКО

1. Опыты по отклонению к востоку свободно падающих тел в принципе могли бы служить экспериментальным доказательством неинерциальности земной системы отчета и приближенной инерциальности системы Коперника. Однако постановка таких опытов затруднительна, а их точность невелика. Для этой цели более подхо-

дящим является *маятник Фуко*. Так называется массивный шар, подвешенный на достаточно длинной нити и совершающий малые колебания около положения равновесия. Отклоним маятник из положения равновесия, а затем предоставим его самому себе. Если бы Земля была инерциальной системой отсчета, то на маятник действовали бы только «настоящие силы»: сила веса $m\mathbf{g}$ и сила натяжения нити $\mathbf{F}$ (силами трения и сопротивления воздуха пренебрегаем). Обе эти силы лежат в вертикальной плоскости. Поэтому если маятнику не сообщен толчок в боковом направлении, то он все время будет колебаться в одной и той же вертикальной плоскости, неподвижной относительно Земли. Опыты показали, что это не так: плоскость качаний маятника в земной системе отсчета медленно поворачивается вокруг вертикали рассматриваемого места и притом в том же направлении, в каком совершают суточное вращение Солнце и звезды на небесной сфере. Это доказывает, что земная система отсчета не является инерциальной.

Чтобы объяснить вращение плоскости качаний маятника, предположим, что Земля равномерно вращается относительно неизвестной нам инерциальной системы отсчета с угловой скоростью $\boldsymbol{\omega}$. В земной системе отсчета к «настоящим силам», действующим на маятник, добавятся еще силы инерции: центробежная и кориолисова. Движение маятника будет описываться уравнением (65.3). Кориолисова сила $2m[\mathbf{v}\boldsymbol{\omega}]$ перпендикулярна к плоскости качаний маятника. Она-то и вызывает вращение этой плоскости.

2. Допустим сначала, что опыт произведен на полюсе Земли. Тогда в уравнении (65.3) вектор $\boldsymbol{\omega}$ будет направлен вдоль вертикали. Но результат легко предсказать, если рассмотреть качания маятника в инерциальной системе отсчета. В этой системе нет никаких сил инерции — действуют только сила веса $m\mathbf{g}$ и сила натяжения нити $\mathbf{F}$. Поэтому в инерциальной системе плоскость качаний маятника будет сохранять неизменное положение. Земля же будет вращаться относительно этой неподвижной плоскости с угловой скоростью $\boldsymbol{\omega}$. Иными словами, плоскость качаний маятника будет вращаться относительно Земли с той же угловой скоростью $\boldsymbol{\omega}$, но в противоположном направлении. Разумеется, результат предсказаний не может зависеть от способа рассмотрения (если только способ правильный). Поэтому к тому же результату мы пришли бы, если бы с самого начала рассматривали задачу в земной системе отсчета с помощью уравнения относительного движения (65.3). Это замечание позволяет легко разобраться в вопросе, как будет вести себя плоскость качаний маятника, если опыт произведен в любом месте земной поверхности (а не только на полюсе).

3. Допустим, что опыт произведен в точке земной поверхности с географической широтой ϑ. Разложим вектор угловой скорости $\boldsymbol{\omega}$ на

две составляющие: вертикальную $\boldsymbol{\omega}_в$ и горизонтальную $\boldsymbol{\omega}_г$: $\boldsymbol{\omega} = \boldsymbol{\omega}_в + \boldsymbol{\omega}_г$. Горизонтальную составляющую в свою очередь разложим на две составляющие: $\boldsymbol{\omega}_\parallel$ и $\boldsymbol{\omega}_\perp$, из которых $\boldsymbol{\omega}_\parallel$ лежит в плоскости качаний маятника, а $\boldsymbol{\omega}_\perp$ к ней перпендикулярна (рис. 189). Тогда уравнение (65.3) представится в виде

$$m\mathbf{a} = m\mathbf{g} + 2m[\mathbf{v}\boldsymbol{\omega}_в] + 2m[\mathbf{v}\boldsymbol{\omega}_\perp] + 2m[\mathbf{v}\boldsymbol{\omega}_\parallel] + \mathbf{F}.$$

Составляющая силы Кориолиса $2m[\mathbf{v}\boldsymbol{\omega}_\perp]$ направлена вдоль нити маятника. Она слегка меняет натяжение нити, а с ним и период колебаний маятника. На положение плоскости качаний маятника эта составляющая не оказывает влияния. В задаче о вращении плоскости качаний маятника ее можно отбросить. Вторая составляющая силы Кориолиса $2m[\mathbf{v}\boldsymbol{\omega}_в]$ в нашей задаче наиболее важна. Она перпендикулярна к плоскости качаний маятника и вызывает вращение этой плоскости. Третья составляющая $2m[\mathbf{v}\boldsymbol{\omega}_\parallel]$ тоже перпендикулярна к плоскости качаний маятника, а потому она также оказывает влияние на эту плоскость. Однако при малых колебаниях маятника эта составляющая мала в силу малости угла α. Кроме того, при колебаниях маятника она периодически меняет направление. Когда маятник движется от центра O вправо или влево, составляющая $2m[\mathbf{v}\boldsymbol{\omega}_\parallel]$ направлена за плоскость рисунка (рис. 189). Когда маятник из крайних положений приближается к центру O, она направлена противоположно, т. е. к читателю. Поэтому сила $2m[\mathbf{v}\boldsymbol{\omega}_\parallel]$ не приводит к систематическому вращению плоскости колебаний маятника, а вызывает лишь малые колебания ее относительно среднего положения. Эту силу можно также отбросить. В результате уравнение относительного движения примет вид

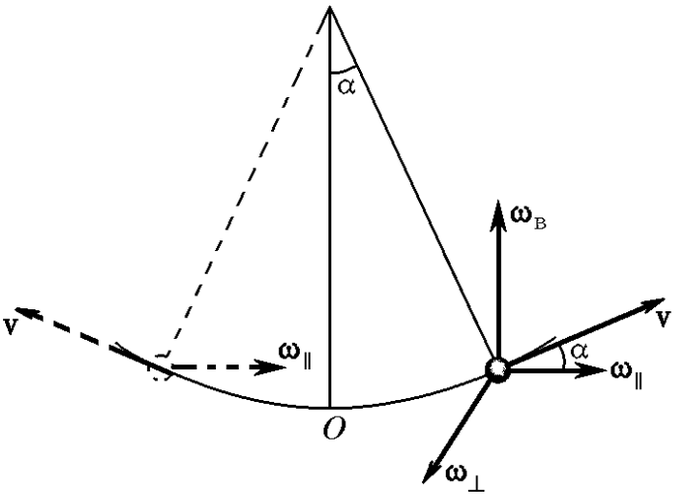


Рис. 189

$$m\mathbf{a} = m\mathbf{g} + 2m[\mathbf{v}\boldsymbol{\omega}_в] + \mathbf{F}. \qquad (68.1)$$

Из уравнения выпала горизонтальная составляющая угловой скорости $\boldsymbol{\omega}$. Уравнение приняло такой же вид, как и на полюсе. Вся разница только в том, что вместо полной угловой скорости в него вошла ее вертикальная составляющая $\boldsymbol{\omega}_в$, т. е. маятник будет вести себя так же, как и на полюсе. Но плоскость качаний его будет вращаться с меньшей угловой скоростью

$$\omega_в = \omega \sin \vartheta. \qquad (68.2)$$

Полный оборот плоскость качаний маятника совершит за время

$$\tau = \frac{2\pi}{\omega \sin \vartheta} = \frac{T}{\sin \vartheta}, \qquad (68.3)$$

где T — период вращения Земли относительно инерциальной системы отсчета.

Реальный опыт впервые был произведен Фуко в Парижской обсерватории в 1850 г. и повторен в 1851 г. в Пантеоне. Маятник имел длину 67 м и состоял из металлического шара массой $m = 28$ кг. Опыт показал, что относительно Земли плоскость качаний маятника вращается вокруг вертикали рассматриваемого места в соответствии с формулами (68.2) и (68.3), если только вращение самой Земли относить к системе Коперника. Это доказывает, что земная система отсчета не инерциальна, а система Коперника — инерциальна. Конечно, последнее заключение не может быть столь же категоричным, каким является первое. Лучше сказать, что опыт Фуко не противоречит предположению об инерциальности коперниковой системы отсчета.

4. Исследуем более детально форму траектории маятника Фуко при его колебаниях относительно земной системы отсчета. Как уже выяснено, можно отвлечься от горизонтальной составляющей угловой скорости $\boldsymbol{\omega}$ и считать, что Земля вращается вокруг вертикали с угловой скоростью $\boldsymbol{\omega}_в$. Иначе говоря, можно рассуждать так, как если бы опыт Фуко был произведен на полюсе, но Земля вращалась бы с меньшей угловой скоростью $\boldsymbol{\omega}_в$. Пусть вектор угловой скорости $\boldsymbol{\omega}_в$ направлен перпендикулярно к плоскости рисунка к читателю (рис. 190). Кориолисова сила $2m[\mathbf{v}\boldsymbol{\omega}_в]$, действующая на маятник при его колебаниях, перпендикулярна к его траек-

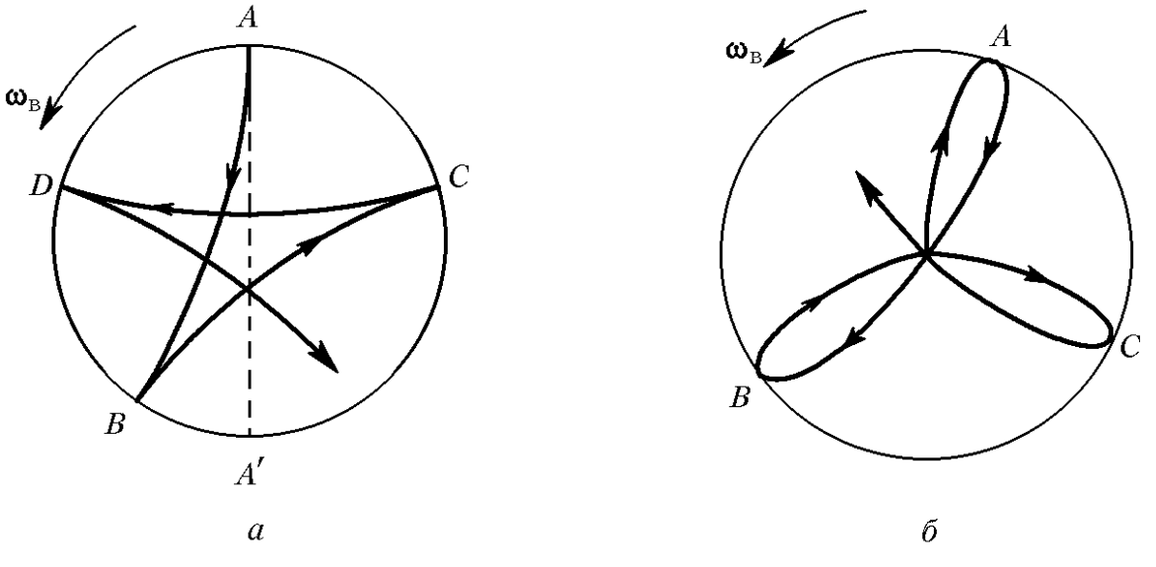


Рис. 190

тории и направлена вправо по ходу движения маятника. Эта сила искривляет траекторию маятника. Допустим сначала, что маятник отклонен в крайнее положение A, а затем отпущен без начальной скорости. Если бы не было кориолисовой силы, то маятник пришел бы в диаметрально

противоположную точку A'. Кориолисова сила отклонит маятник в сторону, и он придет в точку B, расположенную правее. В этой точке скорость маятника обратится в нуль, а затем изменит направление. Изменит направление и кориолисова сила. Она по-прежнему будет изгибать траекторию маятника вправо (так как наблюдатель также должен повернуться, чтобы все время смотреть в направлении движения маятника). Затем маятник будет последовательно приходить в новые точки поворота C, D, ... В результате получится сложная кривая с угловыми точками, изображенная схематически на рис. 190 *а*.

Несколько иной характер траектории получится в том случае, когда маятнику сообщен толчок из положения равновесия. Траектория по-прежнему будет изгибаться вправо. Но в крайние точки A, B, C, ... (рис. 190 *б*) маятник будет приходить с отличными от нуля азимутальными скоростями, которые он приобрел под действием кориолисовой силы, когда двигался от центра. В результате в местах поворота получатся не точки заострения, а плавные закругления, как это изображено на рисунке. Вследствие медленности вращения Земли наблюдатель не замечает искривления плоскости качания маятника. В обоих случаях ему кажется, что плоскость качаний маятника вращается вокруг вертикали с угловой скоростью $\omega_в = \omega \sin \vartheta$.

ЗАДАЧА

Один из маятников Фуко был установлен в Ленинграде в Исаакиевском соборе. Длина маятника $l = 98$ м, линейная амплитуда колебаний шара маятника (т. е. наибольшее отклонение его из положения равновесия) $x_0 = 5$ м. Маятник отпускался из крайнего положения без начального толчка. Определить боковое отклонение шара маятник от положения равновесия в момент прохождения его через среднее положение. Географическая широта Ленинграда $\vartheta = 60°$.

Р е ш е н и е. Эта задача решается проще, если движение рассматривать в неподвижной системе отсчета (точнее, в системе отсчета, вращающейся относительно Земли вокруг вертикали рассматриваемого места с угловой скоростью $\boldsymbol{\omega}_в$). В этой системе уравнение малых колебаний математического маятника имеет вид $\ddot{\mathbf{r}} + \Omega^2 \mathbf{r} = 0$, где $\Omega^2 = g/l$, а $\mathbf{r}$ — смещение маятника из положения равновесия. В начальный момент маятник, вращаясь вместе с Землей, имеет боковую скорость $\omega_в x_0$. Поместим начало координат O в положение равновесия маятника. Ось X направим из точки O к точке $(x = x_0,\ y = 0)$, в которой маятник находился в начальный момент. Для движения вдоль оси Y имеем $\ddot{y} + \Omega^2 y = 0$. Решая это уравнение при начальных условиях $y_{t=0} = 0$, $\dot{y}_{t=0} = \omega_в x_0$, получим

$$y = \frac{\omega_в x_0}{\Omega} \sin \Omega t.$$

В среднем положении $\Omega t = \pi/2$, для бокового отклонения в этом положении наша формула дает

$$y = \frac{\omega_в x_0}{\Omega} = \frac{\omega x_0}{\Omega} \sin \vartheta \approx 1 \text{ мм}.$$

Читателю рекомендуется получить тот же результат, рассматривая движение в земной системе отсчета.

§ 69. ПРИЛИВЫ

1. У берегов океанов и морей дважды в сутки наблюдается поднятие (*прилив*) морской воды до некоторого максимального уровня (*полная вода*). После этого начинается опускание ее (*отлив*) до минимального уровня (*малая вода*). Разность уровней большой и малой воды называется *амплитудой прилива*. Время между следующими друг за другом положениями полной (или малой) воды составляет 12 ч 25 мин. Это время точно совпадает с половиной промежутка времени, в течение которого Луна в своем видимом движении совершает полный оборот вокруг Земли. Поэтому уже давно причину приливов и отливов связывали с положением Луны на небесном своде. Однако научное объяснение этого явления впервые было дано Ньютоном.

Приливы и отливы объясняются *неоднородностью поля тяготения Луны и отчасти Солнца*. Если бы внешнее гравитационное поле было однородно, то в земной системе отсчета оно полностью компенсировалось бы поступательной силой инерции, связанной с ускоренным движением центра масс Земли (где мы помещаем начало координат этой системы). На самом деле гравитационное поле неоднородно, и полная компенсация имеет место только в центре масс Земли. В остальных точках полной компенсации нет. Остаются нескомпенсированные силы, которые и вызывают приливы. Влияние Луны более существенно, чем Солнца. Хотя лунное поле тяготения и слабее солнечного, но оно *более неоднородно*, так как Луна примерно в 400 раз ближе к Земле, чем Солнце. Рассмотрим сначала, как выглядело бы явление приливов, если бы Солнца не было, а Земля подвергалась бы воздействию гравитационного поля одной только Луны.

2. Для простоты будем считать Землю твердым недеформированным шаром, покрытым океаном постоянной глубины. Будем считать также, что Луна движется в плоскости земного экватора. Рассмотрим точки океана, расположенные вдоль экватора. Земля и Луна вращаются вокруг их общего центра масс, как бы непрерывно падая на него. Но точка A (рис. 191), для которой Луна находится в *зените*, расположена ближе к Луне, чем центр Земли O. Последний в свою очередь, ближе к Луне, чем диаметрально противоположная точка B, для которой Луна находится в *надире*. Поэтому гравитационное поле Луны в точке A сильнее, а в точке B слабее, чем в центре Земли. Под влиянием гравитационного притяжения Луны частицы воды в точке A будут приближаться к Луне с бóльшим ускорением, чем центр Земли O, а частицы воды в точке B — с меньшим ускорением. Начиная с этого места, большинство авторов по примеру Ньютона рассуждает неточно. Заключения, касающиеся ускорений частиц, переносятся на их скорости и перемещения. Говорят, что частицы воды в A будут приближаться к Луне быстрее, чем центр Земли O, а потому они будут опере-

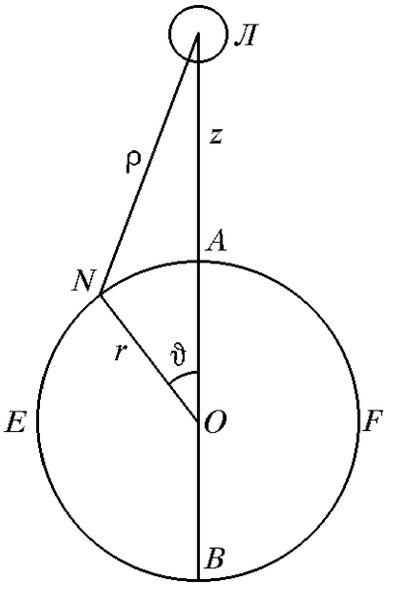


Рис. 191

жать последний. Напротив, частицы воды вблизи точки B будут отставать от центра Земли. По этой причине на поверхности океана образуются два диаметрально противоположных горба или выступа с центрами в точках A и B (рис. 192 *а*). Центры горбов все время обращены к Луне и от нее. Вследствие осевого вращения Земли они бегут по поверхности океана, непрерывно следуя за движением Луны. Вот почему два последовательных прилива (или отлива) отделены друг от друга промежутком времени в 12 ч 25 мин. Согласно приведенному объяснению полная вода должна наблюдаться в моменты времени, когда Луна находится в *верхней* или *ниж-*

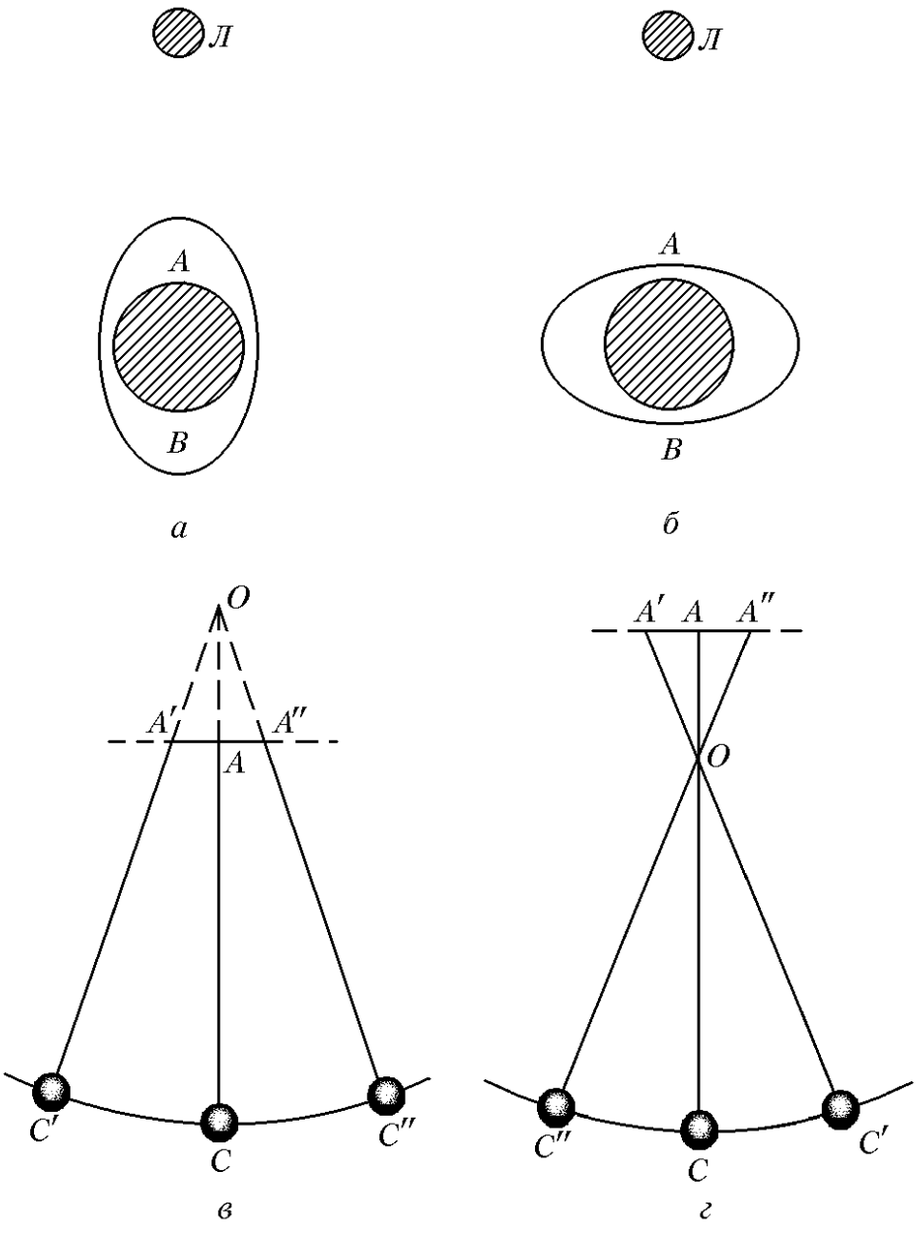


Рис. 192

ней кульминации (в зените или надире), а малая вода — когда они находятся в *квадратуре*. Наблюдения не согласуются с этим заключением. Скорее, справедлива обратная закономерность: *полная вода наблюдается в квадратурах, а малая — в кульминациях Луны* (рис. 192 *б*). Во всяком случае, между кульминацией Луны и последующей полной водой проходит значительный промежуток времени, составляющий несколько часов. В службе портов среднее значение этого промежутка называется *прикладным часом*. Такое расхождение между теорией и наблюдениями связано

прежде всего с неточностью в рассуждениях, отмеченной выше. Смещение и скорости частиц воды определяются не только ускорениями, но и их *начальными значениями*. Если бы в какой-либо один и тот же момент времени (который можно принять за начальный) частицы воды находились, например, в состоянии покоя, то рассуждение было бы верным. Но это условие на Земле как раз и не выполняется. На этом вопросе мы остановимся несколько ниже.

3. Задачу построения теории приливов можно разделить на две части. Одна, более простая, заключается в нахождении *приливообразующих сил*, действующих на воду океана в различных точках земного шара. Вторая, несравненно более трудная, состоит в том, чтобы определить *вынужденное движение* воды, которое установится под действием этих сил. Остановимся на первой части задачи.

Силы, действующие на частицы воды, в земной системе отсчета складываются из сил тяготения и сил инерции. Силы притяжения самой Земли, а также центробежные силы, возникающие из-за вращения Земли вокруг ее центра масс, в вопросах образования приливов роли не играют. Их результирующую напряженность мы будем обозначать буквой **g** (ускорение свободного падения). *Вектор* **g** *в каждой точке земной поверхности остается постоянным. Он определяет форму свободной поверхности океана в состоянии равновесия.* Эта поверхность всюду перпендикулярна к вектору **g**. В теории приливов нас интересуют *отклонения от этой равновесной формы*, связанные с действием переменных приливообразующих сил. При определении этих отклонений равновесную форму поверхности воды в океане можно считать шаровой. Кориолисову силу инерции мы не будем принимать во внимание, потому что воду в океане в отсутствие возмущающих приливообразующих сил мы будем считать покоящейся. Кориолисовы силы, возникающие из-за движений воды, вызванных приливами и отливами, пренебрежимо малы. Таким образом, при вычислении приливообразующих сил надо учесть только силы тяготения внешних сил (Луны), а также силы инерции, связанные с ускоренным движением центра Земли. Такие силы инерции по нашей терминологии называются *поступательными силами инерции* (см. § 64).

4. Приливообразующую силу мы будем относить к единице массы, на которую она действует, и обозначать буквой **f**. Проще сначала вычислить не самый вектор **f**, а соответствующий ему потенциал $\varphi_{\text{пр}}$, т. е. потенциальную энергию единицы массы, находящейся под действием силы **f**. Он складывается из потенциала сил тяготения Луны ($\varphi_{\text{Л}}$) и потенциала поступательных сил инерции ($\varphi_{\text{ин}}$). Опуская эти вычисления (они приводятся в задаче к этому параграфу), приведем окончательные результаты. Приливообразующий потенциал определяется выражением

$$\varphi_{\text{пр}} = -\frac{3}{4}\frac{GM_{\text{Л}}}{R_{\text{ЗЛ}}^3} r^2 \cos 2\vartheta, \tag{69.1}$$

где $M_{\text{Л}}$ — масса Луны, $R_{\text{ЗЛ}}$ — расстояние между центрами Земли и Луны, ϑ — зенитное расстояние Луны в рассматриваемый момент времени, r — расстояние от центра Земли до точки наблюдения N (см. рис. 191). Величины r и ϑ являются полярными координатами точки наблюдения. Приливообразующая сила найдется дифференцированием по-

тенциала $\varphi_{пр}$. Она содержит вертикальную ($f_в$) и горизонтальную ($f_г$) составляющие

$$f_в = -\frac{\partial \varphi_{пр}}{\partial r},\quad f_г = -\frac{1}{r}\frac{\partial \varphi_{пр}}{\partial \vartheta}. \tag{69.2}$$

(За положительные приняты направления возрастания величин r и ϑ.) Дифференцируя и вводя ускорение свободного падения $g = G\frac{M_З}{r^2}$, получим

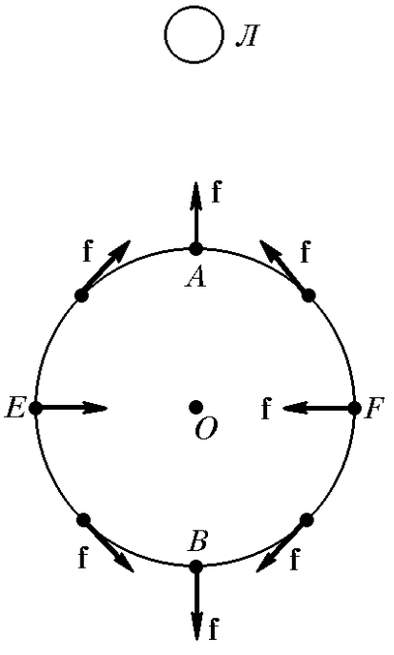


Рис. 193

$$f_в = \frac{3}{2}\frac{M_Л}{M_З}\left(\frac{r}{R_{ЗЛ}}\right)^2 g\cos 2\vartheta, \tag{69.3}$$

$$f_г = \frac{3}{2}\frac{M_Л}{M_З}\left(\frac{r}{R_{ЗЛ}}\right)^2 g\sin 2\vartheta. \tag{69.4}$$

Распределение приливообразующих сил вдоль экватора показано стрелками на схематическом рис. 193. Полная приливообразующая сила будет

$$f = \sqrt{f_в^2 + f_г^2} = \frac{3}{2}\frac{M_Л}{M_З}\left(\frac{r}{R_{ЗЛ}}\right)^3 g. \tag{69.5}$$

Положив здесь $M_Л/M_З = 1/81$, $r/R_{ЗЛ} = 1/60$, получим $f/g = 8{,}57\cdot 10^{-8}$. Следует заметить, что Луна обращается вокруг Земли по эллиптической орбите. В перигее она бывает на расстоянии 57 земных радиусов, в апогее — на расстоянии 63,7. Это сказывается на величине приливообразующей силы Луны. Для Луны отношение f/g меняется в пределах от $7{,}2\cdot 10^{-8}$ (в апогее) до 10^{-7} (в перигее). Конечно, приведенные формулы полностью справедливы и для приливообразующих сил, вызываемых Солнцем. В этом случае $f/g = 3{,}8\cdot 10^{-8}$, т. е. в 21/4 раза меньше, чем для Луны при ее среднем удалении от Земли. Величина приливообразующей силы Солнца меняется в течение года примерно на 10%.

5. Приведенные результаты показывают, насколько ничтожны приливообразующие силы по сравнению с обычной силой тяжести на Земле. То обстоятельство, что эти силы вызывают такое грандиозное явление природы, как приливы и отливы, связано с тем, что они не постоянны, а *периодически меняются во времени*. Если бы приливообразующие силы менялись от точки к точке на земной поверхности, но оставались постоянными во времени, то они лишь слегка изменили бы равновесную форму свободной поверхности воды в океане. Но эта форма не менялась бы с течением времени, т. е. не было бы никаких приливов и отливов. Такая ситуация осуществлялась бы, если бы Луна вращалась по круговой орбите с угловой скоростью, равной угловой скорости осевого вращения Земли. В действительности, как показывают формулы (69.3)—(69.5), в каждой точке земного шара остается неизменной лишь модуль приливообразующей силы, но не ее направление. Обе составляющие приливообразующей силы $f_в$ и $f_г$ в каждой точке земного шара периодически меняются во времени из-за суточных изменений зенитного расстояния Луны ϑ. Отвле-

каясь от второстепенных обстоятельств, можно положить $\vartheta = \omega t$, где ω — угловая скорость осевого вращения Земли (относительно прямой Земля—Луна). Поэтому $f_в \sim \cos 2\omega t$, $f_г \sim \sin 2\omega t$. Когда сила $f_в$ проходит через максимум, сила $f_г$ обращается в нуль, и наоборот. Это вызывает *периодические изменения направления отвеса* в каждой точке земного шара, что и является непосредственной причиной приливов и отливов.

6. Теперь мы должны обратиться ко второй части задачи о приливах, а именно определить воздействие заданных приливообразующих сил на воду в океане. Первая — *статическая* — теория приливов была разработана Ньютоном. Эта теория определяла мгновенную форму свободной поверхности океана, как если бы приливообразующие силы были постоянными, т. е. не менялись во времени. Согласно законам гидростатики свободная поверхность жидкости в состоянии равновесия в каждой точке перпендикулярна к (постоянным) действующим силам. Отсюда следует, что вдоль свободной поверхности жидкости потенциал всех действующих сил φ не должен меняться. Очевидно, $\varphi = \varphi_0 + \varphi_{пр}$, где φ_0 — потенциал всех сил, определяющих ускорение свободного падения g в отсутствие приливообразующих сил. Таким образом, по статической теории приливов уравнение свободной поверхности воды в океане должно иметь вид $\varphi_0 + \varphi_{пр} = \text{const}$, или более подробно

$$\varphi_0 - \frac{3}{4}\frac{GM_л}{R_{зл}^3} r^2 \cos 2\vartheta = \text{const}. \tag{69.6}$$

Применим это уравнение к точкам A и E на поверхности океана (см. рис. 193). Полагая сначала $\vartheta = 0$, а затем $\vartheta = \pi/2$, получим

$$\varphi_0(A) - \frac{3}{4}\frac{GM_л}{R_{зл}^3} r_A^2 = \varphi_0(E) + \frac{3}{4}\frac{GM_л}{R_{зл}^3} r_E^2.$$

Но $\varphi_0(A) - \varphi_0(E) = gH$, где $H = r_A - r_E$ — амплитуда прилива. У оставшихся двух членов знаки A и E можно опустить, полагая $r_A = r_E = r$. Замечая еще, что $g = G\frac{M_з}{r^2}$, и используя формулу (69.5), получим

$$H = \frac{f}{g} r. \tag{69.7}$$

По этой формуле находим для амплитуды лунных приливов $H = 0{,}55$ м, а для амплитуды солнечных приливов $H = 0{,}24$ м. Таким образом, по статической теории картина приливов и отливов должна соответствовать рис. 192 *а*, а не рис. 192 *б*. В этом основной недостаток статической теории приливов.

7. Правильная полная теория приливов должна быть *динамической*. Надо определить *вынужденное движение* воды в океане под действием заданных *переменных* приливообразующих сил. Важный принципиальный момент, который должна учесть теория, состоит в том, что вода в океане представляет собой механическую систему, которой, подобно маятнику, свойственны определенные *собственные частоты свободных колебаний*. Чтобы простейшим образом пояснить суть дела, вообразим вместе с Эйри, что на Земле вдоль ее экватора прорыт канал постоянной глубины, заполненный водой и опоясывающий весь земной шар. Если в каком-либо месте канала возникло возмущение,

то оно будет распространяться вдоль него с определенной скоростью. Пренебрежем силами трения, действующими в жидкости. В этом случае, как доказывается в гидродинамике, скорость распространения длиноволновых возмущений (т. е. таких возмущений, длины волн которых очень велики по сравнению с глубиной канала h) определяется формулой $u = \sqrt{gh}$. Возьмем в качестве h среднюю глубину воды в океане ($h = 3{,}5$ км). Тогда нетрудно подсчитать, что возмущение обежит вокруг Земли за 60 ч. При рассмотрении явления приливов играет роль время, вдвое меньшее. Дело в том, что в этом случае возмущение состоит из двух одинаковых горбов A и B, расположенных в диаметрально противоположных точках земного шара (рис. 192 *а, б*). По истечении 30 ч горб A перейдет в положение B, а горб B — в положение A, и первоначальная форма поверхности воды в канале восстановится. Значит, воде в канале свойствен *собственный период колебаний* $T_0 = 30$ ч. Он больше периода колебаний приливообразующей силы $T = 12$ ч 25 мин. Из элементарной теории колебаний известно, что в этом случае (при отсутствии сил трения) *внешняя сила и возбужденные ею вынужденные колебания находятся в противоположных фазах*. Наоборот, при $T_0 < T$ колебания совершаются в *одинаковых фазах*. Так, если привести в колебания точку подвеса A математического маятника, то шарик маятника C также придет в колебание (рис. 192 *в, г*). При малых частотах колебаний точки подвеса она и шарик в каждый момент времени будут двигаться в одинаковых, при больших — в противоположных направлениях. Поскольку в разбираемом нами вопросе $T_0 > T$, картина приливов должна соответствовать рис. 192 *б*, а не рис. 192 *а*. Статическая теория приливов качественно верно описывала бы явление приливов, если бы было $T_0 < T$. Но для этого нужно, как нетрудно подсчитать, чтобы глубина h превышала 20 км.

8. Солнечные приливы накладываются на приливы лунные. Если при этом наложении они усиливают друг друга, то приливы получаются особенно сильными. Это происходит тогда, когда Солнце и Луна находятся на одной прямой с Землей, т. е. в полнолуние и новолуние. Наступающие тогда приливы называются *большими* (*сизигийными*) *приливами*. Наоборот, когда Луна находится в первой или последней четверти, лунный прилив ослабляется солнечным. Тогда говорят о *малом*, или *квадратурном, приливе*.

Полная теория приливов, отвечающая всем требованиям практики, еще не создана. Это и понятно. На характере приливов существенно сказывается сложный рельеф дна океанов и морей, наличие материков и островов, очертания берегов, трение, морские течения и ветры, деформации самой Земли под действием приливообразующих сил и множество других трудно учитываемых факторов.

На открытых островах в океане амплитуда прилива в полнолуние и новолуние обычно бывает порядка 1 м. Это находится в согласии с тем, что дает статическая теория приливов. У берегов океана амплитуда приливов обычно около 2 м. Мест с амплитудой в 3 м уже немного, а с амплитудой более 6 м очень мало. Все они находятся либо в узких приливах, либо в глубине длинных заливов. Наиболее значительные приливы наблюдаются в заливе Фунди, на восточном берегу Канады. Этот залив расположен между материком и полуостровом Новая Шотландия. Амплитуда от 4 м при входе нарастает до 12—16 м в глубине залива. Во время сизигийных приливов здесь наблюдались амплитуды свыше 20 м.

ЗАДАЧА

Вывести формулы (69.1), (69.3), (69.4).

Решение. Как выяснено в тексте, $\varphi_{\text{пр}} = \varphi_{\text{Л}} + \varphi_{\text{ин}}$. Направим ось Z в сторону Луны (см. рис. 191). Пусть w — ускорение, с которым центр Земли O приближается к центру масс Земля—Луна. Соответствующая сила инерции будет $-mw = -w$. Считая ее однородной, имеем $\varphi_{\text{ин}} = wz = wr\cos\vartheta$. Потенциал силы тяготения Луны равен

$$\varphi_{\text{Л}} = -G\frac{M_{\text{Л}}}{\rho}.$$

С помощью рис. 191 находим $\rho^2 = R_{\text{ЗЛ}}^2 = 2R_{\text{ЗЛ}}r\cos\vartheta + r^2$. Применяя формулу бинома Ньютона и пренебрегая кубами и высшими степенями r, получим

$$\varphi_{\text{Л}} = -\frac{GM_{\text{Л}}}{R_{\text{ЗЛ}}}\left(1 - \frac{2R_{\text{ЗЛ}}r\cos\vartheta - r^2}{R_{\text{ЗЛ}}^2}\right)^{-1/2} =$$

$$= -\frac{GM_{\text{Л}}}{R_{\text{ЗЛ}}}\left[1 + \frac{2R_{\text{ЗЛ}}r\cos\vartheta - r^2}{2R_{\text{ЗЛ}}^2} + \frac{3}{8}\left(\frac{2r\cos\vartheta}{R_{\text{ЗЛ}}}\right)^2\right].$$

Постоянный член $-GM_{\text{Л}}/R_{\text{ЗЛ}}$, как и всякую постоянную в выражении потенциала, можно отбросить. Линейный по r член компенсируется потенциалом $\varphi_{\text{ин}}$, так как $w = GM_{\text{Л}}/R_{\text{ЗЛ}}^2$. Далее, из потенциала $\varphi_{Л}$ можно исключить все члены, зависящие только от r, но не зависящие от угла ϑ. Они вносят одну и ту же радиальную добавку в действующую силу во всех точках земной поверхности. Эту добавку можно включить в g. К образованию приливов она не имеет отношения. С учетом этих замечаний нетрудно получить

$$\varphi_{\text{пр}} = -\frac{3}{4}\frac{r^2}{R_{\text{ЗЛ}}}w\cos 2\vartheta = -\frac{3}{4}\frac{GM_{\text{Л}}}{R_{\text{ЛЗ}}^3}r^2\cos 2\vartheta.$$

§ 70. ГРАВИТАЦИОННАЯ МАССА И ОБОБЩЕННЫЙ ЗАКОН ГАЛИЛЕЯ

1. Понятие массы было введено нами с помощью закона сохранения импульса. В основе этого понятия лежат *инерционные свойства тел.* Поэтому так определенную массу называют *инертной массой* и иногда обозначают через $m^{(i)}$. Однако *тела обладают не только свойствами инерции, но и способностью возбуждать в окружающем пространстве гравитационные поля.* В этом отношении они аналогичны электрически заряженным телам, создающим вокруг себя электрическое поле. Инерция тел и их способность возбуждать в окружающем пространстве гравитационные поля не должны априори рассматриваться как взаимосвязанные и тем более тождественные свойства тел. Можно думать, что тела являются источниками гравитационных полей не потому, что они обладают инертными массами, а потому, что они несут особые заряды, аналогичные

электрическим зарядам. Такие заряды называются *гравитационными зарядами* или *гравитационными массами. Силы взаимодействия гравитационных масс*, как показывает опыт, *изменяются обратно пропорционально квадрату расстояния между ними.* Для количественного определения гравитационных масс можно поступить так же, как поступают с электрическими зарядами в электростатике. Именно, обозначим гравитационные массы взаимодействующих точечных тел через $m_1^{(g)}$ и $m_2^{(g)}$. Тогда для силы их гравитационного притяжения можно написать

$$F = C\frac{m_1^{(g)}m_2^{(g)}}{r^2}, \tag{70.1}$$

где C — числовой коэффициент, зависящий только от выбора единиц. Этому коэффициенту можно приписать произвольную размерность и произвольное числовое значение. Тогда, считая единицы для r и F установленными, мы установим также единицу гравитационной массы и ее размерность, а формула (70.1) даст принципиальный способ измерения гравитационных масс.

Пропорциональность силы гравитационного взаимодействия тел их гравитационным массам не является физическим законом. Мы так вводим понятие гравитационной массы, что указанная пропорциональность соблюдается *по определению. Физический закон, установленный Ньютоном, состоит в том, что сила гравитационного взаимодействия тел пропорциональна их инертным массам.* Отсюда следует, что *инертная масса тела пропорциональна его гравитационной массе.* Единицы этих масс можно выбрать так, чтобы они были не только пропорциональны, но и численно равны между собой. Поэтому этот фундаментальный физический закон называется *законом равенства или эквивалентности инертной и гравитационной масс.* Посмотрим, каковы его опытные основания и физические следствия.

2. Рассмотрим сначала свободное падение тел в поле тяжести Земли. По второму закону Ньютона $m^{(i)}\mathbf{a} = \mathbf{F}$, где $\mathbf{F}$ — сила тяжести. По смыслу под $m^{(i)}$ следует понимать инертную массу тела. Сила же тяжести может быть представлена в виде $\mathbf{F} = m^{(g)}\mathbf{g}$, где $m^{(g)}$ — гравитационная масса того же тела. Заметим, что сейчас мы рассматриваем движение относительно инерциальной системы отсчета и поэтому не вводим никаких сил инерции. Все силы являются «реальными» в ньютоновском смысле. В частности, сила тяжести $\mathbf{F}$ в нашем теперешнем рассмотрении есть сила только гравитационного притяжения между телом и Землей (центробежная сила в нее не входит). Второй закон Ньютона дает $m^{(i)}\mathbf{a} = m^{(g)}\mathbf{g}$, откуда

$$\mathbf{a} = \frac{m^{(g)}}{m^{(i)}}\mathbf{g}. \tag{70.2}$$

Так как инертная и гравитационная массы равны, то $\mathbf{a} = \mathbf{g}$. Все тела в поле тяжести падают с одним и тем же ускорением. Это экспериментальный факт, установленный впервые Галилеем, является подтверждением закона о равенстве инертной и гравитационной масс. Он справедлив и для любого гравитационного поля. *В одном и том же гравитационном поле все тела при свободном падении приобретают одинаковое ускорение.* Этим положением под названием обобщенного закона Галилея мы широко пользовались, начиная с § 65. Мы видим, что обобщенный закон *Галилея по своему содержанию совершенно эквивалентен принципу равенства инертной и гравитационной масс.*

Опыты Галилея имели малую точность. Значительно большей точности достигли Ньютон, а затем Бессель (1784—1846) в опытах с колебаниями маятника. Для периода малых колебаний математического маятника мы вывели формулу

$$T = 2\pi\sqrt{\frac{l}{g}}. \tag{70.3}$$

Если бы инертная и гравитационная массы не были равны между собой, то в этой формуле величину g следовало бы заменить на ускорение a, определяемое выражением (70.2). Тогда мы получили бы

$$T = 2\pi\sqrt{\frac{l}{g}\frac{m^{(i)}}{m^{(g)}}}. \tag{70.4}$$

Только при $m^{(i)} = m^{(g)}$ формула (70.4) переходит в формулу (70.3). В опытах Ньютона и Бесселя было установлено, что период колебаний математического маятника не зависит от материала, из которого он изготовлен. Это подтверждает закон равенства инертной и гравитационной масс. Относительная точность, с какой это равенство было установлено в опытах Бесселя, составляет 1/60 000.

3. Однако рекордными по точности долгое время оставались исследования венгерского физика Роланда Этвеша (1848—1919), начатые в 1887 г. и продолжавшиеся до конца его жизни. Этвеш установил равенство инертной и гравитационной масс с относительной точностью $5\cdot10^{-9}$. По сравнению с опытами Ньютона точность была повышена примерно в сто тысяч, а по сравнению с опытами Фридриха Бесселя (1784—1846) — более чем в десять тысяч раз. Идея опытов Этнеша заключается в следующем. Вес тела складывается из двух различных сил: силы гравитационного притяжения Земли и центробежной силы инерции. Первая сила пропорциональна гравитационной массе, вторая равна $m^{(i)}\omega^2\mathbf{r}_\perp$, т. е. пропорциональна инертной массе $m^{(i)}$. Если бы инертная и гравитационная массы не были строго пропорциональны друг другу, то направление отвеса зависело бы от материала тела. Опыты Этвеша имели целью обнаружение этого эффекта. С указанной выше точностью они привели к отрицательному результату, что является доказательством справедливости закона равенства инертной и гравитационной масс. Чтобы достигнуть такой точности, надо было

оценивать изменения направления отвеса в $1{,}5 \cdot 10^{-6}$ дуговой секунды. Под таким углом был бы виден земному наблюдателю предмет в 3 мм, лежащий на поверхности Луны. Такой точности Этвешу и его сотрудникам удалось достигнуть при помощи *крутильных весов* и *гравитационных вариометров*. Хотя основные опыты были выполнены с гравитационными вариометрами, но мы опишем (конечно, схематически) опыты с крутильными весами, так как в идейном отношении они более просты.

На длинной тонкой нити подвешивался стержень, к концам которого можно было прикреплять грузы *1* и *2* (рис. 194 *б*), изготовленные из различных материалов, например из платины и меди. Стержень устанавливался перпендикулярно к меридиану рассматриваемого места. Пусть $\mathbf{g}$ означает напряженность земного гравитаци-

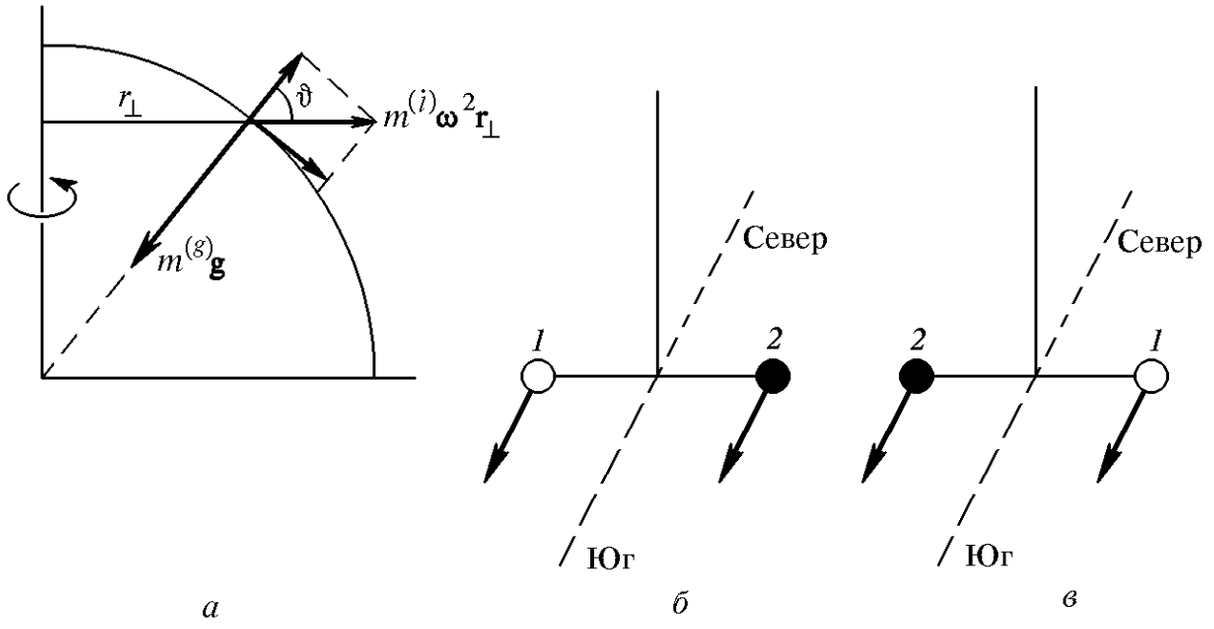


Рис. 194

онного поля, т. е. силу, с которой это поле действует на единицу гравитационной массы. На груз будут действовать две силы: гравитационная $m^{(g)}\mathbf{g}$ и центробежная $m^{(i)}\omega^2\mathbf{r}_\perp$. Последняя имеет вертикальную составляющую $m^{(i)}\omega^2\mathbf{r}_\perp \cos\vartheta$ (рис. 194 *а*), где ϑ — географическая широта рассматриваемого места. Поэтому если стержень (рычаг) равноплечий, то одно из условий равновесия грузов будет

$$m_1^{(g)}g - m_1^{(i)}\omega^2 r_\perp \cos\vartheta = m_2^{(g)}g - m_2^{(i)}\omega^2 r_\perp \cos\vartheta,$$

или

$$m_1^{(i)}(\alpha_1 g - \omega^2 r_\perp \cos\vartheta) = m_2^{(i)}(\alpha_2 g - \omega^2 r_\perp \cos\vartheta),$$

где α_1 и α_2 — отношение гравитационных масс к инертным для грузов *1* и *2* соответственно. Если бы $\alpha_1 \neq \alpha_2$, то из полученного соотношения следовало бы, что $m_1^{(i)} \neq m_2^{(i)}$. В этом случае центробежные силы, действующие на грузы, а с ними и их горизонтальные состав-

ляющие, направленные к югу (рис. 194 *б* и *в*), не были бы одинаковыми; появился бы вращающий момент

$$M_1 = (m_1^{(i)} - m_2^{(i)}) \frac{l}{2} \omega^2 r_\perp \sin \vartheta$$

(l — длина стержня), стремящийся закрутить нить. В состоянии равновесия угол кручения $\varphi_1 = (1/f)M_1$, где f — модуль кручения. Если весь прибор повернуть на 180°, т. е. перейти из положения *б* в положение *в* (рис. 194), то вращающий момент и угол кручения изменят знаки ($M_2 = -M_1$, $\varphi_2 = -\varphi_1$). При этом нить закрутится на угол $\varphi = \varphi_2 - \varphi_1 = -(2/f)M_1$. Опыт Этвеша привел к отрицательному результату, т. е. он показал, что $\varphi = 0$, каковы бы ни были материалы, из которых изготовлены грузы. Следовательно, $\alpha_1 = \alpha_2$, что и доказывает равенство инертной и гравитационной масс.

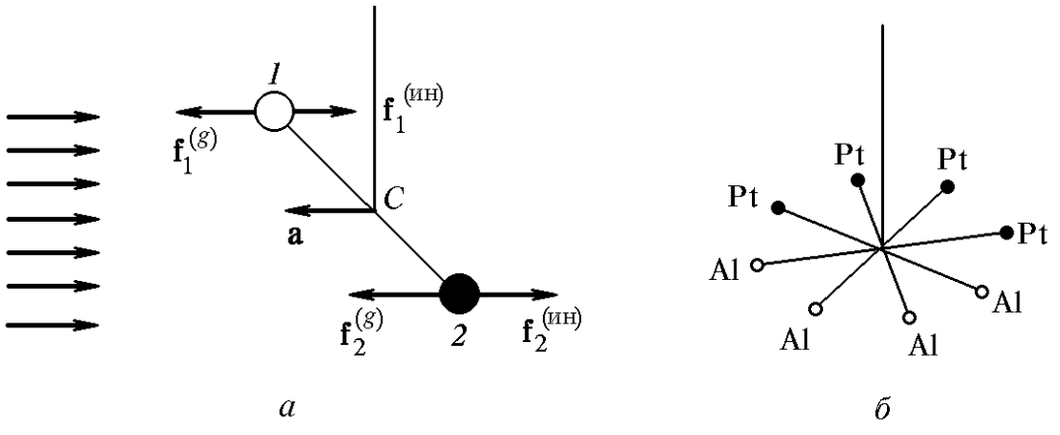
4. *Одним из фундаментальных следствий теории относительности является связь между энергией и массой* $E = mc^2$. Здесь m означает инертную массу. Таким образом, *всякая энергия обладает инертной массой*. Закон эквивалентности инертной и гравитационной масс позволяет распространить это утверждение и на гравитационную массу. *Всякая энергия должна обладать также и гравитационной массой*. Высокая чувствительность опыта Этвеша позволила подвергнуть это заключение экспериментальной проверке. С этой целью Саузернс повторил опыт Этвеша с радиоактивными веществами. Опыт дал тот же результат: никакого различия между гравитационной и инертной массами обнаружено не было. Так как при радиоактивных превращениях энергия и инертная масса уменьшаются, то отсюда следует, что пропорционально уменьшается также и гравитационная масса. Таким образом, равенство инертной и гравитационной масс все время соблюдается.

5. Опыт Этвеша в усовершенствованном виде был повторен американским физиком Робертом Дикке (р. 1916) и его сотрудниками в 1961—1964 гг. Им удалось повысить точность результатов Этвеша более чем в 100 раз. Сравнивались грузы из меди и свинца, из золота и алюминия. С относительной точностью $3 \cdot 10^{-11}$ авторы констатировали равенство коэффициентов пропорциональности между гравитационной и инертной массами для этих материалов.

В идейном отношении опыт Дикке проще опыта Этвеша. В опыте Этвеша речь шла об эффектах, определяющихся совокупным действием гравитационного притяжения Земли и сил инерции, возникающих из-за ее осевого вращения. В опытах Дикке вместо Земли использовалось Солнце. Сравниваемые грузы *1* и *2* по-прежнему закреплялись на концах прямолинейного коромысла, подвешенного на тонкой нити (рис. 195 *а*). Для максимального уменьшения влияния посторонних возмущающих факторов это устройство помещалось в сосуд с высоким вакуумом. Прибор устанавливался в глубокой термостатированной шахте, удаленной от зданий. После установки

прибора шахта запечатывалась, а прибор контролировался дистанционно из удаленной приборной будки на протяжении нескольких месяцев подряд.

Сила гравитационного притяжения Земли и центробежная сила, возникающая из-за вращения Земли вокруг своего центра, в опытах Дикке принципиальной роли не играют. От этих сил можно отвлечься. Они в рассматриваемой точке земного шара постоянны и определяют лишь положение равновесия, в котором стремится установиться коромысло. Для опыта имеют значение сила гравитационного притяжения Солнца и поступательная сила инерции, связанная с ускоренным движением центра Земли по направлению к Солнцу (влиянием Луны можно пренебречь). Обозначим это ускорение через **a**. По самому определению гравитационной массы сила гравитационного притяжения Солнца, отнесенная к единице такой массы, для всех тел одна и та же. Это есть напряженность гравитационного поля Солнца, зависящая только от самого Солнца. Обозначим ее через **g**. Но если бы нарушался закон эквивалентности инертной и гравитационной масс, то сила гравитационного притяжения Солнца, отнесенная к единице инертной массы, была бы разной для различных тел. В этом случае возник бы вра-



а *б*

Рис. 195

щающийся момент, стремящийся закрутить нить, на которой подвешено коромысло. Если h_1 и h_2 — плечи коромысла, а последнее подвешено за центр масс C (рис. 195 *а*), то вращающий момент относительно точки C будет

$$M = (m_1^{(g)}g - m_1^{(i)}a)h_1 + (m_2^{(i)}a - m_2^{(g)}g)h_2.$$

По определению центра масс (точнее, следовало бы сказать центра инертных масс) $m_1^{(i)}h_1 = m_2^{(i)}h_2$. Поэтому, используя ранее введенные обозначения α_1 и α_2, получим

$$M = m_1^{(i)}h_1g\alpha_1 - m_2^{(i)}h_2g\alpha_2 = m_1^{(i)}h_1g(\alpha_1 - \alpha_2).$$

Плечи коромысла периодически меняются из-за видимого движения Солнца по небесному своду. Поэтому момент M также будет пери-

одически изменяться и притом с периодом в одни сутки. В результате возникли бы вынужденные колебания коромысла с таким же периодом, которые можно было бы обнаружить с помощью чувствительной аппаратуры. На фоне неизбежных случайных толчков, которым подвержена система, такие колебания обнаружены не были. Отсюда следует, что в пределах ошибок измерений $\alpha_1 = \alpha_2$, т. е. соблюдается закон эквивалентности.

6. Опыт Дикке был повторен в усовершенствованном виде В. Б. Брагинским и В. И. Пановым в 1971 г. Вместо одного коромысла применялся крутильный маятник, эквивалентный четырем коромыслам, соединенным вместе, как указано на рис. 195 *б*. Точность опыта была повышена еще примерно в 30 раз. Сравнивались платина и алюминий. Равенство коэффициентов пропорциональности между гравитационной и инертной массами для этих веществ было подтверждено с относительной точностью 10^{-12}. Это то же самое, как если бы мы взвесили корабль водоизмещением в десять тысяч тонн вместе с грузом с точностью до одной сотой грамма.

7. Дорелятивистская физика не придавала существенного значения равенству инертной и гравитационной масс, рассматривая это равенство как случайное совпадение. Основополагающее значение закона эквивалентности инертной и гравитационной масс было принято Эйнштейном. Закон эквивалентности послужил для Эйнштейна отправным пунктом при построении *общей теории относительности*, называемой иначе *релятивистской теорией гравитации*. Этот закон является главным опытным фактом, на котором основана общая теория относительности. Последняя была бы неверна и от нее следовало бы отказаться, если бы было обнаружено малейшее нарушение закона эквивалентности инертной и гравитационной масс. Вот почему повышение и без того исключительной точности, с которой проверяется этот закон, имеет важное принципиальное значение, а не является просто спортивным увлечением с целью побития рекорда и установления нового.

§ 71. ПРИНЦИП ЭКВИВАЛЕНТНОСТИ ГРАВИТАЦИОННЫХ СИЛ И СИЛ ИНЕРЦИИ

1. Мы уже неоднократно отмечали, что все тела, независимо от их масс и химического состава, получают в данном гравитационном поле одинаковые ускорения. Поэтому в таком поле они движутся совершенно одинаково, если только одинаковы начальные условия. Тем же свойством обладают свободно движущиеся тела, если их движение рассматривать относительно какой-либо неинерциальной системы отсчета. Иначе говоря, указанным свойством обладают также силы инерции. Эта аналогия между силами тяготениями и силами инерции явилась отправной точкой при построе-

нии общей теории относительности, или релятивистской теории гравитации, Эйнштейна.

Рассмотрим по примеру Эйнштейна, что происходит в движущемся лифте. Допустим сначала, что лифт неподвижно висит на тросе или движется равномерно относительно Земли. Все тела в лифте подвергаются действию земного поля тяготения. Пассажир в лифте ощущает вес собственного тела, оказывает давление на пол лифта, подвергается со стороны поля равному и противоположно направленному противодавлению. Груз, подвешенный на пружине, растягивает ее силою своего веса. Все тела, предоставленные самим себе, свободно падают относительно лифта с одним и тем же ускорением $\mathbf{g}$ и т. д.

Вообразим теперь, что лифт настолько удален от Земли и прочих небесных тел, что он практически не подвергается с их стороны никаким гравитационным воздействиям. Пусть кто-то тянет за трос лифта, сообщая последнему постоянное ускорение $\mathbf{a} = -\mathbf{g}$. Гравитационного поля в лифте нет, зато есть сила инерции $-m\mathbf{a} = m\mathbf{g}$. Под действием таких сил все тела в лифте, если их ничем не удерживать, начнут «падать» с прежним ускорением $\mathbf{g}$. Груз, подвешенный на пружине, растянет ее, как если бы она обладала весом $m\mathbf{g}$. Пассажир в лифте будет оказывать на пол такое же давление, как и в предыдущем случае. Короче говоря, *все механические явления и движения в лифте будут в точности такими же, что и в неподвижном лифте, висящем в поле тяжести.* Эйнштейн распространил это утверждение не только на механические, но и на любые *физические явления*, как это он сделал также с галилеевским принципом относительности. Для такой гипотезы имеются веские основания. *В природе нет чисто механических явлений.* В основе каждого «механического» явления лежит громадное множество разнообразных *других явлений*, относящихся к различным разделам физики. Так, столкновение бильярдных шаров обычно рассматривают как типично механическое явление. Но существование самих шаров и их внутренняя структура определяются *квантовыми законами*, а упругие силы, развивающиеся во время столкновения, сводятся к силам *электростатического взаимодействия* заряженных частиц, из которых построены тела.

Итак, *все физические явления в равномерно ускоренном лифте будут происходить в точности так же, как и в неподвижном лифте, висящем в однородном поле тяжести.* Между тем дорелятивистская физика рассматривала оба случая как существенно разные. В первом случае явления объяснялись действием гравитационного поля, во втором — действием сил инерции. В первом случае лифт является инерциальной системой отсчета, в которой есть однородное поле тяготения. Во втором случае поля тяготения нет, зато есть силы инерции, так как лифт является инерциальной системой отсчета.

Если лифт в однородном поле тяжести движется вверх или вниз с ускорением $\mathbf{a}$, то на тело в лифте действует сила тяжести $m\mathbf{g}$ и

сила инерции $-m\mathbf{a}$. Результирующая сила $m(\mathbf{g}-\mathbf{a})$ состоит из двух слагаемых, совершенно различных по своей физической природе. Между тем все явления внутри лифта будут происходить так, как если бы в нем действовало однородное гравитационное поле с напряженностью $\mathbf{g}'=\mathbf{g}-\mathbf{a}$. В частности, когда лифт падает свободно, $\mathbf{g}'=0$, т. е. наступает «состояние невесомости». Допустим, что пассажир в лифте имеет возможность производить опыты только над телами внутри лифта и лишен возможности наблюдать внешний мир. Замечая, что все тела падают в лифте с одним и тем же ускорением, он не может на основании одного только этого факта решить, чем вызвано это ускорение: однородным гравитационным полем, ускоренным поступательным движением самого лифта, или, наконец, и тем и другим. Никакие опыты по свободному падению тел в лифте не могут отделить однородное гравитационное поле от однородного поля сил инерции. По предположению Эйнштейна, это невозможно сделать и с помощью любых физических опытов. Это предположение Эйнштейн возвел в постулат и выдвинул *принцип эквивалентности гравитационных сил и сил инерции.*

Согласно этому принципу *все физические явления в гравитационном поле происходят совершенно так же, как и в соответствующем поле сил инерции, если напряженности обоих полей в соответствующих точках пространства совпадают, а начальные условия одинаковы для всех тел замкнутой системы.*

2. *Принцип эквивалентности вовсе не утверждает, что всякое гравитационное поле может быть имитировано силами инерции, т. е. создано надлежащим ускоренным движением системы отсчета. Он не утверждает также, что любые силы инерции во всем пространстве можно заменить гравитационными. Оба эти утверждения верны, вообще говоря, только для однородных полей, т. е. таких полей, напряженность которых одна и та же во всех точках пространства.* Для пояснения этого вернемся к прежнему примеру с лифтом. Пусть лифт неподвижно висит в поле тяжести Земли. Располагая точным гравитометром, пассажир в лифте заметит, что направления отвеса в различных точках кабины лифта не совсем параллельны, их продолжения пересекаются приблизительно в центре Земли. Далее, он найдет, что земное гравитационное поле возрастает в направлении к центру Земли. Короче, пассажир в неподвижном лифте может установить, что земное гравитационное поле *неоднородно.* Напротив, поле сил инерции, возникающее в лифте при его ускоренном поступательном движении, *однородно.* Оно не может во всех точках пространства внутри кабины лифта подменить неоднородное ньютоново поле земной тяжести. Можно создать внутри кабины лифта и неоднородное поле сил инерции, приведя лифт во вращение. Однако такое поле возрастало бы при удалении от оси вращения, т. е. вело бы себя совсем иначе, чем гравитационное поле Земли. Ньютоново гравитационное поле точечной массы убывает обратно пропорционально квадрату расстояния от

нее. Поле центробежных сил $\omega^2\mathbf{r}_\perp$, напротив, возрастает пропорционально расстоянию до оси вращения. Ясно, что *ньютоново гравитационное поле точечной массы не может быть получено никаким вращением системы отсчета.*

Однако в небольших объемах пространства, в которых гравитационное поле может считаться практически однородным, оно может быть приближенно имитировано ускоренным движением системы отсчета. Если хотят отметить это обстоятельство, то говорят, что принцип эквивалентности имеет *локальный характер.*

3. Кроме того, между ньютоновыми силами гравитационного притяжения и силами инерции имеется существенное различие, когда последние возникают во вращающихся системах отсчета. *Ньютоновы гравитационные силы не зависят от скоростей тел, на которые они действуют. Тем же свойством обладают поступательные и центробежные силы инерции, а также вообще все переносные силы инерции.* По своим физическим действиям переносные силы инерции совершенно эквивалентны ньютоновым гравитационным силам. Невозможно однозначно отделить ньютоно гравитационное поле от поля переносных сил инерции. Напротив, *силы Кориолиса ведут себя существенно иначе, чем ньютоновы гравитационные силы. На покоящиеся* (в рассматриваемой системе отсчета) *тела они не действуют. Они возникают только при движении тела и пропорциональны его скорости.* Тем не менее эквивалентность инертной и гравитационной масс делает целесообразным объединить гравитационное поле и поле всех сил инерции в *единое поле.* Это и делается в общей теории относительности. Для поля, получающегося в результате такого объединения, сохранено прежнее название — *гравитационное поле.* Сила инерции является частным случаем сил гравитационного поля, понимаемого в таком расширенном смысле. Общая теория относительности, или релятивистская теория гравитации, устанавливает уравнения гравитационного поля. Они называются *уравнениями Эйнштейна.* Закон всемирного тяготения Ньютона содержится в уравнениях Эйнштейна и верен только приближенно. Приближенный характер закона всемирного тяготения, впрочем, следует уже из того, что в основе этого закона лежит представление о мгновенном распространении взаимодействий. А такое представление имеет ограниченную область применимости.

4. В свете изложенного вернемся еще раз к вопросу об инерциальных системах отсчета. Пусть тело A настолько удалено от Солнечной системы, что ее гравитационным полем можно пренебречь. Тогда еще нельзя утверждать, что оно не подвержено действию никаких гравитационных полей. Мы не можем утверждать, что во Вселенной нет удаленных тел, создающих в месте нахождения тела A гравитационное поле $\mathbf{g}$ конечной напряженности. Убывание гравитационного поля из-за расстояния до этих тел может быть компенсировано возрастанием их масс. Однако если изучаются явления в

ограниченной области пространства S, то при не слишком больших размерах ее поле $\mathbf{g}$ может считаться однородным. Тогда если тело A свободно падает в гравитационном поле $\mathbf{g}$, то это поле будет *полностью компенсировано поступательными силами инерции.* Если тело A и не вращается (относительно удаленных масс), то оно не будет подвержено действию и остальных сил инерции. Система отсчета, связанная с таким невращающимся свободно падающим телом A, и будет инерциальной системой отсчета. Во всякой системе отсчета A', вращающейся или движущейся ускоренно относительно системы A появятся силы инерции. Но это движение есть не движение в «абсолютном пространстве», а *движение относительно удаленных тел Вселенной.* С этой точки зрения, принадлежащей австрийскому физику Эрнесту Маху (1838—1916), силы инерции возникают из-за вращений и ускоренных движений координатных систем относительно удаленных тел Вселенной. Это утверждение известно под названием *принципа Маха.*

Одно время точка зрения Маха казалась очень привлекательной. Ее в принципе разделял в первоначальных работах Эйнштейн. Однако в дальнейшем он от нее отошел. В современных космологических теориях принцип Маха не используется. Ввиду конечной скорости распространения взаимодействий всякое движение или изменение удаленных тел Вселенной может повлиять на процессы вблизи Земли не раньше чем через время, затрачиваемое светом на прохождение расстояния от этих тел до Земли. Принцип Маха этого не учитывает. Он не принимает во внимание силовые поля, существующие в галактическом и межгалактическом пространствах, а также наличие открытого в 1965 г. реликтового радиоизлучения, т. е. равновесного излучения, возникшего на ранней стадии развития Вселенной, а затем остывшего до температуры около 2,7 К и увеличившего длины волн из-за космологического расширения Вселенной. Здесь преждевременно обсуждать эти сложные и далеко еще не решенные вопросы.

§ 72. ГРАВИТАЦИОННОЕ СМЕЩЕНИЕ СПЕКТРАЛЬНЫХ ЛИНИЙ

1. В качестве примера применения принципа эквивалентности гравитационных сил и сил инерции рассмотрим явление *гравитационного смещения спектральных линий,* теоретически предсказанное Эйнштейном. Будем исходить из представления, что свет есть волны, которые в вакууме распространяются со скоростью $c \approx 300\,000$ км/с. Свет определенной спектральной линии характеризуется определенной частотой, или числом колебаний в секунду, которую мы будем обозначать через ν. Такой свет называется *монохроматическим,* т. е. *одноцветным.* Пусть монохроматический свет приходит к нам от какого-либо удаленного источника, причем в пространстве, через которое он распространяется, гравитационного поля нет. Обозначим через через ν_0 частоту световой волны, которую воспри-

нимает наблюдатель, покоящийся в какой-либо инерциальной системе отсчета. Если наблюдатель начнет двигаться навстречу световым лучам с постоянным ускорением a (рис. 196 $а$), то частота воспринимаемого света увеличится (эффект Доплера).

Простой расчет показывает, что с точностью до членов порядка $(v/c)^2$ относительное изменение воспринимаемой частоты определяется формулой

$$\frac{\nu-\nu_0}{\nu_0}=\frac{v}{c},$$

где v — скорость наблюдателя. За положительные направления $\mathbf{v}$ и $\mathbf{a}$ мы принимаем направления против распространения света. Если наблюдатель двигался в течение времени t, то $\mathbf{v}=\mathbf{a}t$. За это время свет проходит расстояние $l=ct=cv/a$, а потому изменение частоты за то же время определится формулой

$$\frac{\nu-\nu_0}{\nu_0}=\frac{al}{c^2}.$$

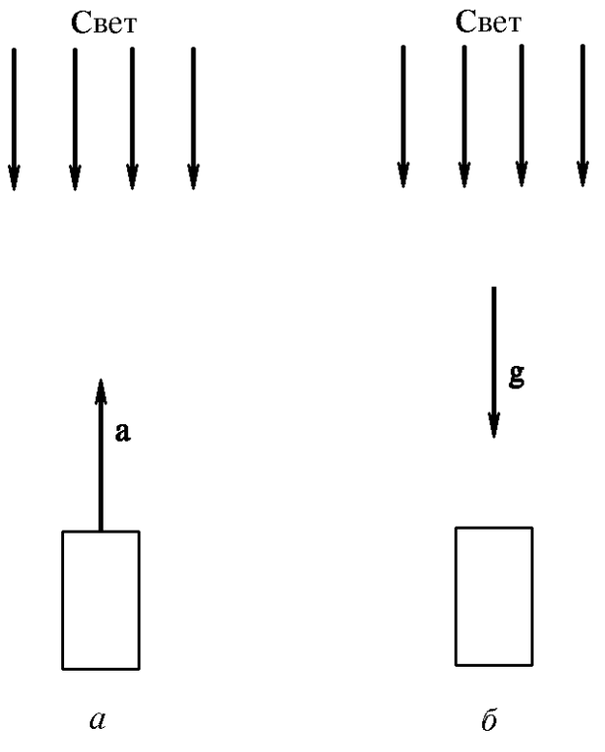


Рис. 196

2. Допустим теперь, что наблюдатель в инерциальной системе отсчета неподвижен, но в ней имеется однородное гравитационное поле с напряженностью $\mathbf{g}$ (рис. 196 $б$). Если величину $\mathbf{g}$ подобрать равной $-\mathbf{a}$ ($\mathbf{g}=-\mathbf{a}$), то по принципу эквивалентности гравитационное поле вызовет в точности такое же изменение частоты света, что и в предыдущем случае. *При распространении света по направлению гравитационного поля* $\mathbf{g}$ *частота световой волны будет возрастать, а при распространении в противоположном направлении — убывать.* В этом и состоит явление гравитационного смещения спектральных линий, предсказанное Эйнштейном. Значение смещения определяется формулой

$$\frac{\nu-\nu_0}{\nu_0}=\frac{gl}{c^2}, \tag{72.1}$$

где l — расстояние, проходимое светом в поле тяготения.

При выводе формулы (72.1) предполагалось, что поле постоянно и однородно. Результат нетрудно обобщить на случай произвольного постоянного неоднородного гравитационного поля. С этой целью разобьем путь светового луча на бесконечно малые участки $d\mathbf{r}$. На протяжении каждого из таких участков гравитационное поле может считаться однородным. Если $d\nu$ — изменение частоты светового луча при прохождении участка $d\mathbf{r}$, то по формуле (72.1)

$$\frac{d\nu}{\nu}=\frac{\mathbf{g}\,d\mathbf{r}}{c^2},$$

так как составляющая вектора $\mathbf{g}$, перпендикулярная к направлению распространения света, на изменение частоты не оказывает влияния. Если свет проходит конечный путь из начального положения 1 в конечное положение 2, то изменение частоты на этом пути найдется интегрированием полученного выражения, т. е. по формуле

$$\ln \frac{\nu_2}{\nu_1} = \frac{1}{c^2} \int \mathbf{g}\, d\mathbf{r}. \tag{72.2}$$

Интегрирование не обязательно проводить вдоль пути, по которому распространяется свет. Можно взять произвольный путь, соединяющий начальную точку 1 с конечной точкой 2. Гравитационные силы постоянных полей являются силами консервативными, так что интеграл от формы пути не зависит. Интеграл имеет смысл работы, которую совершили бы силы гравитационного поля над единичной массой при ее перемещении из положении 1 в положение 2. Эта работа называется *разностью гравитационных потенциалов* $\varphi_1 - \varphi_2$ между точками 1 и 2. В этих обозначениях

$$\ln \frac{\nu_2}{\nu_1} = \frac{\varphi_1 - \varphi_2}{c^2}. \tag{72.3}$$

При малой разности потенциалов, когда

$$|\varphi_1 - \varphi_2| \ll c^2,$$

формула переходит в

$$\frac{\nu_2 - \nu_1}{\nu_1} = \frac{\varphi_1 - \varphi_2}{c^2}. \tag{72.4}$$

При распространении света от высшего гравитационного потенциала к низшему его частота увеличивается, при распространении в противоположном направлении — уменьшается.

В 1960 г. американским физикам Роберту Паунду (р. 1919) и Дж. Ребке удалось с уверенностью наблюдать с использованием так называемого *эффекта Мёссбауэра* (по имени немецкого физика—экспериментатора Рудольфа Мессбауэра (р. 1929 г.)) гравитационное смещение спектральных линий при распространении света даже в поле тяжести Земли. Проходимый путь (сверху вниз) составляет всего 20 м. В этом случае ожидаемое смещение $(\nu - \nu_0)/\nu_0 \approx 2 \cdot 10^{-14}$.

Измерения дали такой же результат. Это является подтверждением принципа эквивалентности гравитационных сил и сил инерции.

ГЛАВА X

МЕХАНИКА УПРУГИХ ТЕЛ

§ 73. ИДЕАЛЬНО УПРУГИЕ ТЕЛА

1. Все реальные тела *деформируемы*. Под действием приложенных сил они меняют свою форму или объем. Такие изменения называются *деформациями*. В случае твердых тел различают два предельных случая: *деформации упругие и деформации пластические. Упругими* называют деформации, исчезающие после прекращения действия приложенных сил. *Пластическими* или *остаточными деформациями* называют такие деформации, которые сохраняются в теле, по крайней мере частично, и после прекращения действия внешних приложенных сил. На пластических деформациях основана *холодная обработка металлов* — штамповка, ковка и пр. Является ли деформация упругой или пластической зависит не только от материала тела, но и от *приложенных сил.* Если сила (точнее, сила, отнесенная к единице площади, т. е. напряжение) не превосходит известной величины, называемой *пределом упругости*, то возникающая деформация будет упругой. Если же она превосходит этот предел, то возникающая деформация будет пластической. Предел упругости имеет различные значения для разных материалов. Она является не вполне четкой определенной величиной. Разделение тел на упругие и пластические также в какой-то степени условно. Строго говоря, все деформации после прекращения действия внешних сил исчезают не полностью, а поэтому являются пластическими. Однако если остаточные деформации малы, то во многих случаях их можно не принимать во внимание. Как велика должна быть остаточная деформация, чтобы можно было так поступать, зависит от конкретных условий. В некоторых случаях, например, можно пренебречь остаточными деформациями, если они не превосходят 0,1 % от максимальных значений, достигавшихся под действием приложенных сил. В других случаях этот предел должен быть снижен до 0,01 % и т. д.

2. В настоящей главе мы ограничимся изучением только упругих деформаций. При этом мы остановимся только на механике, но не на физике явлений. Механика описывает упругие свойства тел посредством некоторых эмпирически вводимых *упругих постоянных*, различных для различных тел и зависящих от их физического состояния (например, от температуры). Более глубоким является физический подход, рассматривающий явление деформаций с атомистической точки зрения. Этим занимается *теория твердого тела.* Она позволяет в принципе не только вывести основные уравнения механики деформируемых тел с атомистической точки зрения, но и

установить связь между упругими постоянными вещества и другими его физическими свойствами.

Тела мы будем считать *идеально упругими*. Так называются идеализированные тела, которые могут претерпевать только упругие, но не пластические деформации. Такими идеализациями можно пользоваться, когда силы, приложенные к реальным телам, не превосходят предела упругости. *Для идеально упругих тел существует однозначная зависимость между действующими силами и вызываемыми ими деформациями*. В случае пластических деформаций такой однозначной связи не существует. Это видно хотя бы из того, что до и после пластической деформации тело имеет различную форму, хотя в обоих случаях оно не подвергается действию внешних сил. Мы ограничимся изучением только *малых деформаций*. *Малыми* называются упругие деформации, подчиняющиеся закону Гука. Это приближенный закон, согласно которому *деформации пропорциональны силам, их вызывающим*.

3. Твердые тела разделяются на *изотропные* и *анизотропные*. *Изотропными* называются тела, свойства которых по всем направлениям одинаковы. *Анизотропными* называются тела, свойства которых в разных направлениях не одинаковы. Типичными представителями анизотропных тел являются *кристаллы*. Приведенные определения отличаются некоторой неопределенностью, поскольку в них явно не указано, о каких физических свойствах идет речь. Дело в том, что *тела могут вести себя как изотропные по отношению к одним свойствам и как анизотропные — по отношению к другим*. Так, все кристаллы *кубической системы* ведут себя как изотропные, если речь идет о распространении света в них. Однако они будут анизотропными, если интересоваться их упругими свойствами. В настоящей главе нас интересует изотропия или анизотропия тел по отношению к их упругим свойствам. Но мы ограничимся простейшим случаем, когда тела являются изотропными. Металлы обычно имеют *поликристаллическую структуру*, т. е. состоят из мельчайших беспорядочно ориентированных кристалликов. Каждый из таких кристалликов есть тело анизотропное. Но кусочек металла, содержащий множество их, ведет себя как изотропное тело, если всевозможные ориентации кристалликов представлены с одинаковой вероятностью. В результате пластической деформации хаотичность в ориентации кристалликов может нарушиться. Тогда после пластической деформации металл становится анизотропным. Такое явление наблюдается, например, при вытягивании или кручении проволоки.

§ 74. УПРУГИЕ НАПРЯЖЕНИЯ

1. Различные части деформированного тела взаимодействуют между собой на поверхностях раздела, вдоль которых они граничат

друг с другом. Рассмотрим произвольное деформированное тело или среду. Мысленно разделим его на две части: тело I и тело II, граничащие между собой вдоль поверхности AB (рис. 197). Так как тело I деформировано, то оно действует на тело II с некоторой силой. По той же причине тело II действует на тело I с такой же, но противоположно направленной силой. Однако для определения возникающих деформаций недостаточно знать суммарные силы, действующие в сечении AB. Надо еще указать, как эти силы распределены по этому сечению. Возьмем на поверхности AB бесконечно малую площадку dS. Пусть $d\mathbf{F}$ — сила, с которой на этой площадке тело II действует на тело I. *Сила, отнесенная к единице площади*, т. е. $\frac{d\mathbf{F}}{dS}$, *называется напряжением*, действующим в соответствующей точке на границе AB тела I. Напряжение, действующее в той же точке на границе тела II, будет таким же, но его направление противоположно.

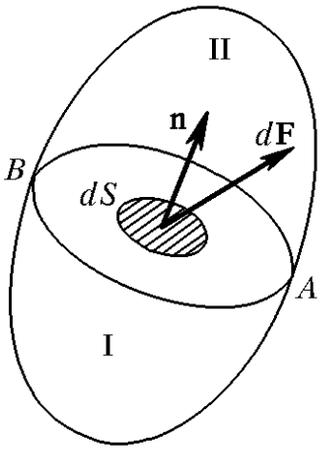


Рис. 197

2. Ориентацию площадки dS можно задать, указав направление нормали к ней. Условимся эту нормаль проводить *наружу* от поверхности тела, на которое действует сила $d\mathbf{F}$. Обозначим через $\mathbf{n}$ единичный вектор такой нормали, а через σ_n — соответствующее напряжение. Тогда $\boldsymbol{\sigma}_{-n}$ будет означать напряжение на поверхности AB тела II, с которым граничит тело I. В силу равенства действия и противодействия $\boldsymbol{\sigma}_n = -\boldsymbol{\sigma}_{-n}$. Вектор $\boldsymbol{\sigma}_n$ можно разложить на составляющую вдоль нормали $\mathbf{n}$ и составляющую, лежащую в касательной плоскости к площадке dS. Первая составляющая называется *нормальным*, а вторая — *тангенциальным* напряжениями, действующими на площадке dS. Как и всякий вектор, напряжение $\boldsymbol{\sigma}_n$ можно характеризовать тремя составляющими его вдоль координатных осей X, Y, Z прямоугольной системы координат. Эти составляющие будем обозначать соответственно через σ_{nx}, σ_{ny}, σ_{nz}. Первый индекс указывает направление внешней нормали к поверхности тела, на которой лежит площадка dS, а второй — направление оси, на которую проецируется напряжение $\boldsymbol{\sigma}_n$. В частности, $\boldsymbol{\sigma}_x$ означает напряжение на площадке, внешняя нормаль к которой параллельна положительному направлению оси X. Величины σ_{xx}, σ_{xy}, σ_{xz} означают проекции вектора $\boldsymbol{\sigma}_x$ на координатные оси.

3. Для того чтобы определить напряжение в среде на произвольно ориентированной площадке в какой-либо точке ее, достаточно задать напряжение на трех взаимно перпендикулярных площадках, проходящих через эту точку. Это справедливо как для покоящейся среды, так и для среды, движущейся с произвольным ускорением. Для доказательства поместим начало координат в рассматриваемую

точку среды и выделим из нее бесконечно малый элемент объема $OABC$, ограниченный координатными плоскостями и пересекающей их плоскостью ABC (рис. 198). Пусть $\mathbf{n}$ — внешняя нормаль к плоскости треугольника ABC. Тогда сила, действующая на грани ABC на выделенный элемент со стороны окружающей среды, будет $\boldsymbol{\sigma}_n S$, где S — площадь этой грани. Аналогично силы, действующие на трех боковых гранях, будут $\boldsymbol{\sigma}_{-x} S_x$, $\boldsymbol{\sigma}_{-y} S_y$, $\boldsymbol{\sigma}_{-z} S_z$, где S_x, S_y, S_z — площади этих граней. Помимо этих сил на выделенный элемент могут действовать *массовые* или *объемные* силы, например сила тяжести. Обозначим равнодействующую таких сил через $\mathbf{f}$. Сила $\mathbf{f}$ пропорциональна объему выделенного элемента. Если масса элемента m, а ускорение $\mathbf{a}$, то

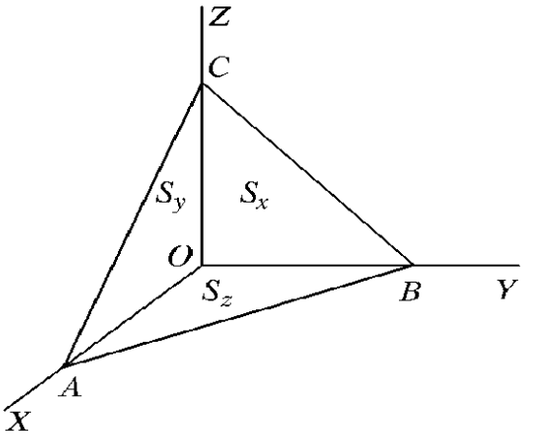


Рис. 198

$$m\mathbf{a} = \mathbf{f} + \boldsymbol{\sigma}_n S + \boldsymbol{\sigma}_{-x} S_x + \boldsymbol{\sigma}_{-y} S_y + \boldsymbol{\sigma}_{-z} S_z.$$

Выполним в этом соотношении предельный переход, стягивая элемент $OABC$ в точку. При таком предельном переходе члены $m\mathbf{a}$ и $\mathbf{f}$ можно отбросить. Они пропорциональны объему элемента $OABC$ и, следовательно, являются бесконечно малыми *высшего порядка* по сравнению с остальными членами, пропорциональными поверхности элемента. Как известно из геометрии, проекции площади S на координатные плоскости выражаются соотношениями

$$S_x = Sn_x, \quad S_y = Sn_y, \quad S_z = Sn_z.$$

Учтем далее, что $\boldsymbol{\sigma}_{-x} = -\boldsymbol{\sigma}_x$, $\boldsymbol{\sigma}_{-y} = -\boldsymbol{\sigma}_y$, $\boldsymbol{\sigma}_{-z} = -\boldsymbol{\sigma}_z$. Тогда в результате предельного перехода получится

$$\boldsymbol{\sigma}_n = \boldsymbol{\sigma}_x n_x + \boldsymbol{\sigma}_y n_y + \boldsymbol{\sigma}_z n_z. \tag{74.1}$$

Так как координатные оси X, Y, Z можно выбрать произвольно, то это соотношение и доказывает теорему.

Таким образом, *напряжение в каждой точке упруго деформированного тела можно характеризовать тремя векторами* $\boldsymbol{\sigma}_x$, $\boldsymbol{\sigma}_y$, $\boldsymbol{\sigma}_z$ *или девятью их проекциями*

$$\begin{matrix} \sigma_{xx}, & \sigma_{xy}, & \sigma_{xz}, \\ \sigma_{yx}, & \sigma_{yy}, & \sigma_{yz}, \\ \sigma_{zx}, & \sigma_{zy}, & \sigma_{zz}. \end{matrix} \tag{74.2}$$

Совокупность этих девяти величин называется *тензором упругих напряжений.*

Вообще говоря, эти величины меняются от точки к точке среды, т. е. являются функциями координат. Только в статике в отсутствие массовых сил тензор упругих напряжений остается одним и тем же во всех точках среды.

4. *Тензор упругих напряжений симметричен*, т. е.

$$\sigma_{ij} = \sigma_{ji} \quad (i, j = x, y, z). \tag{74.3}$$

Для доказательства рассмотрим элементарный параллелепипед вещества со сторонами dx, dy, dz (рис. 199). Момент сил M_z относительно оси Z, действующий на этот параллелепипед, равен

$$M_z = (\sigma_{xy}\, dy\, dz)\, dx - (\sigma_{yx}\, dx\, dz)\, dy = (\sigma_{xy} - \sigma_{yx})\, dV,$$

где dV — объем рассматриваемого элементарного параллелепипеда. По уравнению моментов

$$(\sigma_{xy} - \sigma_{yx})\, dV = I_z \frac{d\omega_z}{dt},$$

где I_z и ω_z — момент инерции и угловая скорость относительно оси Z. Но момент инерции I_z пропорционален произведению массы на квадрат линейных размеров рассматриваемого параллелепипеда, т. е. является бесконечно малой величиной *более высокого порядка*, чем объем параллелепипеда dV. Поэтому при стягивании параллелепипеда в точку правая часть $I_z \frac{d\omega_z}{dt}$ будет быстрее обращаться в нуль, чем левая. В пределе мы получим $\sigma_{xy} = \sigma_{yx}$. Аналогично доказываются и остальные два соотношения: $\sigma_{yz} = \sigma_{zy}$ и $\sigma_{zx} = \sigma_{xz}$.

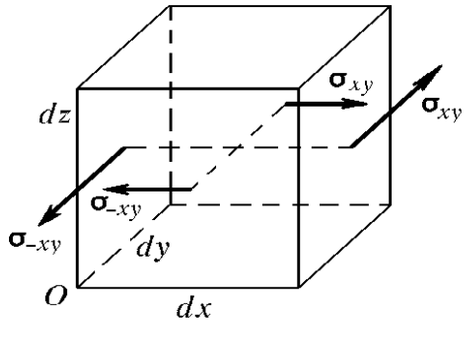


Рис. 199

5. Можно доказать, что *координатную систему* X, Y, Z *можно выбрать так, чтобы в этой системе обратить в нуль все недиагональные элементы тензора упругих напряжений, т. е.* $\sigma_{ij} = 0$ при $i \neq j$. Не останавливаясь на доказательстве, заметим только, что это можно сделать потому, что тензор упругих напряжений σ_{ij} симметричен. Таким образом, в этой системе координат упругие напряжения в каждой точке тела характеризуются только величинами σ_{xx}, σ_{yy} и σ_{zz}. В целях краткости их можно обозначать с помощью одного индекса, т. е. σ_x, σ_y и σ_z. Соответствующие координатные оси называют *главными осями* тензора упругих напряжений.

§ 75. РАСТЯЖЕНИЕ И СЖАТИЕ СТЕРЖНЕЙ

1. Возьмем однородный стержень и приложим к его основаниям растягивающие или сжимающие силы F (рис. 200 *а* и *б*). Стержень будет деформирован, т. е. растянут или сжат. Мысленно проведем произвольное сечение C, перпендикулярное к оси стержня. Для равновесия стержня AC необходимо, чтобы на его нижнее основание C действовала сила $F_1 = F$. Это есть сила, с которой нижняя часть стержня BC тянет верхнюю или давит на нее. Такая сила возникает потому, что нижняя часть стержня деформирована. Верхняя часть стержня также деформирована и действует на нижнюю с силой, равной F_1 и противоположно направленной. Такие силы действуют в любом поперечном сечении растянутого или сжатого стержня. Таким образом, деформация стержня связана с возникновением упругих сил, с которыми каждая часть стержня действует на другую, с которой она граничит. Силу, отнесенную к единице площади поперечного сечения стержня, мы назвали напряжением. В рассматриваемом случае напряжение перпендикулярно к поперечному сечению стержня. *Если стержень растянут, то это напряжение называется натяжением* и определяется выражением

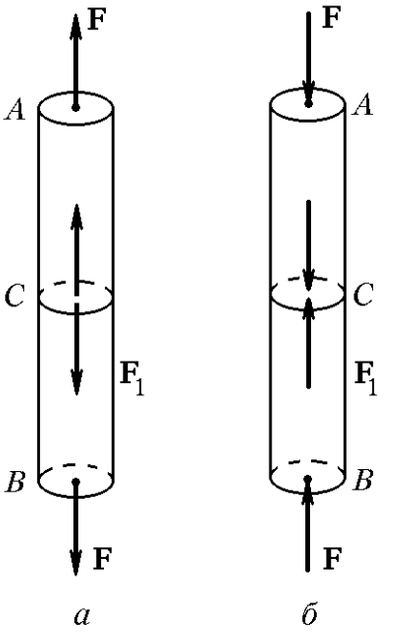


Рис. 200

$$T = \frac{F}{S}, \tag{75.1}$$

где S — площадь поперечного сечения стержня. *Если же стержень сжат, то напряжение называется давлением* и численно определяется той же формулой

$$P = \frac{F}{S}. \tag{75.2}$$

Давление можно рассматривать как отрицательное натяжение и наоборот, т. е.

$$P = -T. \tag{75.3}$$

Это замечание освобождает нас от необходимости рассматривать отдельно растяжение и сжатие.

2. Пусть l_0 — длина недеформированного стержня. После приложения силы F его длина получает приращение Δl и делается равной $l = l_0 + \Delta l$. Отношение

$$\varepsilon = \frac{\Delta l}{l_0} \tag{75.4}$$

называется *относительным удлинением стержня*. В случае растягивающих сил оно положительно, в случае сжимающих сил — отрицательно. Относительное удлинение, взятое с противоположным знаком, называется *относительным сжатием*. Таким образом, по определению относительным сжатием называется величина $-(\Delta l)/l_0$. Она положительна в случае сжимающих сил и отрицательна в случае растягивающих.

Опыт показывает, что *для не слишком больших упругих деформаций натяжение T (или давление P) пропорционально относительному удлинению (или относительному сжатию)*

$$T = E\frac{\Delta l}{l_0} \quad \text{или} \quad P = -E\frac{\Delta l}{l_0}, \tag{75.5}$$

где E — постоянная, зависящая только от материала стержня и его физического состояния. Она называется *модулем Юнга* по имени английского ученого Томаса Юнга (1773—1829). Формулы (75.5) выражают *закон Гука для деформаций растяжения и сжатия стержней*. Это приближенный закон. Для больших деформаций он может не оправдываться. Деформации, для которых приближенно выполняется закон Гука, называются *малыми деформациями*. Если в формуле (75.3) положить $\Delta l = l_0$, то получится $T = E$. Поэтому модуль Юнга часто определяют как натяжение, которое надо приложить к стержню, чтобы его длина удвоилась, если бы при такой деформации закон Гука оставался еще верным. Недостаток этого определения состоит в том, что при таких больших деформациях закон Гука почти для всех тел становится недействительным: тело либо разрушается, либо нарушается пропорциональность между деформацией и приложенным напряжением.

3. Более общим, чем закон Гука, является утверждение, что в *случае упругих деформаций натяжение T является однозначной функцией относительного удлинения ε: $T = T(\varepsilon)$*. Эта функция должна обращаться в нуль при $\varepsilon = 0$, так как с исчезновением деформации ε исчезает и напряжение T. Поэтому в разложении функции $T(\varepsilon)$ в ряд по степеням ε должен отсутствовать нулевой член. Это разложение должно иметь вид

$$T = E\varepsilon + A\varepsilon^2 + B\varepsilon^3 + \ldots,$$

причем коэффициенты E, A, B, ... являются постоянными, зависящими только от материала стержня и его физического состояния. Если относительное удлинение ε мало, то высшими степенями ε можно пренебречь. Тогда мы придем к закону Гука (75.5). При этом мы делаем относительную ошибку порядка $A\varepsilon^2/E\varepsilon \sim \varepsilon$. Эти общие соображения показывают, что *закон Гука и основанные на нем расчеты верны с относительной ошибкой порядка* ε. Поэтому во всех таких расчетах мы не только можем, но и должны отбросить слагаемые, которые по сравнению с основными членами являются

величинами порядка ε, ε^2 и т. д. Например, относительное удлинение ε можно определить не только выражением (75.4), но и выражением $(\Delta l)/l$. Дело в том, что разность этих двух выражений

$$\frac{\Delta l}{l_0} - \frac{\Delta l}{l} = \frac{(l - l_0)\Delta l}{l l_0} = \frac{(\Delta l)^2}{l l_0} \sim \varepsilon^2$$

второго порядка малости по ε, а потому ею следует пренебречь. Таким образом, закон Гука (75.5) можно также представить в виде

$$T = E\frac{\Delta l}{l}, \quad P = -E\frac{\Delta l}{l}, \tag{75.6}$$

или

$$\frac{\Delta l}{l} = \frac{T}{E}, \quad \frac{\Delta l}{l} = -\frac{P}{E}. \tag{75.7}$$

Это замечание, касающееся точности вычислений, разумеется, относится не только к деформациям растяжения и сжатия, но и ко всем малым деформациям, о которых будет идти речь ниже.

4. Пусть в стержне создано натяжение T_1. Оно вызовет относительное удлинение $\frac{\Delta_1 l}{l_0} = \frac{T_1}{E}$, и длина стержня сделается равной $l_1 = l_0 + \Delta_1 l$. Свойства материалов при деформациях, вообще говоря, изменяются. Поэтому можно было бы ожидать, что изменится и модуль Юнга. Однако если деформации малы (а только для таких деформаций и имеет смысл говорить о модуле Юнга), то с такими изменениями можно не считаться. Действительно, обозначим через E_1 модуль Юнга деформированного стержня. Если к деформированному стержню приложить дополнительное натяжение T_2, то его длина получит дополнительное приращение $\Delta_2 l$, причем $\frac{\Delta_2 l}{l_1} = \frac{T_2}{E_1}$. Принимая во внимание точность, с которой справедлив закон Гука, можно считать, что $\frac{\Delta_2 l}{l_1} = \frac{\Delta_1 l}{l_0}$. Имея еще в виду, что полное удлинение равно $\Delta l = \Delta_1 l + \Delta_2 l$, получим

$$\frac{\Delta l}{l_0} = \frac{T_1}{E} + \frac{T_2}{E_1}.$$

Но Δl не обязательно разлагать на составные части $\Delta_1 l$ и $\Delta_2 l$. Эту величину можно рассматривать как единое удлинение под действием результирующего натяжения $T = T_1 + T_2$. Поступая так, можно написать на основании закона Гука

$$\frac{\Delta l}{l_0} = \frac{T}{E} = \frac{T_1}{E} + \frac{T_2}{E}.$$

Сравнивая с предыдущим выражением, получим $E = E_1$, что и требовалось доказать.

Приведенное рассуждение справедливо не только для деформаций растяжения и сжатия, но и для любых малых деформаций. *Если деформации малы, то упругие постоянные тел не изменяются при деформациях.* Отсюда следует, что *если на тело действует несколько сил, то для вычисления результирующей деформации можно вычислить сначала деформации, вызываемые каждой силой в отдельности* (как если бы остальных сил не было вовсе), а *затем полученные деформации сложить.* Это важное положение называется *принципом суперпозиции малых деформаций.*

5. Для того чтобы деформировать тело, над ним надо совершить работу. В свою очередь, деформированное тело само может совершать работу. Оно обладает запасом потенциальной энергии. Эта энергия называется *упругой.* Она равна работе сил, затраченной на деформацию тела, при том существенном условии, что вся эта работа тратится только на приращение упругой энергии тела и не расходуется на увеличение кинетической энергии. Для того чтобы кинетическая энергия при деформации не возникала, надо деформацию производить достаточно медленно, постепенно увеличивая внешние силы, чтобы в любой момент времени каждая часть тела практически находилась в состоянии равновесия. Иначе говоря, при деформации внешние силы все время должны уравновешиваться возникающими при этом силами внутренних напряжений. Если это условие выполнено, то говорят, что тело совершает *квазистатический процесс.*

Возьмем для иллюстрации спиральную пружину, которая может служить моделью деформированного тела. Повесим ее за верхний конец. К нижнему концу подвесим груз, удерживая его рукой, чтобы пружина не растягивалась. Если груз внезапно отпустить, то возникнут колебания. Работа силы веса груза идет не только на растяжение пружины, но расходуется также на увеличение кинетической энергии груза и пружины. Это — процесс не квазистатический. Для вычисления упругой энергии пружины такой процесс не годится. Прикрепим теперь к нижнему концу пружины легкую чашечку и будем очень медленно нагружать ее песком. Колебания не возникают, пружина медленно и непрерывно удлиняется по мере увеличения нагрузки. Вся работа силы тяжести идет на увеличение потенциальной энергии деформируемой пружины. Такой процесс является квазистатическим, и им можно воспользоваться для вычисления упругой энергии пружины.

6. После этих замечаний легко вычислить упругую энергию растянутого стержня. Приложим к стержню растягивающую силу $f(x)$ и будем непрерывно и медленно увеличивать ее от начального значения $f = 0$ до конечного значения $f = F$. При этом удлинение стержня будет меняться от $x = 0$ до конечного значения $x = \Delta l$. По закону Гука $f(x) = kx$, где k — *коэффициент упругости*, который легко выразить через модуль Юнга. Вся работа в

рассматриваемом процессе пойдет на приращение упругой энергии U, а потому

$$U = \int_0^{\Delta l} f(x)\, dx = k \int_0^{\Delta l} x\, dx = \frac{1}{2} k(\Delta l)^2. \tag{75.8}$$

Так как в конечном состоянии $x = \Delta l$, то $F = f(\Delta l) = k\Delta l$. Учитывая это, получим

$$U = \frac{1}{2} F\Delta l. \tag{75.9}$$

Если бы к недеформированному стержню мы сразу приложили постоянную силу F, то при удлинении его на Δl была бы совершена вдвое большая работа $A = F\Delta l$. Так как запас упругой потенциальной энергии в стержне получился бы тем же самым, то ясно, что только половина работы A расходуется на приращение упругой энергии стержня. Вторая половина этой работы тратится на кинетическую энергию упругих колебаний и волн, которые всегда возбуждаются в стержне при неквазистатическом воздействии на него. При квазистатическом воздействии колебания и волны не возникают. Вот почему в формулах (75.8) и (75.9) появился числовой коэффициент $1/2$.

Найдем *объемную плотность упругой энергии*, т. е. *упругую энергию, и приходящуюся на единицу объема растянутого* (*или сжатого*) *стержня*. Она найдется делением выражения (75.9) на объем стержня $V = Sl$. Это дает

$$u = \frac{1}{2} \frac{F}{S} \frac{\Delta l}{l} = \frac{1}{2} T\varepsilon. \tag{75.10}$$

Если воспользоваться законом Гука, то эту формулу нетрудно привести к виду

$$u = \frac{1}{2} E\varepsilon^2 = \frac{T^2}{2E} = \frac{P^2}{2E}. \tag{75.11}$$

7. Опыт показывает, что *под действием растягивающей или сжимающей силы F изменяются не только продольные, но и поперечные размеры стержня. Если сила F — растягивающая, то поперечные размеры стержня уменьшаются. Если она сжимающая, то они увеличиваются.* Пусть a_0 — толщина стержня до деформации, a — после деформации. За толщину можно принять для круглого стержня его диаметр, для прямоугольного — одну из сторон его прямоугольного основания и т. д. Если сила F — растягивающая, то величина $-\frac{\Delta a}{a_0} \approx -\frac{\Delta a}{a}$ называется *относительным поперечными сжатием* стержня ($\Delta a = a - a_0$). Отношение относительного поперечного сжатия к соответствующему относительному

продольному удлинению называется *коэффициентом Пуассона* по имени французского ученого Симеона Пуассона (1781—1840):

$$\mu = -\frac{\Delta a}{a} \Big/ \frac{\Delta l}{l} = -\frac{\Delta a}{\Delta l}\frac{l}{a}. \qquad (75.12)$$

Коэффициент Пуассона зависит только от материала тела и является одной из важных постоянных, характеризующих его упругие свойства. Случай сжимающих сил не обязательно выделять особо, так как сжимающую силу можно рассматривать как растягивающую, взятую с противоположным знаком.

Модуль Юнга E и коэффициент Пуассона μ полностью характеризуют упругие свойства изотропного материала. Все прочие упругие постоянные могут быть выражены через E и μ.

8. Заметим, наконец, что все модули и коэффициенты упругости, с которыми мы имели и будем иметь дело, следовало бы для точности называть *изотермическими модулями* и *коэффициентами*. Они характеризуют деформации тел в предположении, что температура их поддерживается постоянной. Это обычно имеет место в случае статических деформаций. Но если деформации динамические (например, волны в упругих средах), то они могут происходить настолько быстро, что разности температур, возникшие при деформации, не успевают выравниваться в результате теплообмена. Важнейшим является предельный случай, когда между различно нагретыми частями среды теплообмен совсем не происходит. Соответствующие процессы, модули и коэффициенты упругости называются *адиабатическими*. Соотношения между изотермическими и адиабатическими модулями упругости будут рассмотрены в т. II.

ЗАДАЧИ

1. Найти относительное удлинение вертикально подвешенного стержня под действием собственного веса P. Площадь поперечного сечения равна S.

Ответ. $(l - l_0)/l_0 = P/(2SE)$.

2. Упругий стержень массой m, длиной l и площадью поперечного сечения S движется в продольном направлении с ускорением a (одинаковым для всех точек стержня). Найти упругую энергию деформации, возникающую вследствие ускоренного движения.

Ответ. $U = m^2a^2l/(6ES)$.

3. Какой максимальной кинетической энергией может обладать маховик, объем которого $V = 1$ м^3, если прочность материала на разрыв $T = 10^{10}$ дин/см^2. Всю массу маховика считать сосредоточенной в его ободе (тонком по сравнению с радиусом маховика). Показать, что при неизменной прочности материала маховика максимальная кинетическая энергия зависит только от объема, но не от массы маховика.

Ответ. $K = {}^1/_2\, VT = 5 \cdot 10^8$ Дж.

4. Тонкий стержень длиной $2l$ равномерно вращается вокруг перпендикулярной к нему оси, проходящей через центр стержня, с угловой

скоростью ω. Показать, что натяжение T, возникающее в стержне при таком вращении, удовлетворяет уравнению

$$\frac{dT}{dx} = -\rho\omega^2 x,$$

где ρ — плотность материала, а x — расстояние от оси вращения. Интегрируя это уравнение, найти распределение натяжения в стержне. В каком месте стержня натяжение максимально и чему оно равно? Показать, что максимальная кинетическая энергия, которую можно сообщить стержню при неизменной прочности его материала, зависит только от объема стержня V, но не от его массы. Вычислить максимальную кинетическую энергию для $V = 3 \cdot 10^4$ см^3, если максимальное натяжение, которое может выдержать стержень, равно $T_{\text{макс}} = 10^{10}$ дин/см^2.

Ответ. $T = 1/2\ \rho\omega^2(l^2 - x^2)$. Натяжение максимально в центре и равно $T_{\text{макс}} = 1/2\ \rho\omega^2 l^2$.

$$K_{\text{макс}} = \frac{1}{3} VT_{\text{макс}} = 10^7 \text{ Дж}.$$

(Ср. с задачей 3 к § 19.)

5. Стержень поперечного сечения S растягивается силой F, параллельной его оси. Под каким углом α к оси наклонено сечение, в котором тангенциальное напряжение τ максимально? Найти это напряжение.

Ответ. $\alpha = 45°$, $\tau = F/2S$.

6. Резиновый цилиндр с высотой h, весом P и площадью основания S поставлен на горизонтальную плоскость. Найти энергию упругой деформации цилиндра, возникающей под действием его собственного веса. Во сколько раз изменится энергия упругой деформации рассматриваемого цилиндра, если на верхнее основание его поставить второй такой же цилиндр?

Ответ. $U = P^2h/(6ES)$. Во втором случае упругая энергия увеличится в 7 раз.

§ 76. ДЕФОРМАЦИИ ПРЯМОУГОЛЬНОГО ПАРАЛЛЕЛЕПИПЕДА ПОД ДЕЙСТВИЕМ ТРЕХ ВЗАИМНО ПЕРПЕНДИКУЛЯРНЫХ СИЛ

1. Допустим, что однородное изотропное тело имеет форму прямоугольного параллелепипеда, к противоположным граням которого приложены силы F_x, F_y, F_z, нормальные к этим граням. Соответствующие им натяжения обозначим через T_x, T_y, T_z (рис. 201). Определим деформации, которые возникнут под действием этих сил. Будем предполагать деформации малыми. Тогда для решения задачи можно воспользоваться принципом суперпозиции малых деформаций.

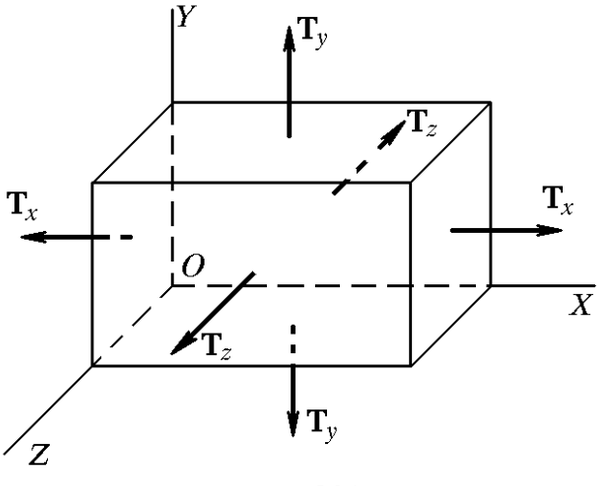


Рис. 201

Направим координатные оси параллельно ребрам параллеле-

пипеда. Пусть x, y, z — длины этих ребер. Если бы действовала только сила F_x, то ребро x получило бы приращение $\Delta_1 x$, определяемое соотношением $\frac{\Delta_1 x}{x} = \frac{T_x}{E}$. Если бы действовала только сила F_y, то размеры параллелепипеда, перпендикулярные к оси Y, сократились бы. В частности, ребро x при этом получило бы отрицательное приращение $\Delta_2 x$, которое можно вычислить по формуле $\frac{\Delta_2 x}{x} = -\mu \frac{T_y}{E}$. Наконец, относительное приращение ребра x под действием одной только силы F_z было бы равно $\frac{\Delta_3 x}{x} = -\mu \frac{T_z}{E}$. Если все силы действуют одновременно, то согласно принципу суперпозиции малых деформаций результирующее удлинение ребра x будет равно $\Delta x = \Delta_1 x + \Delta_2 x + \Delta_3 x$. Аналогично вычисляются удлинения параллелепипеда и вдоль остальных двух направлений Y и Z. В результате для удлинений всех трех ребер параллелепипеда можно написать

$$\varepsilon_x \equiv \frac{\Delta x}{x} = \frac{T_x}{E} - \frac{\mu}{E}(T_y + T_z),$$

$$\varepsilon_y \equiv \frac{\Delta y}{y} = \frac{T_y}{E} - \frac{\mu}{E}(T_z + T_x), \qquad (76.1)$$

$$\varepsilon_z \equiv \frac{\Delta z}{z} = \frac{T_z}{E} - \frac{\mu}{E}(T_x + T_y).$$

2. При квазистатическом растяжении параллелепипеда вдоль оси X совершается работа $A_1 = \frac{1}{2} S_x T_x \Delta x$, где $S_x = yz$ — площадь грани, перпендикулярной к оси X. Эту работу можно представить в виде $A_1 = \frac{1}{2} xyz \cdot T_x \frac{\Delta x}{x} = \frac{1}{2} V T_x \varepsilon_x$, где $V = xyz$ — объем параллелепипеда. Аналогично запишутся работы при квазистатических растяжениях в направлениях координатных осей Y и Z. Сложив эти три работы и разделив результат на объем параллелепипеда, получим следующее выражение для плотности упругой энергии в рассматриваемом случае:

$$u = \frac{1}{2}(T_x \varepsilon_x + T_y \varepsilon_y + T_z \varepsilon_z). \qquad (76.2)$$

С помощью формул (76.1) это выражение приводится к виду

$$u = \frac{1}{2E}[T_x^2 + T_y^2 + T_z^2 - 2\mu(T_x T_y + T_y T_z + T_z T_x)]. \qquad (76.3)$$

Если из трех натяжений T_x, T_y, T_z только одно отличается от нуля, то эти формулы переходят в более простые формулы (75.10) и (75.11). Согласно формулам (75.11) *плотность упругой энергии и пропорциональна квадрату натяжения* T (*или давления* P). В общем случае, как показывает формула (76.3), плотность упругой энергии является квадратичной однородной функцией T_x, T_y, T_z (или P_x, P_y,

P_z). При заданных натяжениях (или давлениях) она обратно пропорциональна модулю упругости E. Чем жестче пружина, тем меньше при неизменном натяжении ее упругая сила. Идеально твердые тела (для которых $E = \infty$) совершенно не обладают упругой энергией, какие бы силы натяжения и давления на них не действовали.

Натяжения T_x, T_y, T_z выражаются через ε_x, ε_y, ε_z линейно, как это следует из формул (76.1). Поэтому плотность упругой энергии является квадратичной однородной функцией деформаций ε_x, ε_y, ε_z. В частном случае ($\varepsilon_x = \varepsilon$, $\varepsilon_y = \varepsilon_z = 0$) она пропорциональна *квадрату деформации*. При заданных деформациях ε_x, ε_y, ε_z плотность упругой энергии *и* пропорциональна модулю упругости E. Чем жестче пружина, тем больше ее упругая энергия (при неизменной деформации).

ЗАДАЧА

Определить относительное изменение объема полого латунного шара радиусом $R = 5$ см, в который накачан воздух до давления 11 атм (наружное давление 1 атм). Толщина сферической оболочки $d = 1$ мм. Модуль Юнга латуни $E = 10^{12}$ дин/см^2, коэффициент Пуассона $\mu = 0{,}3$.

Р е ш е н и е. В силу симметрии касательное напряжение τ, действующее в оболочке, одно и то же и одинаково во всех направлениях. Возьмем малый элемент оболочки, имеющий форму прямоугольника. При вычислении относительного изменения площади этого элемента под действием касательных напряжений τ можно отвлечься от кривизны элемента, приняв его за плоскую прямоугольную пластинку. Тогда вычисление дает

$$\frac{\Delta S}{S} = 2(1-\mu)\frac{\tau}{E}$$

(изменение площади, вызванным нормальным давлением, пренебрегаем). Поскольку поверхность шара S пропорциональна $V^{2/3}$, где V — объем шара, относительное изменение объема будет $\frac{\Delta V}{V} = \frac{3}{2}\frac{\Delta S}{S}$. Так как поверхность искривлена, то натяжение τ создаст разность нормальных давлений. Для нее нетрудно получить $2\tau d/R$ (см. *формулу Лапласа* в учении о поверхностном натяжении, том II). Эта разность должна быть уравновешена разностью давлений газа ΔP по разные стороны оболочки. В результате получим

$$\frac{\Delta V}{V} = \frac{3}{2}\frac{(1-\mu)R}{Ed}\Delta P \approx 5\cdot 10^{-4}.$$

§ 77. ВСЕСТОРОННЕЕ И ОДНОСТОРОННЕЕ РАСТЯЖЕНИЕ И СЖАТИЕ

1. Рассмотрим частный случай, когда все натяжения T_x, T_y, T_z равны и отрицательны. В этом случае на параллелепипед со всех сторон действует постоянное давление $P = -T_x = -T_y = -T_z$. Как видно из

формул (76.1), все три относительные деформации ε_x, ε_y, ε_z равны между собой и определяются выражением

$$\varepsilon_x = \varepsilon_y = \varepsilon_z = -\frac{P}{E}(1 - 2\mu). \tag{77.1}$$

Их легко выразить через относительное изменение объема параллелепипеда при деформации. Действительно, взяв логарифмические производные от обеих частей равенства $V = xyz$, получим

$$\frac{\Delta V}{V} = \frac{\Delta x}{x} + \frac{\Delta y}{y} + \frac{\Delta z}{z},$$

или

$$\frac{\Delta V}{V} = \varepsilon_x + \varepsilon_y + \varepsilon_z. \tag{77.2}$$

Поэтому формулу (77.1) можно представить в виде

$$\frac{\Delta V}{V} = -\frac{P}{K}, \tag{77.3}$$

где постоянная K определяется выражением

$$K = \frac{E}{3(1 - 2\mu)}. \tag{77.4}$$

Эта постоянная называется *модулем всестороннего сжатия*. Формула (77.3) применима к телам произвольной, а не только прямоугольной формы. Для доказательства достаточно заметить, что произвольное тело можно мысленно разделить на малые части, каждая из которых имеет форму прямоугольного параллелепипеда. Эти части находятся под постоянным внешним давлением. Относительные изменения их объемов, а следовательно, и относительное изменение объема всего тела одинаковы и определяются формулой (77.3).

Выражение (76.3) для плотности упругой энергии в случае *деформации всестороннего сжатия* переходит в

$$u = \frac{3(1 - 2\mu)}{2E} P^2 = \frac{P^2}{2K}. \tag{77.5}$$

Так как величина u существенно положительна, то должно быть $1 - 2\mu > 0$, т. е.

$$\mu < \frac{1}{2}. \tag{77.6}$$

2. Рассмотрим другой важный случай — *деформацию одностороннего растяжения или сжатия*. Пусть однородный стержень может свободно растягиваться или сжиматься в направлении его оси (которую мы примем за координатную ось X), а его поперечные размеры изменяться не могут. Этот случай имеет важное значение в теории распространения продольных волн в неограниченной упру-

гой среде (см. § 83). Можно мысленно вырезать часть среды, имеющую форму стержня, направленного вдоль распространения волн. Такой «стержень» может сжиматься или расширяться в продольном направлении. Однако изменениям его поперечных размеров препятствует окружающая среда. Форма поперечного сечения стержня не имеет значения. Возьмем стержень с прямоугольным поперечным сечением, чтобы можно было воспользоваться формулами (76.1). Пусть вдоль стержня действует постоянное натяжение T_x. Поперечные напряжения T_y и T_z найдутся из условия неизменности размеров стержня в направлениях координатных осей Y и Z. Полагая в формулах (76.1) $\Delta y = \Delta z = 0$, получим

$$T_y - \mu(T_z + T_x) = 0, \quad T_z - \mu(T_x + T_y) = 0.$$

Отсюда

$$T_y = T_z = \frac{\mu}{1-\mu} T_x, \tag{77.7}$$

$$\varepsilon_x = \frac{T_x}{E}\left(1 - \frac{2\mu^2}{1-\mu}\right). \tag{77.8}$$

Введем обозначение

$$E' = E\,\frac{1-\mu}{1-\mu-2\mu^2} = E\,\frac{1-\mu}{(1+\mu)\,(1-2\mu)}, \tag{77.9}$$

или

$$E' = \left[\frac{2}{3(1+\mu)} + \frac{1}{3(1-2\mu)}\right] E. \tag{77.10}$$

Тогда

$$\frac{\Delta x}{x} = \frac{T_x}{E'}. \tag{77.11}$$

Это соотношение аналогично соотношениям (75.7). Постоянная E' называется *модулем одностороннего растяжения.*

ЗАДАЧА

Прямоугольная пластинка зажата между вертикальными плоскостями, перпендикулярными к оси X, так что в направлении этой оси частицы пластинки смещаться не могут (рис. 202). В направлении оси Z пластинка подвергается равномерному одностороннему давлению P. Определить давление P_x, которому подвергается пластинка со стороны плоскостей, между которыми она зажата. Найти выражение для

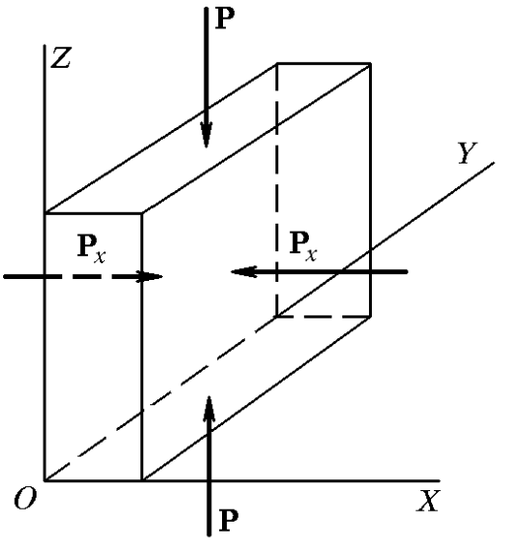


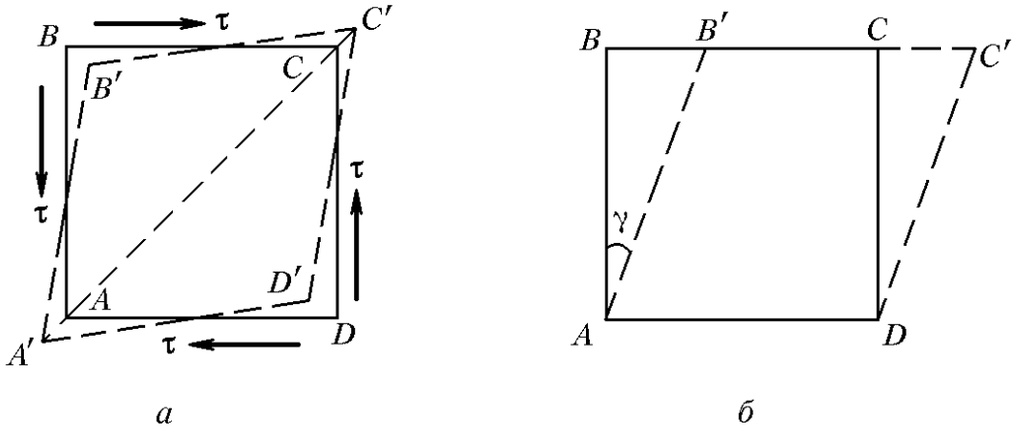
Рис. 202

плотности упругой энергии u, а также относительное сжатие пластинки в направлении оси Z и относительное расширение в направлении оси Y.

Ответ. $P_x = \mu P$, $\frac{\Delta y}{y} = \frac{\mu P}{E}(1+\mu)$, $\frac{\Delta z}{z} = -\frac{P}{E}(1-\mu^2)$, $u = \frac{P^2}{2E}(1-\mu^2)$.

§ 78. СДВИГ

1. Возьмем куб из однородного и изотропного вещества. Приложим к противоположным граням его AD и BC равные и противоположно направленные касательные силы (рис. 203 *а*). Они образуют пару сил, под действием которых куб начинает вращаться. Для устранения вращения приложим такие же касательные силы к граням AB и CD. Тогда куб вращаться не будет, а будет только деформироваться. Необходимость приложения касательных напряжений к



а *б*

Рис. 203

граням AB и CD непосредственно следует также из симметрии тензора упругих напряжений (см. § 74).

Опыт показывает, что под действием приложенных напряжений квадрат $ABCD$ переходит в ромб $A'B'C'D'$. При этом длина диагонали AC увеличивается, а диагонали BD — уменьшается. Объем тела, как будет показано ниже, при такой деформации практически изменяться не будет. Относительные изменения объема будут величинами более высокого порядка малости, чем относительные изменения длин диагоналей AC и BD. В теории малых деформаций такими изменения пренебрегают. Высшего порядка малости будут и изменения длин сторон квадрата $ABCD$. Поэтому куб после деформации можно повернуть так, чтобы новое основание $A'D'$ совместилось с прежним основанием AD (рис. 203 *б*). Отсюда видно, что рассматриваемая деформация состоит в том, что все слои куба, параллельные основанию AD, сдвигаются в одном и том же направлении, параллельном тому же основанию. Поэтому эта деформация называется *сдвигом*. Сдвиг пропорционален расстоянию сдвигаемого слоя от основания AD. Угол γ между гранью AB до деформации и той же гранью AB' после деформации назы-

вается *углом сдвига*. Конечно, ту же деформацию можно получить путем сдвига параллельно грани AB или CD на тот же угол γ. Мы предполагаем, конечно, что угол γ мал ($\gamma \ll 1$) и пользуемся законом Гука. Для деформации сдвига этот закон можно записать в виде

$$\tau = G\gamma, \tag{78.1}$$

где τ — касательное напряжение, действующее на гранях куба. Постоянная G называется *модулем сдвига* и зависит от материала, из которого изготовлен куб.

2. Найдем выражение для плотности упругой энергии при деформации сдвига. Закрепив неподвижно основание AD (рис. 203 *б*), будем производить сдвиг квазистатически. Тогда вся работа, затрачиваемая на сдвиг, пойдет на увеличение упругой энергии тела. Совершаемая работа, очевидно, пойдет на увеличение упругой энергии тела. Совершаемая работа, очевидно, равна $A = \frac{1}{2}\,\tau S\,\Delta x$, где Δx — смещение грани BC при сдвиге, а S — площадь этой грани. Если a — длина ребра куба, то $\Delta x = a\gamma$, а потому $A = \frac{1}{2}\,\tau S\,a\gamma = \frac{1}{2}\,V\tau\gamma$, где V — объем куба. Таким образом, объемная плотность упругой энергии выражается формулой

$$u = \frac{1}{2}\,\tau\gamma = \frac{\tau^2}{2G}. \tag{78.2}$$

3. Тангенциальные напряжения, действующие параллельно граням куба, можно свести к совокупности натяжения и давления, равных по модулю и действующих во взаимно перпендикулярных направлениях. Действительно, проведем диагональное сечение куба AC (плоскость, перпендикулярной к плоскости рис. 203 *а*). Сила F, действующая на часть куба ACD на плоскости AC, будет нормальна к этой плоскости и направлена внутрь рассматриваемой части. Это и есть сила нормального давления. Определим это давление. Если длина ребра куба есть a, то сила F, очевидно, равна

$$F = a^2(\tau \sin 45^\circ + \tau \cos 45^\circ) = \sqrt{2}\,a^2\tau.$$

Площадь диагонального сечения AC есть $a^2\sqrt{2}$. Разделив F на эту площадь, получим искомое давление $P = \tau$. Итак, в диагональном сечении AC и во всякой плоскости, ему параллельной, напряжение сводится к нормальному давлению, численно равному τ. Рассуждая аналогично, можно доказать, что в диагональном сечении BD и во всякой плоскости, параллельной ему, действует нормальное натяжение T, также численно равное τ.

4. На основании изложенного ясно, что *сдвиг эквивалентен растяжению тела в некотором направлении и сжатию в перпендикулярном направлении*. Вырежем, например, мысленно из нашего куба прямоугольный параллелепипед с поперечным сечением $PQRS$ (рис. 204). В направлении диагонали куба AC он будет растянут натяжением $T = \tau$, в перпендикулярном направлении BD — сжат дав-

лением $P = \tau$. В направлении, перпендикулярном к плоскости рисунка, размеры параллелепипеда останутся неизменными. Направим ось X параллельно ребрам PQ и SR, а ось Y — параллельно ребрам QP и PS. Тогда, подставляя в формулы (76.1) $T_x = \tau$, $T_y = \tau$, $T_z = 0$, получим $\varepsilon_z = 0$, $\varepsilon_x + \varepsilon_y = 0$. В силу соотношения (77.2) $\Delta V = 0$. Деформация не сопровождается изменением объема тела — утверждение, которое упоминалось выше без доказательства.

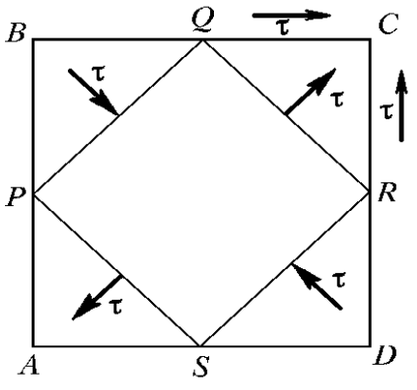


Рис. 204

5. Таким же путем из формулы (76.3) получаем для плотности упругой энергии при сдвиге

$$u = \frac{1+\mu}{E}\tau^2. \tag{78.3}$$

Эта величина должна совпадать с (78.2), так как значение u не может зависеть от способа вычисления. Сравнивая оба выражения, получим

$$G = \frac{E}{2(1+\mu)}. \tag{78.4}$$

Эта формула устанавливает связь между модулем Юнга E, коэффициентом Пуассона μ и модулем сдвига G. Используя ее, а также формулы (77.10) и (77.4), получим

$$E' = K + \frac{4}{3}G. \tag{78.5}$$

§ 79. КРУЧЕНИЕ

1. Деформации, о которых шла речь до сих пор, были деформациями *однородными*, т. е. такими, когда все бесконечно малые элементы тела деформированы одинаково. *Деформации кручения и изгиба*, к изучению которых мы обращаемся, *являются деформациями неоднородными*. Это значит, что в этих случаях деформации внутри тела меняются от точки к точке.

Возьмем однородную проволоку, закрепим ее верхний конец, а к нижнему концу приложим закручивающие силы, создающие вращающий момент M относительно продольной оси проволоки. Проволока закрутится — каждый радиус нижнего основания ее повернется вокруг продольной оси на угол φ. Такая деформация называется *кручением*. Закон Гука для деформации кручения записывается в виде

$$M = f\varphi, \tag{79.1}$$

где f — постоянная для данной проволоки величина, называемая ее *модулем кручения*. В отличие от ранее введенных модулей E, K,

E', G и коэффициента μ, *модуль кручения зависит не только от материала, но и от геометрических размеров проволоки.*

2. Выведем выражение для модуля кручения f. Сначала сделаем это для цилиндрической трубки радиусом r и длиной l, предполагая, что толщина δr стенки трубки очень мала по сравнению с радиусом r. Площадь основания трубки есть $2\pi r\delta r$. Момент сил, действующий на это основание, будет $M = 2\pi r\delta r \cdot \tau r$, где τ — касательное напряжение в том же основании. При квазистатическом закручивании проволоки на угол φ совершается работа $A = \frac{1}{2} M\varphi = M^2/2f$. Разделив ее на объем трубки $V = 2\pi rl\delta r$, найдем плотность упругой энергии при деформации кручения

$$u = \frac{\pi\tau^2 r^3 \delta r}{fl}. \qquad (79.2)$$

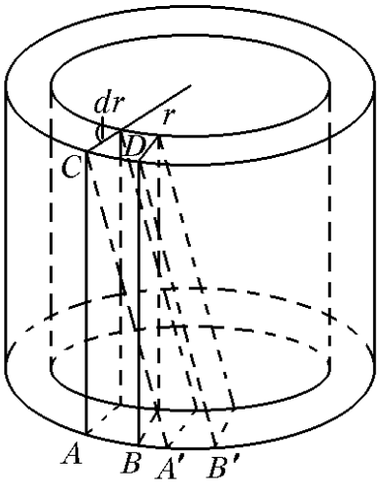


Рис. 205

Ту же величину можно выразить иначе. Вырежем мысленно из трубки бесконечно короткую часть, изображенную на рис. 205. В результате деформации кручения бесконечно малый элемент трубки $ABDC$ перейдет в положение $A'B'DC$. Это есть сдвиг. Таким образом, деформацию кручения можно рассматривать как неоднородный сдвиг. Плотность упругой энергии при сдвиге дается выражением (78.2). Приравнивая его выражению (79.2), находим искомое соотношение

$$f = \frac{2\pi G}{l} r^3 \delta r. \qquad (79.3)$$

Если стенка трубки имеет конечную толщину, то модуль f найдется интегрированием последнего выражения по r. Это дает

$$f = \frac{\pi G}{2l}(r_2^4 - r_1^4), \qquad (79.4)$$

где r_1 — внутренний радиус трубки, а r_2 — наружный. Для сплошной проволоки радиуса r

$$f = \frac{\pi G}{2l} r^4. \qquad (79.5)$$

3. Экспериментально модуль кручения можно измерить, наблюдая крутильные колебания тяжелого тела, подвешенного к нижнему концу проволоки. Эти колебания будут гармоническими с периодом

$$T = 2\pi\sqrt{\frac{I}{f}} \qquad (79.6)$$

(см. § 42). Если момент инерции тела I известен, то, измерив период колебаний T, можно вычислить по этой формуле модуль кручения f.

ЗАДАЧИ

1. Две проволоки одинаковой длины сделаны из одного и того же материала, но диаметр второй из них вдвое больше, чем первой. В одном из опытов нижнее основание каждой проволоки было закручено относительно ее верхнего основания на один и тот же угол. В другом опыте проволоки были сварены своими основаниями так, что ось одной из них стала продолжением оси другой; затем нижнее основание получившейся составной проволоки было закручено относительно верхнего на некоторый угол. Найти отношения упругих энергий проволок в обоих случаях.

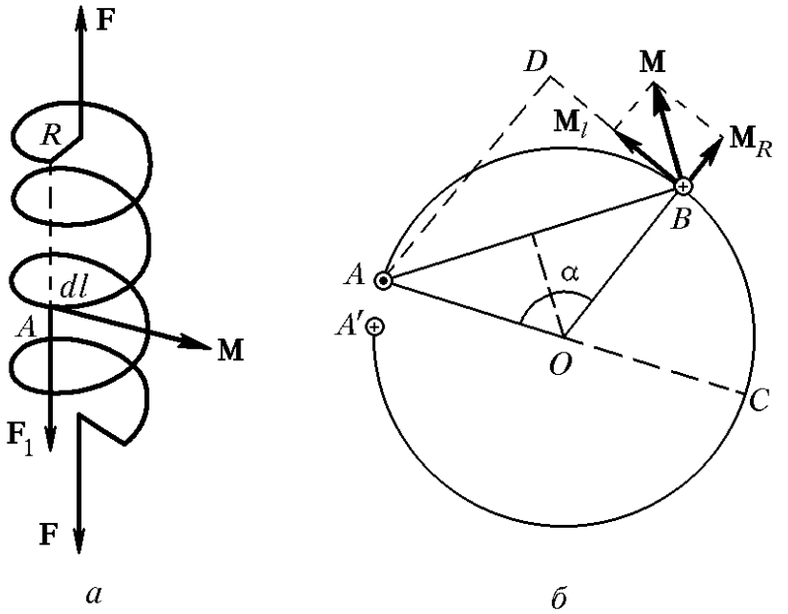
Ответ. 1) $U_1/U_2 = 1/16$. 2) $U_1/U_2 = 16$.

2. Шар, подвешенный на проволоке, совершает крутильные колебания с периодом T вокруг вертикальной оси. Найти период колебаний того же шара T', если проволоку, на которой он был подвешен, заменить цилиндрической трубкой той же длины и массы с внешним радиусом R и внутренним радиусом r и изготовленной из того же материала.

Ответ. $T' = T\sqrt{\dfrac{R^2-r^2}{R^2+r^2}}$.

3. Определить удлинение спиральной пружины, если растягивающие силы действуют вдоль ее оси. Шаг спирали считать пренебрежимо малым по сравнению с радиусом витка R. Модуль кручения проволоки, из которой изготовлена спираль, считать известным.

Решение. Произведем мысленный разрез проволоки пружины в произвольной точке A плоскостью, проходящей через ось пружины (рис. 206 *а*). Пусть $\mathbf{F}_1$ — сила, с которой нижняя часть пружины действует на верхнюю в месте разреза. Для равновесия необходимо, чтобы $\mathbf{F}_1 = -\mathbf{F}$, где $\mathbf{F}$ — растягивающая сила, действующая на верхнюю часть пружины. Так как силы $\mathbf{F}$ и $\mathbf{F}_1$ образуют пару, то момент этой пары не зависит от выбора точки, относительно которой он берется. Этот момент перпендикулярен к плоскости разреза и равен $M = FR$. Из-за малости шага витка можно считать, что момент в точке A направлен вдоль оси проволоки. Чтобы рассматриваемая часть пружины находилась в равновесии, необходимо,



а *б*

Рис. 206

чтобы возникло кручение проволоки вокруг ее оси, компенсирующее момент **M**. Когда растягивающие силы F действуют вдоль оси пружины, модуль момента M не меняется вдоль проволоки, а потому кручение ее будет равномерным. Пусть dl — элемент длины проволоки. Под действием момента M он закрутится на угол $d\varphi = M/f_1$, где f_1 — модуль кручения рассматриваемого элемента. Обозначим через f модуль кручения всей проволоки (если ее выпрямить). Так как модуль кручения обратно пропорционален длине проволоки l_0, то $f = f_1 \frac{dl}{l_0}$, а потому $d\varphi = \frac{M}{f}\frac{dl}{l_0}$. В результате закручивания элемента dl на угол $d\varphi$ нижний конец проволоки опустится на $dx = Rd\varphi = \frac{MR}{f}\frac{dl}{l_0} = \frac{FR^2}{f}\frac{dl}{l_0}$. Интегрируя по всей длине проволоки, найдем удлинение пружины

$$x = \frac{FR^2}{f}. \tag{79.7}$$

Введем жесткость пружины по формуле $F = kx$. Тогда

$$k = \frac{f}{R^2}. \tag{79.8}$$

4. Рассмотреть ту же задачу для случая, когда растягивающие силы действуют не вдоль оси пружины, а вдоль одной из образующих цилиндрической поверхности, на которую она намотана.

Р е ш е н и е. Мысленно выделим произвольный участок проволоки AB (рис. 206 *б*). Силы, действующие на его концы, перпендикулярны к плоскости рисунка (параллельны продольной оси пружины). Каждая из этих сил равна внешней силе F, приложенной к пружине. Силу, направленную к нам, изобразим точкой, от нас — крестиком. Момент **M** сил, приложенных к выделенному участку, перпендикулярен к хорде AB и равен $M = 2FR \sin(\alpha/2)$. Разложим этот момент на составляющую M_l вдоль проволоки и составляющую M_R, перпендикулярную к ней. Если пружина содержит много витков, то составляющую M_R можно не учитывать. Она вызывает изгиб проволоки вокруг оси, параллельной радиусу OB. Но легко видеть, что такой момент изгибает участок AC в одну сторону, а участок CA' — в противоположную, так что в целом при большом числе витков изгиб практически не влияет на удлинение пружины. Момент M_l равен $M \sin(\alpha/2) = 2FR \sin^2(\alpha/2)$. Элемент проволоки длиной dl он закрутит на угол $d\varphi = \frac{M_l}{f_1} = \frac{M_l}{f}\frac{dl}{l_0}$ и сместит свободный конец пружины на величину

$$dx = AD \cdot d\varphi = AB \sin\frac{\alpha}{2} \cdot d\varphi = \frac{4FR^2}{fl_0}\sin^4\frac{\alpha}{2}\,dl = \frac{4FR^3}{fl_0}\sin^4\frac{\alpha}{2}\,d\alpha.$$

В отличие от предыдущего случая, кручение проволоки неоднородно. Интегрируя по всей длине пружины и считая число витков целым, получим

$$x = \frac{3}{4}\frac{FR^2}{f}.$$

§ 80. ИЗГИБ

1. Рассмотрим изгиб однородного бруса (балки) произвольного поперечного сечения, которое, однако, должно оставаться одинаковым на протяжении всей длины бруса. Пусть до деформации брус имел прямолинейную форму. Проведя сечения AB и $A'B'$, нормальные к оси бруса, мысленно вырежем из него бесконечно малый элемент $AA'B'B$ (рис. 207 *а*), длину которого обозначим через l_0. Ввиду бесконечной малости выделенного элемента можно считать, что в результате изгиба прямые AA', NN', BB' и все прямые, им параллельные, перейдут в дуги окружности с центрами, лежащими на оси O, перпендикулярной к плоскости рисунка (рис. 207 *б*). Эта ось называется *осью изгиба*. Наружные волокна, лежащие выше линии NN', при изгибе удлиняются, волокна, лежащие ниже линии NN', укорачиваются. Длина линии NN' остается неизменной. Эта линия называется *нейтральной линией*. Проходящее через нее сечение (недеформированного) бруса плоскостью, перпендикулярной к плоскости рис. 207 *а*, называется *нейтральным сечением*. Таким образом, все наружные волокна будут растянуты,

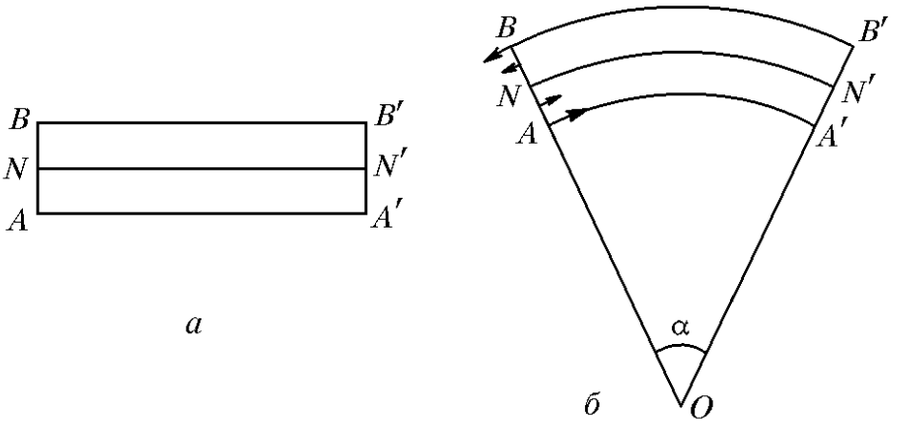


Рис. 207

все внутренние — сжаты. Пусть R — радиус кривизны нейтральной линии NN'. Тогда $l_0 = R\alpha$, где α — центральный угол, опирающийся на дугу NN'. Рассмотрим волокно бруса, находящегося на расстоянии ξ от нейтрального сечения. (Величина ξ положительна, если волокно находится выше нейтрального сечения (рис. 207 *б*), и отрицательна, если оно находится ниже.) Если брус не слишком толст, так что $|\xi| \ll R$, то длина рассматриваемого волокна будет $l = (R + \xi)\alpha$, а удлинение $\Delta l = l - l_0 = \xi\alpha$. Следовательно, натяжение, действующее вдоль рассматриваемого волокна, $\tau = E\frac{\Delta l}{l_0} = E\xi\frac{\alpha}{l_0}$, или

$$\tau = E\frac{\xi}{R}. \tag{80.1}$$

Натяжение, таким образом, меняется линейно с расстоянием ξ. Ниже нейтрального сечения оно отрицательно, т. е. является давлением. Сумма сил натяжения и давления, действующих в сечении AB, может быть и отличной от нуля. Однако в этом случае на изгиб бруса будет накладываться растяжение или сжатие его, одинаковое для всех волокон. Оно может быть учтено особо и исключено из рассмотрения, когда речь идет об изгибе в чистом виде. Поэтому мы будем считать, что сумма всех сил натяжения, действующих в каждом нормальном сечении бруса, равна нулю, т. е. $\int \tau\, dS = 0$,

или в силу (80.1) $\int \xi\, dS = 0$, где dS — элемент площади рассматриваемого поперечного сечения. Интегрирование ведется по всему поперечному сечению бруса. Отсюда видно, что нейтральная линия и нейтральное сечение проходят через центр тяжести поперечного сечения бруса. Из соотношения $\int \tau\, dS = 0$ следует, что момент сил натяжения M_τ, действующих на сечение AB, не зависит от того, относительно какой оси он берется. Для вычисления M_τ проще всего взять ось, перпендикулярную к плоскости рисунка и проходящую через точку N. Очевидно,

$$M_\tau = \int \xi\tau\, dS = \frac{E}{R}\int \xi^2\, dS,$$

или

$$M_\tau = \frac{E}{R} I, \tag{80.2}$$

где введено обозначение

$$I = \int \xi^2\, dS. \tag{80.3}$$

Величина I называется *моментом инерции* поперечного сечения бруса по аналогии с соответствующей величиной, вводимой при рассмотрении вращения тела вокруг неподвижной оси. Однако в отличие от последней величины, имеющей размерность массы, умноженной на квадрат длины, (80.3) есть чисто геометрическая величина с размерностью четвертой степени длины.

Можно воспользоваться формулами для моментов инерции, выведенными в § 36, заменив всюду массу m на площадь поперечного сечения S. Если поперечное сечение бруса имеет форму прямоугольника с шириной a и высотой b, то

$$I = \frac{ab^3}{12}. \tag{80.4}$$

Для кругового поперечного сечения радиуса r

$$I = \frac{\pi}{4} r^4. \tag{80.5}$$

Для цилиндрической трубы с внутренним диаметром r_1 и наружным r_2

$$I = \frac{\pi}{4}(r_2^4 - r_1^4). \tag{80.6}$$

Направим ось X вдоль нейтральной линии недеформированного бруса. Ось Y направим к ней перпендикулярно и расположим в плоскости изгиба. Тогда уравнение нейтральной линии изогнутого бруса можно представить в виде $y = y(x)$. По известной формуле

$$\frac{1}{R} = \frac{y''}{(1+y'^2)^{3/2}}.$$

Если изгиб мал ($y' \ll 1$), то квадратом производной можно пренебречь. В этом приближении

$$M_\tau = EIy''. \tag{80.7}$$

2. Вырежем произвольную (конечную или бесконечную) часть бруса, мысленно проведя в нем два нормальных сечения. В состоянии равновесия

момент упругих сил натяжения, действующих на торцах вырезанной части, должен быть уравновешен противоположно направленными моментами всех прочих внешних сил, действующих на рассматриваемую часть. Это дает метод решения задач на изгиб. Он иллюстрируется примерами, приводимыми ниже, а также задачами к этому параграфу.

Пример 1. Однородный стержень AB лежит на двух симметрично расположенных опорах C и D (рис. 208). К концам стержня A и B приложены равные и одинаково направленные силы F. Определим форму стержня в состоянии равновесия, пренебрегая весом самого стержня. Из симметрии ясно, что в равновесии опоры C и D будут давить на стержень с одинаковыми силами F_1 и F_1, каждая из которых равна F. Проведем в стержне нормальное сечение через произвольно расположенную точку O. Достаточно рассмотреть равновесие одной из частей стержня, например части OB. Упругие натяжения в сечении O создают вращающий момент M_τ, определяемый формулой (80.2). Пара сил F_1 и F создает противоположно направленный момент $M = Fa$, где a — расстояние между линиями действия сил F_1 и F. Как и M_τ, момент M не зависит от положения оси, относительно которой он берется. Кроме того, момент M не зависит от положения точки O. Он одинаков вдоль всего стержня. Уравнение равновесия $M_\tau = M$ принимает вид $IE/R = Fa$. Из него следует, что радиус кривизны R одинаков во всех точках нейтральной линии стержня. Следовательно, в состоянии равновесия стержень будет иметь форму дуги окружности, как это изображено на рис. 208 штриховыми линиями.

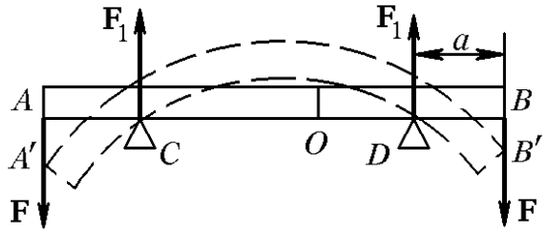


Рис. 208

Для демонстрации можно взять деревянную доску и вколотить в нее гвозди в точках A', B', C, D. Если между этими гвоздями заложить гибкую стальную линейку, то она примет форму дуги окружности. Это дает практический способ черчения окружностей, когда обычный чертежный циркуль оказывается непригодным (например, в случае окружностей очень большого радиуса).

Пример 2. Определим стрелу прогиба балки, жестко закрепленной в стене одним из своих концов (рис. 209). На другой конец балки действует сосредоточенная сила F. Весом самой балки будем пренебрегать. Стрелой прогиба мы называем смещение свободного конца балки под действием приложенной силы F.

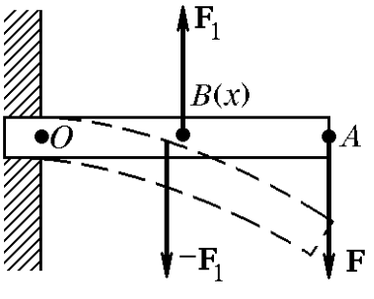


Рис. 209

Поместим начало координат в точке O, в которой нейтральная линия балки пересекается с плоскостью стены. Через произвольную точку $B(x)$ (с координатой $x = OB$) проведем нормальное сечение. Для равновесия необходимо, чтобы сила F_1, действующая на часть BA со стороны части OB, была направлена вверх и равнялась F. Вместе с F она образует пару сил с моментом $M = F(l - x)$, где l — длина балки. Момент сил натяжения возьмем в приближенном виде (80.7), считая, что прогиб мал. Это приводит к уравнению

$$EIy'' = F(l - x).$$

(Ось y направлена в сторону вогнутости, т. е. вниз. При таком условии вторая производная y'' положительна, и обе части последнего соотношения имеют одинаковые знаки.) Интегрируя это уравнение один раз, получим

$$y' = \frac{F}{EI} x \left(l - \frac{x}{2}\right) + C.$$

Постоянная интегрирования C равна нулю, так как при $x = 0$, т. е. в точке O, касательная к нейтральной линии горизонтальна. Интегрируя вторично и учитывая, что в точке O (т. е. при $x = 0$) $y = 0$, найдем

$$y = \frac{Fx^2}{2EI}\left(l - \frac{x}{3}\right). \tag{80.8}$$

Полагая здесь $x = l$, находим стрелу прогиба

$$\lambda = \frac{Fl^3}{3El}. \tag{80.9}$$

Пример 3. Определим стрелу прогиба центра балки, лежащей на двух опорах, если к ее середине O приложена сосредоточенная сила F, направленная вниз (рис. 210). Весом балки, как и в предыдущем примере, пренебрежем. Вследствие симметрии сила F распределится между опорами поровну. Поместим начало координат в точку A нейтральной линии, расположенную над левой опорой. Отсечем мысленно слева часть балки, проведя нормальное сечение через произвольную точку $C(x)$ (с координатой x), расположенную левее центра O ($x < l/2$, где l — длина балки). Справа на отсеченную часть балки будет действовать сила $F/2$, направленная вниз. Момент внешних сил, действующих на отсеченную часть, будет $M = (F/2)x$. Уравнение равновесия принимает вид

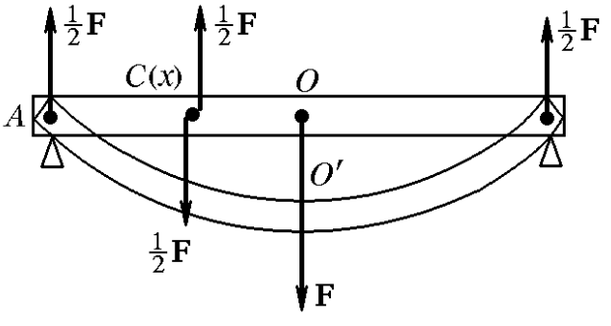


Рис. 210

$$EIy'' = -\frac{F}{2} x \left(x \leqslant \frac{l}{2}\right). \tag{80.10}$$

Теперь ось Y направлена вниз, т. е. в сторону выпуклости балки. Производная y'' отрицательна. По этой причине правая часть взята со знаком минус. Интегрируя полученное уравнение и учитывая, что $y' = 0$ при $x = l/2$, $y = 0$ при $x = 0$, найдем

$$y = \frac{Fx}{48EI}(3l^2 - 4x^2) \quad \left(x \leqslant \frac{l}{2}\right). \tag{80.11}$$

Полагая здесь $x = l/2$, находим стрелу прогиба

$$\lambda = \frac{Fl^3}{48EI}. \tag{80.12}$$

Этот результат можно также получить из формулы (80.9). Действительно, в точке O' касательная к нейтральной линии изогнутой балки горизонтальна (рис. 210). Если мысленно разрезать балку нормальным сечением, проведенным через O', на две равные части, то каждая половина будет эквивалентна балке, жестко закрепленной одним концом в точке O' и подверженной на свободном конце действию сосредоточенной силы $F/2$, направ-

ленной вверх. Следовательно, стрела прогиба центра балки найдется из формулы (80.9), если в ней сделать замену $F \to F/2$, $l \to l/2$. Это дает

$$\lambda = \frac{1}{3EI}\frac{F}{2}\left(\frac{l}{2}\right)^3 = \frac{Fl^3}{48EI},$$

т. е. прежний результат (80.12).

Пример 4. Определим стрелу прогиба центра однородной балки с жестко закрепленными концами под действием сосредоточенной силы F, приложенной к ее середине (рис. 211). Снова будем пренебрегать весом балки. Когда балка свободно лежала на двух опорах (рис. 210), влияние последних сводилось к силам $F/2$ и $F/2$, с которыми точечные опоры давили на балку. В случае балки с жестко закрепленными концами результирующая сил реакций опоры, действующих на какой-либо конец балки, по-прежнему равна $F/2$. Но помимо этого силы реакции создают вращающий момент M, действующий на балку. Поэтому вместо уравнения (80.10) надо писать

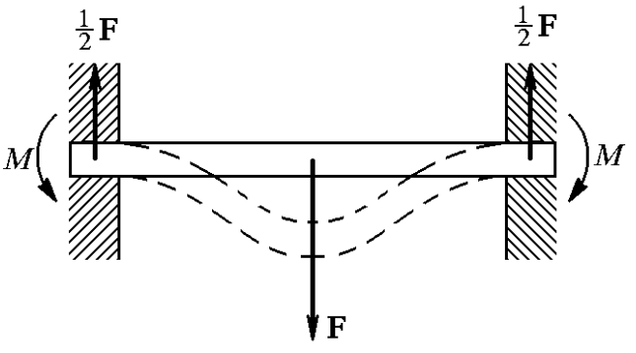


Рис. 211

$$EIy'' = -\frac{F}{2}x + M \quad \left(x \leqslant \frac{l}{2}\right), \tag{80.13}$$

считая вращательный момент M неизвестным и подлежащим определению. Уравнение надо решить при условиях: 1) $y' = 0$ при $x = 0$; 2) $y' = 0$ при $x = l/2$; 3) $y = 0$ при $x = 0$. Это дает

$$M = \frac{1}{8}Fl,$$

$$y = \frac{1}{16}\frac{Fx^2}{EI}\left(l - \frac{4}{3}x\right), \tag{80.14}$$

$$\lambda = \frac{1}{192}\frac{Fl^3}{EI}.$$

Пример 5. Рассмотрим теперь изгиб балки под действием собственного веса P, предполагая, что один конец ее закреплен в стене, а другой свободен (рис. 212). Для равновесия всей балки необходимо, чтобы стена действовала на конец балки O с силой, направленной вверх и равной ее весу P. Проведем нормальное сечение через произвольную точку $B(x)$ нейтральной линии (с координатой $OB = x$). В примере 2 при решении аналогичной задачи мы исходили из условия равновесия части балки BA. Так же можно было бы поступить и при решении рассматриваемой задачи. Однако мы хотим теперь воспользоваться условием равновесия другой части балки, OB, чтобы показать, как поступать в этом случае. Пусть F — сила, действующая на правый конец рассматриваемой части балки OB со стороны части BA. Вес части

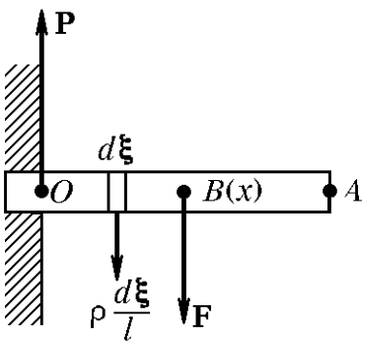


Рис. 212

OB равен Px/l. Для равновесия этой части необходимо условие $P = F + Px/l$ или $F = P(1 - x/l)$. На элемент балки $d\xi$ действует сила веса $P\frac{d\xi}{l}$. Момент M_1 всех вертикальных сил, действующих на часть OB, не зависит от положения оси, относительно которой он берется. Возьмем в качестве таковой ось, проходящую через конец O. Получим

$$M_1 = Fx + \int_0^x P\xi \frac{d\xi}{l} = Px - P\frac{x^2}{2l}.$$

Сюда надо добавить еще момент горизонтальных сил упругих напряжений, действующих на закрепленный конец O. Обозначая этот момент через M_2, для полного момента сил, действующих на часть OB, может написать

$$M = Px - P\frac{x^2}{2l} + M_2. \tag{80.15}$$

Постоянную M_2 можно найти из условия равновесия всей балки OA. На ее свободном конце не действуют никакие силы и упругие напряжения. Поэтому, полагая в (80.15) $x = l$, мы найдем полный момент сил, действующих на всю балку. В равновесии он должен равняться нулю, т. е. $Pl - P\frac{l^2}{2l} + M_2 = 0$. Отсюда $M_2 = -Pl/2$. Это дает

$$M = Px - P\frac{x^2}{2l} - P\frac{l}{2}. \tag{80.16}$$

Уравнение равновесия части балки OB принимает вид

$$EIy'' = -Px + P\frac{x^2}{2l} + P\frac{l}{2}.$$

Решая его при условиях: 1) $y' = 0$ при $x = 0$; 2) $y = 0$ при $x = 0$, получим

$$y = \frac{P}{4EI} lx^2 - \frac{P}{6EI} x^3 + \frac{P}{24EI}\frac{x^4}{l}. \tag{80.17}$$

Полагая здесь $x = l$, находим стрелу прогиба свободного конца балки

$$\lambda = \frac{P}{8EI} l^3. \tag{80.18}$$

Если на свободный конец балки действует еще внешняя сила F, направленная вниз, то вместо формул (80.16) и (80.18) нетрудно получить

$$M = F(x - 1) + Px - P\frac{x^2}{2l} - \frac{Pl}{2}, \tag{80.19}$$

$$\lambda = \frac{l^3}{EL}\left(\frac{F}{3} + \frac{P}{8}\right). \tag{80.20}$$

Результирующий прогиб, таким образом, равен сумме прогибов, получающихся при раздельном и независимом действии сил F и P. Этот результат справедлив для любых малых деформаций, а не только для деформаций изгиба, что непосредственно следует из принципа суперпозиции.

Пример 6. Упругий стержень AB длиной l сдавливается с концов двумя равными и противоположно направленными силами F, действующими вдоль одной прямой (рис. 213). Концы стержня, закрепленные в шарнирах,

могут свободно перемещаться по направлению действия сил. При известной нагрузке F стержень начинает изгибаться в сторону. Это показывает, что помимо сжатого состояния стержня возможны и другие равновесные состояния его. На рис. 213 изображен стержень в таком изогнутом состоянии. Направим ось X вдоль продольной оси недеформированного стержня, а ось Y — вбок в сторону изгиба. Уравнение равновесия изогнутого стержня имеет вид

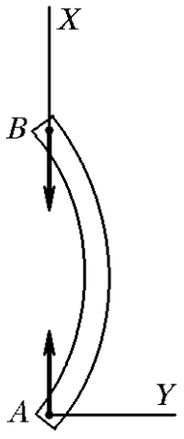


Рис. 213

$$y'' + k^2 y = 0, \tag{80.21}$$

где введено обозначение

$$k^2 = \frac{F}{EI}. \tag{80.22}$$

Общее решение этого уравнения есть

$$y = C \cos kx + D \sin kx,$$

где C и D — постоянные интегрирования. Если начало координат поместить на одном из концов стержня, то должно быть $C = 0$. Это следует из того, что при $x = 0$ ордината должна обращаться в нуль. На другом конце, т. е. при $x = l$, ордината y также равна нулю. Значит, должно также быть $D \sin kl = 0$ *). Если $\sin kl \neq 0$, то $D = 0$, а потому $y = 0$. В этом случае стержень может быть только сжат, но не изогнут. Если же $\sin kl = 0$, т. е. $kl = \pi, 2\pi, 3\pi, \ldots$, то прямолинейная равновесная форма стержня хотя и является теоретически возможной, но, как легко доказать, она будет неустойчивой. Стержень принимает форму дуги синусоиды в соответствии с уравнением $y = D \sin kx$, причем постоянная D зависит от прогиба, т. е. в конце концов от нагрузки. Значения l и F, соответствующие наименьшему корню ($kl = \pi$),

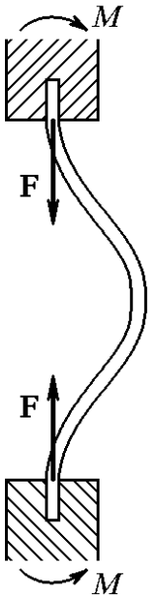


Рис. 214

$$l = \frac{\pi}{k} = \pi\sqrt{\frac{EI}{F}} \quad \text{и} \quad F = \frac{\pi^2 EI}{l^2}, \tag{80.23}$$

называются соответственно *критической длиной и предельной нагрузкой* при продольном изгибе. Эти величины можно рассматривать как предельные значения длины или нагрузки, при которых стержень начнет изгибаться в сторону, если только до этого он не был разрушен действующими силами.

Если оба конца стержня жестко закреплены (рис. 214), то надо учесть дополнительные моменты сил, действующие на концах стержня, подобно тому, как это делалось в примере 4. Вместо уравнения (80.21) теперь надо решать уравнение

$$y'' + k^2 y = k^2 C,$$

где C — постоянная, подлежащая определению. Общее решение этого уравнения:

$$y = A \cos kx + B \sin kx + C.$$

*) Строго говоря, под l здесь следует понимать не длину стержня, а расстояние вдоль прямой между концами изогнутого стержня. Это расстояние, очевидно, меняется с изменением нагрузки и является величиной, подлежащей определению. Однако при малых деформациях такое уточнение несущественно, и под l можно понимать длину самого стержня.

Условие $y = 0$ при $x = 0$ дает $A + C = 0$. Из второго условия $y' = 0$ при $x = 0$ получаем $B = 0$, так что

$$y = A\,(\cos kx - 1).$$

Надо еще потребовать, чтобы y и y' обращались в нуль и на другом конце стержня. Это даст два новых условия: 1) $\cos kl = 1$; 2) $\sin kl = 0$. Из них получаем $kl = 2\pi, 4\pi, \ldots$ Критическая длина в этом случае в два, а предельная нагрузка в четыре раза больше, чем в предыдущем:

$$l = \frac{2\pi}{k} = 2\pi\sqrt{\frac{EI}{F}},\quad F = \frac{4\pi^2 EI}{l^2}. \tag{80.24}$$

Если один конец стержня жестко закреплен, а другой закреплен в шарнире, то для тех же величин получаем

$$l = \frac{1}{2}\pi\sqrt{\frac{EI}{F}},\quad F = \frac{\pi^2 EI}{4l^2}. \tag{80.25}$$

ЗАДАЧИ

1. Определить стрелу прогиба центра однородной балки под действием собственного веса P, если балка лежит своими концами на двух опорах.

Ответ. $\lambda = \frac{5}{384}\frac{Pl^3}{EI}$.

2. То же для балки, обоими концами жестко закрепленной в стене.

Ответ. $\lambda = \frac{1}{384}\frac{Pl^3}{EI}$.

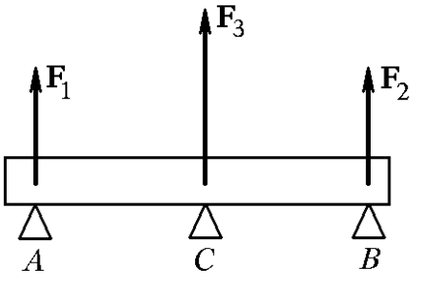


Рис. 215

3. Определить распределение веса P балки, лежащей на трех опорах A, B, C (рис. 215). Средняя опора C расположена посередине между крайними опорами A и B и смещена на величину λ вниз относительно горизонтальной плоскости, в которой лежат крайние опоры.

Решение. При равновесии $F_1 + F_2 + F_3 = P$, причем в силу симметрии $F_1 = F_2$. Мысленно уберем опоры, заменив их силами F_1, F_2, F_3, с которыми они давили на балку. Кроме того, закрепим балку посередине. От этого деформации балки не изменятся. Воспользуемся формулами (80.9) и (80.18). Под действием силы F_1 левый конец балки поднимется относительно средней опоры на величину $y_1 = \frac{F_1}{3EI}\left(\frac{l}{2}\right)^3$. Под действием собственного веса тот же конец опустится вниз на $y_2 = \frac{P}{2}\frac{1}{8EI}\left(\frac{l}{2}\right)^3$. Общее поднятие вверх будет $y_1 - y_2$. По условию оно равно λ. В результате получим

$$F_1 = F_2 = \frac{3}{16}P + \frac{24EI\lambda}{l^3},\quad F_3 = \frac{5}{8}P - \frac{48EI\lambda}{l^3}.$$

Когда все три опоры находятся на одной высоте, то

$$F_1 = F_2 = \frac{3}{16}P,\quad F_3 = \frac{5}{8}P.$$

В этом случае распределение веса балки между тремя опорами не зависит от ее упругих свойств, хотя без учета последних задача становится неопределенной (ср. § 44). Эта зависимость объясняется тем, что мы не учитывали деформацию самих опор.

4. Та же задача (см. рис. 215), но опора C не находится посередине между опорами A и B ($AC = a$, $CB = b$).

Решение. Поместим начало координат в нейтральном сечении над опорой A, направив ось X вправо, а ось Y — вниз. Написав уравнение равновесия для частей $AC(x \leqslant a)$ и $CB(x \geqslant a)$ и интегрируя их при условиях $y = 0$ при $x = 0$ и $x = a + b \equiv l$, а также $y = \lambda$ при $x = a$, получим

$$y = \frac{\lambda}{a} x + \frac{F_1 x}{6EI} (a^2 - x^2) - \frac{Px}{24EIl} (a^3 - x^3) \quad (x \leqslant a),$$

$$y = \frac{\lambda}{b} (l - x) + \frac{F_2 (l-x)}{6EI} [b^2 - (l-x)^2] - \frac{P(l-x)}{24EIl} [b^3 - (l-x)^3] \quad (x \geqslant a).$$

Далее, надо потребовать, чтобы в точке C у балки не было излома, т. е. чтобы первые производные обоих выражений в точке $x = a$ совпадали. Наконец, надо учесть, что при равновесии суммы всех внешних сил и их моментов, действующих на балку в целом, равны нулю. В результате получим

$$F_1 = \frac{3EI}{a^2 b} \lambda + \frac{P}{8} \frac{3a^2 + ab - b^2}{a(a+b)},$$

$$F_2 = \frac{3EI}{ab^2} \lambda + \frac{P}{8} \frac{3b^2 + ab - a^2}{b(a+b)},$$

$$F_3 = -\frac{3EI(a+b)}{a^2 b^2} \lambda + \frac{P}{8} \frac{3ab + a^2 + b^2}{ab}.$$

5. Цилиндрический стержень и трубка одинаковой длины и массы, изготовленные из одного и того же материала, лежат своими концами на двух опорах и прогибаются под действием собственного веса. Определить отношение их стрел прогиба λ_1/λ_2, если радиус стержня равен r, а наружный радиус трубки R.

Ответ. $\lambda_1/\lambda_2 = (2R^2 - r^2)/r^2$.

6. Если на две опоры положить концами бумажный лист, то он прогибается и падает под действием собственного веса. Если же лист скатать в сплошной цилиндр или свернуть в трубку, склеив его края, то получившиеся тела ведут себя как твердые. Их можно даже нагружать без заметного прогиба. Вычислив моменты инерции I_1, I_2, I_3 соответствующих поперечных сечений, объяснить явление. Длина листа бумаги (расстояние между опорами) l, ширина a, толщина h.

Ответ. $I_1 = \frac{1}{12} ah^3$, $I_2 = \frac{1}{4\pi} (ah)^2$, $I_3 = \frac{1}{8\pi^2} a^3 h$.

7. Из круглого бревна диаметром D требуется изготовить балку прямоугольного поперечного сечения, чтобы ее изгиб был минимальным. Определить ширину a и толщину h такой балки.

Ответ. $a = \frac{D}{2}$, $b = \frac{\sqrt{3}}{2} D$. Задача сводится к исследованию экстремума выражения ab^3 при дополнительном условии $a^2 + b^2 = \text{const}$.

§ 81. СКОРОСТЬ РАСПРОСТРАНЕНИЯ ПРОДОЛЬНЫХ УПРУГИХ ВОЗМУЩЕНИЙ В СТЕРЖНЯХ

1. Если в каком-либо месте упругой среды возникла деформация, то по прекращении внешних воздействий она не остается на месте, а распространяется в среде во всех направлениях. В таких случаях говорят о распространении в среде *упругих возмущений* или *волн*. Примерами могут служить звуковые волны в твердых телах, жидкостях или газах. Закрепим, например, в горизонтальном положении длинный железный стержень. Если ударить молотком по одному концу стержня, то на этом конце возникает деформация сжатия, которая начинает распространяться вдоль стержня с большой скоростью. Чтобы обнаружить такую деформацию, наденем на стержень проволочную катушку, концы которой присоединим к осцил-

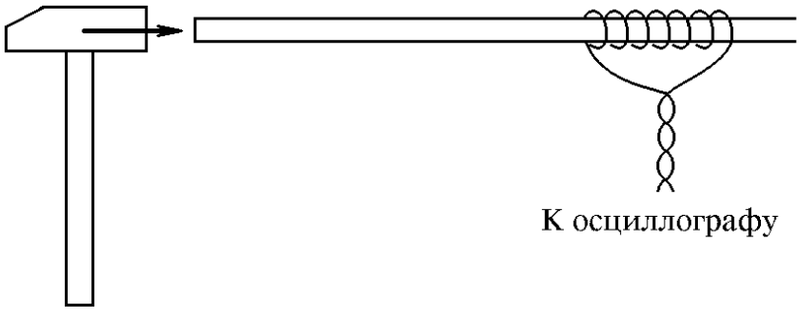


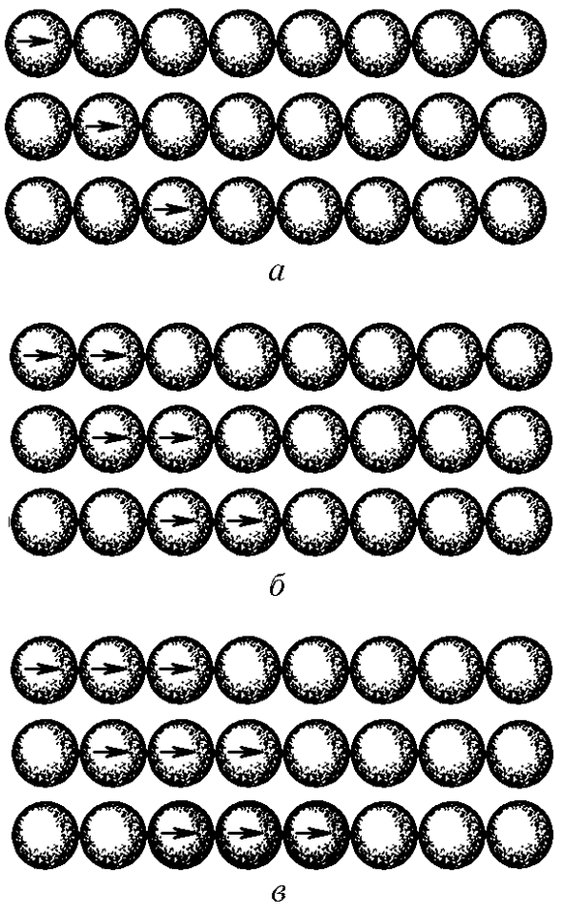
Рис. 216

лографу (рис. 216). Железный стрежень всегда намагничен, хотя бы потому, что он находится в магнитном поле Земли. Пока нет возмущения, магнитный поток через катушку остается постоянным, и электрический ток через нее не идет. Но если возмущение достигает той части стержня, на которую надета катушка, то магнитный поток через нее изменяется. Возникает индукционный электрический ток, фиксируемый осциллографом.

Проследить за распространением упругого возмущения вдоль стержня довольно затруднительно из-за большой скорости распространения и малости самого возмущения. Но это легко сделать на модели, взяв вместо стержня длинную спиральную пружину из мягкой проволоки, подвешенную горизонтально на нескольких нитях. Если по одному концу пружины нанести легкий удар, то видно, как деформация сжатия распространяется вдоль пружины. Если же конец пружины был оттянут, то возникает *деформация растяжения*, также распространяющаяся с определенной скоростью вдоль пружины.

2. Важным является вопрос о *скорости распространения упругих возмущений*. Рассмотрим этот вопрос сначала для упругих возмущений, распространяющихся вдоль стержня. Начнем с модели. Пусть имеется прямолинейный ряд, состоящий из одинаковых твердых идеально упругих шаров, соприкасающихся между собой. Ряд таких шаров неограниченно простирается вправо (рис. 217).

Модель не предназначена непосредственно для решения вопроса о скорости распространения упругих возмущений в стержне. Но она позволяет простейшим образом составить представление о распределении скорости движения вещества в стержне, когда в нем распространяется возмущение, возникшее в результате действия определенной силы. Нанеся удар по первому шару, сообщим ему некоторую скорость v (рис. 217 *а*). Первый шар ударится о второй. При упругом ударе шары просто обмениваются скоростями: первый шар остановится, а второй придет в движение с той же скоростью v (см. § 28). Затем второй шар передаст движение третьему, а сам остановится и т. д. Движение будет передаваться от шара к шару. В результате возникает возмущение, распространяющееся вдоль ряда шаров. Скорость распространения такого возмущения обозначим через c. Ее нельзя смешивать со скоростью v того шара, который в рассматриваемый момент движется. Изменим теперь постановку опыта. В тот момент, когда при столкновении со вторым шаром первый шар остановится, нанесем по нему второй удар, чтобы он приобрел прежнюю скорость v. Тогда в этот момент первые два шара будут иметь одну и ту же общую скорость v. Затем при ударе о третий шар второй шар передаст ему свою скорость, а сам остановится. Первый шар при столкновении со вторым сделает то же самое. В результате движение перейдет от первых двух шаров ко второму и третьему. Затем оно будет передано третьему и четвертому шарам и т. д. Короче говоря, вдоль ряда шаров побежит возмущение, в котором в каждый момент движутся какие-то два шара, соприкасающиеся между собой, а остальные покоятся (рис. 217 *б*). Допустим теперь, что всякий раз, как первый шар передает свое движение второму шару, он получает удар, в результате которого его скорость v восстанавливается. Состояние движения представлено на схематическом рис. 217 *в*. Все шары, расположенные левее некоторой границы, движутся с одной и той же скоростью v, а шары, расположенные правее этой границы, находятся в состоянии покоя. Сама граница перемещается вправо со скоростью c, так что в движение вовлекаются все новые и новые шары.



а

б

в

Рис. 217

Очевидно, ничто не изменится, если вместо шаров взять прямолинейный ряд, состоящий из упругих цилиндриков, соприкасающихся между собой своими основаниями (рис. 218). Это замечание позволяет легко выполнить предельный переход к сплошной среде. Допустим, что длины цилиндриков неограниченно уменьшаются, а число их неограниченно растет. Вместе с тем удары, которым подвергается первый цилиндрик, становятся все чаще и чаще, а сила каждого удара — все слабее и слабее. В пределе получится сплошной стержень, на свободный конец которого действует постоянная сила F (рис. 219). От реального стержня наша модель отличается тем, что она не оказывает сопротивления на разрыв. Но это несущественно, когда рассматривается вопрос о распространении возмущения сжатия, поскольку сопротивлением на сжатие модель обладает. Можно было бы усовершенствовать модель, введя между цилиндриками пружинки пренебрежимо малой массы, связывающие их между собой. Но при рассмотрении возмущений сжатия в этом нет необходимости. Мгновенное состояние движения стержня, возникшее под действием постоянной силы F, может быть охарактеризовано следующим образом. Вещество стержня, находящееся левее некоторой границы B, движется с одной и той же постоянной скоростью v, а вещество правее этой границы находится в покое. Сама граница B перемещается вправо с постоянной скоростью c. В акустике, как правило, имеют дело с так называемыми *малыми возмущениями*. В этих случаях скорость вещества v бывает очень мала по сравнению со скоростью распространения возмущения c. Нарушение этого условия наблюдается только в случае очень сильных возмущений, называемых *ударными волнами*, которые здесь рассматриваться не будут. Мы ограничимся исследованием распространения только малых возмущений.

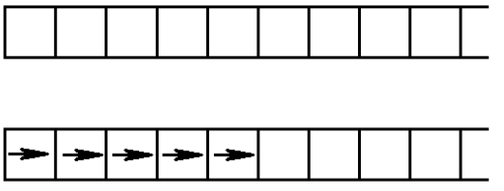


Рис. 218

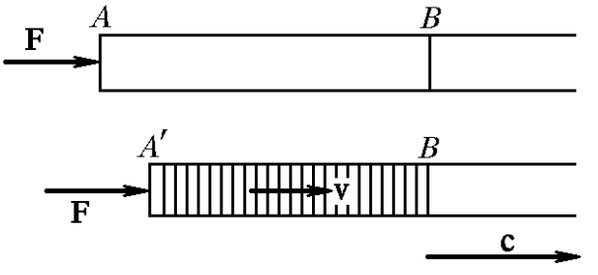


Рис. 219

3. Вычислим скорость распространения *малых продольных возмущений* в стержне, возникших в результате действия постоянной силы давления **F**, приложенной в некоторый момент к его свободному концу (рис. 219). Этот момент в дальнейшем принимается за нулевой, т. е. за начало отсчета времени. В возмущенной области стержня все вещество в любой момент времени t движется с постоянной скоростью v, а сам стержень в указанной области всюду деформирован одинаково. Если m — масса деформированной части стержня в момент t, то его импульс в тот же момент будет mv. При-

ращение импульса стержня за время dt, т. е. $d(mv)$, равно импульсу силы $F\,dt$ за то же время. Это дает

$$\frac{d(mv)}{dt} = F\ ^{*)}. \tag{81.1}$$

За время t возмущение проходит путь $l = ct$, так что масса возмущенной области стержня будет $m = \rho Sct$, где S — площадь поперечного сечения стержня, а ρ — его плотность. Строго говоря, под S и ρ в этом выражении следовало бы понимать значения этих величин для *невозмущенного стержня*. Однако в пределах принятой здесь точности расчета в соотношениях подобного рода нет необходимости учитывать разницу между значениями ρ, S и аналогичных величин в возмущенном и невозмущенном состояниях. Это необходимо делать только при рассмотрении сильных возмущений. Подставив в формулу (81.1) $m = \rho Sct$, $F = PS$, где P — давление в возмущенной области стержня, получим

$$P = \rho cv. \tag{81.2}$$

Давление P связано с относительным сжатием стержня соотношением $P = E\varepsilon$. Для нахождения ε заметим, что к моменту времени t правый конец сжатой области стержня B еще не успел переместиться, тогда как левый свободный конец его A' двигался в течение времени t и переместился на расстояние vt. В результате длина возмущенной области стержня по сравнению со своей исходной длиной укоротится на $\Delta l = vt$. Поэтому

$$\varepsilon = \frac{\Delta l}{l} = \frac{v}{c}, \tag{81.3}$$

$$P = \frac{Ev}{c}. \tag{81.4}$$

Исключая P из формул (81.2) и (81.4), получим

$$c = \sqrt{\frac{E}{\rho}}. \tag{81.5}$$

Этой формулой и определяется скорость распространения упругих возмущений в рассматриваемом случае.

4. Работа, совершаемая силой F за время t, равна $A = Fvt = PS\varepsilon ct = P\varepsilon V$, где V — объем возмущенной части стержня. С другой стороны, потенциальная энергия, запасенная при сжатии,

*) Если раскрыть производную, то получится

$$m\frac{dv}{dt} = F - v\frac{dm}{dt}. \tag{81.1a}$$

Это соотношение является частным случаем уравнения (21.2). Достаточно заметить, что возмущенную часть стержня можно рассматривать как тело с переменной массой, причем $v_{\text{отн}} = -v$. Формулу (81.2), которая выводится ниже, можно получить и из уравнения (81.1а), заметив, что в рассматриваемом случае

$$\frac{dv}{dt} = 0, \quad \frac{dm}{dt} = S\rho c.$$

$U = \frac{1}{2} P\varepsilon V$. Таким образом, $U = \frac{1}{2} A$. Только половина работы идет на увеличение потенциальной энергии стержня. Другая половина тратится на приращение кинетической энергии. *В каждый момент времени кинетическая энергия равна потенциальной.* Этим свойством, как будет показано в следующем параграфе, обладает любое малое возмущение, распространяющееся в *одном направлении.*

5. Если в некоторый момент времени сила F прекратит свое действие, то в стержне образуется возмущенная область, ограниченная с обеих сторон. Это нетрудно понять, воспользовавшись прежней моделью из прямолинейного ряда соприкасающихся упругих шаров (см. рис. 217) и выполнив затем предельный переход к сплошному стержню. Таким же путем нетрудно убедиться, что обе границы возмущенной области должны распространяться в одном направлении и с одной и той же скоростью. Последняя определяется формулой (81.5). Для доказательства достаточно в возмущенной области провести произвольное сечение (рис. 220), все время состоящее из одних и тех же частиц. Очевидно, такое сечение будет двигаться вправо со скоростью вещества v. Оно играет роль свободного конца стержня. На него оставшаяся часть деформированного стержня, расположенная левее, давит с силой $F = PS$. Поэтому к части стержня, расположенной правее рассматриваемого сечения, полностью применимо наше прежнее рассуждение. Из него следует, что граница возмущений области B будет распространяться вправо со скоростью c, определяемой формулой (81.5).

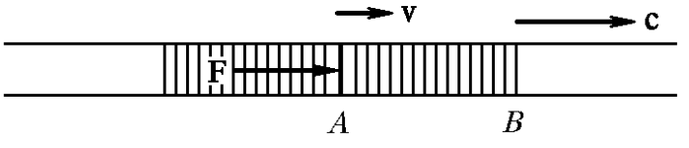


Рис. 220

6. Рассуждение не меняется существенно, если вместо постоянной силы давления к концу стержня приложить в некоторый момент времени постоянную силу натяжения. Разница состоит только в том, что по стержню вместо возмущения сжатия побежит *возмущение разрежения.* Скорость распространения такого возмущения по-прежнему будет определяться формулой (81.5). Модель, состоящая из соприкасающихся упругих шаров, в этом случае, конечно, неприменима. Но ее можно заменить моделью, в которой соприкасающиеся шары связаны между собой бесконечно короткими пружинками пренебрежимо малой массы.

7. В предыдущих рассуждениях предполагалось, что возмущение в стержне вызывается *постоянной силой*, приложенной к его концу в какой-то момент времени. Обобщение на случай *переменной силы* не представляет труда. Обратимся к нашей прежней модели, состоящей из ряда упругих шаров, но скрепленных пружинками пренебрежимо малой массы. Если по первому шару наносить удары различной силы в определенный моменты времени, то и сообщаемые ему скорости будут различными. В соответствии с этим распределение скоростей можно представить прежними схематическими рисунками (см. рис. 217). Однако скорость v будет меняться от шара к шару. Выполнив предель-

ный переход к непрерывному стержню, получим возмущение, распространяющееся в определенном направлении, в котором скорость вещества непрерывно меняется от точки к точке. Может изменяться даже направление скорости v, если сила, приложенная к концу стержня, меняет свое направление. Возмущенная область будет ограничена с обеих сторон, если возбуждающая сила действует ограниченное время. Докажем, что для рассматриваемого возмущения остаются справедливыми формулы (81.2) и (81.3), а следовательно, и формула (81.5). На рис. 221 возмущенная область заштрихована и представлена в два бесконечно близких момента времени t и $t + dt$. За время dt возмущенная область перемещается на расстояние $c\,dt$. Проведем в возмущенной области произвольное сечение A, состоящее из одних и тех же частиц

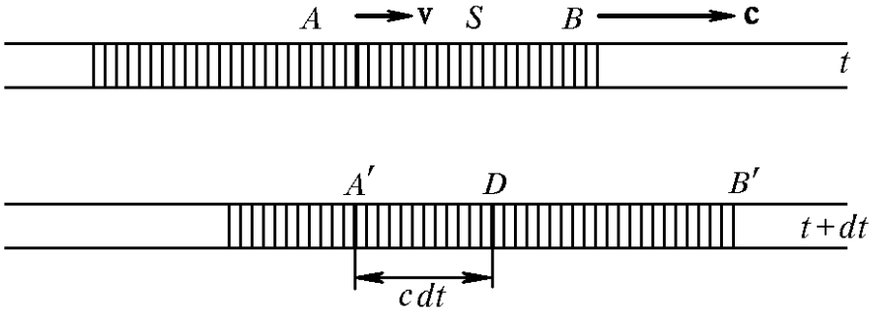


Рис. 221

вещества. Оно перемещается вправо со скоростью v, которую имеют частицы вещества в сечении A в момент времени t. За время dt частицы переместятся в A', пройдя малое расстояние $v\,dt$, которым мы пренебрегаем. Само возмущение переместится на много большее расстояние $c\,dt$. Найдем приращение импульса вещества, расположенного правее выделенного сечения A. Возмущение из точки A переместится в точку D, пройдя расстояние $c\,dt$. Вещество, расположенное правее D, в момент $t + dt$ будет обладать в точности таким же движением, каким обладало в момент t вещество, расположенное правее A. Поэтому ясно, что искомое приращение импульса будет равно импульсу, локализованному между сечениями A' и D, т. е. $Sc\,dt\,\rho v$. Оно равно импульсу сил давления $PS\,dt$, действующих в течение времени dt в сечении A. Приравнивая оба выражения, получаем формулу (81.2). Так же легко получить формулу (81.3). Рассмотрим бесконечно малую возмущенную область $A'D$ (рис. 221). Ее первоначальная длина была равна $l = c\,dt$. Но возмущение достигло сечения A' на время dt раньше, чем сечения D. Благодаря этому путь, пройденный веществом, связанным с сечением A', будет на $v\,dt$ длинее пути, пройденного веществом, связанным с сечением D. Значит, укорочение области $A'D$ в результате деформации равно $\Delta l = v\,dt$. Разделив Δl на l, получим формулу (81.3).

Плотность кинетической энергии в возмущенной области $w_{\text{кин}} = \frac{1}{2}\rho v^2$. Плотность потенциальной энергии $w_{\text{пот}} = \frac{E}{2}\varepsilon^2 =$

$= \frac{E}{2}\frac{v^2}{c^2}$. Подставив сюда выражение для c из формулы (81.5), получим $w_{\text{пот}} = 1/2\ \rho v^2$. Таким образом, $w_{\text{кин}} = w_{\text{пот}}$. *Во всяком бегущем упругом возмущении, т. е. возмущении, распространяющемся в определенном направлении, полная энергия распределяется поровну между кинетической и потенциальной.*

§ 82. ПРИМЕНЕНИЯ ПРИНЦИПА СУПЕРПОЗИЦИИ

1. Мы пользовались уже *принципом суперпозиции в статике.* Но этому принципу подчиняется также *распространение малых возмущений.* Пусть в среде распространяется какое-либо возмущение. Смещение какой-либо частицы среды из положения равновесия в этом возмущении обозначим через $\mathbf{s}_1(\mathbf{r}_0, t)$. Вектор $\mathbf{r}_0$ означает радиус-вектор рассматриваемой точки в состоянии покоя, т. е. до того момента, когда до нее дошло возмущение. Пусть $\mathbf{s}_2(\mathbf{r}_0, t)$ — смещение во втором возмущении в той же среде. Какое возмущение возникнет в среде, если в ней возбудить оба эти возмущения? Принцип суперпозиции утверждает, что результирующее смещение будет

$$\mathbf{s}(\mathbf{r}_0, t) = \mathbf{s}_1(\mathbf{r}_0, t) + \mathbf{s}_2(\mathbf{r}_0, t).$$

Это означает, что *всякое возмущение, существующее в среде, не влияет на распространение другого возмущения. Каждое возмущение распространяется так, как если бы других возмущений в среде не было.* Примером могут служить волны на поверхности воды. Если на спокойную поверхность пруда бросить два камня, то из точек падения будут распространяться круговые волны. Там, где они накладываются одна на другую, возникает довольно сложное результирующее движение. Но каждая волна после прохождения через область наложения остается в точности такой же, какой она была бы при отсутствии другой волны. Разумеется, принцип суперпозиции справедлив не только для двух, но для произвольного числа возмущений, накладывающихся друг на друга. Принцип суперпозиции в том виде, в каком он сформулирован выше, следовало бы назвать *принципом суперпозиции смещений.* Но он справедлив и для *скоростей частиц,* поскольку скорости получаются дифференцированием смещений по времени. Он верен и для *упругих напряжений,* поскольку последние линейно выражаются через деформации, т. е. смещения. Принцип суперпозиции можно рассматривать как опытный факт. Он является также следствием *линейности уравнений* (относительно смещений), которым описываются малые возмущения. *Для сильных возмущений принцип суперпозиции не справедлив.*

2. В предыдущем параграфе было показано, что полная энергия *бегущего возмущения* распределяется поровну между кинетической

и потенциальной. Необходимость такого результата выступает особенного отчетливо, если к вопросу подойти с точки зрения принципа суперпозиции. Для определенности рассмотрим возмущения, распространяющиеся вдоль стержня, хотя наши рассуждения имеют общий характер. Пусть в начальный момент времени некоторая область стержня деформирована, но все вещество внутри этой области находится в покое. Вся начальная энергия стержня будет чисто потенциальной. Обозначим ее через E. Если убрать внешние силы, создавшие начальную деформацию, то из возмущенной области вдоль стержня в противоположных направлениях побегут два возмущения. Если первоначальное возмущение было симметрично, то, очевидно, полная энергия E разделится поровну между обоими возмущениями, возникшими из него. Покажем теперь, что в каждом из этих двух бегущих возмущений кинетическая энергия равна потенциальной. Для этого рассмотрим оба возмущения в начальный момент времени, когда они полностью перекрываются. Если P_1 и P_2 — давления, а v_1 и v_2 — скорости вещества в обоих возмущениях, то по принципу суперпозиции в начальный момент $P_1 + P_2 = P$, $v_1 + v_2 = 0$, где P — давление в возмущенной области в тот же момент времени. В силу симметрии $P_1 = P_2 = \frac{1}{2} P$. Такое же соотношение между давлением в соответствующих точках сохранится и в каждый последующий момент времени. В частности, оно останется справедливым и тогда, когда оба возмущения разойдутся, т. е. перестанут накладываться друг на друга. Тогда уже имеет смысл говорить о разделении полной энергии между возмущениями, возникшими из начальной возмущенной области. Так как потенциальная энергия пропорциональна квадрату давления, то потенциальная энергия в каждом из бегущих возмущений будет $E/4$, а потенциальная энергия обоих возмущений вместе — $E/2$. Для сохранения энергии необходимо, чтобы другая половина полной энергии перешла в кинетическую. Понятно, что и кинетическая энергия распределится поровну между обоими бегущими возмущениями. Таким образом, в каждом бегущем возмущении кинетическая и потенциальная энергии будут одинаковы и равны $E/4$.

3. Приведенное рассуждение, поскольку оно основано на соображениях симметрии, не вызывает возражений, если начальное распределение деформации само обладает требуемой симметрией. Но рассуждение остается применимым и в тех случаях, когда это условие не выполняется. Чтобы убедиться в этом, достаточно мысленно разбить начальную возмущенную область на бесконечно малые области. Внутри каждой из таких бесконечно малых областей давление можно считать постоянным, а его распределение можно изобразить в виде бесконечно узкого прямоугольника. Таким образом, начальное распределение давления в каждой из бесконечно малых возмущенных областей будет обладать требуемой симметрией. По принципу суперпозиции возмущения, исходящие из каждой беско-

нечно малой области, совершенно не зависят от того, возмущены или нет другие бесконечно малые области. Поэтому к этим возмущениям полностью применимы рассуждения, приведенные выше. За время t возмущения из рассматриваемой бесконечно малой области

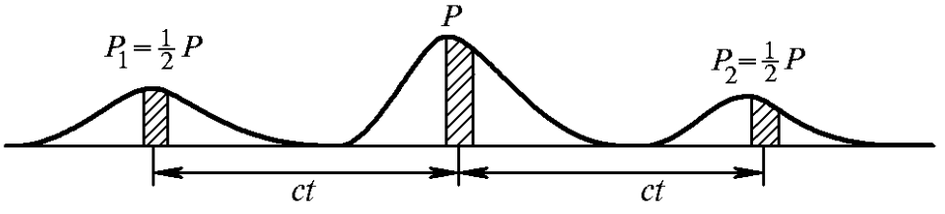


Рис. 222

распространяются на расстояние ct. Если возмущения, возникшие из всей возмущенной области в момент t уже не перекрываются, то не будут перекрываться и возмущения, возникшие из отдельных бесконечно малых возмущенных областей (рис. 222). Для них остается справедливым соотношение $P_1 - P_2 = \frac{1}{2} P$. Отсюда следует, что *в каждом из бегущих возмущений, возникших из возмущенной области, равны не только полные кинетические и потенциальные энергии, но и их плотности.*

4. В приведенном рассуждении предполагалось, что оба бегущих возмущения возникли из начальной деформированной области, находившейся в *состоянии покоя.* Те же рассуждения, понятно, можно было бы провести и для возмущений, возникших из недеформированных областей, частицам которых в начальный момент времени сообщены скорости, произвольным образом распределенные по этим областям (см. задачу 1 к этому параграфу).

5. Итак, *для того чтобы возмущение было бегущим, необходимо, чтобы плотности кинетической и потенциальной энергий в нем были одинаковы.* Вопрос в том, в какую сторону будет распространяться возмущение, легко решается с помощью энергетических соображений. Пусть, например, возмущенная область AB распространяется вправо (рис. 221). Проведем в ней произвольное сечение S. Чтобы возмущение распространялось вправо, необходимо, чтобы часть стержня AS совершала *положительную работу* над частью SB, т. е. должно быть $Pv > 0$, если условиться считать скорости частиц стержня положительными, когда она направлены вправо. Если $v > 0$, то должно быть $P > 0$, т. е. напряжение в сечении S должно иметь *характер давления.* Если же $v < 0$, то должно быть $P < 0$, т. е. напряжение в сечении S должно сводиться к *натяжению* $T = -P$. Чтобы возмущение распространялось влево, необходимо выполнение условия $Pv < 0$.

Если равенство кинетической и потенциальной энергий в возмущении не имеет места, то возмущение разделится на *два возмущения*, распространяющихся в противоположных направлениях. В общем случае эти возмущения будут уносить разные энергии. Например, если всюду в начальной возмущенной области $Pv > 0$, то

энергия, уносимая вправо, будет больше энергии, уходящей влево. При $Pv < 0$ соотношение между этими энергиями будет обратным. Если же $Pv = 0$, то оба возмущения унесут одинаковые энергии.

6. Из изложенного следует, что в бегущей волне сжатия частицы стержня движутся в *том же направлении*, в каком распространяется само возмущение. Если же возмущение носит характер растяжения, то эти направления *противоположны*. Предположим сначала, что возмущение является сжатием и распространяется вдоль стержня слева направо. Исследуем, что произойдет, когда оно достигнет правого конца стержня. Будем предполагать, что правый конец стержня свободен, т. е. не закреплен. Тогда с приходом возмущения частицы на конце стержня приобретут скорости, направленные вправо. Так как конец стержня свободен, то остановиться эти частицы могут лишь тогда, когда со стороны стержня на них подействуют силы, направленные влево. А для этого стержень у правого конца должен оказаться растянутым. Сжатие на конце стержня переходит *в растяжение*. Последнее вызовет в стержне возмущение растяжения, которое будет распространяться в нем влево (рис. 223). Все происходит так, как если бы в некоторый момент времени был оттянут свободный конец стержня и в нем создана деформация растяжения. В возмущении, распространяющемся налево, поскольку оно является возмущением растяжения, частицы среды должны иметь скорости, направленные вправо. Эти скорости частицы приобретают под влиянием сил натяжения, с которыми на них действуют растянутые части стержня, лежащие правее. Мы видим, что *от свободного конца стержня возмущение сжатия отражается и переходит в возмущение растяжения. Аналогично ведет себя и возмущение растяжения. Оно также отражается от свободного конца и переходит в возмущение сжатия* (рис. 224). В обоих случаях при отражении от свободного конца стержня знак деформации меняется на противоположный, тогда как знак скорости вещества v сохраняется неизменным.

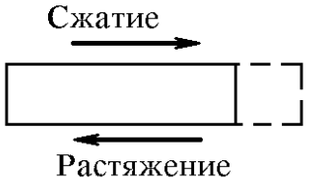


Рис. 223

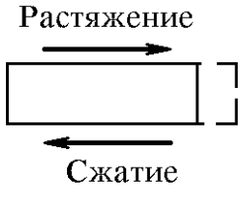


Рис. 224

Иначе ведет себя возмущение при отражении от закрепленного конца стержня. В общем случае возмущение распадается на два: одно возвращается назад в виде отраженного возмущения, другое проходит в среду, с которой граничит закрепленный конец стержня. Только в предельном случае, когда эта среда бесконечно жесткая, возмущение отражается целиком. Рассмотрим этот предельный случай. Когда возмущение достигает границы, то сжатие (растяжение) продолжает оставаться сжатием (растяжением), так как конец стержня закреплен и смещаться не может. Но силы, действующие на этот конец со стороны среды, с которой он граничит, меняют направление скоростей частиц на противоположные. Знаки деформа-

ций при отражении сохраняются, а знаки скоростей изменяются. В результате возмущение сжатия отражается также в виде возмущения сжатия, а возмущение растяжения — в виде возмущения растяжения.

ЗАДАЧИ

1. В упругом стержне создана такая начальная деформация сжатия, что скорости всех частиц в деформированной области направлены в одну сторону (например, вправо), причем в каждой точке плотность потенциальной энергии в α раз превосходит плотность кинетической энергии. Определить, какая доля первоначальной энергии будет унесена возмущением, распространяющимся вправо, а какая доля — возмущением, распространяющимся влево.

Р е ш е н и е. Для простоты введем такие единицы, чтобы плотности кинетической и потенциальной энергий выражались формулами $w_{\text{кин}} = v^2$, $w_{\text{пот}} = P^2$. Представим начальные значения P и v в виде

$$P = P_1 + P_2,\ v = v_1 + v_2.$$

Пусть каждое из начальных возмущений P_1, v_1 и P_2, v_2 порождает возмущение, бегущее в одном направлении. Тогда $P_1^2 = v_1^2$, $P_2^2 = v_2^2$. Если первое возмущение бежит вправо, а второе — влево, то $P_1 v_1 > 0$, $P_2 v_2 < 0$. Учитывая это, получаем

$$P_1 = v_1,\ P_2 = -v_2$$

и далее

$$P_1 = v_1 = \frac{P+v}{2},\ P_2 = -v_2 = \frac{P-v}{2}.$$

Отношение энергий, уносимых возмущениями, равно

$$\frac{E_1}{E_2} = \left(\frac{P_1}{P_2}\right)^2 = \left(\frac{P+v}{P-v}\right)^2 = \frac{1+2\frac{P}{v}+\left(\frac{P}{v}\right)^2}{1-2\frac{P}{v}+\left(\frac{P}{v}\right)^2},$$

или

$$\frac{E_1}{E_2} = \frac{1+2\sqrt{\alpha}+\alpha}{1-2\sqrt{\alpha}+\alpha} = \left(\frac{\sqrt{\alpha}+1}{\sqrt{\alpha}-1}\right)^2.$$

Если в начальном возмущении вся энергия либо кинетическая ($\alpha = 0$), либо потенциальная ($\alpha = \infty$), то $E_1 = E_2$.

2. Стальной цилиндр $l = 10$ м, движущийся вдоль своей оси со скоростью v, сталкивается торцом с таким же неподвижным цилиндром, ось которого является продолжением оси первого цилиндра. Рассматривая упругие возмущения, возбуждаемые при ударе, определить время соударения цилиндров. При каких значениях скорости v наступают пластические деформации цилиндра или их разрушение? Для стали $E = 2 \cdot 10^{12}$ дин/см^2, $\rho = 7{,}8$ г/см^3, предел упругости $P_0 = 2 \cdot 10^9$ дин/см^2.

Р е ш е н и е. В момент соприкосновения цилиндр A движется со скоростью v, цилиндр B покоится, оба цилиндра не деформированы (рис. 225, положение *1*). После того как произойдет удар, от места удара в обе стороны побегут волны сжатия со скоростью c относительно цилиндров (положение *2*). Частицы обоих цилиндров в области сжатия движутся в одну и ту же сторону со скоростью $v/2$. Это следует из закона сохранения импульса. Когда возмущения дойдут до концов цилиндров, все вещество будет двигаться с общей скоростью $v/2$ (положение *3*). Масса движущегося вещества удвоилась, скорость уменьшилась вдвое, так что закон сохранения импульса соблюдается. Кинетическая энергия по сравнению с начальной уменьшилась вдвое. Половина энергии перешла в потенциальную — оба цилиндра равномерно сжаты и прижимаются друг к другу. Затем начинается отражение возмущений от свободных концов цилиндра (положение *4*). Возмущения сжатия переходят в возмущения разрежения. При этом на левом конце давление со стороны смежных областей останавливает частицы вещества, а на правом — ускоряет. Слева возникает недеформированная область, в которой вещество покоится, справа — недеформированная область, в которой вещество движется вправо со скоростью v. Чтобы убедиться в этом, перейдем в систему отсчета, движущуюся вправо со скоростью $v/2$. В начальный момент (положение *3*) оба цилиндра в этой системе отсчета покоятся и равномерно сжаты. При отражении на обоих концах возникают возмущения разрежения: от левого конца разрежение пойдет вправо со скоростью c, от правого — влево с той же скоростью. У свободных концов стержней образуются недеформированные области. Скорости вещества в этих областях (относительно движущейся системы отсчета) должны быть направлены наружу, так как движение в них возникает под действием сил сжатия, направленных в те же стороны. В силу симметрии скорости вещества в обеих недеформированных областях одинаковы по модулю, но направлены противоположно. Обозначим через v' скорость вещества в правой недеформированной области. (Очевидно, она положительна.) Тогда скорость вещества в левой недеформированной области будет $-v'$. Чтобы найти v', перейдем снова в неподвижную системы отсчета. Относительно этой системы скорости вещества в недеформированных областях будут $v/2 - v'$ и $v/2 + v'$. Когда возмущения встретятся в месте соприкосновения цилиндров, деформации исчезнут, и оба цилиндра будут двигаться как целые со скоростями $v/2 - v'$ и $v/2 + v'$. Кинетическая энергия этого движения будет

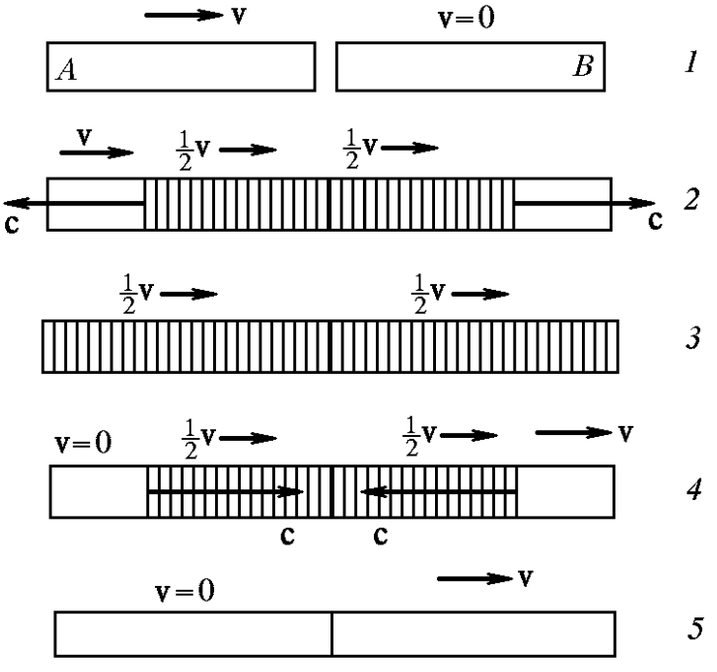


Рис. 225

$$\frac{m}{2}\left(\frac{v}{2}-v'\right)^2+\frac{m}{2}\left(\frac{v}{2}+v'\right)=\frac{mv^2}{4}+mv'^2.$$

Но эта величина должна быть равна $mv^2/2$. Отсюда следует, что $v'^2 = v^2/4$, а потому $v' = v/2$. Таким образом, когда обе волны разрежения сойдутся в центре, первый цилиндр остановится и деформирован не будет, второй будет двигаться вправо со скоростью v также в недеформированном состоянии (положение 5). Как и следовало ожидать, цилиндры обменялись скоростями. Начиная с этого момента контакт между цилиндрами прекратится. Поэтому время соударения цилиндров найдется как промежуток времени, затрачиваемый на прохождение возмущения по одному из цилиндров (любого) туда и обратно:

$$\tau_{уд} = \frac{2l}{c} = 2l\sqrt{\frac{\rho}{E}} \approx 4 \cdot 10^{-5} \text{ с}.$$

Найдем теперь относительное сжатие цилиндров при деформации. После соприкосновения левый конец цилиндра B приобрел скорость $v/2$, правый конец продолжал покоится в течение времени $^1/_2\ \tau_{уд}$. За это время левый конец переместился на расстояние $x = {}^1/_4\ \tau_{уд} v$. Относительное сжатие цилиндра будет

$$\frac{x}{l} = \frac{v}{2c},$$

а давление $P = E\frac{v}{2c}$. Чтобы не возникало пластических деформаций или разрушений, должно быть $P < P_0$, т. е.

$$v < \frac{2cP_0}{E} = \frac{2P_0}{\sqrt{E\rho}} \approx 10 \text{ м/с}.$$

§ 83. СКОРОСТИ РАСПРОСТРАНЕНИЯ ПРОДОЛЬНЫХ И ПОПЕРЕЧНЫХ ВОЗМУЩЕНИЙ В НЕОГРАНИЧЕННОЙ СРЕДЕ

1. Возмущения в стержне, рассмотренные в § 81, мы назвали продольными. Это не совсем точно. Каждая деформация сжатия стержня сопровождается увеличением поперечных размеров его. В случае деформации растяжения поперечные размеры стержня сокращаются. Для количественного описания этих явлений был введен коэффициент Пуассона. Следовательно, частицы в стержне движутся не совсем параллельно его оси: наряду с продольной составляющей скорости они имеют и *поперечную составляющую*. Чтобы сделать возмущение чисто *продольным*, надо лишить частицы стержня возможности перемещаться в поперечных направлениях, т. е. «закрепить» боковую поверхность стержня. Такой случай осуществляется в неограниченной среде при распространении в ней продольных возмущений. Если в такой среде мысленно вырезать произвольный «стержень» с осью, параллельной направлению распространения возмущения (которое в случае продольных возмущений параллельно смещениям частиц), то частицы, находящиеся на боковой поверхности его, удерживаемые соседними частями среды, не будут претерпевать никаких боковых смещений. Все смещения будут происходить только параллельно оси «стержня». Рассуждения,

проведенные в предыдущих параграфах, применимы и в рассматриваемом случае. Надо только модуль Юнга E заменить *модулем одностороннего растяжения* E'. В результате для скорости распространения продольных возмущений в неограниченной среде получится выражение

$$c_{\parallel} = \sqrt{\frac{E'}{\rho}}, \tag{83.1}$$

или в силу соотношений (77.9) и (78.5)

$$c_{\parallel} = \sqrt{\frac{1-\mu}{(1+\mu)(1-2\mu)}\frac{E}{\rho}} = \sqrt{\frac{K+4/3\,G}{\rho}}. \tag{83.2}$$

2. В неограниченной твердой среде наряду с продольными могут распространяться также *поперечные возмущения.* Так называются возмущения, в которых частицы среды смещаются *перпендикулярно* к направлению распространения возмущения. Скорость распространения поперечных возмущений может быть найдена совершенно так же, как и соответствующая скорость для продольных возмущений. Для этого в среде мысленно вырежем произвольный «стержень», ось которого параллельна направлению распространения возмущения, т. е. перпендикулярна к направлениям смещения частиц (рис. 226). Если к основанию такого «стержня» в начальный момент времени приложить постоянное касательное напряжение τ, то в стержне возникнет *деформация сдвига*, распространяющаяся со скоростью, которую мы обозначим $c_{\perp}$. Рассуждая так же, как и в § 81, найдем, что касательное напряжение τ связано с $c_{\perp}$ и скоростью частиц стержня v соотношением

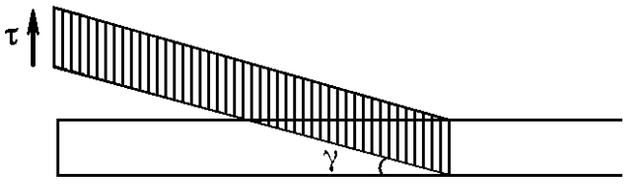


Рис. 226

$$\tau = \rho c_{\perp} v. \tag{83.3}$$

Здесь $\tau = G\gamma$, где γ — угол сдвига. Последний легко найти из следующих соображений. За время t свободный конец стержня перемещается на расстояние vt, в то время как само возмущение проходит путь $c_{\perp}t$. Поскольку $v \ll c_{\perp}$, отсюда следует

$$\gamma = \frac{v}{c_{\perp}}. \tag{83.4}$$

Из этих соотношений легко получить

$$c_{\perp} = \sqrt{\frac{G}{\rho}}. \tag{83.5}$$

3. Поперечные возмущения, если они малы, подчиняются принципу суперпозиции. Поэтому в поперечном возмущении, распространяющемся в определенном направлении, плотности кинетической и потенциальной энергий одинаковы. Вопрос о направлении распро-

странения поперечного возмущения решается с помощью энергетических соображений совершенно так же, как и для продольных возмущений.

4. Так как $K > 0$, то из формул (83.2) и (83.5) следует

$$c_{\parallel} > c_{\perp}. \tag{83.6}$$

Поэтому если в неограниченной среде возникло какое-либо возмущение, то, вообще говоря, оно разделится на продольное и поперечное, причем продольное возмущение придет в точку наблюдения быстрее поперечного. Необходимость такого разделения непосредственно следует из принципа суперпозиции малых возмущений, согласно которому продольное и поперечное возмущения должны распространяться независимо *друг от друга.*

Вообще, скорости продольных и поперечных возмущений в неограниченной среде и скорость «продольных» возмущений в стержне связаны соотношением

$$c_{\parallel} > c > c_{\perp}. \tag{83.7}$$

Неравенство $c > c_{\perp}$ непосредственно следует из формул (81.5), (83.5) и (78.4), так как в силу последней формулы $G < E$. Неравенство же $c_{\parallel} > c$ может быть получено без всяких вычислений. Действительно, если стержень удлиняется только продольно направленной силой, то происходит также его поперечное сжатие. Для устранения последнего к боковой поверхности стрежня надо приложить нормальные поперечные натяжения. Тогда в стержне будут распространяться только чисто продольные возмущения со скоростью $c_{\parallel}$. Но приложенные боковые натяжения будут уменьшать продольное удлинение. При наличии боковых натяжений стержень становится как бы жестче по отношению к деформации растяжения. Это и ведет к увеличению скорости распространения продольных возмущений.

В качестве примера вычислим скорости распространения упругих возмущений в железе или стали. Из опытов найдено: $E = 21{,}2 \cdot 10^{10}$ Н/м^2, $G = 8{,}2 \cdot 10^{10}$ Н/м^2, $\mu = 0{,}29$, $\rho = 7{,}8 \cdot 10^3$ кг/м^3. Используя эти данные, получим

$$c = \sqrt{E/\rho} = 5{,}2 \cdot 10^3 \text{ м/с},$$

$$c_{\parallel} = c\sqrt{\frac{1-\mu}{(1+\mu)(1-2\mu)}} = 6 \cdot 10^3 \text{ м/с},$$

$$c_{\perp} = \sqrt{G/\rho} = 3{,}4 \cdot 10^3 \text{ м/с}.$$

ЗАДАЧИ

1. Показать, что скорость распространения крутильных колебаний вдоль стержня совпадает со скоростью поперечных возмущений $c_{\perp}$.

Р е ш е н и е. Для общности будем считать, что стержень представляет собой цилиндрическую трубку с внутренним радиусом r_1 и наружным радиусом r_2. Пусть к основанию трубки приложены постоянные касательные напряжения, создающие вращающий момент M относительно ее геометрической оси. В трубке возникнет деформация кручения, скорость распространения которой обозначим через c. В возмущенной области вещество будет вращаться с постоянной угловой скоростью ω. Если момент M действовал в течение времени t, то, очевидно,

$$Mt = I\omega,$$

где I — момент инерции возмущенной области. С другой стороны, $M = f\varphi = f\omega t$. Это дает $ft^2 = I$. Подставляя сюда $I = \frac{1}{2}\pi\rho l(r_2^4 - r_1^4)$, $t = l/c$ (l — длина возмущенной области) и пользуясь соотношением (79.4), получим

$$\rho c^2 = G.$$

2. Найти выражение для скорости продольных звуковых возмущений, распространяющихся в безграничной двумерной тонкой пластинке (см. задачу к § 77). Показать, что эта скорость меньше скорости продольных возмущений в неограниченной среде и больше скорости распространения «продольных» возмущений в стержне.

О т в е т. $c' = \sqrt{\frac{E}{\rho(1-\mu^2)}}$. Неравенство $c' > c$ можно доказать без вычисления совершенно так же, как это было сделано для неравенства $c_{\parallel} > c$.

§ 84. СКОРОСТЬ РАСПРОСТРАНЕНИЯ ПОПЕРЕЧНЫХ ВОЗМУЩЕНИЙ В НАТЯНУТОМ ШНУРЕ

1. Возможность распространения поперечных возмущений в твердых телах обусловлена присущей им поперечной упругостью, т. е. способностью тел сопротивляться всякому изменению формы, происходящему без изменения объема. Поперечная упругость может быть создана искусственно и в случае таких тел, у которых в естественном состоянии она отсутствует. Примером может служить гибкий шнур или веревка. Если шнур не натянут, то поперечные возмущения в нем распространяться не могут. Если же закрепить один конец шнура, а к другому подвесить груз, перекинув шнур через блок, то в шнуре возникнет постоянное натяжение, обозначаемое в дальнейшем T. Такой шнур обладает *упругостью формы*, и в нем могут распространяться поперечные возмущения. Скорость таких возмущений можно вычислить по формуле (83.5). Но для этого надо решить вопрос, какая величина в натянутом шнуре играет роль модуля сдвига G. Рассмотрим небольшой участок AB натянутого и изо-

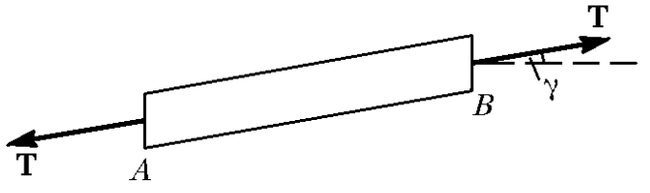


Рис. 227

гнутого шнура (рис. 227). Будем предполагать, что деформации натянутого шнура, связанные с поперечными смещениями его частиц, малы. Тогда можно пренебречь изменениями натяжения T, обусловленными изгибом шнура при таких малых деформациях. В этом приближении натяжения T, действующие на концы участка AB вдоль его оси, одни и те же. Их составляющие, касательные к основаниям участка AB, равны $T \sin\gamma \approx T\gamma$. Поэтому на основаниях рассматриваемого участка будут действовать касательные напряжения $\tau = (T/S)\gamma$, где S — площадь поперечного сечения шнура. Деформацию участка AB можно рассматривать как сдвиг под действием таких касательных напряжений. Сравнивая поэтому предыдущее выражение с формулой $\tau = G\gamma$, находим, что роль модуля сдвига играет величина $G = T/S$. Подставим это выражение в формулу (83.5) и введем обозначение $\delta = \rho S$. Тогда для скорости распространения поперечных возмущений в шнуре получим

$$c = \sqrt{\frac{T}{\delta}}. \qquad (84.1)$$

Величина δ равна массе, приходящейся на единицу длины шнура. Она называется *линейной плотностью* шнура.

2. Если возмущение в шнуре распространяется в одном направлении, то в таком возмущении плотности кинетической и потенциальной энергий в любой момент времени, конечно, будут одинаковы. Направление распространения возмущения можно определить из энергетических соображений. Для этого помимо формы шнура в рассматриваемый момент времени надо еще задать скорость каждой его точки. Так, например, возмущение, представленное на рис. 228, распространяется вправо. Вертикальными стрелками обозначены скорости частиц шнура в рассматриваемый момент времени.

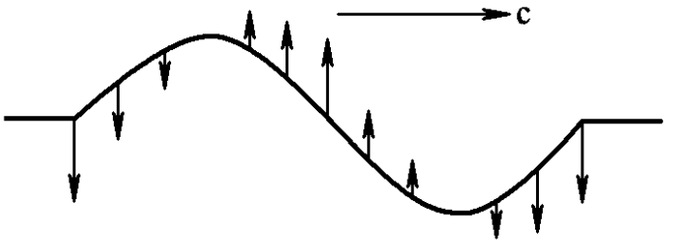


Рис. 228

Если мысленно провести в шнуре какое-либо поперечное сечение, то угол между силой натяжения, действующей на правую часть шнура, и ее скоростью в рассматриваемом сечении будет острым. Напротив, сила натяжения, действующая на левую часть шнура, составляет с соответствующей скоростью тупой угол. Это значит, что над правой частью шнура сила натяжения совершает положительную, а над левой — отрицательную работу. Поэтому-то возмущение и распространяется вправо. Если изменить на противоположные направления скоростей всех частиц, то возмущение пойдет влево.

3. Формулу (84.1) можно получить также следующим, очень поучительным способом. Пусть в шнуре возбуждено поперечное возмущение, распространяющееся, например, вправо (рис. 228) со скоростью c. Рассмотрим явление в системе отсчета, равномерно движущейся вправо со скоростью c. В этой системе отсчета возмущение будет стоять на месте, а весь шнур — двигаться влево со скоростью c.

В возмущенной области на это движение будут накладываться малые поперечные колебания частиц шнура. Ось шнура является траекторией движущихся частиц, находящихся на этой оси. Если на шнур надеть надлежащим образом изогнутую цилиндрическую трубку, неподвижную в рассматриваемой движущейся системе отсчета, то наличие такой трубки никак не отразится на движении шнура. Шнур будет просто протягиваться через трубку, нигде не касаясь ее стенок. Для того чтобы это имело место, необходимо тянуть шнур с вполне определенной скоростью c. При малых возмущениях скорости поперечных движений частиц шнура v малы по сравнению с c. В выражении для полной скорости частиц $\sqrt{c^2+v^2}$ квадратом малой величины v можно пренебречь. В этом приближении полная скорость частиц считается одной и той же на протяжении всей его длины и равной c. Однако в области трубки, где шнур изогнут, его частицы движутся ускоренно. Их ускорения направлены нормально к траектории и определяются выражением $a = c^2/R$. Для создания таких ускорений нужна сила, действующая нормально к траектории. Она возникает из-за изгиба шнура. Найдем ее значение. Выделим мысленно бесконечно малый элемент изогнутого шнура AB, длину которого обозначим через s (рис. 229). Его можно рассматривать как бесконечно малую дугу окружности радиуса R. На концы этого элемента действуют продольные натяжения T_1 и T_2. Их абсолютные величины в пределах принятой точности расчета одинаковы ($T_1 = T_2 = T$). Но направления немного отличаются друг от друга. Благодаря этому и появляется результирующая сила, направленная нормально к элементу AB. Она равна

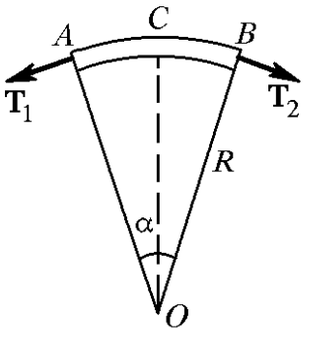


Рис. 229

$$F = 2T \sin\frac{\alpha}{2} \approx T\alpha = T\frac{s}{R}.$$

Приравнивая эту силу массе элемента AB, умноженное на его ускорение, получим

$$T\frac{s}{R} = s\delta\frac{c^2}{R},$$

откуда снова получается формула (84.1).

§ 85. СКОРОСТЬ РАСПРОСТРАНЕНИЯ ЗВУКА В ЖИДКОСТЯХ И ГАЗАХ

1. Жидкости и газы обладают только *объемной упругостью*, но не упругостью формы. Поэтому в них могут распространяться только продольные возмущения, но не могут распространяться возмущения поперечные. Скорость распространения продольных возмущений в жидкой или газообразной среде можно вычислить по формуле

(81.5). Но для этого надо решить, что́ в этом случае играет роль модуля Юнга E. Вообразим, что жидкая или газообразная среда заключена в гладкую прямолинейную трубу постоянного поперечного сечения. Трением между средой и стенками трубы пренебрежем. Стенки трубы будут препятствовать поперечному движению среды, нисколько не мешая продольному движению. Газ или жидкость в такой трубе можно рассматривать как стержень, вдоль которого распространяются продольные возмущения. Отличие от твердых тел состоит в том, что газы могут существовать *только под давлением.* При отсутствии такового всякий газ неограниченно расширился бы. Поэтому необходимо предполагать, что в невозмущенном состоянии давление внутри газа отлично от нуля. Обозначим его через P_0. Так же будем поступать в случае жидкости. Если давление внутри газа получит приращение и сделается равным $P = P_0 + \Delta P$, то изменится и объем рассматриваемой массы газа.

Определим, как изменение объема газа ΔV связано с приращением его давления ΔP. При этом мы будем предполагать, что ΔP мало́ по сравнению с P_0: $\Delta P \ll P_0$. Если газ заключен в трубе, один из концов которой закрыт неподвижным поршнем, то при изменении давления на поршень на величину ΔP длина газового столба изменится на Δl. Величина $-(\Delta l/l)$ есть относительное сжатие столба газа. При малых сжатиях

$$\Delta P = -A\frac{\Delta l}{l},$$

где A — постоянная. С другой стороны, формулу (75.5) для стержня можно переписать в виде $\Delta P = -E\frac{\Delta(\Delta l)}{l}$, где $\Delta(\Delta l)$ — приращение длины стержня при изменении давления на ΔP. По смыслу оно совпадает с тем, что в случае газового столба мы обозначили через Δl. Поэтому, меняя обозначение, модуль Юнга можно определить также с помощью формулы

$$\Delta P = -E\frac{\Delta l}{l}. \tag{85.1}$$

Из нее видно, что в случае газового столба $A = E$. Длина столба газа l пропорциональна его объему V, и предыдущую формулу можно записать в виде

$$\Delta P = -E\frac{\Delta V}{V}. \tag{85.2}$$

В этом виде формула сохраняет смысл для любой формы сосуда, в котором заключен газ, тогда как формула (85.1) относится только к газам в сосудах цилиндрической формы.

Будем считать, что давление газа зависит только от его объема V. Тогда для малых изменений объема

$$\Delta P = \frac{dP}{dV}\Delta V$$

или

$$\Delta P = \left(-V\frac{dP}{dV}\right)\left(-\frac{\Delta V}{V}\right).$$

Сравнивая эту формулу с предыдущей, видим, что в газах (и жидкостях) роль модуля Юнга играет величина

$$E = -V\frac{dP}{dV}. \tag{85.3}$$

Вместо объема тела V удобнее ввести плотность ρ. Величина $V\rho$ есть масса тела, остающаяся постоянной при всех изменениях. Из соотношения $V\rho = \text{const}$ путем дифференцирования находим

$$\frac{dV}{V} = -\frac{d\rho}{\rho},$$

а потому

$$E = \rho\frac{dP}{d\rho}. \tag{85.4}$$

Подставляя это выражение в формулу (81.5), получаем для скорости звука в газах и жидкостях

$$c = \sqrt{\frac{dP}{d\rho}}. \tag{85.5}$$

2. Применим формулу (85.5) к вычислению скорости звука в газах. Впервые это было сделано Ньютоном. Он принял, что изменения давления и плотности газа в звуковой волне подчиняются *закону Бойля—Мариотта* (Роберт Бойль (1627—1691) — знаменитый английский физик и химик, Эдм Мариотт (1620—1684) — французский физик): $P = A\rho$, где $A = \text{const}$. Отсюда $dP/d\rho = A = P/\rho$. В результате получается *формула Ньютона*

$$c_{\text{Н}} = \sqrt{\frac{P}{\rho}}. \tag{85.6}$$

Здесь скорость звука обозначена через $c_{\text{Н}}$, чтобы подчеркнуть, что речь идет о скорости звука, вычисляемой по формуле Ньютона.

Преобразуем формулу (85.6) к другому виду, более удобному в численных расчетах. Как известно, объем, давление и абсолютная температура T идеальных газов связаны соотношением

$$PV = RT, \tag{85.7}$$

где R — постоянная. Если газ взят в количестве одного моля, то постоянная R будет иметь одно и то же числовое значение для всех газов. Она называется *универсальной газовой постоянной* и равна $R = 8{,}31 \cdot 10^7$ эрг$\cdot K^{-1}\cdot$моль$^{-1}$. Напомним, что *молем* называется количество вещества, масса которого в граммах численно равна молекулярной массе этого вещества μ. Отсюда следует, что плотность ρ связана с объемом V моля идеального газа соотношением $\mu = \rho V$. В результате получаем

$$P = \frac{RT}{\mu}\rho, \tag{85.8}$$

$$c_{\text{Н}} = \sqrt{\frac{RT}{\mu}}. \tag{85.9}$$

Вычислим по этой формуле скорость звука в воздухе при 0 °C ($T = 273$ K). Воздух есть смесь различных газов, основными частями которой являются азот ($\mu = 28$) и кислород ($\mu = 32$). Среднюю молекулярную массу такой смеси примем равным $\mu = 28{,}8$. Подставляя в формулу (85.9) числовые значения, получим $c_{\text{н}} = 280$ м/с. Опыт дает $c = 330$ м/с. Налицо значительное расхождение между теорией и опытом. Причина этого расхождения долгое время оставалась непонятной. Она была установлена Лапласом (1749—1827) лишь в начале XIX века. Закон Бойля—Мариотта относится к таким изменениям давления и объема газа, при которых его *температура остается постоянной*. Между тем звуковая волна состоит из следующих друг за другом *сжатий* и *разрежений* газа. Над сжатыми областями производится внешняя работа, которая идет на *повышение их температуры*. Разреженные области сами совершают внешнюю работу и благодаря этому охлаждаются. Так как сжатия и разрежения совершаются очень быстро, то температуры между ними не успевают выравниваться: сжатые области всегда теплее разреженных. Наличие этой разности температур повышает перепад давления между сжатиями и разрежениями и ведет к увеличению скорости звука в газах. Это обстоятельство и не было учтено формулой Ньютона. Ньютон при вычислении скорости звука подставил в формулу (81.5) *изотермический модуль* упругости E, а надо было пользоваться *адиабатическим модулем* (см. § 79). Количественное исследование вопроса будет дано в томе II нашего курса.

ГЛАВА XI

МЕТОДЫ ПОДОБИЯ И РАЗМЕРНОСТИ

§ 86. РАЗМЕРНОСТЬ И СИСТЕМЫ ЕДИНИЦ

1. До сих пор мы ничего не говорили о *размерности физических величин*. Мы пользовались этим понятием, предполагая, что читатель имеет некоторое представление об относящихся сюда вопросах. В задачах, которые мы рассматривали, этого было достаточно. *Метод размерности* весьма эффективен в более сложных вопросах, например в гидродинамике, где полная теоретическая трактовка затруднительна. С привлечением добавочных соображений весьма общего характера или опытных данных он приводит, и притом быстро и просто, к важным результатам, дающим предварительную, хотя и неполную, ориентировку в рассматриваемом круге явлений. Поэтому необходимо познакомиться с этим методом.

Понятие размерности возникает в связи с построением систем единиц. В принципе можно было бы (так и поступали раньше) для каждой физической величины установить свою единицу, никак не связанную с единицами других величин. Но тогда в уравнения, выражающие физические законы, вошло бы множество числовых коэффициентов. Их значения не укладывались бы ни в какую простую и легко запоминаемую схему, а определялись бы случайным выбором единиц. Такое множество числовых коэффициентов весьма сильно усложняло бы формулы. Запоминание их было бы нелегкой и в сущности бесполезной нагрузкой для памяти. Во избежание этого в физике уже давно отказались от независимого выбора единиц для всех физических величин, а стали применять системы единиц, построенные по определенному принципу.

2. Принцип этот заключается в следующем. Некоторые физические величины условно принимаются за *основные* или *первичные*, т. е. такие, для которых единицы устанавливаются произвольно и независимо. Так, например, в механике применяется система LMT, в которой за основные величины принимаются длина (L), масса (M) и время (T). *Выбор основных величин и их число произвольны.* Это — вопрос соглашения. Например, в технической механике до недавнего времени применялась система LFT. Основными величинами в ней были длина (L), сила (F) и время (T). В так называемой *Международной системе единиц* (сокращенно СИ) за основные приняты семь величин: длина, масса, время, температура, сила электрического тока, сила света и количества вещества. Величины, не являющиеся основными, называются *производными* или *вторичными*. Для них едини-

цы устанавливаются из требования, чтобы числовые коэффициенты, входящие в физические законы или формулы, служащие определением рассматриваемых величин, принимали определенные, заранее выбранные значения. Например, скорость равномерно движущейся материальной точки есть величина особого рода, пропорциональная пройденному пути s и обратно пропорциональная времени t, затрачиваемому на прохождение этого пути. При независимом выборе единиц для s, t и v следует писать $v = Cs/t$, где C — числовой коэффициент, значение которого определяется выбором единиц. Если фиксировать значение этого коэффициента, то единицы для s, t и v перестанут быть независимыми. Для простоты полагают $C = 1$ и пишут $v = s/t$. Если за основные величины принять путь s и время t, то скорость v становится величиной производной. За единицу скорости мы обязаны принять скорость такого равномерного движения, когда за единицу времени проходится единица длины. Говорят, что скорость имеет размерность длины, деленной на время. Символически это записывается так: $[v] = \mathrm{LT}^{-1}$. Аналогично пока единицы выбираются независимо для ускорения a можно написать $a = C\frac{dv}{dt}$. Полагая $C = 1$, мы делаем ускорение a величиной производной, имеющей размерность скорости, деленной на время, или размерность длины, деленной на квадрат времени. После этого за единицу ускорения мы обязаны принять ускорение такого равномерно ускоренного движения, когда за каждую единицу времени скорость возрастает на единицу. В произвольных единицах второй закон Ньютона пишется в виде $F = Cma$. Фиксируя числовой коэффициент C, мы делаем силу F величиной производной и устанавливаем для нее единицу. Например, при $C = 1$ получаем $F = ma$. После этого сила получает размерность массы, умноженной на ускорение: $[F] = [ma] = \mathrm{MLT}^{-2}$. Формула $F = ma$ обязывает нас за единицу силы принять такую силу, которая массе в одну единицу сообщает ускорение, равное единице.

3. Размерность физической величины еще не определяет, как велика ее единица. Она устанавливает только связь между единицами различных физических величин. *Размерность дает правило, позволяющее определить, как меняется единица производной физической величины при изменении масштабов основных величин.* Это правило, выраженное в виде математической формулы, называется *формулой размерности.* Допустим, например, что за единицу длины принят километр, а за единицу времени — минута. Единицей ускорения в такой системе единиц будет км/мин^2. Спрашивается, как изменится единица ускорения, если за единицу длины принять сантиметр, а за единицу времени — секунду. Формула размерности позволяет быстро ответить на этот вопрос. Мы пишем прежде всего 1 км $= 10^5$ см, 1 мин $= 60$ с и далее

$$1\ \text{км/мин}^2 = \frac{10^5\ \text{см}}{60^2\ \text{с}^2} = \frac{1000}{36}\ \text{см/с}^2.$$

Отсюда видно, что единица ускорения 1 км/мин^2 крупнее единицы 1 см/с^2 в 1000/36 раз. В соответствии с этим числовое значение ускорения, измеренное в км/мин^2, окажется меньше числового значения того же ускорения в 1000/36 раз, если его измерить в см/с^2.

§ 87. ФОРМУЛА РАЗМЕРНОСТИ

1. Термин «система единиц» употребляется в двух смыслах. В широком смысле система единиц характеризуется выбором основных величин и формулами, определяющими производные величины через основные, причем *масштабы основных величин не фиксируются.* Примером может служить система LMT, в которой основными величинами являются длина, масса и время. Другим примером является применявшаяся ранее электротехническая система LMTI, в которой за основные величины принимаются длина, масса, время и сила электрического тока I. Система единиц в узком смысле дополнительно характеризуется также *определенным выбором масштабов основных единиц.* Примерами могут служить системы СГС и МКСА. Первая есть частный случай системы LMT, когда за единицы длины, массы и времени приняты сантиметр, грамм и секунда. Вторая является частным случаем электротехнической системы LMTI. В ней единицами длины, массы, времени и силы тока являются соответственно метр, килограмм, секунда и ампер. В теории размерности термин «система единиц» понимается в широком смысле.

Понятие размерности возникает в связи с требованием, чтобы в одной и той же системе единиц количественные соотношения между различными физическими величинами выражались *одними и теми же* формулами, независимо от того, как велики единицы основных физических величин. Этим требованием определяется общий вид «*формул размерности*» физических величин. Допустим, что имеется несколько физических величин, связанных между собой. Для простоты можно ограничиться случаем двух величин, одна из которых принимается за основную, а другая — за производную. Числовые значения их x и y связаны уравнением $y = f(x)$. Определим общий вид функции $f(x)$. Если единицу основной величины x уменьшить в α раз, то числовое значение этой величины увеличивается в такое число раз и сделается равным $X = \alpha x$. При этом единица производной величины y уменьшится, а ее числовое значение увеличится в β раз и станет равным $Y = \beta y$. Мы требуем, чтобы числовые значения X и Y были связаны тем же уравнением, что и числа x и y, т. е. $Y = f(X)$ или $\beta y = f(\alpha x)$. Этому условию можно удовлетворить при любых значениях α, если надлежащим образом подобрать β. Задача сводится к нахождению β как функции аргумента α. На этот вопрос и отвечает «формула размерности».

Прежде чем его решать, изменим постановку вопроса. Пусть две физические величины связаны соотношением $y = f(x)$. Будем менять сами физические величины, оставляя их единицы неизменными. Допустим, что величины x и y увеличились соответственно в α и β раз и сделались равными $X = \alpha x$ и $Y = \beta y$. Спрашивается, какому условию должны удовлетворять числа α и β, чтобы связь между новыми значениями физических величин X и Y была та же, что и между старыми значениями x и y, т. е. $Y = f(X)$. На этот вопрос отвечает *теория подобия*. Вопрос опять сводится к исследованию уравнения $\beta y = f(\alpha x)$. Мы видим, что теории размерности и подобия отличаются друг от друга только формой постановки вопроса. По существу они не отличаются одна от другой. Теория подобия позволяет исследовать количественные соотношения между различными параметрами реальных физических систем на их уменьшенных или увеличенных моделях. Так поступают, например, в авиационной технике, помещая в аэродинамические трубы уменьшенные копии реальных летательных аппаратов. Изучив поведение моделей реальных систем, можно с помощью теории подобия или размерности сделать выводы о поведении самих систем в реальных условиях. Теория размерности сводит вопрос о подобии физических явлений в указанном выше смысле к *анализу размерностей* физических величин.

2. После этих предварительных замечаний установим общий вид формулы размерности. Как разъяснено выше, мы должны требовать, чтобы из уравнения $y = f(x)$ вытекало уравнение $Y = f(X)$, где $X = \alpha x$, $Y = \beta y$. Аргумент x и параметр α могут независимо принимать любые значения. Задача состоит в том, чтобы по заданному значению α найти значение β. Путем дифференцирования при фиксированных α и β находим

$$\frac{dy}{dx} = f'(x), \quad \frac{dY}{dX} = f'(X).$$

Вторую из этих формул запишем в виде

$$\frac{\beta}{\alpha}\frac{dy}{dx} = f'(X).$$

Поделив ее почленно на первую формулу и заменив α и β выражениями $\alpha = \frac{X}{x}$, $\beta = \frac{Y}{y} = \frac{f(X)}{f(x)}$, получим

$$x\frac{f'(x)}{f(x)} = X\frac{f'(X)}{f(X)}.$$

Слева стоит функция только x, справа — та же функция только X. Обозначив ее через F, имеем $F(x) = F(X)$. Но в силу произвольности и параметра α аргументы x и $X = \alpha x$ могут независимо принимать любые значения. Поэтому равенство $F(x) = F(X)$ должно выполняться тождественно, каковы бы ни были x и X. Это значит, что $F(x)$ есть постоянная. Обозначив эту постоянную че-

рез m, получим дифференциальное уравнение

$$x\frac{f'(x)}{f(x)}=m,$$

или

$$\frac{df(x)}{f(x)}=m\frac{dx}{x}.$$

Отсюда находим

$$f(x)=f_0x^m,$$

где f_0 — постоянная интегрирования. Таким образом, $y=f_0x^m$. Аналогично $Y=f_0X^m$ или $\beta y=f_0(\alpha x)^m$. Исключая почленным делением x и y, находим

$$\beta=\alpha^m. \tag{87.1}$$

Это и есть формула размерности. Мы видим, что требование независимости функциональной связи между x и y от выбора масштаба единицы основной физической величины x может быть удовлетворено только тогда, когда размерность выражается формулой *степенного вида.*

Приведенные рассуждения без труда обобщаются на случай, когда рассматриваемая физическая величина зависит от нескольких основных физических величин. Для этого в рассуждениях надо только фиксировать единицы всех основных физических величин, за исключением одной из них. Таким путем нетрудно показать, что формула размерности должна быть *степенного вида относительно всех основных физических величин.* Допустим, например, что число основных величин выбрано равным трем, и за них приняты длина (L), масса (M) и время (T). Тогда размерность любой физической величины y представится формулой

$$[y]=\mathrm{L}^p\mathrm{M}^q\mathrm{T}^r, \tag{87.2}$$

где p, q, r — постоянные числа. Формула (87.2) означает, что если единицы длины, массы и времени уменьшить соответственно в α, β и γ раз, то единица производной величины y уменьшится в $\alpha^p\beta^q\gamma^r$ раз, а следовательно, ее числовое значение увеличится в такое же число раз.

3. Если посмотреть на размерности физических величин, фактически встречающих в физике, то нетрудно заметить, что во всех случаях числа p, q, r оказываются рациональными. Это не обязательно с точки зрения теории размерности, а является результатом соответствующих определений физических величин. Так, например, скорость вводится по формуле $v=s/t$ и поэтому имеет размерность длины, деленной на время. Для нее $p=1$, $q=0$, $r=-1$. Но в принципе теория размерности допускает введение величин с иррациональными значе-

ниями p, q, r, например величины $(1/t)s^{\sqrt{2}}$. Для такой величины было бы $p = \sqrt{2}$. Подобные величины не вводятся в физику не по каким-то принципиальным соображениям, а просто потому, что в них нет надобности. Теория размерности здесь ни при чем.

4. Часто размерность физической величины отождествляют с ее единицей в соответствующей системе единиц. Так, например, говорят, что скорость имеет размерность см/с, а сила — г·см/с^2. Хотя это и нелогично, но особой беды в этом нет. Всегда, если есть необходимость, единицы такого типа позволяют перейти к формулам размерности, в которых масштабы единиц основных величин не фиксированы.

5. В зависимости от выбора основных величин, а также от вида формул, связывающих эти величины с производными, одна и та же физическая величина получает в разных системах единиц не только различные числовые значения, но и различные размерности. Так, например, в системе LMT размерность силы устанавливается из второго закона Ньютона $f = Cma$, в котором коэффициент C условно считается безразмерным и полагается равным единице. Тогда сила получает размерность LMT^{-2}. Но так поступать не обязательно. Можно коэффициенту C приписать произвольную размерность и придать произвольное числовое значение. Тогда получится новая система единиц, в которой сила будет иметь уже другую размерность. Например (и так иногда делают), в уравнении $f = G\dfrac{m_1 m_2}{r^2}$, выражающем закон всемирного тяготения Ньютона, приравнивают гравитационную постоянную G единице и считают эту величину безразмерной. Тогда сила f получает размерность $\mathrm{M}^2\mathrm{L}^{-2}$, а во втором законе Ньютона $f = Cma$ появляется коэффициент C с размерностью $\mathrm{ML}^{-3}\mathrm{T}^2$.

Разные физические величины могут иметь одинаковые размерности даже в одной и той же системе единиц. Примерами могут служить в механике работа и кинетическая энергия или работа и момент сил (система MLT, а в учении об электричестве и магнетизме — емкость и индуктивность, имеющие в так называемой гауссовой системе единиц размерность длины. В таких случаях и единицам этих физических величин часто дают одинаковые наименования, хотя по существу это совершенно разные вещи. Одинаковая размерность двух различных физических величин в какой-либо системе единиц говорит не об их тождестве, а только о том, что в рассматриваемой системе масштабы единиц этих величин меняются одинаково при изменении масштабов единиц основных физических величин. В других системах единиц размерности тех же физических величин могут и не совпадать.

Несовпадение размерностей одной и той же величины в разных системах единиц иногда истолковывают как некоторое логическое противоречие, требующее объяснения. Оно подало повод к постановке вопроса об «истинной» размерности физических величин. На основании изложенного нет никакой необходимости доказывать, что физи-

ческой величине самой по себе не свойственна никакая размерность. Последняя появляется лишь после установлений той или иной системы единиц, а вопрос об «истинной» размерности физических величин, по меткому замечанию Макса Планка, имеет не более смысла, чем вопрос об «истинном» названии какого-либо предмета.

6. *Безразмерными комбинациями физических величин* называются такие комбинации, которые в рассматриваемой системе единиц имеют нулевую размерность. Их числовые значения не меняются при изменении масштабов единиц основных величин. Легко привести примеры таких комбинаций. Если величина y имеет размерность величины x в степени α, то очевидно, y/x^{α} будет безразмерной комбинацией, составленной из x и y.

Общий метод нахождения безразмерных комбинаций можно разъяснить на примере системы единиц, построенной на основе трех величин: длины (L), массы (M) и времени (T). Пусть n величин $x_1, x_2, \ldots, x_n$ в этой системе имеют размерности соответственно

$$\mathrm{L}^{p_1}\mathrm{M}^{q_1}\mathrm{T}^{r_1},\ \mathrm{L}^{p_2}\mathrm{M}^{q_2}\mathrm{T}^{r_2},\ \ldots,\ \mathrm{L}^{p_n}\mathrm{M}^{q_n}\mathrm{T}^{r_n}.$$

Требуется составить из них безразмерную комбинацию. На основании теоремы, доказанной в п. 2, искомая комбинация должны иметь вид $x_1^{\alpha_1}x_2^{\alpha_2}\ldots x_n^{\alpha_n}$. Ее размерность будет

$$(\mathrm{L}^{p_1}\mathrm{M}^{q_1}\mathrm{T}^{r_1})^{\alpha_1}(\mathrm{L}^{p_2}\mathrm{M}^{q_2}\mathrm{T}^{r_2})^{\alpha_2}\ldots(\mathrm{L}^{p_n}\mathrm{M}^{q_n}\mathrm{T}^{r_n})^{\alpha_n},$$

т. е. $\mathrm{L}^{p}\mathrm{M}^{q}\mathrm{T}^{r}$, где

$$\begin{aligned} p &= p_1\alpha_1 + p_2\alpha_2 + \ldots + p_n\alpha_n, \\ q &= q_1\alpha_1 + q_2\alpha_2 + \ldots + q_n\alpha_n, \\ r &= r_1\alpha_1 + r_2\alpha_2 + \ldots + r_n\alpha_n. \end{aligned} \tag{87.3}$$

Для того чтобы комбинация была безразмерной, необходимо и достаточно, чтобы $p = q = r = 0$. Это приводит к системе трех однородных уравнений

$$\begin{aligned} p_1\alpha_1 + p_2\alpha_2 + \ldots + p_n\alpha_n &= 0, \\ q_1\alpha_1 + q_2\alpha_2 + \ldots + q_n\alpha_n &= 0, \\ r_1\alpha_1 + r_2\alpha_2 + \ldots + r_n\alpha_n &= 0 \end{aligned} \tag{87.4}$$

с неизвестными $\alpha_1, \alpha_2, \ldots, \alpha_n$.

Одно из этих неизвестных всегда можно выбрать произвольно, так как безразмерная комбинация остается безразмерной, если ее возвести в произвольную степень. Фиксируем, например, α_1. Тогда получится три уравнения для определения $n-1$ неизвестных, за которые удобно принять отношения $\alpha_2/\alpha_1, \alpha_3/\alpha_1, \ldots, \alpha_n/\alpha_1$. Если эти уравнения независимы, то $(n-1)-3 = n-4$ отношений

можно выбрать произвольно. Три остальных определятся из уравнений (87.4). В результате найдутся $n-4$ независимых безразмерных комбинаций. Всякая функция этих безразмерных комбинаций будет также безразмерной комбинацией. Если же три уравнения (87.4) не независимы, то число независимых безразмерных комбинаций увеличится. Например, если в системе (87.4) независимы только два уравнения, то независимых безразмерных комбинаций будет $n-3$ и т. д.

§ 88. ПРАВИЛО РАЗМЕРНОСТИ

1. Все применения теории размерностей основаны на двух теоремах. Одна из них выражается формулой (87.2), устанавливающей общий вид размерности физических величин. Другая теорема утверждает, что всякое количественное соотношение между различными физическими величинами может быть выражено в виде функциональной связи между *безразмерными комбинациями этих величин.*

Для доказательства предположим, что между величинами $a, b, c, x_1, x_2, x_3, \ldots$ имеется функциональная связь $f(a, b, c, x_1, x_2, x_3, \ldots) = 0$. Примем величины a, b, c за основные, а остальные величины $x_1, x_2, x_3, \ldots$ — за производные. (Мы взяли число основных величин равным трем, но это несущественно.) Пусть размерности производных величин будут $[x_1] = [a^{p_1} b^{q_1} c^{r_1}]$, $[x_2] = [a^{p_2} b^{q_2} c^{r_2}]$, ... Уменьшим единицы основных величин в α, β, γ раз соответственно. Тогда они примут значения αa, βb, γc, а производные величины — значения $\alpha^{p_1}\beta^{q_1}\gamma^{r_1} x_1$, $\alpha^{p_2}\beta^{q_2}\gamma^{r_2} x_2, \ldots$ Рассматриваемая функциональная связь запишется в виде

$$f(\alpha a, \beta b, \gamma c, \alpha^{p_1}\beta^{q_1}\gamma^{r_1} x_1, \alpha^{p_2}\beta^{q_2}\gamma^{r_2} x_2, \ldots) = 0,$$

причем α, β, γ можно выбирать произвольно. Выберем их так, чтобы $\alpha a = \beta b = \gamma c = 1$. Это означает переход от жестко фиксированных единиц к меняющейся системе единиц, в которой числовые значения основных физических величин в рассматриваемом вопросе принимаются равными единице. При таком выборе

$$f\left(1, 1, 1, \frac{x_1}{a^{p_1} b^{q_1} c^{r_1}}, \frac{x_2}{a^{p_2} b^{q_2} c^{r_2}}, \ldots\right) = 0.$$

Но это уравнение в качестве переменных аргументов содержит только безразмерные комбинации физических величин. Его можно записать в виде

$$F\left(\frac{x_1}{a^{p_1} b^{q_1} c^{r_1}}, \frac{x_2}{a^{p_2} b^{q_2} c^{r_2}}, \ldots\right) = 0, \tag{88.1}$$

где F — новая функция. Теорема доказана.

2. Доказанной теореме можно придать другую форму. Разрешим уравнение (88.1) относительно одного из аргументов, например первого, и результат умножим на знаменатель этого аргумента. Получим

$$x_1 = a^{p_1} b^{q_1} c^{r_1} \varphi\left(\frac{x_2}{a^{p_2} b^{q_2} c^{r_2}}, \ldots\right), \tag{88.2}$$

где φ — какая-то функция безразмерных аргументов. Это означает, что во всяком физическом законе типа $A = B$ размерности обеих частей равенства должны быть *одинаковы*. В таком виде доказанная теорема получила название *правила размерностей*. В равенство типа $A = B$ могут входить в качестве множителей либо постоянные коэффициенты, либо безразмерные комбинации физических величин. Над размерными величинами правило размерности допускает выполнение только степенных математических операций. Все прочие математические операции ($\sin x$, e^x, $\ln x$ и т. п.) могут выполняться только над безразмерными величинами. Правило размерности очень полезно для проверки формул. Если вычисления проводятся в какой-то одной системе единиц, то размерности обеих частей всех полученных равенств должны быть одинаковы. Несовпадение размерностей указывает на наличие ошибки, допущенной при вычислениях.

Из доказанного отнюдь не следует, что невозможны физические законы, выражающиеся в виде равенств между величинами разной размерности. Равенства подобного рода встречаются в физике сплошь и рядом. Например, скорость свободного падения можно выразить приближенной формулой $v = 10t$ (если начальная скорость равна нулю), а гидростатическое давление слоя воды — формулой $P = 1/10h$. Однако подобные формулы справедливы только тогда, когда *точно фиксированы* единицы входящих в них физических величин. В приведенных примерах предполагается, что время t измеряется в секундах, скорость v — в метрах в секунду, толщина слоя воды h — в метрах, давление P — в атмосферах. Изменения масштабов единиц такие формулы не допускают. Но в таком случае нет смысла говорить и о размерности входящих в них физических величин.

3. Теория размерности сама по себе, т. е. без использования добавочных данных, не может привести ни к каким конкретным физическим выводам, поскольку в ее основах не заложены никакие физические законы. Для того чтобы извлечь из этой теории конкретные выводы, нужно установить, *между какими физическими величинами* существуют количественные связи. На этот счет теория размерности не может дать никаких указаний. Это можно сделать только либо опытным путем, либо с помощью каких-то физических законов. Приводимые ниже примеры могут служить иллюстрацией высказанных утверждений.

ЗАДАЧИ

1. Составить все независимые безразмерные комбинации из величин l, m, t, v, a, ρ, E, φ (l — длина, m — масса, t — время, v — скорость, a — ускорение, ρ — плотность вещества, E — модуль Юнга, φ — угол, измеренный в радианах).

Р е ш е н и е. Проще всего поступить следующим образом. Из перечисленных величин угол φ уже является безразмерной величиной. Далее замечаем, что vt имеет размерность длины, at — размерность скорости, ρl^3 — размерность массы, ρv^2 — размерность давления, а следовательно, и размерность модуля Юнга. Поэтому сразу можно написать следующие безразмерные комбинации:

$$\frac{vt}{l},\ \frac{at}{v},\ \frac{\rho l^3}{m},\ \frac{\rho v^2}{E},\ \varphi. \tag{88.3}$$

Этот способ обладает, однако, тем недостатком, что он не дает ответа на вопрос, исчерпываются ли рядом (88.3) все независимые безразмерные комбинации рассматриваемых физических величин. Общий метод, изложенный в § 87, п. 6, свободен от этого недостатка. Поэтому мы приведем решение также по этому методу. При отыскании безразмерных комбинаций угол φ, как величину безразмерную, можно не принимать во внимание. Из оставшихся семи величин составим комбинацию вида

$$l^\alpha m^\beta t^\gamma v^\delta a^\lambda \rho^\mu E^\nu.$$

Если выразить размерности v, a, ρ, E через размерности основных величин l, m, t, то эта комбинация перейдет в

$$l^\alpha m^\beta t^\gamma l^\delta t^{-\delta} l^\lambda t^{-2\lambda} m^\mu l^{-3\mu} m^\nu l^{-\nu} t^{-2\nu},$$

т. е. в комбинацию

$$l^{\alpha+\delta+\lambda-3\mu-\nu} m^{\beta+\mu+\nu} t^{\gamma-\delta-2\lambda-2\nu}.$$

Для того чтобы эта комбинация была безразмерной, должно быть

$$\alpha + \delta + \lambda - 3\mu - \nu = 0,$$
$$\beta + \mu + \nu = 0,$$
$$\gamma - \delta - 2\lambda - 2\nu = 0.$$

Из этих трех уравнений три неизвестных параметра можно выразить через оставшиеся четыре. За независимые параметры проще всего принять δ, λ, μ, ν, так как уравнения фактически уже разрешены относительно оставшихся неизвестных α, β, γ:

$$\alpha = -\delta - \lambda + 3\mu + \nu,$$
$$\beta = -\mu - \nu,$$
$$\gamma = \delta + 2\lambda + 2\nu.$$

Параметры δ, λ, μ, ν могут независимо принимать любые значения. Полагая последовательно

1) $\delta = 1$, $\lambda = \mu = \nu = 0$; 2) $\lambda = 1$, $\delta = \mu = \nu = 0$;
3) $\mu = 1$, $\delta = \lambda = \nu = 0$; 4) $\nu = 1$, $\delta = \lambda = \mu = 0$,

получим

1) $\alpha = -1$, $\beta = 0$, $\gamma = 1$; 2) $\alpha = -1$, $\beta = 0$, $\gamma = 2$;
3) $\alpha = 3$, $\beta = -1$, $\gamma = 0$; 4) $\alpha = 1$, $\beta = -1$, $\gamma = 2$.

Этим значениям соответствуют следующие безразмерные комбинации:

1) $\frac{vt}{l}$, 2) $\frac{at^2}{l}$, 3) $\frac{\rho l^3}{m}$, 4) $\frac{Elt^2}{m}$.

Присоединив к ним угол φ, получим всего пять независимых безразмерных комбинаций. Все они являются функциями безразмерных комбинаций (88.3). Значит, рядом (88.3) исчерпываются все независимые безразмерные комбинации, которые можно составить из рассматриваемых физических величин.

2. Как зависит от высоты h скорость свободного падения тела, если начальная скорость его равна нулю?

Решение. Ускорение свободного падения g постоянно и не зависит от массы, плотности, упругих свойств тел и пр. Поэтому искомая скорость v может зависеть только от g и h. Из безразмерных комбинаций (88.3) можно составить всего одну независимую безразмерную комбинацию v^2/la или v^2/gh, содержащую только длину, скорость и ускорение. Она получается делением первой безразмерной комбинации ряда (88.3) на вторую. Поэтому должно быть $f\left(\frac{v^2}{gh}\right) = 0$, откуда $v^2/gh = C = \text{const}$, или $v^2 = Cgh$. Числовой коэффициент C из теории размерности найти нельзя.

3. Пользуясь соображениями размерности, найти зависимость периода колебаний T физического маятника от его приведенной длины l, ускорения свободного падения g и угловой амплитуды α.

Ответ. $T = \varphi(\alpha)\sqrt{l/g}$. Вид функции $\varphi(\alpha)$ из теории размерности определить нельзя. Если эту функцию разложить в ряд Тейлора и сохранить в нем только нулевой член (что можно делать в случае малых колебаний), то получится $T = C\sqrt{l/g}$, где C — постоянный числовой коэффициент, значение которого из теории размерности определить также нельзя. То обстоятельство, что $C \neq 0$, также не вытекает из теории размерности, а должно быть установлено особо (например, опытным путем).

4. Пользуясь соображениями размерности, определить зависимость скорости распространения v продольных упругих возмущений в стержне от модуля Юнга E и плотности материала ρ.

Ответ. $v = C\sqrt{E/\rho}$. Числовой коэффициент C из размерных соображений найти нельзя.

5. Две невзаимодействующие материальные точки, находящиеся в центральном силовом поле, описывают геометрически подобные траектории. Сила F, действующая на каждую материальную точку, пропорциональна ее массе и меняется с расстоянием r до силового центра как r^n, где n — постоянная. Как связаны длины l_1 и l_2 геометрически подобных дуг траекторий с временами T_1 и T_2, затрачиваемыми материальными точками на прохождение этих дуг?

Решение. Должна существовать связь между длиной дуги траектории l, временем T, затрачиваемым материальной точкой не прохождение этой дуги, а также ускорением a, направленным к силовому центру. Ускорения можно выбрать в произвольных, но обязательно подобно расположенных точках. Из этих трех величин можно составить единственную независимую безразмерную комбинацию, за которую можно принять aT^2/l. Следовательно, должно быть $aT^2/l = \text{const}$. Для ускорения можно написать $a = Ar^n$, где A — постоянная, одинаковая для обеих материальных точек. В силу геометрического подобия траекторий, по которым дви-

жутся материальные точки, можно также написать $a = Bl^n$, где B — другая постоянная, также одинаковая для обеих точек. В результате получим $T^2 l^{n-1} = \text{const}$, а потому $T_1^2 l_1^{n-1} = T_2^2 l_2^{n-1}$. В частных случаях $n = 1$ и $n = -2$ получаем $T = \text{const}$ и $T^2/l^3 = \text{const}$. Первое соотношение означает, что в случае гармонического осциллятора период колебаний или период обращения вокруг силового центра не зависит от амплитуды или размеров орбиты. Второе соотношение выражает третий закон Кеплера. Однако этот закон доказан здесь не в общем виде, а только для частиц, движущихся по геометрически подобным траекториям.

6. Наряду с гравитационной постоянной $G = 6{,}6726 \cdot 10^{-8}$ см$^3 \cdot$г$^{-1} \cdot$с$^{-2}$ и скоростью света в вакууме $c = 2{,}9972438$ см/с в физике особую роль играют две фундаментальные постоянные: постоянная Планка $h = 6{,}6261741 \cdot 10^{-27}$ эрг$\cdot$с и постоянная Больцмана $k = 1{,}380662$ эрг/К. Определение последней см. в т. II (§ 62, п. 3). Планк построил естественную систему единиц, в которой эти четыре фундаментальные постоянные полагаются равными единице. Выразить единицы длины, времени, массы и температуры в планковской системе через соответствующие единицы в системе СГС, дополненной единицей температуры — кельвином (К).

Р е ш е н и е. Удобно считать, что обе системы единиц принадлежат к одному и тому же классу, в котором за основные единицы приняты длина, время, масса и температура. Они отличаются одна от другой только различным выбором единиц этих величин. Поэтому любое равенство двух величин одинаковой размерности в одной системе единиц должно быть справедливо и в другой. В качестве таких равенств можно выбрать следующие:

$$l = ct,\ mc^2 = h/t,\ Gm^2/l = mc^2\ (\text{или } Gm/l = c^2),\ kT = mc^2. \qquad (1)$$

Не требуется, чтобы эти равенства выражали какие-то физические законы. Требуется лишь, чтобы: 1) правые и левые части этих равенств имели одинаковые размерности; 2) числовые значения длины l, времени t, массы m и температуры T были подобраны так, чтобы рассматриваемые равенства выполнялись. Первое требование выполняется: в случае первого равенства это очевидно, остальные содержат только члены размерности энергии. В планковской системе единиц $c = h = G = k = 1$. Но тогда из написанных равенств следует $l = t = m = T = 1$.

Это значит, что значения l, t, m, T являются единицами длины, времени, массы и температуры в планковской системе. Решая уравнения (1), находим искомый ответ

$$l = \sqrt{hG/c^3} = 4{,}051 \cdot 10^{-33} \text{ см};\ t = l/c = 1{,}351 \cdot 10^{-43} \text{ с};$$

$$m = h/(lc) = 5{,}456 \cdot 10^{-5} \text{ г};$$

$$\mathscr{E} = kT = mc^2 = 4{,}904 \cdot 10^{16} \text{ эрг} = 3{,}061 \cdot 10^{19} \text{ ГэВ};\ T = mc^2/k = 3{,}552 \cdot 10^{32} \text{ К}.$$

В теоретической физике вместо постоянной h предпочитают применять постоянную $\hbar = h/(2\pi) = 1{,}0545887$ эрг$\cdot$с. Тогда получается

$$l = \sqrt{\hbar G/c^3} = 1{,}616 \cdot 10^{-33} \text{ см};\ t = l/c = 0{,}5390 \cdot 10^{-43} \text{ с};$$

$$m = \hbar/(lc) = 2{,}177 \cdot 10^{-5} \text{ г};$$

$$\mathscr{E} = kT = mc^2 = 1{,}956 \cdot 10^{16} \text{ эрг} = 1{,}221 \cdot 10^{19} \text{ ГэВ};\ T = mc^2/k = 1{,}417 \cdot 10^{32} \text{ К}.$$

ГЛАВА XII

МЕХАНИКА ЖИДКОСТЕЙ И ГАЗОВ

§ 89. ОБЩИЕ СВОЙСТВА ЖИДКОСТЕЙ И ГАЗОВ

1. В отличие от твердых тел *жидкости и газы в состоянии равновесия не обладают упругостью формы**). *Они обладают только объемной упругостью. В состоянии равновесия напряжение в жидкости и газе всегда нормально к площадке, на которую оно действует.* Касательные напряжения вызывают только изменения формы элементарных объемов тела (сдвиги), но не величину самих объемов. Для таких деформаций в жидкостях и газах усилий не требуется, а потому в этих средах при равновесии касательные напряжения не возникают. С точки зрения механики *жидкости и газы могут быть определены как такие среды, в которых при равновесии касательные напряжения существовать не могут.*

Из этого определения следует, что в состоянии равновесия нормальное напряжение в жидкости или газе не зависит от ориентации площадки, на которую оно действует. Для доказательства возьмем произвольно ориентированную площадку, внешнюю нормаль к которой будем характеризовать единичным вектором $\mathbf{n}$. Так как напряжение нормально к площадке, то его можно представить в виде $\boldsymbol{\sigma}_n = -P\mathbf{n}$. Напряжения на площадках, перпендикулярных к координатным осям, запишутся как $\boldsymbol{\sigma}_x = -P_x\mathbf{i}$, $\boldsymbol{\sigma}_y = -P_y\mathbf{j}$, $\boldsymbol{\sigma}_z = -P_z\mathbf{k}$, где $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ — координатные орты. Подставляя эти значения в формулу (74.1), получим

$$P\mathbf{n} = P_x n_x \mathbf{i} + P_y n_y \mathbf{j} + P_z n_z \mathbf{k}.$$

Умножая скалярно это соотношение последовательно на $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$, найдем

$$P = P_x = P_y = P_z. \tag{89.1}$$

Отсюда делаем вывод, что *в состоянии равновесия нормальное напряжение (давление P) не зависит от ориентации площадки, на которую оно действует.*

2. В случае газов нормальное напряжение всегда направлено внутрь газа, т. е. имеет *характер давления.* В жидкостях, как исключение, могут реализоваться и такие случаи, когда нормальное напря-

---

*) Исключения составляют жидкие пленки и поверхностные слои жидкостей. Однако связанные с ними явления в настоящей главе не рассматриваются. Они рассмотрены в т. II нашего курса.

жение является *натяжением* (отрицательным давлением): жидкости оказывают сопротивление на разрыв. Это сопротивление, вообще говоря, довольно значительно и в однородных жидкостях составляет несколько десятков ньютонов на квадратный миллиметр. Однако обычные жидкости неоднородны. Они содержат мельчайшие пузырьки газов, которые действуют подобно надрезам на натянутой веревке и сильно ослабляют прочность жидкости на разрыв. Поэтому в подавляющем большинстве случаев в жидкостях напряжения также имеют характер давлений. Вот почему для обозначения нормального напряжения мы пользуемся символом $-P\mathbf{n}$ (давление), а не $+T\mathbf{n}$ (натяжение). Если давление переходит в натяжение, т. е. становится отрицательным, то это, как правило, ведет к нарушению плотности жидкости. С отмеченными особенностями связано и то обстоятельство, что газы обладают способностью к неограниченному расширению: газ всегда полностью заполняет объем сосуда, в котором он заключен. Напротив, каждой жидкости свойствен определенный *собственный объем*, лишь незначительно меняющийся с изменением внешнего давления. Жидкости имеют свободную поверхность и могут собираться в капли. Чтобы отметить эти обстоятельства, жидкие среды называют также *капельно-жидкими*. В механике при рассмотрении движений капельных жидкостей и газов газ обычно рассматривают как частный случай жидкости. Таким образом, под жидкостью в обобщенном смысле слова понимают либо капельную жидкость, либо газ. *Отдел механики, занимающийся изучением движения и равновесия жидкостей, называется гидродинамикой.*

3. Давление, существующее в жидкости, обусловлено ее сжатием. А так как касательные напряжения не возникают, то упругие свойства жидкостей по отношению к малым деформациям характеризуются только одной упругой постоянной: *коэффициентом сжимаемости*

$$\gamma = -\frac{1}{V}\frac{dV}{dP}, \tag{89.2}$$

или обратной ему величиной — *модулем всестороннего сжатия*

$$K = -V\frac{dP}{dV}. \tag{89.3}$$

Предполагается, что температура жидкости при сжатии поддерживается постоянной. При рассмотрении деформаций, сопровождающихся изменениями температуры, вместо (89.2) и (89.3) предпочтительнее писать

$$\gamma_T = -\frac{1}{V}\left(\frac{dV}{dP}\right)_{T=\text{const}}, \tag{89.2а}$$

$$K_T = -V\left(\frac{dP}{dV}\right)_{T=\text{const}} \tag{89.3а}$$

и называть γ_T и K_T *изотермическими коэффициентом и модулем всестороннего сжатия.* В быстрых процессах, происходящих практически без теплообмена, особую роль играют *адиабатические коэффициенты* и *модули упругости* (см. § 75, п. 8).

При рассмотрении деформаций твердых тел модуль всестороннего сжатия мы определили формулой (77.3), отличающейся от (89.3) тем, что вместо величины dP в ней стоит просто P. Такое определение было возможно потому, что твердое тело обладает определенным объемом, когда внешнее давление P обращается в нуль, и этот объем меняется мало даже при конечных изменениях P. Формула (89.3) переходит в (77.3), если положить $dP = P - P_0$ и считать, что $P_0 = 0$. Так же можно было поступать и в случае капельных жидкостей. Но в случае газов формула (77.3) не годится. Надо пользоваться более общей формулой (89.3), так как при отсутствии внешнего давления объем газа становится бесконечно большим. Именно так мы поступали в § 85 при рассмотрении вопроса о скорости звука в газах.

Можно также сказать, что некоторое состояние тела с давлением P_0 (и температурой T) мы выбираем за *нормальное* и рассматриваем изменения объема тела по отношению к этому нормальному состоянию. В случае твердых и капельно-жидких тел модуль упругости (89.3) в широких пределах не зависит от значения P_0. По этой причине и можно положить $P_0 = 0$. В случае же газов конкретизация значения P_0 существенна. Приравнивать P_0 к нулю в этом случае нельзя. Так, если воспользоваться законом Бойля—Мариотта $P \sim 1/V$ (при $T = \text{const}$), то из (89.3) легко получить $K = P$. Отсюда видно, что о модуле упругости газа можно говорить лишь тогда, когда указанно его давление (при заданной температуре).

4. Малую сжимаемость капельных жидкостей можно демонстрировать с помощью следующего эффективного опыта. Сосуд из пластмассы наполовину наполняется водой. Если произвести выстрел из мелкокалиберной винтовки, чтобы пуля пролетела выше уровня жидкости, то она оставляет лишь отверстия в стенках сосуда, а самый сосуд остается целым. Если же пуля попадает в сосуд на несколько сантиметров ниже уровня жидкости, то сосуд разлетается вдребезги. Дело в том, что для проникновения пули в воду она должна либо сжать ее на величину своего объема, либо вытеснить наверх. Для вытеснения недостаточно времени. Происходит сжатие — в жидкости развиваются большие давления, которые и разрывают стенки сосуда. Для опыта годятся также деревянные ящики или бумажные коробки, наполненные водой. В последнем случае опыт удается уже с духовым ружьем. Аналогичные явления возникают при разрывах глубинных бомб, применяемых против подводных лодок. Вследствие малой сжимаемости воды при взрыве в воде развиваются громадные давления, которые и разрушают лодку.

Малая сжимаемость жидкостей позволяет во многих случаях вообще полностью пренебречь изменениями их объема. Тогда вводят представление об *абсолютно несжимаемой жидкости.* Это — идеа-

лизация, которой постоянно пользуются. Конечно, и в несжимаемой жидкости давление определяется степенью ее сжатия. Однако даже при очень больших давлениях изменения объема «несжимаемых жидкостей» столь ничтожны, что с ними во многих случаях можно не считаться. Можно сказать, что несжимаемая жидкость — это предельный случай сжимаемой жидкости, когда для получения бесконечно больших давлений уже достаточны бесконечно малые сжатия. Несжимаемая жидкость является такой же абстракцией, как и твердое тело. Деформации твердых тел существенны для выяснения механизма возникновения внутренних напряжений. Но когда деформации малы, можно в ряде случаев заменить реальное тело идеализированным твердым телом. Твердое тело — это предельный случай реального тела, когда для получения бесконечно больших напряжений достаточны бесконечно малые деформации.

Можно или нельзя реальную жидкость заменять идеальной — это зависит не столько от того, насколько мала сжимаемость жидкости, сколько от содержания тех вопросов, на которые надо получить ответы. Так, при рассмотрении звуковых волн, вообще говоря, принципиально невозможно отвлечься от сжимаемости жидкостей. А при рассмотрении воздушных течений, если только перепады давления не слишком велики, воздух часто можно рассматривать как несжимаемую жидкость (см. § 94, п. 5).

5. В состоянии равновесия давление жидкости (или газа) P меняется с изменением ее плотности ρ и температуры T. Оно однозначно определяется значениями этих параметров. Соотношение

$$P = f(\rho, T) \tag{89.4}$$

между давлением, плотностью и температурой в состоянии равновесия называется *уравнением состояния*. Оно имеет разный вид для разных веществ и особой простотой отличается в случае разреженных газов. Вопросы, связанные с уравнением состояния, подробно разбираются во втором томе нашего курса. Здесь мы ограничимся замечанием, что изотермический модуль упругости K_T можно вычислить, зная уравнение состояния, простым дифференцированием. Он в общем случае является функцией плотности и температуры или давления и температуры.

6. *Если жидкость находится в движении, то наряду с нормальными напряжениями в ней могут возникать и касательные силы.* Однако последние определяются не самими деформациями жидкости (сдвигами), а их *скоростями*, т. е. производными деформаций по времени. Поэтому их следует относить к классу *сил трения* или *вязкости*. Они называются *касательными* или *сдвиговыми вязкими силами*.

Наряду с касательными могут существовать и *нормальные* или *объемные* силы вязкости. От обычных сил давления P эти силы отличаются тем, что они также определяются не степенью сжатия жидко-

сти, а *скоростью изменения сжатия во времени*. Эти силы играют существенную роль в быстрых процессах, например при распространении предельно коротких ультразвуковых волн (длина которых приближается к молекулярным размерам и межмолекулярным расстояниям). В предельном случае, когда скорость изменения деформаций в жидкости стремится к нулю, в ней исчезают все силы вязкости, как сдвиговые, так и обусловленные сжатием. Жидкость, в которой при любых движениях не возникают силы вязкости (как касательные, так и нормальные), называется *идеальной*. Иными словами, идеальной называют жидкость, в которой могут существовать только силы нормального давления P, однозначно определяемого степенью сжатия и температурой жидкости. Такие силы могут быть вычислены с помощью уравнения состояния жидкости (89.4) не только тогда, когда жидкость покоится, но и тогда, когда она движется произвольным образом. Конечно, строго идеальных жидкостей не существует. Это — абстракции, которыми можно пользоваться, когда скорости изменения деформаций в жидкости не очень велики.

7. Если к жидкости приложить касательные напряжения, то возникает движение. Оно в конце концов прекращается и переходит в состояние покоя, в котором касательные напряжения отсутствуют. Скорости изменения деформаций жидкости могут меняться в широких пределах. Для таких жидкостей, как вода или спирт, эти изменения происходят весьма быстро; для очень вязких жидкостей, как мед или патока, — весьма медленно. Наконец, есть вещества, которые при быстрых воздействиях на них ведут себя, как твердые тела, а при медленных — как очень вязкие жидкости. Сюда относятся так называемые *аморфные твердые тела*. Например, кусок сапожного вара или асфальта разбивается на мелкие части, если его ударить молотком. На асфальте можно стоять и по нему можно ходить. Но асфальт вытекает из бочки в течение недель или месяцев. Скорость вытекания сильно увеличивается с температурой. Стеклянная палочка, положенная своими концами на две опоры, прогибается, если подождать достаточно длительное время (месяцы или годы), причем ее прогиб не исчезает по прекращении действия силы тяжести. Эти примеры показывают, что нельзя провести резкое разграничение между жидкостями и аморфными твердыми телами. Истинно твердыми телами являются только *кристаллы*. Впрочем, говоря о жидкостях, мы всегда будем иметь в виду жидкости, не обладающие аномально большой вязкостью, когда отличие их от аморфных твердых тел выступает вполне отчетливо.

§ 90. ОСНОВНЫЕ УРАВНЕНИЯ РАВНОВЕСИЯ И ДВИЖЕНИЯ ЖИДКОСТЕЙ

1. Силы, действующие в жидкости, как и во всякой другой сплошной среде, обычно разделяются на силы *массовые* (объемные) и силы *поверхностные*. Массовая сила пропорциональна массе dm,

а с ней и объему dV элемента жидкости, на который она действует. Эту силу можно обозначить как $\mathbf{f}\,dV$, называя $\mathbf{f}$ *объемной плотностью массовых сил.* Важнейшими примерами массовых сил являются сила тяжести и силы инерции (когда движение рассматривают в неинерциальных системах отсчета). В случае силы тяжести $\mathbf{f} = \rho\mathbf{g}$, где ρ — плотность жидкости, а $\mathbf{g}$ — ускорение свободного падения. Поверхностные силы — это такие силы, которым подвергается каждый объем жидкости благодаря нормальным и касательным напряжениям, действующим на его поверхности со стороны окружающих частей жидкости.

2. Рассмотрим случай, когда касательных напряжений нет, а есть только силы нормального давления. В идеальной жидкости это будет всегда, т. е. при любых движениях. В остальных случаях — тогда, когда жидкость покоится, т. е. *в гидростатике.* Определим равнодействующую сил давления, действующих на бесконечно малый элемент объема жидкости dV. Сначала найдем проекцию этой равнодействующей на направление координатной оси X. Возьмем в качестве элемента dV бесконечно малый цилиндр с площадью оснований dS и длиной dx (рис. 230), ориентированный вдоль оси X. Абсциссы оснований цилиндра обозначим соответственно через x и $x + dx$. Сила давления, действующая на первое основание, равна $P(x)\,dS$, на второе — $P(x + dx)\,dS$. В скобках у P указано значение аргумента x, от которого P зависит. Конечно, P может зависеть и от координат y, z, а также от времени t. Но все эти аргументы не меняются при переходе от одного основания цилиндра к другому, а потому в рассматриваемом нами вопросе могут считаться постоянными. При желании поперечные размеры цилиндра можно взять бесконечно малыми высшего порядка по сравнению с длиной dx. А тогда y и z могут рассматриваться постоянными не только при смещениях вдоль цилиндра, но и поперек. Силы давления, действующие на боковую поверхность цилиндра, перпендикулярны к оси X, а потому при вычислении составляющих вдоль этой оси роли не играют. Итак, проекция сил давления на ось X, действующих на рассматриваемый элемент объема жидкости, равна

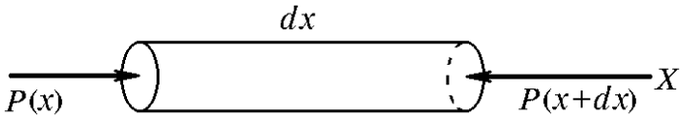


Рис. 230

$$[P(x) - P(x + dx)]\,dS.$$

Бесконечно малую разность в квадратных скобках можно заменить дифференциалом функции P:

$$P(x + dx) - P(x) = dP_{\substack{y=\text{const}\\ z=\text{const}\\ t=\text{const}}} = \left(\frac{dP}{dx}\right)_{\substack{y=\text{const}\\ z=\text{const}\\ t=\text{const}}} dx.$$

Дополнительные условия $y = \text{const}$, $z = \text{const}$, $t = \text{const}$ указывают на то, что при взятии производной $\frac{dP}{dx}$ и дифференциала dP координаты y, z и время t должны рассматриваться как постоянные. Производная функции $P(x, y, z, t)$, взятая при таких дополнительных условиях, как известно, называется *частной производной* и обозначается через $\frac{\partial P}{\partial x}$. Используя это обозначение, получаем для вычисляемой проекции силы

$$-\frac{\partial P}{\partial x}\, dS\, dx = -\frac{\partial P}{\partial x}\, dV,$$

так как $dS\, dx = dV$. Эта проекция, таким образом, пропорциональна элементу объема dV, и ее можно обозначить как $s_x dV$. Величина s_x есть x-составляющая силы, действующей на единицу объема жидкости, которая возникает из-за изменения нормального давления P в пространстве. По самому смыслу она не может зависеть от формы элемента dV. Мы взяли dV в виде цилиндра только потому, что таким путем достигается наибольшая простота и наглядность вычисления. Можно таким же путем найти проекции s_y и s_z, выбирая в качестве dV элементарные цилиндры, ориентированные параллельно координатным осям Y и Z. В результате найдем, что на единицу объема жидкости действует сила s, обусловленная поверхностными силами давления, точнее, их изменениями в пространстве. Ее проекции равны

$$s_x = -\frac{\partial P}{\partial x}, \quad s_y = -\frac{\partial P}{\partial y}, \quad s_z = -\frac{\partial P}{\partial z}. \tag{90.1}$$

Сам вектор $\mathbf{s}$ равен

$$\mathbf{s} = -\frac{\partial P}{\partial x}\mathbf{i} - \frac{\partial P}{\partial y}\mathbf{j} - \frac{\partial P}{\partial z}\mathbf{k}, \tag{90.2}$$

или сокращенно

$$\mathbf{s} = -\operatorname{grad} P. \tag{90.3}$$

Мы ввели обозначение

$$\operatorname{grad} P \equiv \frac{\partial P}{\partial x}\mathbf{i} + \frac{\partial P}{\partial y}\mathbf{j} + \frac{\partial P}{\partial z}\mathbf{k}. \tag{90.4}$$

Этот вектор называется *градиентом скаляра* P (см. также § 29). Таким образом, *объемная плотность* $\mathbf{s}$ *результирующей сил давления, действующих на элементы объема жидкости, равна градиенту* P, *взятому с противоположным знаком*. Мы видим, что сила $\mathbf{s}$ обусловлена не значением давления P, а его *пространственными изменениями*. Величина P также существенна. Она определяет *степень сжатия жидкости* в рассматриваемой точке пространства.

3. В состоянии равновесия сила $\mathbf{s}$ должна уравновешиваться массовой силой $\mathbf{f}$. Это приводит к уравнению

$$\operatorname{grad} P = \mathbf{f}, \tag{90.5}$$

которое является *основным уравнением гидростатики.* В координатной форме оно имеет вид

$$\frac{\partial P}{\partial x} = f_x, \quad \frac{\partial P}{\partial y} = f_y, \quad \frac{\partial P}{\partial z} = f_z. \tag{90.6}$$

Можно написать и основное уравнение гидродинамики идеальной жидкости. В этом случае формула (90.3) также применима, а потому мы получаем

$$\rho \frac{d\mathbf{v}}{dt} = \mathbf{f} - \operatorname{grad} P, \tag{90.7}$$

где $\mathbf{v}$ — скорость, а $\frac{d\mathbf{v}}{dt}$ — ускорение жидкости в рассматриваемой точке. Уравнение (90.7) называется *уравнением Эйлера.*

4. Уравнение (90.5) показывает, что при равновесии жидкости сила $\mathbf{f}$ (точнее, плотность силы или сила, действующая на единицу объема жидкости) должна выражаться градиентом однозначной скалярной функции. Это есть необходимое и достаточное условие того, чтобы сила $\mathbf{f}$ была консервативной (см. § 29). Таким образом, *для равновесия жидкости необходимо, чтобы силовое поле, в котором она находится, было консервативным.* В неконсервативных силовых полях равновесие невозможно.

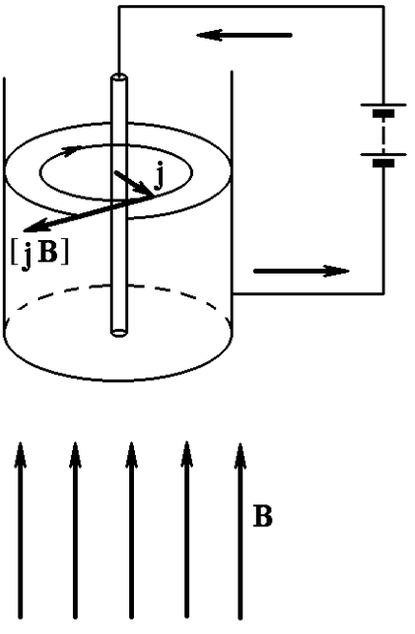


Рис. 231

Примером может служить проводящая жидкость, помещенная в магнитное поле, когда через нее проходит электрический ток. В этом случае со стороны магнитного поля на жидкость действует сила $\mathbf{f} = C[\mathbf{jB}]$, где $\mathbf{B}$ — индукция магнитного поля, $\mathbf{j}$ — плотность тока, а C — числовой коэффициент, значение которого зависит от выбора единиц. Поместим цилиндрический сосуд с раствором электролита (например, $CuSO_4$) над одним из полюсов сильного электромагнита (рис. 231). Вдоль оси цилиндра расположен цилиндрический проводник. Между ним и боковой стенкой сосуда наложим электрическое напряжение в несколько вольт. В электролите вдоль радиусов цилиндра потечет электрический ток. Сила $\mathbf{f} = C[\mathbf{jB}]$ будет направлена по касательным к окружностям с центрами на оси цилиндра. Она вызовет вращение жидкости вокруг указанной оси. Вращение будет ускоряться до тех пор, пока силы, действующие со стороны магнитного поля, не уравновесятся силами вязкости.

§ 91. ГИДРОСТАТИКА НЕСЖИМАЕМОЙ ЖИДКОСТИ

1. Когда $f = 0$, то $\frac{\partial P}{\partial x} = \frac{\partial P}{\partial y} = \frac{\partial P}{\partial z} = 0$. Значит, *если нет массовых сил, то при равновесии давление во всех точках жидкости одина-*

ково. Это — *закон Паскаля* по имени французского ученого Блеза Паскаля (1623—1662). В частности, *при отсутствии массовых сил жидкость может находиться в равновесии только тогда, когда внешнее давление на ее поверхность одно и то же во всех точках этой поверхности.* В противном случае возникнет движение жидкости. *При отсутствии массовых сил одинаковое давление на поверхность жидкости возбуждает такое же давление во всех точках внутри жидкости.*

Если жидкость находится в поле тяжести, то $\mathbf{f} = \rho\mathbf{g}$. Направим ось Z вертикально вверх. Тогда основные уравнения равновесия жидкости примут вид

$$\frac{\partial P}{\partial x} = \frac{\partial P}{\partial y} = 0, \quad \frac{\partial P}{\partial z} = -\rho g. \tag{91.1}$$

Из них следует, что при механическом равновесии давление не может зависеть от x и y. Оно должно оставаться постоянным в каждой горизонтальной плоскости $z = \text{const}$. Горизонтальные плоскости суть *плоскости равного давления.* В частности, свободная поверхность жидкости горизонтальна, поскольку она находится под постоянным давлением атмосферы. Таким образом, при механическом равновесии давление может зависеть лишь от координаты z. Из третьего уравнения (90.1) следует поэтому, что для механического равновесия необходимо, чтобы произведение ρg было функцией только z. Так как g не зависит от x и y (зависимостью g от географической широты и долготы места мы пренебрегаем), то, следовательно, и плотность ρ может меняться только с высотой. В силу уравнения состояния (89.4) давлением P и плотностью ρ определяется температура жидкости T. Итак, *при механическом равновесии давление, температура и плотность жидкости являются функциями только z и не могут зависеть от x и y.*

2. Допустим теперь, что жидкость *однородна* и ее можно рассматривать как *несжимаемую* ($\rho = \text{const}$). Кроме того, будем считать постоянным ускорение свободного падения g, пренебрегая его зависимостью от высоты z. Тогда легко интегрируется и последнее уравнение системы (91.1). В результате такого интегрирования получим

$$P = P_0 - \rho g z. \tag{91.2}$$

Постоянная интегрирования P_0 есть давление жидкости на высоте $z = 0$, т. е. атмосферное давление, если начало координат поместить на свободной поверхности жидкости. Формула (91.2) определяет также давление жидкости на дно и стенку сосуда, а также на поверхность всякого тела, погруженного в жидкость. Она охватывает всю гидростатику, излагаемую в школьных курсах физики.

3. Остановимся теперь на *законе Архимеда* и связанных с ним вопросах. Выделим мысленно из жидкости произвольный объем, ограниченный замкнутой поверхностью S (рис. 232). Если жидкость

находится в механическом равновесии, то, разумеется, должен находиться в равновесии и выделенный объем. Поэтому должны обращаться в нуль равнодействующая и момент внешних сил, действующих на рассматриваемый объем жидкости. Внешние силы — это вес Q выделенного объема жидкости и давление на поверхность S со стороны окружающей жидкости. Значит, равнодействующая F сил гидростатического давления, действующих на поверхность S, должна равняться Q — весу жидкости в объеме, ограниченном поверхностью S. Эта равнодействующая должна быть направлена вверх и проходить через центр масс A выделенного объема жидкости, чтобы полный момент внешних сил, действующих на него был равен нулю. Допустим теперь, что жидкость из выделенного нами объема удалена, и на ее место помещено любое твердое тело. Если это тело удерживается в равновесии, то в состоянии окружающей жидкости *никаких изменений не произойдет*. Не изменится и давление, оказываемое жидкостью на поверхность S. В результате мы приходим к закону Архимеда. *Если тело, погруженное в жидкость, удерживается в механическом равновесии, то со стороны окружающей жидкости оно подвергается выталкивающей силе гидростатического давления, численно равной весу жидкости в объеме, вытесненном телом. Эта выталкивающая сила направлена вверх и проходит через центр масс A жидкости, вытесненной телом.* Точку A будем называть *центром плавучести тела.* Ее положением, как будет показано ниже, определяются *равновесие* и *устойчивость* плавающего тела.

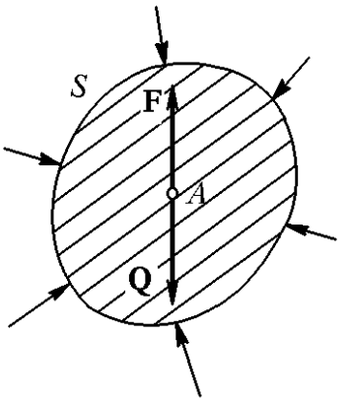


Рис. 232

4. С помощью закона Архимеда решается вопрос о равновесии тел, плавающих в жидкости. *Для равновесия необходимо, чтобы вес тела был равен весу вытесненной им жидкости, а центр плавучести A лежал на одной вертикали с центром масс самого тела.* Что касается устойчивости равновесия, то при решении этого вопроса надо различать два случая.

Случай 1. Плавающее тело погружено в жидкость целиком. В этом случае при любых смещениях и поворотах тела его центр масс C и центр плавучести A сохраняют свое положение относительно тела. *Равновесие устойчиво, если центр масс тела C лежит ниже его центра плавучести A, и неустойчиво, если он лежит выше A.* Действительно, если тело слегка повернуть относительно положения равновесия вокруг горизонтальной оси, то в обоих случаях момент пары сил $\mathbf{Q}$ и $\mathbf{F}$ будет стремиться опустить точку C и поднять точку A (рис. 233). В результате этого тело приходит в положение устойчивого равновесия, в котором точка C расположена ниже точки A.

Случай 2. Плавающее тело погружено в жидкость не целиком, а частично выступает над ее свободной поверхностью. По сравнению

с предыдущим этот случай более сложен, так как при смещении тела из положения равновесия меняется форма вытесняемого им объема жидкости. Вследствие этого положение центра плавучести относительно плавающего тела изменяется, что и усложняет исследование.

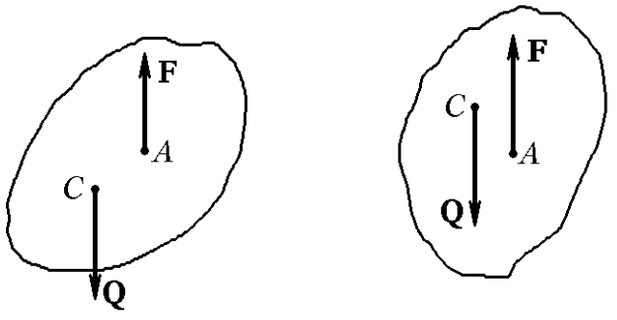
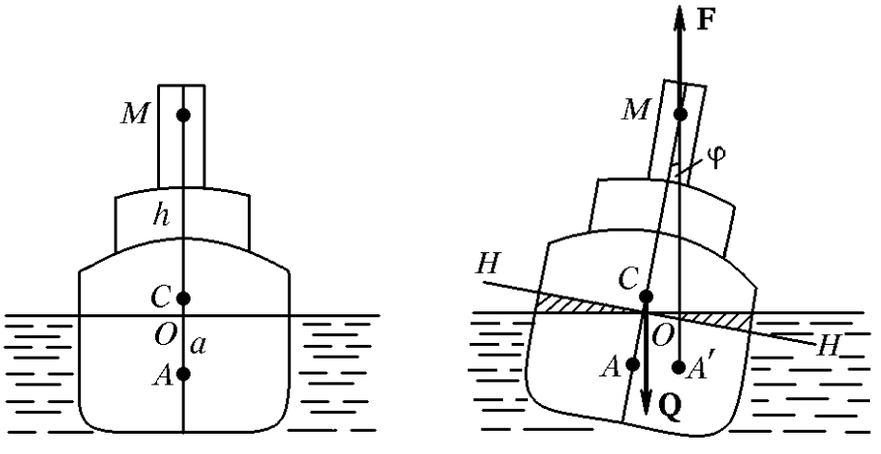


Рис. 233

Рассматриваемый случай представляет основной интерес при исследовании устойчивости плавающих кораблей. На рис. 234 *а* схематически изображен корпус корабля в «килевом» положении, когда центр масс корабля C и центр плавучести A лежат на одной вертикали, совпадающей с вертикальной осью симметрии корабля. При наклоне корабля на малый угол φ (рис. 234 *б*) центр плавучести смещается относительно корабля в точку A', оставаясь практически на прежней высоте. Выталкивающая сила теперь проходит через точку A', и линия ее действия пересекает вертикальную ось симметрии корабля в точке M, называемой *метацентром*. Если ме-



а *б*

Рис. 234

тацентр лежит выше центра масс корабля, то момент пары сил **Q** и **F** будет возвращать корабль в исходное положение. В этом случае равновесие корабля устойчивое. Если же метацентр M лежит ниже центра масс корабля, то пара сил **Q** и **F** будет еще больше отклонять корабль от исходного положения. В этом случае равновесие неустойчиво. Расстояние h между точками C и M называется *метацентрической высотой. Если метацентрическая высота положительна, то равновесие устойчиво, если отрицательна, то неустойчиво.* Чем больше h, тем устойчивее равновесие. Момент пары сил **Q** и **F**, возвращающий корабль в исходное положение, называется *выпрямляющим моментом.* Он, очевидно, равен

$$M = Qh \sin \varphi. \qquad (91.3)$$

Величина h сама зависит от φ, так как при изменении наклона φ меняется и положение метацентра относительно корабля. Найдем это положение и вычислим метацентрическую высоту h в предельном случае бесконечно малых углов наклона φ.

Так как выталкивающая сила проходит через точку A' и направлена вертикально вверх, то ее момент относительно точки A будет $N = Q \cdot AM \sin\varphi$ или (при малых φ) $N = Q(h + a)\varphi$, где a — расстояние между центром масс корабля и его центром плавучести в положении равновесия. Величина a считается положительной, если точка C лежит выше A, и отрицательной, если она лежит ниже A. Момент N, конечно, не зависит от того, в какой точке линии $A'M$ выбрана точка приложения выталкивающей силы $\mathbf{F}$. Разложим силу $\mathbf{F}$ на составляющую $\mathbf{F}_{\parallel}$, параллельную оси корабля AM, и составляющую $\mathbf{F}_{\perp}$, к ней перпендикулярную. Если точку приложения силы $\mathbf{F}$ поместить в A', то составляющая $\mathbf{F}_{\perp}$ не даст момента относительно центра плавучести A, и вычисления упростятся. Тогда полный момент N будет создаваться только составляющей $\mathbf{F}_{\parallel}$. Понятно, что момент этой составляющей будет одним и тем же относительно всех точек, лежащих на оси AM. Из изложенного следует, что величину $N = Q(h + a)\varphi$ можно рассматривать как момент выталкивающих сил давления относительно произвольной точки оси AM, если из этих сил вычесть их составляющие, перпендикулярные к той же оси. Поэтому момент N можно вычислить иначе. Если корабль наклонен на угол φ, то выталкивающие силы давления с правой стороны увеличатся, а с левой — уменьшатся. При этом мы имеем в виду не полные силы, а только их составляющие, параллельные AM. Пусть x — расстояние (координата) произвольной точки плоскости HH от оси Y, проходящей через O перпендикулярно к плоскости рисунка. Тогда увеличение давления в соответствующей точке дна будет $\rho g x \varphi$, а момент N представится выражением

$$N = \rho g \varphi \int x^2\, dS = \rho g I \varphi,$$

где I — *момент инерции* поперечного сечения корабля вдоль ватерлинии относительно оси Y: $I = \int x^2\, dS$ (ср. § 80, п. 1). Сравнивая оба выражения для N, получаем

$$h = \frac{I}{V} - a, \tag{91.4}$$

где $V = Q/(\rho g)$ — *водоизмещение корабля*, т. е. объем вытесняемой им воды.

5. Рассмотрим теперь жидкость в сосуде, равномерно вращающемся вокруг вертикальной оси с угловой скорость $\boldsymbol{\omega}$. Будем предполагать, что жидкость вращается вместе с сосудом, а сам сосуд обладает осевой симметрией, например имеет цилиндрическую форму. Эта задача сводится к статической, если перейти во вращающуюся систему отсчета, в которой жидкость покоится. Теперь в уравнении

(90.5) $\mathbf{f}$ слагается из силы тяжести $\rho\mathbf{g}$ и центробежной силы $\rho\omega^2\mathbf{r}$, где $\mathbf{r}$ — радиус-вектор, проведенный от оси вращения к рассматриваемой точке и перпендикулярный к оси. Если поместить начало координат на оси вращения так, чтобы ось Z совпала с осью вращения, то уравнения (90.6) примут вид

$$\frac{\partial P}{\partial x} = \rho\omega^2 x, \quad \frac{\partial P}{\partial y} = \rho\omega^2 y, \quad \frac{\partial P}{\partial z} = -\rho g. \tag{91.5}$$

Считая ρ постоянной и интегрируя, получим

$$P = \frac{1}{2}\rho\omega^2(x^2 + y^2) - \rho g z + P_0, \tag{91.6}$$

или

$$P = \frac{1}{2}\rho\omega^2 r^2 - \rho g z + P_0. \tag{91.6а}$$

Уравнение свободной поверхности $P = \text{const}$ принимает вид

$$\frac{1}{2}\omega^2(x^2 + y^2) - gz = \text{const}.$$

Это — параболоид вращения, обращенный своей выпуклостью вниз. Если начало координат поместить в вершину параболоида, то постоянная P_0 будет иметь смысл наружного атмосферного давления. Уравнение свободной поверхности жидкости будет

$$\frac{1}{2}\omega^2(x^2 + y^2) = gz.$$

Конечно, рассмотренную задачу можно трактовать и как чисто динамическую. Если жидкость вращается как целое, то при таком движении в ней не возникают силы вязкости. Естественные поверхностные силы, действующие в жидкости, сводятся к силам нормального давления. Поэтому в этом случае можно пользоваться уравнением Эйлера (90.7) независимо от того, является ли жидкость идеальной или вязкой. При равномерном вращении производная $\frac{d\mathbf{v}}{dt}$ сводится к центростремительному ускорению $-\omega^2\mathbf{r}$. Поэтому, полагая в уравнении (90.7) $\mathbf{f} = \rho\mathbf{g}$, получим

$$-\rho\omega^2 r = \rho\mathbf{g} - \text{grad}\, P,$$

а это векторное уравнение эквивалентно трем уравнениям в проекциях (91.5).

Если сосуд имеет плоское дно, то для определения давления на дно надо в формуле (91.6а) положить $z = -h$, где h — высота уровня жидкости над дном на оси вращения (напомним, что ось Z направлена вверх). Получим

$$P - P_0 = \rho g h + \frac{1}{2}\rho\omega^2 r^2. \tag{91.7}$$

Давление в центре, таким образом, минимально и монотонно возрастает к краям. С этим связано, например, следующее явление. Если чайной ложкой привести во вращение воду в стакане, то после прекращения помешивания чаинки и песчинки, имеющиеся в ней, со-

бираются в центре дна. Дело в том, что эти частицы тяжелее воды и опускаются на дно. Здесь их вращение замедляется благодаря силам трения о дно стакана, и под влиянием разности гидростатических давлений частицы перемещаются к центру дна.

Вычислим теперь полную силу давления жидкости на дно сосуда. С этой целью воспользуемся уравнением свободной поверхности жидкости $\frac{1}{2}\omega^2 r^2 = gz$ и перепишем формулу (91.7) в виде $P - P_0 = \rho g(h + z)$. Интегрируя по площади дна, найдем искомую силу

$$F = \rho g \int (h + z)\, dS = \rho g V, \tag{91.8}$$

где V — объем жидкости в сосуде. Как и следовало ожидать, полная сила давления равна весу этого объема жидкости.

ЗАДАЧИ

1. Жидкость налита в сосуд, имеющий форму прямоугольного параллелепипеда. Вычислить момент сил гидростатического давления, действующих на боковую стенку сосуда, относительно ее нижнего основания.

Ответ. $M = \frac{1}{3}\rho g h^2 S$, где h — высота уровня жидкости относительно дна, S — площадь рассматриваемой боковой стенки сосуда.

2. *Гидростатический парадокс.* Сила давления жидкости на дно сосуда не зависит от формы сосуда, а только от площади дна, разности уровней поверхности жидкости и дна, а также от плотности жидкости. Так, эта сила будет одной и той же для всех трех сосудов, изображенных на рис. 235, если они имеют одинаковое дно, а жидкость налита до одного и того же уровня. При взвешивании сосудов с жидкостью весы должны показывать один и тот же вес, поскольку показание весов зависит от силы, с которой дно сосуда давит на чашку весов. Указать, в чем ошибочность приведенного рассуждения. Что в действительности покажут весы?

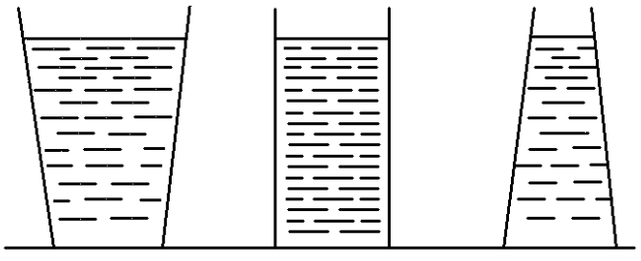


Рис. 235

3. Непосредственным вычислением результирующей сил давления жидкости на поверхность погруженного тела и их моментов убедиться в справедливости закона Архимеда.

Решение. Мысленно разобьем погруженное тело на бесконечно тонкие вертикальные столбики (рис. 236). Допустим для простоты, что каждый столбик пересекает поверхности тела только два раза. (Случай, когда это условие не соблюдается, читателю предлагается разобрать самостоятельно.) Пусть dS_1 и dS_2 — элементарные площадки, вырезаемые одним из столбиков на поверхности тела. Силы, действующие на эти площадки, перпендикулярны к ним и равны соответственно $P_1 dS_1$ и $P_2 dS_2$. Их вертикальные составляющие будут $P_1 dS_1 \cos\alpha_1$ и $P_2 dS_2 \cos\alpha_2$, или $P_1 d\sigma$ и $P_2 d\sigma$, где $d\sigma = dS_1 \cos\alpha_1 = dS_2 \cos\alpha_2$ — площадь нормального сечения столбика. Ре-

зультирующая этих двух сил, направленная вверх, равна $dF_z = (P_2 - P_1)d\sigma = \rho g h d\sigma = \rho g dV$, где h — высота столбика, а $dV = h d\sigma$ — его объем. Интегрируя по всему объему тела, находим выталкивающую силу $F_z = \rho g V$. Теперь надо найти момент вертикальных выталкивающих сил, действующих на столбики, относительно произвольной оси. Если ось вертикальна, то момент, очевидно, равен нулю. Поэтому достаточно ограничиться вычислением момента относительно произвольной горизонтальной оси. Примем таковую за координатную ось X. Искомый момент будет $M_x = \int y\, dF_z = g \int \rho y\, dV = g \int y\, dm$, где dm — масса жидкости, вытесненная соответствующим столбиком тела. Аналогично для момента относительно оси Y: $M_y = g \int x\, dm$. Момент обратится в нуль, когда $\int x\, dm = \int y\, dm = 0$, т. е. когда начало координат помещено на вертикальной оси, проходящей через центр плавучести. Тем самым доказано, что линия действия выталкивающей силы проходит через центр плавучести тела. Для завершения доказательства надо было бы еще исследовать, какие силы давления действуют на поверхность погруженного тела в горизонтальных направлениях. Однако этот вопрос не нуждается в специальном исследовании. Например, когда речь идет о силах, действующих параллельно оси X, то достаточно разбить тело на бесконечно малые столбики, параллельные этой оси, а затем повторить все сказанное выше, с той только разницей, что величину g надо положить равной нулю. Отсюда следует, что равнодействующая горизонтальных сил давления, действующих на погруженное тело, и их момент равны нулю.

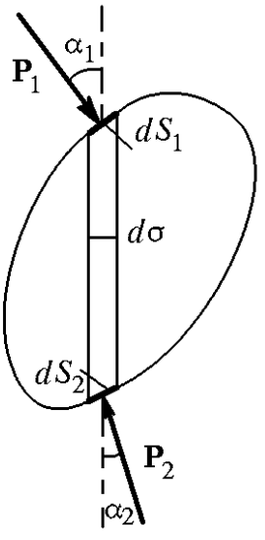


Рис. 236

4. Найти условие устойчивости однородного прямоугольного параллелепипеда, плавающего на поверхности жидкости в положении, когда одно из оснований его горизонтально. Длины сторон горизонтального основания A и B, высота C $(A > B)$. Плотность материала тела относительно жидкости $\rho < 1$.

Ответ. $B^2 > 6\rho(1-\rho)C^2$.

5. Та же задача для однородного цилиндра радиусом r и длиной l, плавающего в вертикальном положении.

Ответ. $r^2 > 2\rho(1-\rho)l^2$.

6. Та же задача для однородного цилиндра радиусом r и длиной l, плавающего в горизонтальном положении.

Ответ. $\frac{l}{r} > \left(2\sin\frac{\alpha}{2}\right)^2$, где угол α определяется из трансцендентного уравнения $\alpha - \sin\alpha = 2\pi\rho$. Например, при $\rho = 1/2$ из него получаем $\alpha = \pi$, и условие устойчивости принимает вид $l > 4r$. При других значениях ρ равновесие может быть устойчивым и при меньших значениях l. Так, при $\alpha = \pi/2$ и $\alpha = 3\pi/2$ получаем соответственно $\rho = 1/4 - 1/(2\pi) \approx 0{,}091$ и $\rho = 3/4 + 1/(2\pi) \approx 0{,}841$. При таких значениях ρ равновесие устойчиво, если $l > 2r$. При $l > 4r$ равновесие устойчиво, каково бы ни было $\rho < 1$.

7. Найти распределение давления внутри земного шара, считая его состоящим из однородной несжимаемой жидкости и пренебрегая осевым вра-

щением Земли. Вычислить в том же приближении давление в центре Земли $P_{ц}$ (см. задачу *5* к § *55*).

Ответ. $P=\frac{\rho g}{2R}(R^2-r^2)$, $P_{ц}=\frac{1}{2}\rho g R$, r — расстояние от центра Земли, R — радиус Земли. Если бы земной шар состоял из несжимаемой воды, то $P_{ц}$ равнялось бы $R/20$ ($P_{ц}$ — в атмосферах, R — в метрах). С учетом плотности Земли ($\rho=5{,}5$)

$$P_{ц}=0{,}275R\approx 1{,}75\cdot 10^6 \text{ атм}.$$

8. Оценить сплюснутость Земли, обусловленную ее осевым вращением, считая Землю однородным несжимаемым жидким шаром.

Решение. Так как фигура Земли мало отличается от шаровой, то ускорение свободного падения внутри земного шара можно считать направленным к центру Земли и пропорциональным расстоянию до ее центра (см. задачу *5* к § *55*). В этом приближении с учетом центробежной силы уравнения гидростатики (90.6) принимают вид

$$\frac{\partial P}{\partial x}=-\rho g\frac{x}{R_0}+\rho\omega^2 x,$$

$$\frac{\partial P}{\partial y}=-\rho g\frac{y}{R_0}+\rho\omega^2 y,$$

$$\frac{\partial P}{\partial z}=-\rho g\frac{z}{R_0},$$

где R_0 — радиус Земли, ω — угловая скорость ее вращения. Начало координат мы поместили в центре Земли, а ось Z направили вдоль оси ее вращения. Интегрируя эти уравнения, получаем

$$P=\frac{\rho}{2}\left(\omega^2-\frac{g}{R_0}\right)(x^2+y^2)-\frac{\rho g}{2R_0}z^2+C,$$

где C — постоянная интегрирования, определяющаяся значением давления P на земной поверхности (его можно считать равным нулю, так как атмосферное давление пренебрежимо мало). Сплюснутость Земли определится из требования постоянства давления на земной поверхности. Выбрав сначала точку на экваторе, а затем на полюсе, можно записать: $P(R_{э}, 0, 0)=P(0, 0, R_{п})$, где $R_{э}$ и $R_{п}$ — экваториальный и полярный радиусы Земли. С учетом явного вида P отсюда получаем

$$\left(\omega^2-\frac{g}{R_0}\right)R_{э}^2=-\frac{g}{R_0}R_{п}^2$$

и далее

$$R_{э}-R_{п}=\frac{\omega^2}{g}\frac{R_{э}^2R_0}{R_{э}+R_{п}}\approx\frac{\omega^2R_0^2}{2g}.$$

Следовательно, для сплюснотости ε земного шара имеем

$$\varepsilon\equiv\frac{R_{э}-R_{п}}{R_0}=\frac{\omega^2R_0}{2g}\approx\frac{1}{580}.$$

Действительное сжатие Земли больше, а именно 1/297. Расхождение объясняется грубостью модели, положенной в основу рассуждений, а также не-

совершенством метода расчета. При строгой постановке задачи надо учитывать, что поле тяготения сплюснутого шара не является центральным*). Тем самым задача сильно усложняется, так как гравитационное поле уже неизвестно заранее, а само оно зависит от неизвестной формы поверхности Земли. Подробное исследование показывает, что задача, сформулированная таким образом, не имеет однозначного решения. Возможно несколько различных форм равновесной поверхности, в том числе и эллипсоид вращения с определенной степенью сжатия.

§ 92. БАРОМЕТРИЧЕСКАЯ ФОРМУЛА

1. Обратимся теперь к *гидростатике сжимаемой жидкости.* Наибольшой интерес представляет равновесие земной атмосферы. Этот случай мы и рассмотрим. Дифференциальные уравнения (90.5) и (91.1) были выведены без использования предположения о несжимаемости жидкости, а потому мы воспользуемся ими здесь. Первые два уравнения системы (91.1) можно не учитывать, так как из них следует лишь, что давление P может зависеть только от z. Оставшееся третье уравнение можно переписать в виде

$$\frac{dP}{dz} = -\rho g, \tag{92.1}$$

так как частная $\frac{\partial P}{\partial z}$ и полная $\frac{dP}{dz}$ производные теперь означают одно и то же. Но одного уравнения (92.1) недостаточно, поскольку в него входят две неизвестные функции — давление P и плотность ρ. Нужно дополнительное соотношение между ними.

Будем предполагать, что состав атмосферы один и тот же на всем ее протяжении. Давление P, плотность ρ и температура T газа в состоянии равновесия связаны уравнением состояния. Если газ не слишком плотный, то таковым является *уравнение Клапейрона* (по имени французского физика Бенуа Клапейрона (1799—1864))

$$P = \frac{RT}{\mu}\rho, \tag{92.2}$$

где μ — молекулярная масса газа, а R — универсальная газовая постоянная. Ее числовое значение равно приближенно

$$R = 8{,}31 \cdot 10^7 \text{ эрг} \cdot \text{К}^{-1} \cdot \text{моль}^{-1} = 8{,}31 \text{ Дж} \cdot \text{К}^{-1} \cdot \text{моль}^{-1}.$$

Соотношение (92.2) позволяет исключить из уравнения (92.1) плотность ρ. В результате получим

$$\frac{dP}{dz} = -\frac{\mu g}{RT} P. \tag{92.3}$$

*) С учетом этого обстоятельства расчет дает $\varepsilon = \frac{5}{4}\omega^2 R_0 g \approx 1/232$.

Понятно, что таким путем мы еще не достигли цели, так как вместо неизвестной плотности ρ ввели новую неизвестную величину — температуру T. Однако последнюю легче измерить на различных высотах. Если T известна как функция z, то уравнение (92.3) уже можно будет проинтегрировать. Следовательно, задача определения давления на различных высотах становится вполне определенной, если задать закон изменения температуры T с высотой.

2. Если отсутствуют ветры и воздушные течения, т. е. атмосфера неподвижна, то говорят, что она находится в *механическом равновесии*. Такое состояние не является еще состоянием *полного равновесия*. Для последнего, кроме того, необходимо, чтобы атмосфера находилась также и в *тепловом равновесии*. Тепловое равновесие означает, что температура T одна и та же на протяжении всей атмосферы. Если это имеет место, то атмосферу называют *изотермической*.

Конечно, изотермическая атмосфера — это идеализации. Но рассмотрение этого идеализированного случая тем не менее представляет большой интерес. При $T = \text{const}$ уравнение (92.3) легко интегрируется. Для этого переписываем его в виде

$$\frac{dP}{P} = -\frac{\mu g}{RT}\,dz$$

и после интегрирования находим

$$\ln\frac{P}{P_0} = -\frac{\mu g z}{RT},$$

или

$$P = P_0 \exp\left(-\frac{\mu g z}{RT}\right). \tag{92.4}$$

По тому же закону меняется и плотность воздуха, а именно

$$\rho = \rho_0 \exp\left(-\frac{\mu g z}{RT}\right). \tag{92.5}$$

Соотношения (92.4) и (92.5) называются *барометрическими формулами*. Постоянные интегрирования P_0 и ρ_0 имеют смысл давления и плотности воздуха на поверхности Земли. Давление и плотность воздуха убывают с высотой по экспоненциальному закону. При поднятии на высоту

$$h = \frac{RT}{\mu g} \tag{92.6}$$

они убывают в e раз. Величина h называется *высотой однородной атмосферы*. Смысл этого названия станет ясным, если поставить следующий вопрос. Какую высоту H должна была бы иметь воображаемая атмосфера постоянной плотности ρ_0, чтобы она производила на поверхность Земли такое же давление P_0, как и действительная атмосфера? Очевидно, искомая величина определится из условия

$P_0 = \rho_0 gH$. Но из уравнения состояния (92.2), если его применить к слою воздуха, прилегающему к поверхности Земли, следует $P_0 = \frac{RT}{\mu}\rho_0$. Используя это соотношение, получаем $H = \frac{RT}{\mu g}$, т. е. $H = h$. Считая среднюю молекулярную массу воздуха равной $\mu = 28{,}8$, находим для высоты однородной атмосферы при нуле градусов Цельсия ($T = 273$ К):

$$h = \frac{8{,}31 \cdot 273}{28{,}8 \cdot 9{,}8} \approx 8000 \text{ м} = 8 \text{ км}.$$

Подставляя h в барометрическую формулу (92.4), можно переписать ее в виде

$$P = P_0 e^{-z/h}. \tag{92.7}$$

В таком виде формула удобна для определения разностей высот двух или нескольких точек земной атмосферы. Для этого нужно знать давление воздуха в этих точках, а также температуру. Последняя в пределах рассматриваемых высот, разумеется, должна быть одной и той же.

3. Сделаем в заключение одно замечание относительно устойчивости механического равновесия атмосферы. Мы не будем вводить ограничения, что температура одна и та же на всех высотах, а будем предполагать, что она может меняться с высотой как угодно. Если нарушено состояние механического равновесия, в результате которого некоторая масса воздуха немного поднялась вверх, то в новом положении она будет подвергаться меньшему внешнему давлению. В результате поднявшаяся масса воздуха расширится, а ее плотность уменьшится, так как вследствие малой теплопроводности воздуха во время поднятия рассматриваемая масса практически не будет получать и отдавать тепло. Если окажется, что в новом положении плотность поднявшейся массы больше плотности окружающего воздуха, то эта масса, как более тяжелая, опустится вниз, и равновесие восстановится. Если же ее плотность окажется меньше плотности окружающего воздуха, то она будет подниматься еще выше, и механическое равновесие окажется неустойчивым. Аналогичные соображения справедливы и для случая, когда нарушение механического равновесия совершается путем небольшого опускания какой-либо массы воздуха. В этом случае опустившаяся масса сжимается внешним давлением. Если в новом положении ее плотность меньше плотности окружающего воздуха, то она начнет подниматься, и равновесие восстановится. Наоборот, если эта плотность окажется больше, то рассматриваемая масса начнет опускаться еще ниже, т. е. равновесие окажется неустойчивым. Эти рассуждения, разумеется, применимы не только к атмосфере, но и к любой неравномерно нагретой сжимаемой жидкости, находящейся в механическом равновесии в поле тяжести. Что касается земной атмосферы, то исследования показали, что изотермическая атмосфера в рас-

сматриваемом смысле устойчива. Еще большая устойчивость получается, когда температура воздуха возрастает с высотой. Если же температура убывает с высотой, то механическое равновесие воздуха возможно лишь тогда, когда это убывание происходит не слишком быстро. При убывании температуры с высотой более чем на один градус на каждые 100 метров высоты атмосфера теряет механическую устойчивость. Появляются восходящие и нисходящие потоки воздуха (конвекция). Во втором томе эти вопросы будут рассмотрены более подробно.

ЗАДАЧА

На какую высоту $H_{1/2}$ надо подняться, чтобы давление (изотермической) атмосферы уменьшилось в 2 раза?

Ответ. $H_{1/2} = h \ln 2 \approx 5{,}55$ км (при 0°C).

§ 93. КИНЕМАТИЧЕСКОЕ ОПИСАНИЕ ДВИЖЕНИЯ ЖИДКОСТИ

1. Для описания движения жидкости можно поступить двояко. Можно проследить за движением *каждой индивидуальной частицы жидкости*, т. е. указать положение и скорость этой частицы в каждый момент времени. Тем самым будут определены и *траектории* всех частиц жидкости. Но можно поступить и иначе. Можно проследить, что происходит с течением времени в *каждой точке пространства.* Точнее, можно указать модули и направления скоростей различных частиц жидкости, которые в различные моменты времени проходят через одну и ту же точку пространства. Если взять всевозможные точки пространства, но фиксировать время t, то при втором способе описания в пространстве получится мгновенная картина распределения скоростей жидкости — *поле скоростей.* В каждой точке пространства будет указан вектор скорости той частицы жидкости, которая проходит через эту точку в рассматриваемый момент времени. Линия, касательная к которой указывает направление скорости частицы жидкости, проходящей в рассматриваемый момент времени через эту точку касания, называется *линией тока.* Если поле скоростей, а следовательно, и соответствующие ему линии тока не меняются с течением времени, то движение жидкости называется *стационарным* или *установившимся.* Если же они меняются во времени, то движение называется *нестационарным* или *неустановившимся.* В случае нестационарного движения при втором способе описания скорость жидкости явно зависит от координат и времени: $\mathbf{v} = \mathbf{v}(\mathbf{r}, t)$. При стационарном движении явной зависимости от времени нет, скорость зависит только от координат: $\mathbf{v} = \mathbf{v}(\mathbf{r})$.

2. В случае нестационарного движения линии тока, вообще говоря, не совпадают с траекториями частиц жидкости. Действительно,

траектория указывает путь *одной и той же частицы* жидкости за все время ее движения. Линия же тока характеризует направления движения *бесконечного множества частиц*, которые в рассматриваемый момент находятся на этой линии. Только *при стационарном течении линии тока совпадают с траекториями частиц.* Для доказательства возьмем траекторию какой-либо произвольной частицы A (рис. 237). Пусть $A(t_1)$ — положение этой частицы в момент времени t_1. Возьмем другую частицу B, которая в момент t_2 занимает то же положение, что и частица A в момент t_1. Так как движение стационарно, то через точку $A(t_1)$ частица A пройдет с той же скоростью, с какой пройдет через нее частица B в момент t_2. Значит, скорость частицы B в положении $A(t_1)$ направлена по касательной к траектории частицы A. Так как момент времени t_2 можно выбрать произвольно, то отсюда следует, что траектория частицы A является также линией тока.

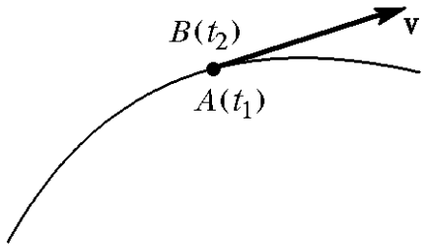


Рис. 237

3. Возьмем произвольный замкнутый контур C и через каждую точку его в один и тот же момент времени проведем линии тока (рис. 238). Они расположатся на некоторой трубчатой поверхности, называемой *трубкой тока.* Так как скорости частиц жидкости направлены по касательным к линиям тока, то при течении жидкость не может пересекать боковую поверхность трубки тока. Трубка тока ведет себя подобно боковой поверхности жесткой трубки, вдоль которой течет жидкость. На такие трубки тока можно разбить все пространство, занимаемое жидкостью. Если поперечное сечение трубки тока бесконечно мало, то можно считать, что скорость жидкости одна и та же во всех точках одного и того же поперечного сечения и направлена вдоль оси трубки тока. Масса жидкости, протекающая за время dt через поперечное сечение трубки, определяется выражением

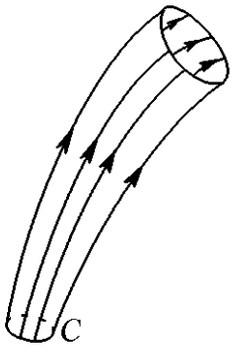


Рис. 238

$$dm = \rho v S\, dt, \tag{93.1}$$

где ρ — плотность жидкости, а S — площадь (нормального) поперечного сечения трубки.

В случае стационарного течения масса dm будет одной и той же для всех сечений трубки тока. Если взять два сечения, площади которых равны S_1 и S_2, то можно написать

$$\rho_1 v_1 S_1 = \rho_2 v_2 S_2. \tag{93.2}$$

Если бы это равенство не соблюдалось, то масса жидкости между сечениями S_1 и S_2 изменялось бы во времени. А это противоречит за-

кону сохранения массы и предположению о стационарности течения. Если жидкость несжимаема, то $\rho_1 = \rho_2$, и соотношение (93.2) принимает вид

$$\frac{v_1}{v_2} = \frac{S_2}{S_1}. \tag{93.3}$$

Скорость жидкости в одной и той же трубке тока тем больше, чем у́же поперечное сечение трубки. Она обратно пропорциональна площади этого поперечного сечения.

§ 94. СТАЦИОНАРНОЕ ДВИЖЕНИЕ ИДЕАЛЬНОЙ ЖИДКОСТИ. УРАВНЕНИЕ БЕРНУЛЛИ

1. Изучение движения реальных жидкостей и газов, вообще говоря, представляет очень сложную задачу. Для ее упрощения сначала полностью пренебрегают силами вязкости. Рассматривают случай идеальной жидкости, в которой при любых движениях не возникают касательные и нормальные силы внутреннего трения (см. § 89, п. 6). Единственные поверхностные силы, которые могут действовать в идеальной жидкости, — это силы нормального давления P. При этом само давление P однозначно определяется плотностью и температурой жидкости. Для упрощения жидкость считается также несжимаемой.

2. Рассмотрим стационарное течение идеальной жидкости в каком-либо консервативном силовом поле, например в поле силы тяжести. Применим к этому течению закон сохранения энергии. При этом будем полностью пренебрегать теплообменом, который может происходить между частями жидкости с окружающей средой. Выделим в жидкости бесконечно узкую трубку тока и рассмотрим часть жидкости, занимающую объем $MNDC$ (рис. 239). Пусть эта часть переместилась в бесконечно близкое положение $M_1N_1D_1C_1$. Вычислим работу A, совершаемую при этом силами давления. Давление, действующее на боковую поверхность трубки тока, перпендикулярно к перемещению и работы не совершает. При перемещении границы MN в положение M_1N_1 совершается работа $A_1 = P_1S_1l_1$, где $l_1 = MM_1$ — величина перемещения. Введя объем $\Delta_1V = S_1l_1$, ее можно представить в виде $A_1 = P_1\Delta V_1$ или $A_1 = P_1\frac{\Delta_1 m}{\rho_1}$, где $\Delta_1 m$ — масса жидкости в объеме MNN_1M_1. При перемещении границы CD в положение C_1D_1 жид-

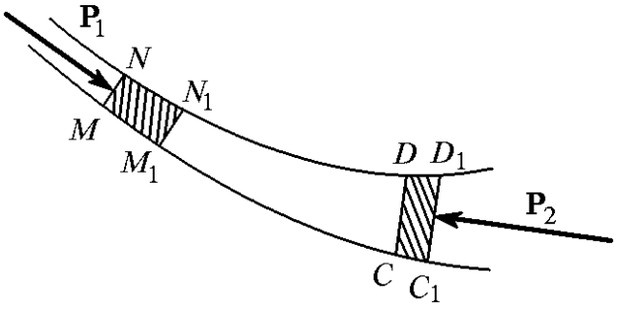


Рис. 239

кость совершает работу против давления P_2 (или давление P_2 совершает над жидкостью отрицательную работу). Для нее, рассуждая аналогично, найдем $A_2 = P_2 \frac{\Delta_2 m}{\rho_2}$, где $\Delta_2 m$ — масса жидкости в объеме CDD_1C_1. Но если движение стационарно, то масса жидкости в объеме M_1N_1DC не изменится, а потому из закона сохранения массы получим $\Delta_1 m = \Delta_2 m$. Опуская индексы у Δm, для работы, совершаемой внешним давлением, окончательно находим

$$A = A_1 - A_2 = \left(\frac{P_1}{\rho_1} - \frac{P_2}{\rho_2}\right) \Delta m.$$

Эта работа должна быть равна приращению ΔE полной энергии выделенной части жидкости. Ввиду стационарности течения энергия жидкости в объеме M_1N_1DC не изменилась. Поэтому величина ΔE равна разности энергий массы жидкости Δm в положениях CDD_1C_1 и MNN_1M_1. Обозначая через ε полную энергию, приходящуюся на единицу массы жидкости, находим $\Delta E = (\varepsilon_2 - \varepsilon_1)\Delta m$. Приравнивая эту величину работе A и сокращая на Δm, получаем

$$\varepsilon_1 + \frac{P_1}{\rho_1} = \varepsilon_2 + \frac{P_2}{\rho_2}. \tag{94.1}$$

Отсюда следует, что *вдоль одной и той же линии тока при стационарном течении идеальной жидкости величина* $\varepsilon + P/\rho$ *остается постоянной*:

$$\varepsilon + \frac{P}{\rho} = B = \text{const}. \tag{94.2}$$

Это соотношение называется *уравнением Бернулли* (по имени математика и физика Даниила Бернулли (1700--1782), который впервые опубликовал его в 1738 г.). При выводе уравнения Бернулли мы нигде не использовали предположения о несжимаемости жидкости. Поэтому оно справедливо и для сжимаемых жидкостей. Требуется только, чтобы жидкость была идеальной, а течение — стационарным. Однако разбор и применения уравнения Бернулли для сжимаемых жидкостей и газов мы отложим до второго тома, так как это требует знания явного выражения для энергии ε. Здесь ограничимся рассмотрением несжимаемых жидкостей, движущихся в поле тяжести Земли. Именно в этих предположениях уравнение (94.2) было установлено самим Бернулли.

Если жидкость несжимаемая, то при течении не меняется та часть полной энергии ε, которая зависит от сжатия жидкости. Эту часть поэтому можно не принимать во внимание. Вся энергия ε складывается из кинетической энергии единицы массы жидкости $v^2/2$ и ее потенциальной энергии gh в поле тяжести. В этом случае уравнение Бернулли принимает вид

$$\frac{v^2}{2} + gh + \frac{P}{\rho} = B = \text{const}. \tag{94.3}$$

3. Подчеркнем еще раз, что постоянная Бернулли B одна и та же *вдоль одной и той же линии тока.* Однако она, вообще говоря, может меняться при переходе от одной линии тока к другой. Но могут быть и такие случаи, когда постоянная Бернулли одна и та же для всего потока жидкости. Отметим один, довольно часто встречающийся случай, когда это имеет место. Допустим, что все линии тока начинаются или оканчиваются в такой области, где жидкость практически находится в состоянии покоя. Возьмем одну из точек линии тока в такой области. Тогда в уравнении (94.3) следует положить $v = 0$, и мы получим $B = gh + P/\rho$. Но во всей области, где жидкость покоится, соблюдается условие равновесия $gh + P/\rho = \text{const}$. Отсюда видно, что постоянная Бернулли B в рассматриваемом случае одна и та же для всего потока жидкости. Более общим является случай, когда в некоторой области пространства несжимаемая идеальная жидкость движется параллельным потоком в любом направлении с постоянной скоростью v_0, а затем параллельность потока нарушается препятствиями, стоящими на его пути, или вследствие расширений или сужений трубы или русла, по которым течет жидкость. В этом случае постоянная Бернулли B также одинакова для всех линий тока. Чтобы убедиться в этом, достаточно перейти к системе отсчета, равномерно движущейся относительно исходной со скоростью v_0.

4. Допустим теперь, что тонкая трубка тока имеет переменное поперечное сечение, а ось ее горизонтальна. (Примером может служить горизонтальная труба переменного сечения, по которой течет жидкость.) Тогда $h = \text{const}$, и уравнение Бернулли принимает вид

$$\frac{v^2}{2} + \frac{P}{\rho} = \text{const}. \qquad (94.4)$$

Отсюда видно, что давление больше там, где меньше скорость v, и наоборот. С другой стороны, согласно соотношению (93.3) скорость v минимальна там, где максимально сечение трубки. Значит, в широких частях трубки давление максимально, а в узких минимально. Такой результат является непосредственным следствием второго закона Ньютона. Действительно, когда жидкость из широкой части течет в узкую (рис. 240), то скорость ее возрастает. Значит, ускорение направлено в сторону течения, т. е. на рис. 240 слева направо. Это ускорение сообщается разностью давлений, действующих на рассматриваемую часть жидкости слева и справа. Следовательно, давление слева, т. е. в более широкой части трубки, должно быть больше, чем справа, где трубка у́же.

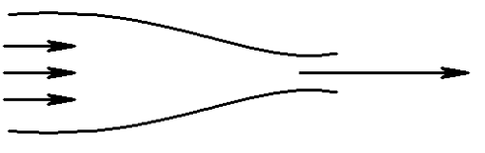


Рис. 240

5. Пользуясь уравнением (94.4), можно дать ответ на вопрос, когда при течении жидкость или газ можно считать несжимаемыми, хотя более строгое доказательство должно основываться на уравнении Бернулли в более общей форме (94.2). Давление и скорость те-

чения в двух точках 1 и 2 на одной и той же линии тока связаны соотношением

$$\Delta P \equiv P_2 - P_1 = \frac{\rho}{2}(v_1^2 - v_2^2).$$

С другой стороны,

$$\Delta\rho = \frac{d\rho}{dP}\Delta P = \frac{1}{c^2}\Delta P,$$

где c — скорость звука (см. § 85, п. 1). Для того чтобы при течении жидкость можно было рассматривать как несжимаемую, необходимо выполнение соотношения $|\Delta\rho| \ll \rho$ при любом выборе точек 1 и 2. Это приводит к условию

$$|v_2^2 - v_1^2| \ll c^2, \tag{94.5}$$

т. е. изменение квадрата скорости течения жидкости должно быть мало по сравнению с квадратом скорости звука. Если течение отнести к системе отсчета, в которой жидкость в какой-либо точке покоится, то условие (94.5) упрощается и принимает вид

$$v^2 \ll c^2, \tag{94.6}$$

т. е. во всем потоке квадрат скорости течения должен быть малым по сравнению с квадратом скорости звука.

Если при течении меняется высота h, то с помощью уравнения (94.3) нетрудно показать, что помимо (94.5) требуется дополнительное условие

$$g\Delta h \ll c^2, \tag{94.7}$$

выполнение которого необходимо, чтобы жидкость или газ могли рассматриваться как несжимаемые.

6. Опишем несколько опытов для иллюстрации уравнения Бернулли. На рис. 241 изображена труба переменного сечения, через которую пропускается воздух. О давлении воздуха в трубе можно судить по уровням воды в

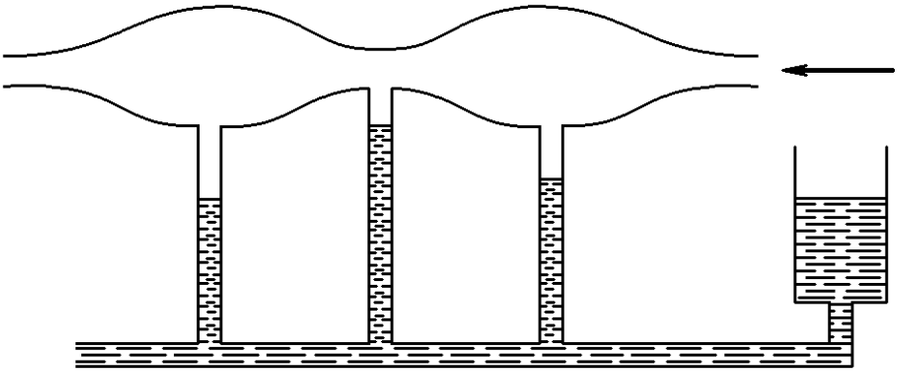


Рис. 241

стеклянных манометрических трубках, соединенных с ней, как показано на рис. 241. Оказывается, что в трубках, соединенных с узкими частями тру-

бы, вода поднимается выше, а соединенных с широкими частями — ниже. Значит, в первом случае давление воздуха в потоке меньше, чем во втором. Так и должно быть согласно уравнению (94.4).

Эта демонстрация может служить для пояснения идеи *водомера*, служащего для измерения *расхода воды*, т. е. массы воды Q, протекающей ежесекундно через поперечное сечение трубы. Труба содержит короткий участок (*трубка Вентури*) с меньшим поперечным сечением. Пусть S_1 и S_2 — площади поперечных сечений широкого и узкого участков трубы, а P_1 и P_2 — давления воды в них, измеряемые с помощью манометров. Тогда по уравнению Бернулли

$$\frac{v_1^2}{2}+\frac{P_1}{\rho}=\frac{v_2^2}{2}+\frac{P_2}{\rho}.$$

Кроме того, $M_1=\rho v_1 S_1=\rho v_2 S_2$. Определив отсюда v_1 и v_2 и подставив их в предыдущее соотношение, получим

$$Q=S_1S_2\sqrt{\frac{2\rho(P_1-P_2)}{S_1^2-S_2^2}}. \qquad (94.8)$$

7. Возьмем резиновую трубку, надетую на суживающийся стеклянный наконечник, и будем продувать через нее воздух (рис. 242, вид сверху). Давление воздуха в узкой части наконечника и в выходящей из него струе будет меньше атмосферного. Поднесем струю сбоку к легкому полому целлулоидному шарику, подвешенному на нити. Шарик втягивается в струю, а затем увлекается ею. Если струю направить вертикально вверх, то втянувшийся в нее шарик можно удерживать в равновесии на определенной высоте. Он ведет себя подобно шарику, помещенному в яму. Привязывать шарик к нити в этом опыте не требуется.

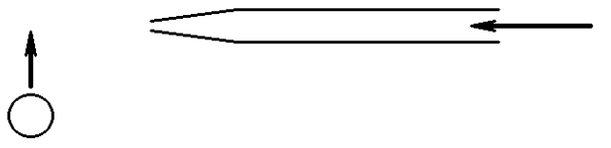
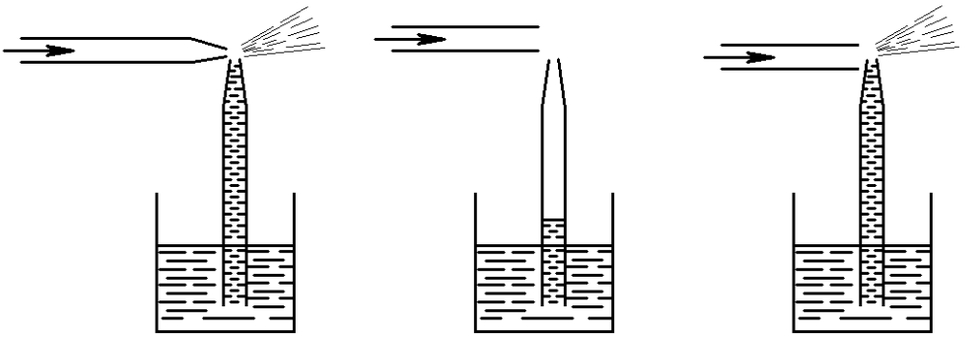


Рис. 242

8. Поднесем теперь струю воздуха к верхнему концу стеклянной трубки, нижний конец которой погружен в воду, а верхний оканчивается узким наконечником (рис. 243 *а*). Вода в стеклянной трубке будет подниматься, разбрызгиваться и увлекаться струей воздуха. На этом принципе основано устройство *пульверизатора*. Если трубка, по которой продувается воздух, не снабжена узким наконечником, а имеет постоянное поперечное сечение



а *б* *в*

Рис. 243

(рис. 243 *б*), то поднятие воды и разбрызгивание не происходит. Если, однако, такую трубку поднести вплотную к наконечнику трубки, погруженной в воду, так чтобы между ними образовался узкий зазор (рис. 243 *в*), то вода опять поднимается и разбрызгивается. Зазор между трубками выполняет роль узкого наконечника, понижающего давление воздуха в струе.

9. Если два легких изогнутых листа твердой бумаги подвесить на горизонтальных проволоках (рис. 244) и продувать между ними воздух, то они притягиваются друг к другу. Дело в том, что давление воздуха P между листами в наиболее узком месте становится меньше атмосферного P_0, и наружное атмосферное давление прижимает листы друг к другу. Можно так же повесить на небольшом расстоянии друг от друга две стеклянные колбы. При продувании воздуха между ними колбы начинают стучать, сталкиваясь друг с другом. Притяжение такого же типа наблюдается между двумя кораблями, когда они идут параллельным курсом на небольшом расстоянии друг от друга. Это легко объяснить, если перейти в систему отсчета, в которой корабли покоятся, а вода между ними течет. Описанное явление не раз было причиной столкновения судов и приводило к авариям.

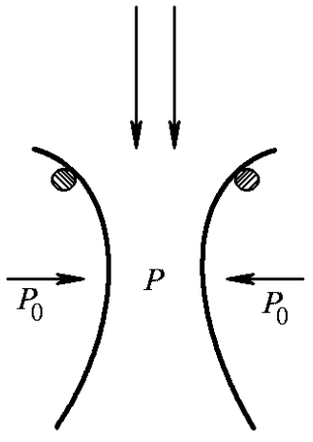


Рис. 244

10. На рис. 245 схематически изображен *прибор Клемана—Дезорма* (по имени французских физиков Н. Клемана-Дезорма (ум. в 1841) и Ш. Дезорма (1777—1862)). Он состоит из латунного диска с отверстием в центре, к краям которого приделана латунная трубка. На эту трубку надета резиновая трубка, через которую продувается воздух. Если диск поднести к листу бумаги, лежащему на столе, то лист притянется диском. Дело в том, что в узком зазоре между диском и листом бумаги образуется расходящийся от центра к краям поток воздуха. Давление в этом зазоре понижается, и лист бумаги прижимается к диску давлением наружного воздуха. Прижатый лист закрывает отверстие AB, течение воздуха через трубку затормаживается, давление его повышается, и снова появляется зазор, через который устремляется поток воздуха. Лист бумаги опять притягивается к диску и все повторяется, пока не прекратится дутье. В результате бумажный лист приходит в быстрые колебания, издавая звук.

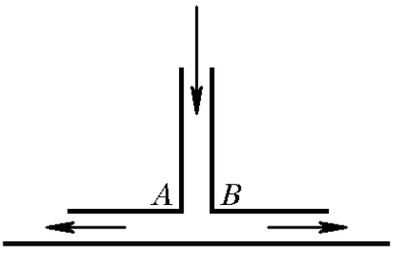


Рис. 245

11. Допустим, что поток жидкости обтекает какое-либо тело (рис. 246). От точки A линии тока расходятся в стороны. В точке A, называемой *критической*, скорость жидкости обращается в нуль, в ней линия тока обрывается. Применяя уравнение Бернулли к линии тока BA, получим

$$P + \frac{1}{2}\rho v^2 = P_0, \qquad (94.9)$$

где P_0 — давление в критической точке, а P — на «бесконечности», откуда жидкость течет. Величина P_0 — это максимальное давление, которое может иметь жидкость на рассматриваемой линии тока. От наличия силы тяжести мы отвлекаемся, пред-

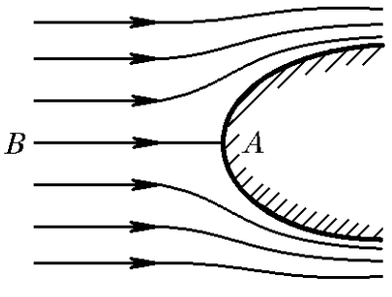


Рис. 246

полагая, что все линии тока плоские и лежат в горизонтальных плоскостях. Величина $\frac{1}{2}\rho v^2$ называется *динамическим* или *скоростным напором*, а сумма $P + \frac{1}{2}\rho v^2$ — *полным напором* жидкости на рассматриваемой линии тока*). Если измерить в отдельности полный и скоростной напоры жидкости в рассматриваемой точке пространства, то по ним легко вычислить и скорость жидкости в той же точке.

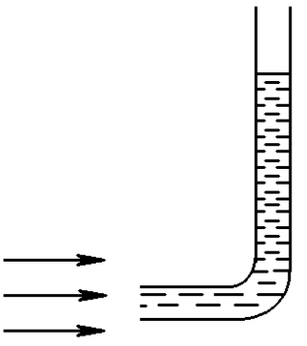


Рис. 247

Для измерения полного напора используется *трубка Пито* (по имени французского математика, физика и гидротехника Анри Пито (1695—1771)). Это небольшая изогнутая манометрическая трубка, обращенная открытым концом навстречу потоку жидкости (рис. 247). Приосевые линии тока, направленные к трубке Пито, заканчиваются внутри трубки, где жидкость покоится. Высота столба жидкости, устанавливающаяся в трубке, является поэтому мерой максимального давления, а следовательно, и полного напора жидкости на рассматриваемой линии тока.

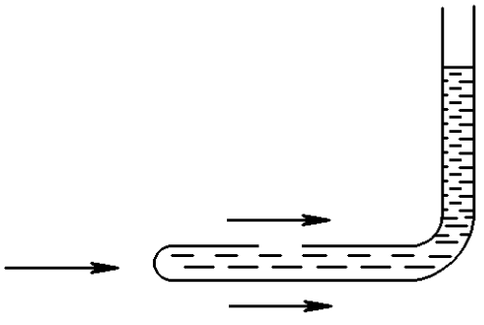


Рис. 248

Если помимо полного напора измерить еще давление P, то по их разности можно найти скоростной напор $\frac{1}{2}\rho v^2$, а затем вычислить скорость v. Измерение P было бы излишним, если бы речь шла о нахождении скорости v, например, в реке, где жидкость имеет открытую поверхность. В этом случае глубина погружения трубки Пито непосредственно давала бы значение искомого давления. Но этот способ не годится, когда жидкость течет, например, в трубке. Он не годится также для измерения скоростей самолетов и т. д. В таких случаях для измерения давления P можно воспользоваться *зондом*. Зонд отличается от трубки Пито тем, что его передняя часть, обращенная навстречу потоку, запаяна, а в боковой стенке имеется небольшое отверстие, как показано на рис. 248. Трубка зонда сильно искажает поток только в непосредственной близости от ее переднего конца, обращенного к потоку. Поток, обтекающий боковую поверхность трубки, практически остается неискаженным. Поэтому в непосредственной близости от отверстия скорость, а с ней и давление жидкости такие же, как и во всех точках линии тока, проходящей вблизи отверстия. Давление в трубке зонда, измеряемое манометром, таким образом, совпадает с давлением обтекающей ее жидкости P. На практике трубку Пито обычно монтируют

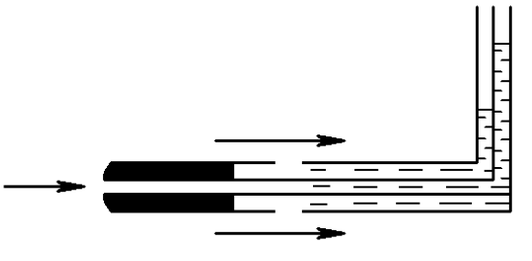


Рис. 249

*) В технической гидродинамике обычно применяется следующая терминология. Величину P называют *статическим давлением*, $\frac{1}{2}\rho v^2$ — *динамическим давлением*, а сумму $P + \frac{1}{2}\rho v^2$ — *полным давлением*. Однако эта терминология, как неоднократно отмечалось многими физиками, нерациональна и может только ввести в заблуждение. Ею мы пользоваться не будем. В жидкости есть лишь единственное давление, которое обусловлено степенью ее сжатия, и таковым является величина P.

вместе с зондом, например, так, как изображено в разрезе на рис. 249. Такая трубка называется *трубкой Прандтля* (по имени немецкого ученого Людвига Прандтля (1875—1953)). Принцип ее действия ясен из рисунка.

§ 95. ПРИМЕРЫ НА ПРИМЕНЕНИЕ УРАВНЕНИЯ БЕРНУЛЛИ. ФОРМУЛА ТОРРИЧЕЛЛИ

1. Рассмотрим истечение идеальной несжимаемой жидкости через малое отверстие в боковой стенке или дне широкого сосуда. Частицы жидкости подходят к отверстию, имея скорости в поперечных направлениях (рис. 250). Из-за инерции это приводит к *сжатию вытекающей струи.* Во избежание этого будем предполагать, что истечение происходит через трубку с закругленными краями (рис. 251). Благодаря этому линии тока перед истечением постепенно меняют направление на параллельные оси трубки, и сжатия

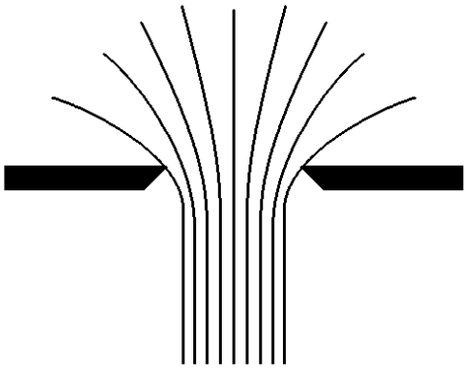


Рис. 250

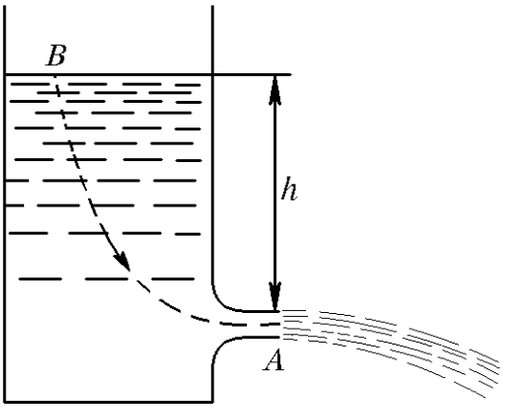


Рис. 251

струи не возникает *). Все линии тока проходят через трубку, начинаясь вблизи свободной поверхности жидкости, где скорость v пренебрежимо мала. Поэтому постоянная Бернулли будет одна и та же у всех линий тока. Применим уравнение Бернулли к точкам B и A какой-либо линии тока (рис. 251). В точке B скорость пренебрежимо мала, ее можно считать равной нулю, скорость в точке A обозначим через v. Уравнение Бернулли дает

$$\frac{P_0}{\rho} + gh = \frac{P_0}{\rho} + \frac{v^2}{2},$$

где P_0 — атмосферное давление, а высота h — отсчитывается от уровня отверстия. Отсюда получаем

$$v = \sqrt{2gh}. \tag{95.1}$$

Это — *формула Торричелли* (по имени итальянского физика и математика Эванджелисты Торричелли (1608—1647)). Она показыва-

*) Это не совсем так, так как остается некоторое сжатие, обусловленное силами поверхностного натяжения.

ет, что при истечении жидкость приобретает такую скорость, какую получило бы тело, свободно падающее с высоты h. Поэтому если изогнуть трубку и направить струю вертикально вверх или под малым углом к вертикали, то в наивысшей своей точке она достигнет уровня жидкости в сосуде. В действительности высота поднятия струи будет несколько меньше из-за трения и сопротивления воздуха, которые при выводе уравнения Бернулли не учитывались.

2. Подсчитаем импульс, уносимый ежесекундно вытекающей струей. Пусть струя вытекает горизонтально через небольшое отверстие в боковой стенке. Если S — площадь отверстия, то ежесекундно вытекает масса жидкости $\rho v S$. Она уносит импульс $mv = \rho v^2 S$, или в силу (95.1) $mv = 2\rho g h S$. Благодаря этому сосуд с жидкостью получает отдачу $F = 2\rho g h S$. Если отверстие закрыть пробкой, то сосуд будет оставаться на месте. Значит, горизонтальные силы давления жидкости, действующие на боковые стенки сосуда, уравновешиваются. Снова откроем отверстие. Тогда из правой боковой стенки будет удален участок площадью S. Если бы состояние жидкости при этом не изменилось, то сила давления жидкости на правую стенку уменьшилось бы на $PS = \rho g h S$. На самом деле ее уменьшение вдвое больше и составляет $2\rho g h S$. Это объясняется *перераспределением давления*, которое происходит при переходе от состояния покоя жидкости к состоянию установившегося движения. Конечно, этот переход совершается не мгновенно. Если мгновенно удалить пробку, то в первый момент сила давления на правую стенку уменьшится только на $\rho g h S$. Затем в процессе установления течения уменьшение давления будет быстро, но непрерывно меняться от $\rho g h S$ до $2\rho g h S$.

ЗАДАЧИ

1. В вертикально стоящий цилиндрический сосуд налита идеальная жидкость до уровня H (относительно дна сосуда). Площадь дна сосуда равна S. Определить время t, за которое уровень жидкости в сосуде опустится до высоты h (относительно дна сосуда), если в дне сосуда сделано малое отверстие площадью σ. Определить также время T, за которое из сосуда выльется вся жидкость.

Ответ. $t = \frac{S}{\sigma}\sqrt{\frac{2}{g}}\ (\sqrt{H} - \sqrt{h})$, $T = \frac{S}{\sigma}\sqrt{\frac{2H}{g}}$.

2. Прямоугольная коробка плавает на поверхности воды, погружаясь под действием собственного веса на глубину h. Площадь дна коробки равна S, высота — H. Через какое время коробка утонет, если в центре дна ее проделать малое отверстие площади σ и с помощью боковых направляющих сохранять неизменной ориентацию коробки?

Ответ. $t = \frac{S}{\sigma}\frac{H-h}{\sqrt{2gh}}$.

3. Через какое время наполнится водой шаровая колба радиусом R, если в центре ее нижнего основания сделано малое отверстие площадью σ? Колба погружена в воду до нижнего основания ее горлышка.

Ответ. $t = \frac{16\pi R^2}{15\sigma}\sqrt{\frac{R}{g}}$.

4. На горизонтальной поверхности стола стоит цилиндрический сосуд, в который налита вода до уровня H (относительно поверхности стола). На какой высоте h (относительно поверхности стола) надо сделать малое отверстие в боковой стенке сосуда, чтобы струя воды встречала поверхность стола на максимальном расстоянии от сосуда? Вычислить это расстояние ($x_{\text{макс}}$).

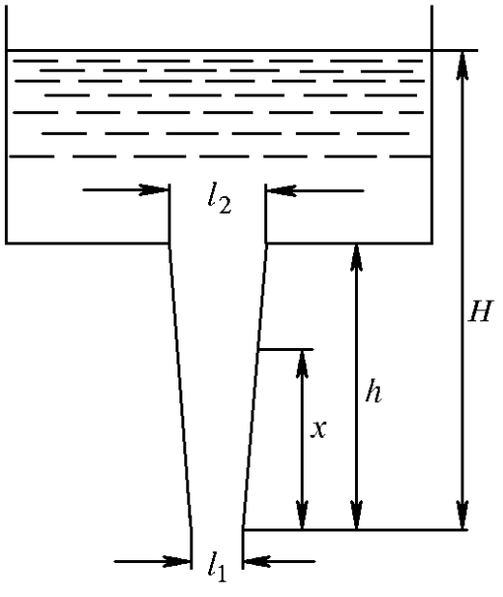


Рис. 252

Ответ. $h = H/2$, $x_{\text{макс}} = H$.

5. Определить форму сосуда, чтобы уровень жидкости в нем опускался с постоянной скоростью, если в центральной точке дна проделать малое отверстие.

Ответ. Площадь горизонтального поперечного сечения сосуда должна быть пропорциональна корню квадратному из расстояния этого сечения от отверстия. Если сосуд обладает осевой симметрией, то он должен иметь форму параболоида вращения четвертого порядка.

6. В широкий сосуд с плоским дном налита идеальная жидкость. В дне сосуда сделана длинная и узкая щель, в которую вставлена насадка, образованная двумя плоскостями, наклоненными друг к другу под малым углом (рис. 252). Расстояние между ними в нижней части насадки равно l_1, а в верхней — l_2. Определить распределение давления жидкости в насадке, если атмосферное давление равно P_0. Длина насадки равна h, расстояние между нижним концом насадки и уровнем жидкости в сосуде равно H.

Ответ. $P = P_0 - \rho g x + \rho g H\left\{1 - \frac{h^2 l_1^2}{[h l_1 + x(l_2 - l_1)]^2}\right\}$, где x — расстояние по вертикали от нижнего конца насадки.

7. Вода вытекает из широкого резервуара через вертикальную коническую трубу, вставленную в его дно. Длина трубы равна l, диаметр верхнего основания d_1, нижнего основания d_2 ($d_1 > d_2$). При каком уровне H воды в резервуаре давление в верхнем сечении трубы будет равно P, если атмосферное давление равно P_0?

Ответ. $H = \frac{(P_0 - P)/\rho g - l(d_2/d_1)^4}{1 - (d_2/d_1)^4}$.

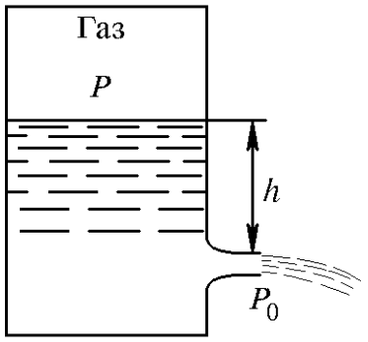


Рис. 253

8. Определить скорость стационарного истечения через малое отверстие струи идеальной несжимаемой жидкости, находящейся под давлением в закрытом сосуде (рис. 253).

Ответ. $v = \sqrt{2(P - P_0)/\rho + 2gh}$ — атмосферное давление.

9. Для того чтобы струя жидкости вытекала из сосуда с постоянной скоростью, применяют устройство, изображенное на рис. 254. Определить скорость истечения струи v в этом случае.

Ответ. Пока уровень жидкости в сосуде выше нижнего конца трубки AB, скорость истечения постоянна и равна $v = \sqrt{2gh}$. После этого скорость истечения начнет уменьшаться.

10. Цилиндрический сосуд с налитой в него идеальной несжимаемой жидкостью вращается вокруг своей геометрической оси, направленной вертикально, с угловой скоростью ω. Определить скорость истечения струи жидкости через малое отверстие в боковой стенке сосуда при установившемся движении жидкости (относительно сосуда).

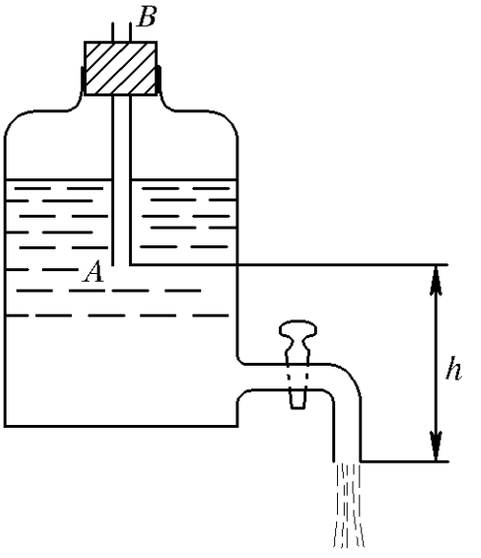


Рис. 254

Решение. Перейдем в систему отсчета, в которой жидкость покоится. В ней добавятся две силы инерции; центробежная и кориолисова. Кориолисова сила не совершает работы. Она лишь искривляет линии тока, но не сказывается на справедливости и форме общего уравнения Бернулли (94.2). Центробежная сила добавляет новый член к потенциальной энергии. Полная потенциальная энергия единицы массы жидкости будет $u = gz - 1/2\ \omega^2 r^2$, так что уравнение (94.2) запишется в виде

$$\frac{1}{2}v^2 + gz - \frac{1}{2}\omega^2 r^2 + \frac{P}{\rho} = B = \text{const}, \qquad (95.2)$$

где v — относительная скорость жидкости (т. е. скорость относительно вращающейся системы отсчета). Постоянная Бернулли B одна и та же для всех линий тока, поскольку все они начинаются вблизи поверхности жидкости, где скорость v пренебрежимо мала. Применим уравнение (95.2) к линии тока AB, начинающейся на поверхности жидкости в точке A (рис. 255). Если начало координат поместить в точке A, то $z_A = r_A = v_A = 0$, $P_A = P_B = P_0$, $v_B = v$, $z_B = -h$, $r_B = R$, и мы получим

$$v = \sqrt{2(gh + \omega^2 R^2)}\,. \qquad (95.3)$$

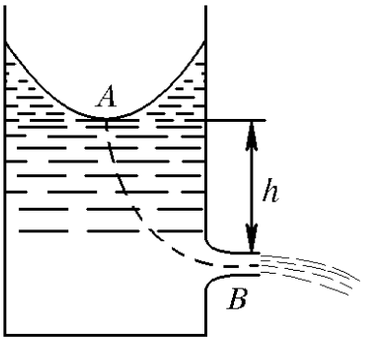


Рис. 255

Здесь h означает высоту наиболее низкой (центральной) точки A уровня жидкости относительно отверстия, а R — радиус цилиндра. Переход к неподвижной системе отсчета не представляет затруднений.

§ 96. ВЯЗКОСТЬ

1. В реальных жидкостях, помимо сил нормального давления, на границах движущихся элементов жидкости действуют еще *касательные силы вязкости*. Убедиться в существовании таких сил можно на простейших примерах. Так, уравнение Бернулли, выводимое в предположении, что силы вязкости отсутствуют, приводит

порционален расстоянию h между пластинками:

$$F = \eta S \frac{v_0}{h}. \tag{96.1}$$

Здесь η — постоянная, называемая *вязкостью жидкости*. Она не зависит от материала пластинок и имеет разные значения для различных жидкостей. Для данной жидкости коэффициент η зависит от параметров, характеризующих ее внутреннее состояние, и в первую очередь от температуры. Таким образом, под вязкостью понимают как явление, так и коэффициент, характеризующий свойства жидкости.

Не обязательно, чтобы пластинка AB покоилась. Обе пластинки могут двигаться равномерно параллельно друг другу. Если скорость пластинки AB равна v_1, а пластинки CD — v_2, то вместо (96.1) можно написать более общую формулу

$$F = \eta S \frac{v_2 - v_1}{h}. \tag{96.2}$$

Чтобы убедиться в этом, достаточно перейти в систему отсчета, в которой пластинка AB покоится.

Заметим далее, что при равномерном движении пластинки CD жидкость должна действовать на нее с силой $-\mathbf{F}$, чтобы полная сила, приложенная к пластинке CD, обращалась в нуль. Значит, сама пластинка CD будет действовать на жидкость с силой $+\mathbf{F}$. Аналогично пластинка AB будет действовать на жидкость с силой $-\mathbf{F}$. Кроме того, исследования показали, что жидкость, обладающая вязкостью, прилипает к поверхности твердого тела, которое она обтекает. Иными словами, скорости частиц жидкости относительно поверхности твердого тела, на которой они находятся, равны нулю.

Поэтому в формуле (96.2) силы $\mathbf{F}$ и $-\mathbf{F}$ можно считать приложенными не к пластинке, а к границам заключенного между ними слоя жидкости. Точно так же v_1 и v_2 можно отождествить со скоростями движения тех же границ жидкого слоя. Тем самым при введении понятия вязкости η надобность в пластинках отпадает.

3. В целях обобщения формулы (96.2) допустим, что жидкость течет в направлении оси X, причем скорость течения зависит только от координаты y:

$$v_x = v_x(y), \quad v_y = v_z = 0.$$

Вырежем мысленно жидкий слой, ограниченный бесконечно близкими плоскостями, перпендикулярными к оси Y. Пусть эти плоскости пересекают ось Y в точках с ординатами y и $y + dy$ (рис. 257). Обозначим через τ_{yx} касательную силу, действующую на единицу площади верхней границы такого слоя со стороны вышележащей жидкости. Первый индекс y указывает направление внешней нормали к верхней границе слоя, а второй — индекс x — направление дейст-

вующей силы (ср. § 74, п. 2). Обобщая формулу (96.2), для касательного напряжения τ_{yx} напишем

$$\tau_{yx} = \eta \frac{\partial v_x}{\partial y}. \tag{96.3}$$

Примем в согласии с опытом, что формула (96.3) справедлива не только для равномерного течения, но и для течения, скорость v_x которого зависит от времени. Касательное напряжение на нижней границе слоя τ_{-yx} направлено в сторону, противоположную τ_{yx}. Оно отличается от τ_{-yx} бесконечно мало в виду бесконечной малости толщины слоя dy ($\tau_{yx} = -\tau_{-yx}$).

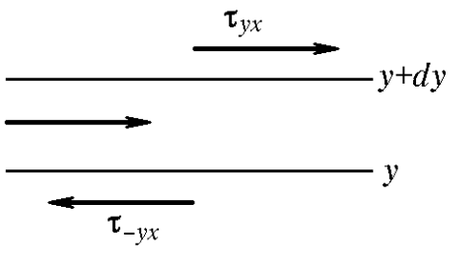


Рис. 257

4. Выделим теперь в том же параллельном потоке жидкости бесконечно малый параллелепипед $ABCD$ с ребрами, параллельными координатным осям (рис. 258). Тензор напряжений, как следует из уравнения моментов, симметричен (см. § 74, п. 4). Поэтому на основаниях параллелепипеда BC и AD, перпендикулярных к потоку, должны также существовать касательные напряжения, причем $\tau_{xy} = \tau_{yx} = \eta \frac{\partial v_x}{\partial y}$. Таким образом, *касательные напряжения действуют не только в плоскостях, параллельных течению, но и в плоскостях, перпендикулярных ему.*

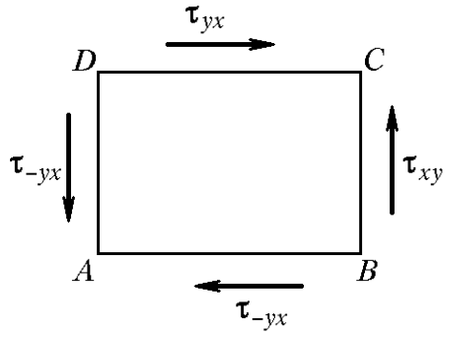


Рис. 258

Допустим теперь, что жидкость течет не параллельным потоком, а произвольным образом. Примем, что касательные составляющие тензора вязких напряжений зависят только от скоростей деформаций жидкости, а не от самих деформаций и их высших производных по времени. Ограничимся *линейным приближением*, т. е. будем пренебрегать квадратами и высшими степенями скоростей деформаций. В этом приближении касательные напряжения являются *линейными однородными функциями* скоростей деформаций $\frac{\partial v_x}{\partial y}, \frac{\partial v_y}{\partial x}, \frac{\partial v_y}{\partial z}, \frac{\partial v_z}{\partial y}, \frac{\partial v_z}{\partial x}, \frac{\partial v_x}{\partial z}$. Если бы из этих шести производных на границе CD была отлична от нуля только производная $\frac{\partial v_x}{\partial y}$, то вдоль оси X на этой границе действовало бы касательное напряжение $\tau'_{yx} = \eta \frac{\partial v_x}{\partial y}$. Если бы отличалась от нуля только производная $\frac{\partial v_y}{\partial x}$, то касательное напряжение имело бы то же направление и было бы равно $\tau''_{yx} = \eta \frac{\partial v_y}{\partial x}$. А если отличны от нуля обе производные $\frac{\partial v_x}{\partial y}$ и

$\frac{\partial v_y}{\partial x}$, то касательное напряжение на границе CD будет $\tau_{yx} = \tau'_{yx} + \tau''_{yx} = \eta\left(\frac{\partial v_x}{\partial y} + \frac{\partial v_y}{\partial x}\right)$. Это непосредственно следует из предположения о линейной однородной зависимости между касательными напряжениями и скоростями деформаций жидкости. Отсюда же следует, что найденное выражение для τ_{yx} сохранит свою силу, каковы бы ни были значения других производных $\frac{\partial v_y}{\partial z}$, $\frac{\partial v_z}{\partial y}$ и т. д. Рассуждая аналогично, найдем выражения и для всех остальных касательных напряжений, действующих на гранях параллелепипеда $ABCD$. Именно,

$$\tau_{xy} = \tau_{yx} = \eta\left(\frac{\partial v_x}{\partial y} + \frac{\partial v_y}{\partial x}\right),$$

$$\tau_{yz} = \tau_{zy} = \eta\left(\frac{\partial v_y}{\partial z} + \frac{\partial v_z}{\partial y}\right), \qquad (96.4)$$

$$\tau_{zx} = \tau_{xz} = \eta\left(\frac{\partial v_z}{\partial x} + \frac{\partial v_x}{\partial z}\right).$$

Если жидкость несжимаема, то этих выражений достаточно для вывода дифференциальных уравнений движения жидкости. Если же жидкость сжимаема, то к ним надо добавить еще выражения для нормальных напряжений. Мы не будем приводить здесь эти выражения, так как они нам не понадобятся.

5. Рассмотрим частный случай, когда вязкая жидкость вращается вокруг неподвижной оси с угловой скоростью ω. Линии тока имеют форму окружностей. Пусть AB — бесконечно малый участок линии тока длиной $rd\varphi$ (рис. 259). Касательное напряжение на цилиндрической поверхности, на которой лежит этот участок, очевидно, направлено в сторону вращения. Его следует обозначить через $\tau_{r\varphi}$. Первый индекс r указывает направление внешней нормали к цилиндрической поверхности, второй индекс φ — положительное направление касательного напряжения. В рассматриваемом случае роль dy играет dr, роль dx — длина $AB = rd\varphi$. Поэтому из (96.4) для касательного напряжения $\tau_{r\varphi}$ получаем

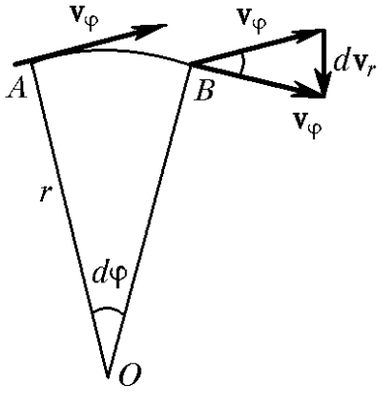


Рис. 259

$$\tau_{r\varphi} = \eta\left(\frac{\partial v_\varphi}{\partial r} + \frac{\partial v_r}{r\,\partial\varphi}\right).$$

В точке A радиальная составляющая скорости v равна нулю. В точке B появляется составляющая скорости вдоль радиуса OA, равная $dv_r = -v_\varphi d\varphi$, так что $\frac{\partial v_r}{\partial\varphi} = -v_\varphi$, а потому

$$\tau_{r\varphi} = \eta\left(\frac{\partial v_\varphi}{\partial r} - \frac{v_\varphi}{r}\right). \qquad (96.5)$$

Подставляя сюда $v_\varphi = \omega r$, получим

$$\tau_{r\varphi} = \eta r \frac{\partial \omega}{\partial r}. \tag{96.6}$$

Вязкие напряжения исчезают, если $\frac{\partial \omega}{\partial r} = 0$, т. е. если жидкость вращается как целое, подобно твердому телу. Этого не получилось бы, если бы в формуле

$$\tau_{yx} = \eta \left(\frac{\partial v_x}{\partial y} + \frac{\partial v_y}{\partial x} \right) \tag{96.7}$$

не было учтено второе слагаемое.

6. В качестве примера на применение формулы (96.5) рассмотрим установившееся движение жидкости между двумя равномерно вращающимися коаксильными цилиндрами. Пусть l — высота цилиндров, R_1 и R_2 — их радиусы, а Ω_1 и Ω_2 — угловые скорости. Величину l будем предполагать очень большой по сравнению с толщиной зазора $R_2 - R_1$ между цилиндрами. Тогда цилиндры можно считать бесконечно длинными и отвлечься от осложняющих обстоятельств, вносимых их краями. Проведем в жидкости произвольную цилиндрическую поверхность радиусом r (рис. 260). Момент сил вязкости, действующих на этой поверхности, относительно оси вращения равен

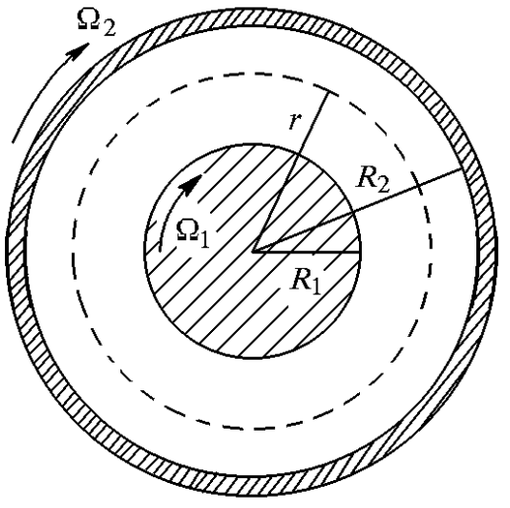


Рис. 260

$$M = 2\pi r^2 l \tau_{r\varphi} = 2\pi \eta l r^3 \frac{\partial \omega}{\partial r}.$$

При установившемся вращении жидкости этот момент не должен зависеть от радиуса r. Только при этом условии момент сил, действующих на жидкость, заключенную между двумя любыми коаксиальными цилиндрическими поверхностями, обращается в нуль, а момент импульса жидкости сохраняется. Таким образом, мы приходим к уравнению

$$r^3 \frac{\partial \omega}{\partial r} = \text{const}.$$

Обозначая входящую сюда постоянную $-2A$ и интегрируя, получим

$$\omega = \frac{A}{r^2} + C,$$

где C — постоянная интегрирования. Постоянные A и C определятся из граничных условий. Так как вязкая жидкость прилипает к поверхности тела, которое она обтекает, то угловая скорость ω при $r = R_1$ должна обращаться в Ω_1, а при $r = R_2$ — в Ω_2. Это приводит к двум уравнениям

$$\frac{A}{R_1^2} + C = \Omega_1, \quad \frac{A}{R_2^2} + C = \Omega_2,$$

решая которые, находим

$$A = \frac{R_1^2 R_2^2}{R_2^2 - R_1^2}(\Omega_1 - \Omega_2), \quad C = \frac{R_2^2 \Omega_2 - R_1^2 \Omega_1}{R_2^2 - R_1^2}$$

и далее

$$\omega = \frac{R_1^2 R_2^2}{R_2^2 - R_1^2} \frac{\Omega_1 - \Omega_2}{r^2} + \frac{R_2^2 \Omega_2 - R_1^2 \Omega_1}{R_2^2 - R_1^2}. \tag{96.8}$$

Момент сил вязкости, действующих на внутренний цилиндр, равен

$$M = 2\pi\eta l(-2A) = 4\pi\eta l \frac{R_1^2 R_2^2}{R_2^2 - R_1^2}(\Omega_2 - \Omega_1). \tag{96.9}$$

Формула (96.9) лежит в основе практического метода измерения вязкости жидкостей и газов. Внутренний цилиндр подвешивается в исследуемой жидкости в вертикальном положении на тонкой нити, а наружный приводится в равномерное вращение с угловой скоростью $\Omega_2 = \Omega$. Измеряется угол закручивания нити φ, при котором внутренний цилиндр находится в равновесии. Это будет тогда, когда момент вязких напряжений M уравновешивается моментом закрученной нити $f\varphi$, где f — модуль кручения. Вязкость рассчитывается по формуле

$$\eta = \frac{f\varphi}{4\pi l} \frac{R_2^2 - R_1^2}{R_1^2 R_2^2 \Omega}. \tag{96.10}$$

ЗАДАЧИ

1. Введя локальную систему координат с началом в рассматриваемой точке пространства, убедиться непосредственным дифференцированием, что формула (96.7) при вращении жидкости переходит в формулу (96.5).

Р е ш е н и е. Проведем через рассматриваемую точку пространства A круговую линию тока. Поместим начало локальной системы координат в точку A, направив координатные оси X и Y, как указано на рис. 261. Для координат и компонентов скорости в точке B получим

$$x = r\sin\varphi,\ y = r\cos\varphi - r_0,$$

$$v_x = v_\varphi\cos\varphi,\ v_y = -v_\varphi\sin\varphi,$$

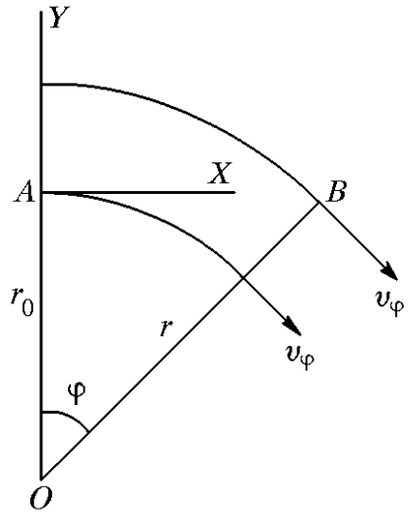


Рис. 261

где r_0 и r — радиусы-векторы точек A и B, а v_φ — скорость жидкости в точке B. Дифференцируя эти соотношения и полагая в окончательных результатах $\varphi = 0$ (точка A), получим в точке A

$$dx = r_0\,d\varphi,\ dy = dr,$$

$$dv_x = dv_\varphi,\ dv_y = -v_\varphi\,d\varphi.$$

Отсюда

$$\frac{\partial v_x}{\partial y} = \frac{\partial v_y}{\partial r},\ \frac{\partial v_y}{\partial x} = -\frac{v_\varphi}{r_0}.$$

После подстановки в (96.7) получаем (96.5).

2. Как изменится формула (96.9) в предельном случае, когда толщина зазора между цилиндрами $h = R_2 - R_1$ пренебрежимо мала по сравнению с радиусами R_1 и R_2?

Ответ. $M = \frac{2\pi\eta l R^3}{h}(\Omega_2 - \Omega_1)$. (96.11)

Эту формулу можно также получить, рассматривая слой жидкости между цилиндрами как плоскопараллельный и используя формулу (96.2). Это рекомендуется сделать читателю.

§ 97. СТАЦИОНАРНОЕ ТЕЧЕНИЕ ЖИДКОСТИ ПО ПРЯМОЛИНЕЙНОЙ ТРУБЕ. ФОРМУЛА ПУАЗЕЙЛЯ

1. Пусть вязкая несжимаемая жидкость течет вдоль прямолинейной цилиндрической трубы радиусом R. Линии тока параллельны оси трубки. Если выделить произвольную бесконечно узкую трубку тока, то из условия несжимаемости следует, что скорость течения v будет одна и та же вдоль всей трубки тока — скорость жидкости не может меняться вдоль трубы. Но она, конечно может изменяться с изменением расстояния r от оси трубы. Таким образом, скорость жидкости v является функцией радиуса r.

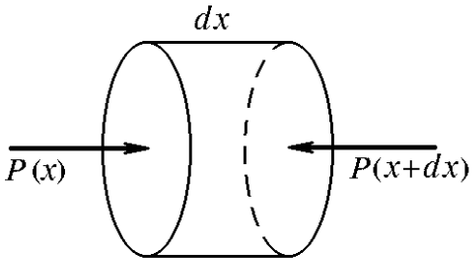


Рис. 262

Примем ось трубы за ось X, направленную в сторону течения. Выделим в трубе произвольную бесконечно короткую цилиндрическую часть длиной dx и радиусом r (рис. 262). На ее боковую поверхность в направлении движения действует касательная сила вязкости $dF = 2\pi r\eta \frac{dv}{dr} dx$. Кроме того, на основания цилиндра в том же направлении действует сила разности давлений $dF_1 = \pi r^2[P(x) - P(x + dx)] = -\pi r^2 \frac{dP}{dx} dx$. При стационарном течении сумма этих двух сил должна обращаться в нуль, а потому

$$2\eta \frac{dv}{dr} = r \frac{dP}{dx}.$$

Скорость $v(r)$, а с ней и производная $\frac{dv}{dr}$ не меняются с изменением x. Поэтому должна быть постоянной и производная $\frac{dP}{dx}$, причем эта производная должна быть равна $(P_2 - P_1)/l$, где P_1 — давление на входе трубы, P_2 — на выходе, а l — длины трубы. В результате приходим к уравнению

$$\frac{dv}{dr} = -\frac{P_1 - P_2}{2\eta l} r. \qquad (97.1)$$

Интегрируя его, получим

$$v = -\frac{P_1 - P_2}{4\eta l} r^2 + C.$$

Постоянная интегрирования C определится из условия, что на стенке трубы, т. е. при $r = R$, скорость v должна обращаться в нуль. Это дает

$$v = \frac{P_1 - P_2}{4\eta l}(R^2 - r^2). \tag{97.2}$$

Скорость v максимальна на оси трубы, где она достигает значения

$$v = \frac{P_1 - P_2}{4\eta l} R^2. \tag{97.3}$$

При удалении от оси скорость v меняется по параболическому закону.

2. Подсчитаем *расход жидкости*, т. е. количество ее, ежесекундно протекающее через поперечное сечение трубы. Масса жидкости, ежесекундно протекающей через кольцевую площадку с внутренним радиусом r и внешним $r + dr$, равна $dQ = 2\pi r\, dr \cdot \rho v$. Подставляя сюда выражение для v и интегрируя, находим искомый расход жидкости

$$Q = \pi\rho \frac{P_1 - P_2}{2\eta l} \int_0^R (R^2 - r^2) r dr,$$

или

$$Q = \pi\rho \frac{P_1 - P_2}{8\eta l} R^4. \tag{97.4}$$

Расход жидкости пропорционален разности давлений $P_1 - P_2$, четвертой степени радиуса трубы и обратно пропорционален длине трубы и вязкости жидкости. Эти закономерности были установлены экспериментально и независимо друг от друга в 1839 г. Гагеном и в 1840 г. французским физиком Жаном Пуазейлем (1799—1869). Гаген исследовал движение воды в трубах, Пуазейль — течение жидкостей в каппилярах. Формула (97.4) называется *формулой Пуазейля*, хотя сам Пуазейль и не выводил ее, он исследовал вопрос только экспериментально. На формуле Пуазейля основан один из экспериментальных методов определения вязкости жидкостей.

Формулу (97.4) можно представить в виде $Q = \rho\pi R^2 \cdot v_0/2$. С другой стороны, можно ввести *среднюю скорость потока* $\bar{v}$, определив ее с помощью соотношения $Q = \rho\pi R^2 \bar{v}$. Сравнивая эти два выражения, получаем

$$\bar{v} = \frac{1}{2} v_0. \tag{97.5}$$

Формула Пуазейля справедлива только для *ламинарных течений* жидкости. *Ламинарным* называется такое течение, когда части-

цы жидкости движутся вдоль прямолинейных траекторий, параллельных оси трубы. (Более общее определение, применимое для любых течений, дается в § 98.) При больших скоростях ламинарное течение становится неустойчивым и переходит в *турбулентное течение*, с которым мы познакомимся в § 98. К турбулентным течениям формула Пуазейля неприменима.

3. Кинетическая энергия, ежесекундно переносимая потоком жидкости через поперечное сечение трубы, определяется выражением

$$K = \int_0^R \frac{\rho v^2}{2} \cdot 2\pi r v \, dr.$$

Подставив сюда значение для v и выполнив интегрирование, получим

$$K = \frac{1}{4} Q v_0^2 = Q(\bar{v})^2. \tag{97.6}$$

Работа, ежесекундно производимая над жидкостью разностью давлений $P_1 - P_2$, определяется выражением $A = \int v\,(P_1 - P_2) \cdot 2\pi r \, dr$, или

$$A = \frac{P_1 - P_2}{\rho} Q. \tag{97.7}$$

Такую же по значению, но противоположную по знаку работу производят силы вязкости, так как при стационарном течении кинетическая энергия жидкости остается неизменной: $A' = -A$. С помощью формулы (97.3) можно исключить разность давлений $P_1 - P_2$ и получить

$$A' = -\frac{4\eta v_0 l}{\rho R^2} Q. \tag{97.8}$$

Полученные формулы позволяют ответить на вопрос, когда при течении жидкости по трубе можно пренебречь силами вязкости и, следовательно, применять уравнение Бернулли. Для этого, очевидно, необходимо, чтобы потеря кинетической энергии жидкости, обусловленная действием сил вязкости, была пренебрежимо мала по сравнению с кинетической энергией самой жидкости, т. е. $|A'| \ll K$. Это приводит к условию

$$\frac{v_0 R^2}{16 \nu l} \gg 1. \tag{97.9}$$

Здесь буквой ν обозначена так называемая *кинематическая вязкость*, т. е. отношение

$$\nu = \frac{\eta}{\rho}. \tag{97.10}$$

Величину η, когда надо отличать ее от ν, называют *динамической вязкостью*.

4. Законы, установленные Пуазейлем, могут быть в общем виде получены методом размерностей. Достоинство этого метода состоит

в том, что он применим к прямолинейным трубам произвольного поперечного сечения, а не только к цилиндрическим трубам. Требуется только, чтобы нормальные поперечные сечения всех труб были геометрически подобраны. Эти сечения могут отличаться друг от друга только размерами. Для каждого поперечного сечения можно установить *характерный размер*. За таковой можно принять, например, его периметр или корень квадратный из площади. Можно также поперечные сечения всех труб геометрически подобно рассечь на две части прямолинейными отрезками. Длины таких отрезков тоже могут служить характерными размерами. Например, в случае трубы эллиптического сечения за характерный размер можно взять длину большой или малой оси соответствующего эллиптического сечения. Но можно взять и другие отрезки, характеризующие размеры эллипса. Заданием характерного размера определяются и все прочие размеры поперечного сечения трубы.

При выводе законов Пуазейля, равно как при исследовании любого вопроса методом размерностей, основной пункт состоит в том, чтобы установить физические величины, связанные между собой функциональной зависимостью. При стационарном ламинарном течении жидкости по трубе силы вязкости уравновешиваются градиентами давлений. В уравнения движения входят эти градиенты, а потому разность давлений $P_1 - P_2$ и длина трубы l могут войти только в комбинации $(P_1 - P_2)/l$. Поскольку жидкость движется без ускорения, характер течения не может зависеть от плотности жидкости. Плотность ρ и расход жидкости Q могут войти лишь в комбинации Q/ρ, так как последняя есть чисто геометрическая величина и равна объему жидкости, ежесекундно протекающему через поперечное сечение трубы. Добавив сюда еще вязкость жидкости η и характерный поперечный размер трубы a, получим четыре величины.

$$\frac{Q}{\rho}, \quad \frac{P_1-P_2}{l}, \quad a, \quad \eta,$$

между которыми должна существовать функциональная связь. Вместо a можно взять площадь поперечного сечения трубы S. Применяя общий метод нахождения безразмерных комбинаций (см. § 87, п. 6), нетрудно убедиться, что из рассматриваемых величин можно составить только одну независимую безразмерную комбинацию, а именно

$$\frac{Q}{\rho}\,\frac{l}{P_1-P_2}\,\frac{\eta}{S^2}.$$

Следовательно, такая комбинация должна быть постоянной. Обозначая эту постоянную через C, получим

$$Q = C\,\frac{P_1-P_2}{l\eta}\,\rho S^2. \tag{97.11}$$

В этой формуле содержатся все законы Пуазейля. Она является обобщением формулы (97.4) на случай прямолинейных труб произ-

вольного поперечного сечения. Постоянная C зависит от формы поперечного сечения трубы и не может быть определена методами теории размерности. Для ее нахождения необходимо обратиться к опыту или к динамическим методам, т. е. к интегрированию уравнений движения.

ЗАДАЧИ

1. Определить стационарное течение вдоль оси и расход несжимаемой жидкости между двумя коаксиальными цилиндрами с внутренним радиусом R_1, внешним R_2 и длиной l.

Р е ш е н и е. Рассмотрим кольцевой слой жидкости с внутренним радиусом r и внешним $r + dr$. Сила вязкости, действующая на него в направлении течения, равна

$$2\pi l\eta\left[\left(r\frac{dv}{dr}\right)_{r+dr} - \left(r\frac{dv}{dr}\right)_r\right] = 2\pi l\eta\frac{d}{dr}\left(r\frac{dv}{dr}\right)dr.$$

(Индексы r и $r + dr$ означают, что величины, заключенные в круглые скобки, должны быть вычислены при значениях радиусов r и $r + dr$ соответственно.) В том же направлении действует сила разности давлений $(P_1 - P_2)\cdot 2\pi r\,dr$. При стационарном течении сумма обеих сил обращается в нуль. Это приводит к уравнению

$$\frac{d}{dr}\left(r\frac{dv}{dr}\right) = -\frac{P_1-P_2}{l\eta}r. \tag{97.12}$$

Решение его, обращающееся в нуль при $r = R_1$ и $r = R_2$, есть

$$v = \frac{P_1-P_2}{4l\eta}\left\{R_2^2 - r^2 + \frac{R_2^2-R_1^2}{\ln(R_2/R_1)}\ln\frac{r}{R_2}\right\}.$$

Расход жидкости

$$Q = \frac{\pi\rho(P_1-P_2)}{8\eta l}\left\{R_2^4 - R_1^4 - \frac{(R_2^2-R_1^2)^2}{\ln(R_2/R_1)}\right\}.$$

2. Показать, что при ламинарном стационарном течении несжимаемой жидкости вдоль прямолинейной трубы с произвольным поперечным сечением и длиной l скорость жидкости v удовлетворяет уравнению

$$\frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial z^2} + \frac{P_1-P_2}{l\eta} = 0. \tag{97.13}$$

(Координатная плоскость YZ перпендикулярна к оси трубы, оси Y и Z взаимно перпендикулярны и ориентированы произвольно.)

У к а з а н и е. Взять произвольный бесконечно тонкий прямоугольный параллелепипед жидкости длиной l с ребрами, параллельными координатным осям, и написать условие обращения в нуль действующих на него сил вязкости и разности давлений, подобно тому как это делалось при выводе уравнения (97.12).

3. Определить скорость течения и расход жидкости в трубе эллиптического сечения.

Р е ш е н и е. Эта задача относится к типу задач, решаемых *методом угадывания*. Угадывается вид решения дифференциального уравнения (97.13), а затем коэффициенты в этом решении подбираются так, чтобы удовлетворить граничному условию на стенке трубы: $v = 0$. Направим координатные оси Y и Z вдоль главных осей нормального эллиптического поперечного сечения трубы и будем искать решение в виде $v = Ay^2 + Bz^2 + v_0$. Это выражение удовлетворяет уравнению (97.13), если

$$2A + 2B = -\frac{P_1 - P_2}{l\eta}.$$

На внутренней поверхности эллиптической трубы $v = 0$, т. е. $Ay^2 + Bz^2 + v_0 = 0$. Это уравнение должно переходить в уравнение эллиптического сечения трубы $\frac{y^2}{a^2} + \frac{z^2}{b^2} - 1 = 0$, а потому

$$A = -\frac{v_0}{a^2}, \quad B = -\frac{v_0}{b^2}.$$

Для определения постоянных A, B, v_0 получилось три линейных уравнения. Решая их, находим

$$v_0 = \frac{P_1 - P_2}{2l\eta} \frac{a^2 b^2}{a^2 + b^2}, \tag{97.14}$$

$$v = v_0 \left(1 - \frac{y^2}{a^2} - \frac{z^2}{b^2}\right). \tag{97.15}$$

Постоянная v_0 есть, очевидно, скорость течения на оси трубы.

Вычислим теперь расход жидкости. Поверхности, на которых скорость v постоянна, суть эллиптические цилиндры

$$\frac{y^2}{a'^2} + \frac{z^2}{b'^2} = 1,$$

полуоси которых определяются соотношениями

$$a'^2 = a^2 \frac{v_0 - v}{v_0}, \quad b'^2 = b^2 \frac{v_0 - v}{v_0}.$$

Возьмем два таких эллиптических цилиндра с бесконечно близкими значениями параметра v. Площадь нормального сечения между ними $dS = d(\pi a' b') = -\pi \frac{ab}{v_0} dv$. Расход жидкости:

$$Q = \rho \int v\, dS = -\rho \frac{\pi ab}{v_0} \int_{v_0}^{0} v\, dv,$$

или

$$Q = \frac{1}{2} \rho \pi ab v_0. \tag{97.16}$$

§ 98. ЗАКОНЫ ГИДРОДИНАМИЧЕСКОГО ПОДОБИЯ

1. Рассмотрим поток жидкости, обтекающий какое-нибудь тело или систему тел. Наряду с этой системой можно ввести бесконечное

множество *подобных* и *подобно расположенных тел*, обтекаемых другими жидкостями. Каким условиям должны удовлетворять параметры потока и постоянные, характеризующие свойства жидкостей (плотность, вязкость и пр.), чтобы оба потока были *механически подобны*? Если подобие имеет место, то, зная картину течения для первой системы тел, можно однозначно предсказать течение жидкости и для другой, геометрически подобной, системы тел. Это имеет важное значение в судостроении и самолетостроении. Вместо реальных кораблей или самолетов испытывают их *уменьшенные геометрически подобные модели*, а затем путем пересчета определяется поведение реальных систем. Простейший метод решения поставленной задачи дает теория размерностей.

Исследуем вопрос в общем виде. Пусть $\mathbf{r}$ и $\mathbf{v}$ — радиус-вектор и скорость жидкости в подобно расположенных точках, l — *характерный размер*, а v_0 — *характерная скорость потока*, например скорость жидкости, с которой она из «бесконечности» натекает на рассматриваемую систему тел. Свойства жидкости характеризуются ее плотностью ρ, вязкостью η и сжимаемостью. Вместо сжимаемости можно пользоваться скоростью звука в рассматриваемой жидкости. Если существенна сила тяжести, то последняя характеризуется ускорением свободного падения g. Если течение не стационарно, то надо ввести какое-то *характерное время* τ, за которое происходит заметное изменение течения. Ввиду наличия уравнений движения между величинами

$$\mathbf{v},\ v_0,\ \mathbf{r},\ l,\ \rho,\ \eta,\ c,\ g,\ \tau$$

должна существовать функциональная связь. Из них можно составить шесть независимых безразмерных комбинаций. Сюда относятся два отношения $\mathbf{v}/v_0$, $\mathbf{r}/l$ и еще четыре безразмерных числа:

$$\mathrm{Re} = \frac{\rho l v_0}{\eta} = \frac{l v_0}{\nu}, \tag{98.1}$$

$$\mathrm{F} = \frac{v_0^2}{gl}, \tag{98.2}$$

$$\mathrm{M} = \frac{v_0}{c}, \tag{98.3}$$

$$\mathrm{S} = \frac{v_0 \tau}{l}. \tag{98.4}$$

Согласно правилу размерности одна из этих безразмерных комбинаций является функцией остальных, например

$$\frac{\mathbf{v}}{v_0} = \mathbf{f}\left(\frac{\mathbf{r}}{l},\ \mathrm{Re},\ \mathrm{F},\ \mathrm{M},\ \mathrm{S}\right) \tag{98.5}$$

или

$$\mathbf{v} = v_0 \mathbf{f}\left(\frac{\mathbf{r}}{l},\ \mathrm{Re},\ \mathrm{F},\ \mathrm{M},\ \mathrm{S}\right) \tag{98.6}$$

Если для двух течений пять из шести безразмерных комбинаций, перечисленных выше, совпадают, то будут совпадать и шестые. Это — общий *закон подобия течений*, а сами течения называются *механически* или *гидродинамически подобными.*

2. Величина (98.1) называется *числом Рейнольдса* (по имени английского физика Осборна Рейнольдса (1842—1912), величина (96.2) — *числом Фруда* (по имени английского ученого Фруда), величина (98.3) — *числом Маха* (по имени австрийского физика Эрнста Маха (1838—1916)), величина (98.4) — *числом Струхаля* (по имени чешского физика В. Струхаля).

Физический смысл чисел Маха и Струхаля не требует пояснений. На физическом смысле чисел Рейнольдса и Фруда необходимо остановится подробнее. При этом само собой станет ясным, что оба эти числа безразмерные. *По порядку величины число Рейнольдса есть отношение кинетической энергии жидкости к потере ее, обусловленной работой сил вязкости на характерной длине.* Действительно, кинетическая энергия жидкости $K \sim \frac{1}{2} \rho v_0^2 l^3$. Силу вязкости найдем, умножая вязкое напряжение $\eta v_0/l$ на характерную площадь l^2. Это дает $\eta v_0 l$. Произведение этой силы на характерную длину определяет по порядку величины работу сил вязкости $A \sim \eta v_0 l^2$. Отношение кинетической энергии K к работе A будет

$$\frac{K}{A} \sim \frac{\rho l v_0}{\eta},$$

а это и есть число Рейнольдса. *Число Рейнольдса, таким образом, определяет относительную роль инерции и вязкости жидкости при течении. При больших числах Рейнольдса основную роль играет инерция, при малых — вязкость.*

Число Рейнольдса, конечно, определено не вполне четко, поскольку оно содержит характерную длину и характерную скорость, которые сами определены не четко. Это число, как и все остальные безразмерные числа в законе подобия, определено лишь по порядку величины. Если размеры тела в разных направлениях примерно одинаковы, то особой неопределенности не возникает. Если же это не так, то в качестве характерной длины могут быть выбраны разные величины, которые могут отличаться друг от друга значительно. Например, при течении жидкости по трубе за характерную длину можно взять длину трубы, ее радиус или какую-либо промежуточную величину. Соответствующие числа Рейнольдса могут отличаться на много порядков. Какое из этих чисел взять — зависит от поставленной задачи. Так, в предыдущем параграфе было выведено условие (97.9), при выполнении которого силами вязкости можно пренебречь. Величину, стоящую слева в формуле (97.9), можно рассматривать как число Рейнольдса, если за характерную длину принять $\frac{1}{16}\frac{R^2}{l}$. В рассматриваемом случае характерный раз-

мер зависит как от длины трубы, так и от ее радиуса. При таком выборе получается условие (97.9), справедливое для всех, а не только геометрически подобных круглых труб (т. е. труб с постоянным отношением R/l). Если труба длинная ($l \gg 16R$), то достаточное условие можно записать в виде

$$v_0 R/\nu \gg 1, \tag{98.7}$$

т. е. за характерную длину можно принять радиус R. Но мы совершили бы ошибку, если бы вместо (98.7) взяли условие $v_0 l/\nu \gg 1$.

Аналогичный смысл имеет *число Фруда* F. *Оно по порядку величины определяет отношение кинетической энергии жидкости к приращению ее, обусловленному работой силы тяжести на пути, равному характерной длине. Чем больше число Фруда, тем больше роль инерции по сравнению с тяжестью, и наоборот.*

3. Для стационарных течений характерное время τ, а с ним и число Струхаля обращаются в бесконечность. Поэтому это число выпадает из соотношения (98.6). То же происходит с чилом Маха в несжимаемых жидкостях, для которых оно обращается в нуль. Таким образом, для стационарных течений несжимаемых жидкостей соотношение (98.6) переходит в

$$\mathbf{v} = v_0 \mathbf{f}\left(\frac{\mathbf{r}}{l}, \mathsf{Re}, \mathsf{F}\right) \tag{98.8}$$

Течения подобны, если они имеют одинаковые числа Рейнольдса и Фруда.

Следует, однако, заметить, что если при испытаниях на моделях применяется та же жидкость, в которой должна двигаться реальная система, то критерии подобия Рейнольдса и Фруда несовместимы друг с другом. В самом деле, запишем эти критерии в виде

$$\frac{l_1 v_1}{\nu_1} = \frac{l_2 v_2}{\nu_2}, \quad \frac{v_1^2}{l_1 g_1} = \frac{v_2^2}{l_2 g_2},$$

где индекс 1 относится к реальной системе тел, а индекс 2 — к ее уменьшенной или увеличенной модели. Перемножая эти соотношения, получим

$$\frac{\nu_2}{\nu_1} = \frac{g_1}{g_2}\left(\frac{v_2}{v_1}\right)^3 = \left(\frac{g_2}{g_1}\right)^{1/2}\left(\frac{l_2}{l_1}\right)^{3/2}. \tag{98.9}$$

Варьирование ускорения свободного падения g принципиально возможно, но практически нереально. Однако и при одинаковых g можно принципиально удовлетворить обоим критериям подобия. Для этого надо применять жидкости с различными кинематическими вязкостями, удовлетворяющими соотношению (98.9). В большинстве случаев это почти невозможно. При испытаниях на моделях практически может выполняться только один критерий подобия: либо Рейнольдса, либо Фруда. В некоторых случаях этого достаточно. Допустим, например, что число Рейнольдса велико, а число Фруда невелико или поряд-

ка единицы. Тогда движение жидкости в основном будет определяться инерцией и тяжестью. Вариации числа Рейнольдса на нем будут сказываться мало. В этом случае для подобия течения необходимо выполнение лишь одного *критерия Фруда*. Напротив, при малых числах Рейнольдса и больших числах Фруда определяющую роль играют инерция и вязкость; влияние тяжести незначительно. Подобие будет иметь место при *равенстве чисел Рейнольдса*.

4. Чтобы исследовать поведение самолета во время полета, например определить действующие на него силы, модель самолета закрепляют в *аэродинамической трубе*, в которой создается равномерный поток воздуха. В основе этого метода лежит *принцип относительности*, согласно которому ход явления может зависеть только от относительного движения самолета и воздуха. Современные аэродинамические трубы представляют грандиозные сооружения, в которых скорость воздуха может быть доведена до сотен метров в секунду. Значительное уменьшение размеров испытываемой модели неосуществимо, и вот почему. Для сохранения аэродинамического подобия необходимо равенство чисел Рейнольдса $\mathrm{Re} = vl/\nu$, так что при уменьшении размеров модели в несколько раз во столько же раз должна быть увеличена скорость потока. Но при больших скоростях начинает существенно сказываться сжимаемость воздуха, нарушающая аэродинамическое подобие. Поэтому при больших скоростях, интересующих современную авиацию, приходится применять модели либо в натуральную величину, либо лишь незначительно уменьшенные. Вот почему сечения аэродинамических труб должны быть очень большими, чтобы в них можно было помещать отдельные части самолета или даже целые самолеты. Для преодоления указанной трудности в принципе можно идти по пути увеличения плотности воздуха, делая аэродинамические трубы геометрическими. Дело в том, что динамическая вязкость газа η практически от плотности газа не зависит (при заданной температуре), а потому кинематическая вязкость $\nu = \eta/\rho$ обратно пропорциональна плотности. Увеличивая плотность ρ, можно сохранить аэродинамическое подобие и для значительно уменьшенных моделей без существенного увеличения и даже без изменения скорости потока v. Несмотря на сложность построения геометрических аэродинамических труб, этот метод все же получил практические применения. Разумеется, он не устраняет трудности, когда скорость потока приближается к скорости звука или превосходит ее, так как в этом случае для сохранения аэродинамического подобия требуется равенство не только чисел Рейнольдса, но и чисел Маха.

ЗАДАЧИ

1. Модель корабля $l_1 = 5$ м приводится в движение мотором с мощностью $P_1 = 5$ л. с. со скоростью $v_1 = 15$ км/ч. Какой мощности P требуется мотор для приведения в движение корабля длиной $l = 80$ м, геометрически подобного модели, если его движение гидродинамически подобно движению модели? Определить скорость корабля v при таких условиях.

Р е ш е н и е. Кинематическая вязкость воды $\nu = 0{,}010$ см^2/с. Вычисляя числа Рейнольдса и Фруда для модели, получаем

$$\mathrm{Re} = \frac{l_1 v_1}{\nu_1} = 2{,}1 \cdot 10^7, \qquad \mathrm{F} = \frac{v_1^2}{g l_1} = 0{,}022.$$

Определяющую роль играет число Фруда, влияние числа Рейнольдса не очень существенно. Из равенства чисел Фруда получаем $v = v_1 (l/l_1)^{1/2} = 60$ км/ч. Далее из соображений размерности находим

$$P = \rho v^2 l^{5/2} g^{1/2} f(\mathrm{Re}, \mathrm{F}) = \rho \mathrm{F} l^{7/2} g^{3/2} f(\mathrm{Re}, \mathrm{F}).$$

Отсюда, если пренебречь влиянием числа Рейнольдса

$$P = P_1 (l/l_1)^{7/2} \approx 80\,000 \text{ л. с.}$$

2. Во сколько раз следует изменить угловую скорость вращения вертикального винта вертолета и мощность его двигателя, чтобы подъемная сила осталась неизменной при замене винта и самого корпуса вертолета геометрически подобными им, но с линейными размерами, увеличенными в α раз?

Р е ш е н и е. Из соображений размерности следует, что подъемная сила должна выражаться формулой

$$F = \rho l^4 \omega^2 f_1 (l^2 \omega \rho / \eta),$$

а мощность — формулой

$$P = \rho l^5 \omega^3 f_2 (l^2 \omega \rho / \eta).$$

Поскольку плотность воздуха и его вязкость в обоих случаях одинаковы, подъемная сила не изменится, если не изменятся значения функции f_1 и коэффициента при ней. Условием этого является $l_1^2 \omega_1 = l_2^2 \omega_2$, откуда

$$\frac{\omega_2}{\omega_1} = \left(\frac{l_1}{l_2}\right)^2 = \frac{1}{\alpha^2}$$

и далее

$$\frac{P_2}{P_1} = \frac{l_2^5 \omega_2^3}{l_1^5 \omega_2^3} = \frac{l_2 \omega_2}{l_1 \omega_1} = \frac{l_1}{l_2} = \frac{1}{\alpha}.$$

§ 99. ТУРБУЛЕНТНОСТЬ И ГИДРОДИНАМИЧЕСКАЯ НЕУСТОЙЧИВОСТЬ

1. До сих пор при изучении движений жидкости мы имели в виду так называемые *ламинарные* (слоистые) течения жидкостей и газов. Особенностью ламинарного течения является его *регулярность*. Течение при сохранении ламинарности может изменяться лишь вследствие изменения сил, действующих на жидкость, или внешних условий, в которых она находится. Так, при ламинарном течении в прямолинейной трубе постоянного поперечного сечения частицы жидкости движутся вдоль прямолинейных траекторий, параллельных оси трубы. Однако при достаточно больших скоростях ламинарное течение оказывается *неустойчивым* и переходит в так называемое *турбулентное течение*. Турбулентное течение — это такое течение, гидродинамические характеристики которого (скорость, давление, а для газов — плотность и температура) быстро и нерегулярно изменяются во времени (флуктуируют). Частицы жидкости совершают нерегулярные, неустановившиеся движения по сложным траекториям, что

приводит к интенсивному перемешиванию между слоями движущейся жидкости. Примерами могут служить движение воды в бурном горном потоке, водопаде или за кормой быстроплывущего корабля, движение дыма, выходящего из фабричной трубы, и т. п. Такие быстрые и нерегулярные изменения происходят не из-за изменений действующих сил или внешних условий, а вследствие неустойчивости ламинарных течений при определенных условиях. Неустойчивость ламинарных течений и возникновение турбулентности — очень сложные вопросы, еще далекие от окончательного решения. Рассмотрение их далеко выходит за рамки нашего курса. Тем не менее имеет смысл привести простейший пример, когда вопрос об устойчивости ламинарного течения решается элементарно.

2. Таким примером может служить установившееся ламинарное движение жидкости между двумя вращающимися коаксиальными цилиндрами при больших числах Рейнольдса (см. § 96). При больших числах Рейнольдса вязкостью жидкости можно пренебречь, считая жидкость идеальной. Для идеальных жидкостей из-за отсутствия тангенциальных напряжений зависимость скорости от расстояния до оси вращения может быть произвольной: $v = v(r)$. Но уже ничтожной вязкости достаточно, чтобы спустя некоторое время после начала движения установилось вполне определенное распределение скоростей вдоль радиуса, а именно (96.8).

Однако для последующих рассуждений конкретизация вида функции $v = v(r)$ не обязательна. В невозмущенном потоке частицы жидкости движутся по окружностям с определенной угловой скоростью $\omega(r) = \frac{v(r)}{r}$. Рассмотрим какой-либо элемент жидкости, вращающийся по окружности радиуса r_0. На него действует центростремительная сила $F_0 = m\omega^2(r_0)r_0$, создаваемая разностью давлений окружающей жидкости. Введя момент импульса $L(r) = mr^2\omega$, запишем выражение для силы в виде $F_0 = \frac{L^2(r_0)}{mr_0^3}$. Допустим теперь, что под влиянием какого-то бесконечно малого случайного толчка рассматриваемый элемент жидкости сместился в новое положение, находящееся на расстоянии r от оси вращения. Можно предполагать, что толчок был совершен в направлении от или к оси вращения, так как если движение жидкости неустойчиво по отношению к возмущениям специального вида, то оно неустойчиво вообще. Момент силы такого толчка относительно оси вращения равен нулю. Результирующая сил давления окружающей жидкости также не дает момента, поскольку она направлена к оси вращения. Поэтому при смещении элемента момент его импульса сохранится, т. е. и в новом положении будет $L(r_0)$. Чтобы сместившийся элемент равномерно вращался по окружности радиуса r, на него должна действовать центростремительная сила $F_0' = \frac{L^2(r_0)}{mr^3}$. Между тем единственная сила, которой он подвержен, есть сила давления окружающей жидкости, а она равна $F = \frac{L^2(r)}{mr^3}$. Если эта сила не равна F_0', то элемент жидкости не удержится на новой круговой орбите, куда он попал. Он будет либо возвращаться к исходной орбите, либо удаляться от нее.

В первом случае движение жидкости устойчиво, во втором — неустойчиво. Допустим, например, что $r > r_0$. Если $F > F_0'$, т. е. $L^2(r) > L^2(r_0)$, то давление окружающей жидкости больше того, которое требуется для удержания сместившегося элемента жидкости на окружности радиуса r. Сместившийся элемент вернется на исходную окружность — движение устойчиво. Если же $F < F_0'$, т. е. $L^2(r) < L^2(r_0)$, то силы давления окружающей жидкости недостаточно, чтобы удержать элемент на окружности радиуса r. Элемент жидкости будет уходить еще дальше — движение неустойчивое. Если $r < r_0$, то, рассуждая аналогично, найдем, что при $L^2(r) < L^2(r_0)$ движение устойчиво, а при $L^2(r) > L^2(r_0)$ — неустойчиво. В обоих случаях критерий устойчивости можно выразить неравенством

$$\frac{dL^2}{dr} > 0, \tag{99.1}$$

или

$$\frac{d}{dr}(r^4\omega^2) > 0. \tag{99.2}$$

3. Таким образом, для устойчивости необходимо, чтобы величина $r^4\omega^2$ монотонно возрастала при удалении от оси вращения. Если цилиндры вращаются в противоположные стороны, то это невозможно. Действительно, в этом случае на поверхностях цилиндров угловая скорость ω имеет противоположные знаки. Так как ω — непрерывная функция r, то она должна обращаться в нуль в какой-то промежуточной точке. В этой точке величина $r^4\omega^2$ равна нулю, т. е. достигает минимума. По разные стороны от нее производная $\frac{d}{dr}(r^4\omega^2)$ имеет противоположные знаки, т. е. условие (99.2) не может выполняться. Значит, если цилиндры вращаются в противоположные стороны, то движение жидкости неустойчиво. Оно будет неустойчивым и в том случае, когда внутренний цилиндр вращается, а наружный покоится. Действительно, на поверхности наружного цилиндра $r^4\omega^2 = 0$, а на поверхности внутреннего $r^2\omega^2 > 0$. Поэтому с увеличением r величина $r^4\omega^2$ не может монотонно возрастать, и движение неустойчиво. Если же вращается наружный цилиндр, а внутренний покоится, то установившееся вращение жидкости будет устойчивым. В этом случае с удалением от оси вращения угловая скорость ω возрастает, а потому тем более будет возрастать $r^4\omega^2$. Теперь становится понятным, почему при измерении вязкости по методу, описанному в конце § 96, должен вращаться наружный, а не внутренний цилиндр. В противном случае вращение жидкости между цилиндрами было бы неустойчивым.

4. Приведенное исследование было выполнено без учета вязкости жидкости. Силы вязкости, уменьшая кинетическую энергию жидкости, всегда препятствуют развитию неустойчивостей. Область неустойчивости ламинарного течения сужается. Ограничимся этим общим замечанием о роли сил вязкости, так как нашей целью было только показать на простейшем примере, что ламинарное течение жидкости не всегда устойчиво.

5. При возрастании скорости течения ламинарное движение переходит в турбулентное. Скорость, при которой это происходит, называется *критической*. Вместо скорости лучше пользоваться безразмерной величиной — числом Рейнольдса. Действительно, соображения о подобии, изложенные в предыдущем параграфе, относятся к турбулентным течениям, а также к переходу от ламинарного режима течения к турбулентному. Поэтому *в геометрически по-*

добных системах переход от ламинарного режима течения к турбулентному должен происходить при одних и тех же значениях числа Рейнольдса. Этот закон был установлен Рейнольдсом из соображений теории размерности. Граничное значение числа Рейнольдса, при котором ламинарный режим течения сменяется турбулентным, называется *критическим числом Рейнольдса* и обозначается как $\mathsf{Re}_{кр}$. Значение $\mathsf{Re}_{кр}$ зависит от конфигурации тел, обтекаемых жидкостью, а также от степени возмущенности самого ламинарного течения. Так, при течении по прямолинейной трубе с круглым сечением $\mathsf{Re}_{кр} = \bar{v}a/\nu \approx 1100$, если трубы непосредственно присоединены к водопроводу и не приняты специальные предосторожности для уменьшения возмущенности воды у края трубы (a — радиус трубы, $\bar{v}$ — средняя скорость течения). Начальную возмущенность можно уменьшить, применяя трубы с гладкими стенками и закругленными краями. Кроме того, следует присоединять их к большому баку с водой и подождать, пока вода в нем не успокоится. Таким путем удается добиться затягивания ламинарного режима в трубах до значительно больших $\mathsf{Re}_{кр}$, например до $\mathsf{Re}_{кр} \approx 25\,000$.

6. Законы Пуазейля, как уже указывалось, относятся только к ламинарным течениям жидкости по трубе. Предположение о ламинарности было явно использовано при выводе формул (97.4) и (97.16). Но не столь очевидно, где используется это предположение при выводе формулы Пуазейля (97.11) методом теории размерности. Разберем этот вопрос, а также выясним, какой формулой должна быть заменена формула Пуазейля при турбулентном течении. При турбулентном течении частицы жидкости движутся с *ускорениями*, а потому существенную роль должна играть *плотность жидкости* ρ. Она не обязательно должна входить в комбинации Q/ρ, как было при ламинарном течении. Напротив, величины Q и ρ могут входить независимо. Функциональная связь должна существовать между пятью величинами

$$Q,\ \rho,\ \frac{P_1-P_2}{l},\ S,\ \eta,$$

а не между четырьмя, как было при ламинарном течении. Из этих пяти величин можно составить две независимые безразмерные комбинации, например

$$\frac{Q}{\rho}\,\frac{l}{P_1-P_2}\,\frac{\eta}{S^2} \quad \text{и} \quad \mathsf{Re} \equiv \frac{\bar{v}a}{\nu},$$

где $\bar{v}$ — средняя скорость течения, определяемая соотношением $Q = \rho\bar{v}S$, a — радиус трубы, $\nu = \eta/\rho$ — кинематическая вязкость. Согласно правилу размерности одна из этих безразмерных комбинаций является функцией другой. Это приводит к соотношению

$$Q = C(\mathsf{Re})\,\frac{P_1-P_2}{l\eta}\,\rho S^2. \tag{99.3}$$

При ламинарном течении коэффициент C есть постоянная, зависящая только от формы поперечного сечения трубы. При турбулентном течении этот коэффициент становится функцией числа Рейнольдса. Формулу (99.3) нетрудно преобразовать к виду

$$\frac{P_1-P_2}{l} = \frac{\lambda(\mathsf{Re})}{a}\,\frac{\rho\bar{v}^2}{2}, \tag{99.4}$$

в каком ее обычно пишут в гидравлике. Коэффициент λ связан с C соотношением

$$\lambda(\mathsf{Re}) = \frac{2}{\pi C(\mathsf{Re})\,\mathsf{Re}}.$$

Он называется *коэффициентом сопротивления трубы*. При ламинарном течении коэффициент сопротивления обратно пропорционален числу Рейнольдса. При турбулентном течении вид функции $\lambda(\mathrm{Re})$ устанавливается эмпирически.

По поводу приведенного вывода формул (99.3) и (99.4) необходимо сделать следующее замечание. Турбулентное течение есть нестационарное течение. На регулярное движение накладываются нерегулярные колебания и вращения — *пульсации*, которым свойственны определенные периоды во времени. Таким образом, речь идет о нестационарном движении с определенным *характерным временем* и даже несколькими *характерными временами*. Поэтому, казалось бы, в формулах (99.3) и (99.4) коэффициенты C и λ должны были бы зависеть не только от числа Рейнольдса, но и от чисел Струхаля. Однако при установившейся турбулентности число Струхаля само является функцией числа Рейнольдса, а потому нет никакого смысла вводить его в формулы (99.3) и (99.4).

ЗАДАЧА

Так как в идеальной жидкости при любых движениях не могут возникать касательные силы, то возможны *разрывные течения*, в которых касательные составляющие скорости жидкости претерпевают разрыв на некоторой поверхности (неподвижной или движущейся). Такие течения называются *тангенциальными разрывами*. Показать, что тангенциальные разрывы в несжимаемой жидкости гидродинамически неустойчивы.

Р е ш е н и е. Понятно, что давление по разные стороны от поверхности разрыва должно быть одинаково. При стационарном течении поверхность тангенциального разрыва неподвижна в пространстве. Поэтому на ней лежат линии тока. Пусть AB — одна из них (рис. 263 *а*). Допустим, что в результате какого-то бесконечно малого возмущения на линии тока AB возник бугор (рис. 263 *б*). Тогда со стороны I расстояния между линиями тока уменьшатся, а скорость жидкости увеличится. Напротив, со стороны II расстояния между линиями тока будут больше, и скорость жидкости уменьшится. Согласно закону Бернулли давление со стороны II возрастет, а со стороны I упадет. Под влиянием возросшей разности давлений бугор будет увеличиваться еще больше, т. е. движение является гидродинамически неустойчивым. Такой неустойчивостью объясняется развевание флагов на ветру.

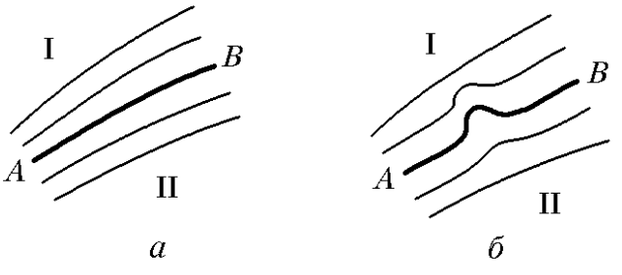


Рис. 263

§ 100. ПАРАДОКС ДАЛАМБЕРА. РАЗРЫВНЫЕ ТЕЧЕНИЯ

1. Оставшиеся параграфы этой главы будут посвящены силовым действиям потока жидкости на находящиеся в ней тела. Ввиду относительности движения эта проблема эквивалентна проблеме нахождения сил, действующих на тела, движущиеся в неподвижной

жидкости. Проблема эта очень обширна и сложна. Во всем объеме она разбирается в специальных курсах гидродинамики и аэродинамики. В общем курсе физики на ней можно остановиться очень кратко, ограничиваясь в основном качественным рассмотрением.

Силу, действующую на тело со стороны потока жидкости, можно разложить на две составляющие: в направлении потока $\mathbf{F}_x$ и перпендикулярную к потоку $\mathbf{F}_y$. Сила $\mathbf{F}_x$ называется *лобовым сопротивлением*, сила $\mathbf{F}_y$ — *подъемной силой*. Подъемная сила действует на крылья летящего самолета. С ней связано представление о силе, направленной вверх. Но подъемная сила может быть направлена и вниз в зависимости от ориентации самолета относительно направления полета. Лобовое сопротивление $\mathbf{F}_x$ слагается из двух различных сил: силы разности давлений на переднюю и заднюю поверхности тела и из вязких сил трения. При больших скоростях (точнее, при больших числах Рейнольдса) преобладающую роль играют разности давлений, при малых — силы вязкости.

2. Рассмотрим прежде всего стационарное течение идеальной несжимаемой жидкости. Допустим, что в отсутствие внешних тел жидкость течет параллельным потоком. Поместим в него какое-либо тело K (рис. 264). Оно исказит поток. Но на достаточно больших расстояниях от тела K (в «бесконечности») поток останется параллельным. По истечении некоторого времени движение жидкости установится. К этому установившемуся течению и относятся последующие рассуждения. Для конкретности будем считать, что жидкость течет в прямолинейной трубе. Вдали от тела K линии тока параллельны стенкам трубы и вследствие несжимаемости жидкости скорость ее в этих участках трубы одна и та же. А в силу уравнения Бернулли будет одинаково и давление P. Рассмотрим часть жидкости $ABDC$, внутри которой находится тело K. Предполагается, что сечения AB и CD находятся далеко от тела K, так что через них жидкость течет параллельным потоком. Спустя короткое время выделенная часть жидкости перейдет в положение $A'B'D'C'$. При этом ее импульс останется без изменения. Действительно, в начальном положении импульс жидкости представляется суммой

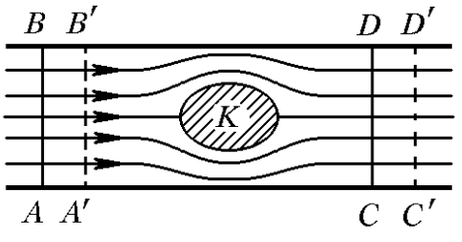


Рис. 264

$$\mathbf{I}_1 = \text{импульс жидкости в объеме } A'B'DC + \\ + \text{импульс жидкости в объеме } ABB'A',$$

а в конечном положении:

$$\mathbf{I}_2 = \text{импульс жидкости в объеме } A'B'DC + \\ + \text{импульс жидкости в объеме } CDD'C'.$$

Но в силу стационарности течения импульс жидкости в объеме $A'B'DC$ один и тот же в обоих случаях. А вследствие одинаковости скорости течения на «бесконечности» импульсы жидкости в объемах $ABB'A'$ и $CDD'C'$ также одинаковы. Итак, при обтекании тела K импульс жидкости не изменяется. Следовательно, полная сила, действующая на рассматриваемый объем жидкости в направлении потока, равна нулю. Но эта сила слагается из сил давления на основаниях AB и CD из силы $\mathbf{F}'_x$, с которой действует на жидкость тело K. (Давление стенок можно не принимать во внимание, так как оно не дает слагающей в направлении потока.) Силы давления на основаниях AB и CD уравновешивают друг друга, а потому $\mathbf{F}'_x = 0$. Следовательно, обращается в нуль и лобовое сопротивление $\mathbf{F}_x$.

Допустим теперь, что труба берется все шире и шире. Наш вывод остается справедливым для сколь угодно широкой трубы. Он остается верным и в пределе, когда трубы совсем нет, а поток во всех поперечных направлениях простирается до бесконечности. Итак, при стационарном течении идеальной несжимаемой жидкости или при равномерном движении тела в ней лобовое сопротивление равно нулю. Этот вывод в свое время казался неожиданным. Он получил название *парадокса Даламбера* (1717—1783). Наличие этого парадокса указывает на то, что при определении лобового сопротивления, испытываемого телом при равномерном движении в жидкости, последнюю нельзя рассматривать как идеальную.

3. Наш вывод относится только к лобовому сопротивлению $\mathbf{F}_x$, но не к подъемной силе $\mathbf{F}_y$, и моменту сил $\mathbf{M}$, с которым поток жидкости действует на тело. Момент $\mathbf{M}$ относительно центра масс равен нулю в тех случаях, когда тело симметрично и симметрично расположено относительно потока. Если такое условие не выполнено, то это, вообще говоря, не так. При обтекании тела происходит смещение всего потока жидкости вбок, т. е. в направлении, перпендикулярном к направлению невозмущенного потока. Это вызывает изменение момента импульса жидкости и ведет к появлению момента сил $\mathbf{M}$, действующего на тело. В результате момент $\mathbf{M}$ поворачивает тело, пока он не обратится в нуль и течение жидкости в окрестности тела K вновь станет стационарным. Что касается подъемной силы $\mathbf{F}_y$, то к этому вопросу вернемся в § 103.

4. Если тело движется неравномерно, то парадокс Даламбера не возникает. Дело в том, что с движущимся телом всегда связана какая-то масса жидкости, увлекаемая им. Она называется *присоединенной массой*. При ускорении тела ускоряется и присоединенная масса жидкости. Поэтому для сообщения ускорения телу в жидкости требуется большая сила, чем для сообщения такого же ускорения при отсутствии жидкости. Это и значит, что жидкость оказывает сопротивление телу, движущемуся в ней ускоренно.

5. Парадокс Даламбера легко уяснить, если рассмотреть картину линий тока. На схематическом рис. 265 изображены линии тока при

стационарном обтекании цилиндра или шара идеальной жидкостью. Линии тока *совершенно симметричны* по отношению к направлению вперед и назад (*зеркальная симметрия*). А скорости частиц жидкости в соответствующих точках перед и за телом равны по модулю и отличаются только направлением. Но в уравнение Бернулли (94.4) скорость v входит в квадрате. Поэтому распределения давления в потоке перед и за телом *совершенно одинаковы*. Давление на переднюю поверхность тела уравновешивается давлением на заднюю поверхность, а следовательно, лобовое сопротивление равно нулю.

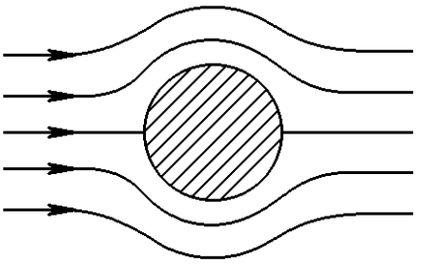


Рис. 265

Если тело, а следовательно, и поток жидкости не обладают симметрией, то рассуждение осложняется. Однако и в этом случае ввиду отсутствия потерь энергии стационарное течение идеальной жидкости обладает следующим свойством. Если в некоторый момент времени изменить на противоположные направления движения всех частиц жидкости, то они будут двигаться по тем же линиям тока с теми же по модулю, но противоположными по направлению скоростями. Так как в уравнение Бернулли скорость течения входит в квадрате, то при таком обращении направления течения распределение давления в жидкости не изменится. Не изменится также модуль и направление силы $\mathbf{F}$, с которой жидкость действует на обтекаемое тело. В частности, не меняется лобовое сопротивление $\mathbf{F}_x$. С другой стороны, опыт показывает, что сила $\mathbf{F}_x$ всегда направлена по течению *), а потому при обращении течения сила $\mathbf{F}_x$ должна изменить знак. Отсюда непосредственно следует, что $\mathbf{F}_x = 0$. К подъемной силе эти соображения неприменимы, так как нет оснований утверждать, что при обращении направления потока должна менять направление и подъемная сила.

6. Во всем изложенном предполагалось, что поток жидкости является *непрерывным*. Однако уравнения гидродинамики допускают и такие стационарные течения, в которых скорость жидкости претерпевает разрыв непрерывности. На эту возможность обратил внимание немецкий физик Густав Кирхгоф (1824—1887). Представим себе, что к телу K прикреплена бесконечно тонкая эластичная перегородка $MCDN$ (рис. 266). Пусть пространство $MCDN$, ограниченное этой перегородкой, заполнено неподвижной жидкостью, находящейся под постоянным давлением P_0. Пусть эту систему тел обтекает идеальная несжимаемая жидкость. Тогда при стационарном течении граница $MCDN$ будет вести себя как поверхность твердого тела, и часть линий тока расположится вдоль этой поверхности. Ширина бесконечно тонких трубок тока в окрестности поверхности

*) Теоретически это не обязательно (см. сноску на с. 524). Сила F_x могла бы быть направлена и против течения. Однако эта возможность представляет чисто умозрительный интерес.

$MCDN$ будет изменяться по такому закону, чтобы обеспечить постоянство скорости жидкости вдоль всей поверхности $MCDN$. Тогда, согласно уравнению Бернулли, будет постоянно и давление жидкости на этой поверхности. Если убрать эластичную перегородку, то характер течения жидкости не изменится. Действительно, поверхность $MCDN$ останется поверхностью постоянного нормального давления, а тангенциальные силы появиться не могут из-за идеальности жидкости. Получилось стационарное течение жидкости с тангенциальным разрывом на поверхности $MCDN$ (см. задачу к § 98). Оно характеризуется тем, что на некоторой линии обтекаемого тела происходит *отрыв течения от тела*. Таких течений, очевидно, можно представить бесконечное множество. Они отличаются друг от друга положением *линии отрыва* CD и формой поверхности тангенциального разрыва $MCDN$. Давление в *области застоя* (т. е. области, где жидкость покоится) P_0, очевидно, равно давлению на линии отрыва CD. Последнее же меньше давления в критической точке B. Это приводит к тому, что равнодействующая сил давления, действующих на переднюю поверхность тела, превышает соответствующую силу, действующую на заднюю сторону его. В результате появляется лобовое сопротивление $\mathbf{F}_x$*).

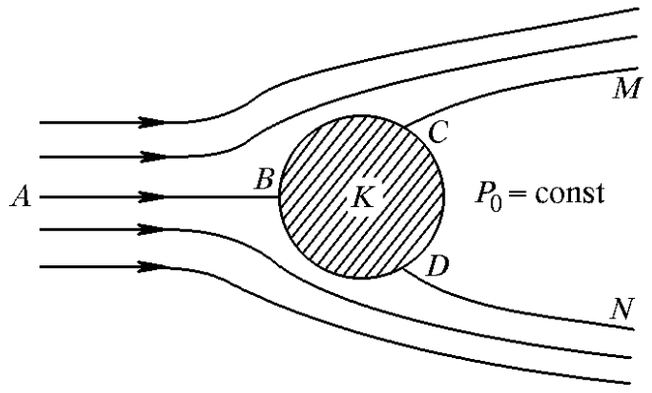


Рис. 266

7. Тангенциальные разрывы гидродинамически неустойчивы (см. задачу к § 98). Поверхности разрыва распадаются в *вихри*. Тем не менее идеальные разрывные течения с отрывом от обтекаемого тела, рассмотренные выше, не совсем лишены интереса. Они в известном смысле могут рассматриваться как предельные случаи реальных течений вязкой жидкости. Силы вязкости мало существенны вдали от обтекаемого тела, где они малы. Их влияние проявляется главным образом в тонком *пограничном слое* вблизи поверхности тела, где они велики. Они приводят к отрыву течения от тела. При этом вместо области застоя за телом возникает область интенсивного турбулентного движения. Наличие такой области и ведет к возникновению лобового сопротивления. При этом силы вязкости автоматически устраняют неоднозначность в положении линии отрыва, характерную для разрывных течений идеальных жидкостей. Чем у́же область отрыва, тем меньше лобовое сопротивление. С целью

*) Если обратить направление течения, то модуль и направление силы $\mathbf{F}_x$ не изменятся. В обращенном течении сила $\mathbf{F}_x$ направлена против течения, т. е. «лобовое сопротивление» отрицательно. Это случай, который имелся в виду в сноске на с. 523 как теоретически возможный.

уменьшения лобового сопротивления самолетам, судам, автомобилям и прочим быстроходным самодвижущимся аппаратам придают «обтекаемую форму».

§ 101. ПРИМЕНЕНИЕ ТЕОРИИ РАЗМЕРНОСТИ

1. Отвлечемся на время от механизма возникновения силы $\mathbf{F}$, с которой стационарный поток несжимаемой жидкости действует на неподвижное тело, и применим к этой проблеме методы теории размерности. Сила $\mathbf{F}$ зависит от формы и размеров тела, от ориентации его по отношению к потоку, от скорости потока $\mathbf{v}$ (на «бесконечности»), а также от свойств жидкости. Ориентацию крыла самолета принято характеризовать *углом атаки*, т. е. углом между плоскостью крыла и направлением полета. Не будем вводить явно такие параметры, предполагая, что мы имеем дело с телами не только геометрически подобными, но и подобно расположенными. Свойства жидкости характеризуются ее плотностью ρ и вязкостью η. Таким образом, должна существовать функциональная связь между величинами $\mathbf{F}$, v, ρ, η, S, где S — характерная площадь поперечного сечения тела. Корень квадратный из нее $\sqrt{S} = l$ может служить характерным линейным размером тела. Из этих пяти величин можно составить две независимые безразмерные комбинации. За таковые можно принять $\frac{\mathbf{F}}{\rho v^2 S}$ и число Рейнольдса $\mathrm{Re} = \frac{\rho l v}{\eta}$. Согласно правилу размерности одна из этих комбинаций является функцией другой. В результате получим

$$\mathbf{F} = \frac{\rho v^2}{2} SC(\mathrm{Re}), \tag{101.1}$$

или

$$F_x = \frac{\rho v^2}{2} SC_x(\mathrm{Re}), \tag{101.2}$$

$$F_y = \frac{\rho v^2}{2} SC_y(\mathrm{Re}). \tag{101.3}$$

Безразмерные коэффициенты $C_x(\mathrm{Re})$ и $C_y(\mathrm{Re})$ называются соответственно коэффициентами *лобового сопротивления* и *подъемной силы*. Оба они являются функциями числа Рейнольдса и зависят от формы тела и его ориентации по отношению к потоку. Теоретическое вычисление этих коэффициентов затруднительно, они обычно определяются опытным путем.

2. При больших числах Рейнольдса лобовое сопротивление F_x обусловлено почти исключительно разностью давлений. Если обтекаемое тело имеет сзади заостренные края, то отрыв течения за телом происходит в одном и том же месте, положение которого не зависит от скорости потока. (Примером может служить пластинка,

поставленная перпендикулярно к направлению потока. Отрыв течения происходит на ее краях.) В этих случаях коэффициент лобового сопротивления приблизительно постоянен, а само лобовое сопротивление пропорционально квадрату скорости v. Понять это проще всего, если воспользоваться идеализированной картиной разрывного течения (рис. 265). Действительно, если при всех скоростях отрыв течения происходит в одном и том же месте, то характерная площадь поперечного сечения S не зависит от скорости. С другой стороны, разность давлений перед и за телом по закону Бернулли равна $\frac{1}{2}\rho v^2$. Отсюда и получается формула (101.2) с постоянным коэффициентом C_x. При больших скоростях v порядка скорости звука и выше коэффициенты C_x и C_y зависят не только от числа Рейнольдса Re, но и от числа Маха M.

3. Рассмотрим теперь случай малых чисел Рейнольдса. В этом случае основной интерес представляет сила лобового сопротивления F_x. Инерция, а с ней и плотность жидкости не играют существенной роли, сила F_x определяется почти исключительно вязкостью. Поэтому плотность ρ должна выпадать из формулы (101.2). Это будет тогда и только тогда, когда коэффициент лобового сопротивления обратно пропорционален числу Рейнольдса, т. е.

$$C_x = \frac{A}{\mathrm{Re}},$$

где A — безразмерная постоянная. Подставляя выражение для Re, получим

$$F_x = A\eta l v. \tag{101.4}$$

Эта формула справедлива при малых числах Рейнольдса ($\mathrm{Re} \ll 1$), так как она выведена в предположении, что влияние инерции жидкости пренебрежимо мало по сравнению с влиянием вязкости. Коэффициент A зависит от формы тела и его ориентации относительно потока. Его теоретическое вычисление довольно кропотливо и требует интегрирования уравнений движения вязкой жидкости. Джорджем Стоксом (1819—1903) было показано, что $A = 6\pi$, если за характерный размер l принять радиус шара a. Таким образом, получается *формула Стокса*

$$F_x = 6\pi\eta a v. \tag{101.5}$$

Так как формула (101.5) получила широкие применения в очень важных физических опытах (определение заряда электрона методом Милликена, броуновское движение и пр.), то имеет смысл более подробно выяснить на конкретных примерах границы ее применимости.

В опытах американского физика-экспериментатора Роберта Милликена (1868—1953) по определению заряда электрона формула Стокса (101.5) применялась к капелькам масла, падавшим в воздухе под действием силы тяжести. Если m — масса капли, то при установившемся равномерном па-

дении вес капли mg должен уравновешиваться силой вязкости $6\pi\eta av$, а потому $mg = 6\pi\eta av$ (архимедовой подъемной силой пренебрегаем). Если ρ_0 — плотность масла, то масса капли $m = 4/3\ \pi a^3\rho_0$. Используя это значение, находим сначала скорость капли v, а затем и число Рейнольдса

$$\mathrm{Re} = \frac{\rho a v}{\eta} = \frac{2}{9}\frac{a^3\rho\rho_0 g}{\eta^2},$$

где ρ — плотность воздуха. Условие применимости формулы Стокса $\mathrm{Re} \ll 1$ дает

$$a^3 \ll \frac{9}{2}\frac{\eta^2}{\rho\rho_0 g}.$$

Подставляя сюда $\eta = 1{,}8\cdot 10^{-4}$ г/(с·см), $\rho = 1{,}29\cdot 10^{-3}$ г/см^3, $\rho_0 = 0{,}9$ г/см^3, найдем, что для применимости формулы Стокса должно выполняться условие $a \ll 0{,}05$ мм. Формулу можно применять для мельчайших капелек тумана. Однако о применении ее к каплям дождя, даже самым мелким, не может быть и речи.

В качестве второго примера возьмем капельки ртути, падающие в жидкости под действием собственного веса. По скорости установившегося падения капли можно вычислить вязкость жидкости. Это дает практический метод измерения вязкости. В рассматриваемом случае надо учитывать архимедову выталкивающую силу. Если ρ_0 — плотность ртути, ρ и η — плотность и вязкость исследуемой жидкости, то для применимости формулы Стокса необходимо выполнение условия

$$a^3 \ll \frac{9}{2}\frac{\eta^2}{(\rho_0-\rho)\rho g}.$$

Для воды $\eta = 0{,}010$ г/(с·см), и мы получаем $a \ll 0{,}15$ мм.

§ 102. ПОТЕНЦИАЛЬНЫЕ И ВИХРЕВЫЕ ДВИЖЕНИЯ

1. Все движения жидкостей подразделяются на *потенциальные* и *вихревые*. Рассмотрим поле скоростей жидкости $\mathbf{v}(\mathbf{r})$ в какой-то фиксированный момент времени. Возьмем в жидкости произвольный замкнутый контур C и на нем установим положительное направление обхода (рис. 267). Пусть $\boldsymbol{\tau}$ — единичный вектор касательной, а $d\mathbf{s}$ — элемент длины контура, проведенные в положительном направлении. Интеграл

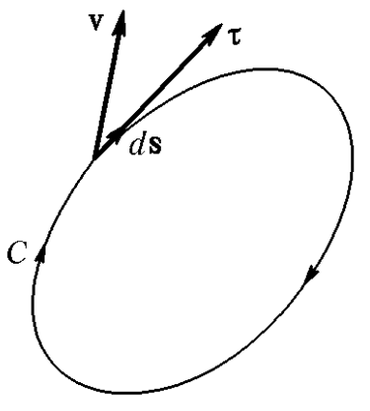


Рис. 267

$$\Gamma = \oint_C v_\tau\, ds = \oint_C (\mathbf{v}\, d\mathbf{s}) \qquad (102.1)$$

называется *циркуляцией вектора скорости* по контуру C. Если циркуляция скорости по любому замкнутому контуру обращается в нуль, то движение жидкости называется потенциальным. В противном случае движение называется *вихревым*.

При этом предполагается, что область пространства, в которой течет жидкость, *односвязна*. Это значит, что любой замкнутый контур в такой области непрерывной деформацией может быть стянут в точку, не пересекая при этом обтекаемые тела. Если же область не односвязна (например, жидкость, обтекающая тор), то приведенные определения необходимо дополнить следующими замечаниями. В качестве C следует брать не все контуры, а только произвольные замкнутые контуры, которые непрерывной деформацией могут быть стянуты в точку, не выходя при этом за границы жидкости. Важным случаем может служить так называемое *плоское течение*, являющееся идеализацией действительных течений. Пусть обтекаемое тело является бесконечно длинным цилиндром с произвольным поперечным сечением, а жидкость течет перпендикулярно к оси этого цилиндра. Тогда достаточно ограничиться рассмотрением течения в одной из плоскостей, перпендикулярных к той же оси. Течение в этой плоскости и называется плоским. Оно будет потенциальным, если циркуляция скорости обращается в нуль по любому замкнутому контуру, *не охватывающему обтекаемый цилиндр*, например контур C_1 (рис. 268). Но циркуляция по контуру C, окружающему цилиндр, может и не обращаться в нуль. Нетрудно показать, что при потенциальном течении циркуляция Γ будет одной и той же для всех замкнутых контуров, обходящих вокруг цилиндра один раз. Если $\Gamma \neq 0$, то говорят о *потенциальном течении с циркуляцией*.

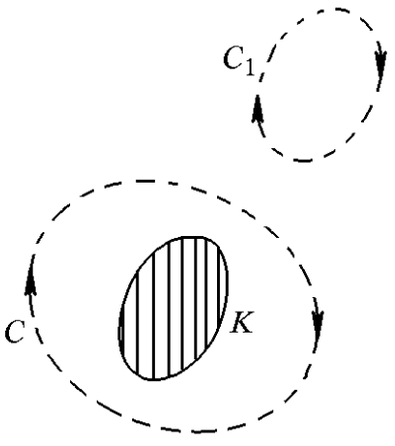


Рис. 268

2. Определение потенциального течения совершенно аналогично определению консервативных сил (см. § 24). Поэтому при потенциальном течении линейный интеграл $\int\limits_{AB} (\mathbf{v}d\mathbf{s})$, взятый вдоль незамкнутой кривой, соединяющей точки A и B, зависит только от положения крайних точек этой кривой A и B, но не зависит от формы самой кривой AB. Рассуждая так же, как в случае потенциальной энергии, можно ввести функцию координат φ, через которую скорость $\mathbf{v}$ выражается формулой

$$\mathbf{v} = \operatorname{grad} \varphi \tag{102.2}$$

(см. § 29). Функция φ называется *потенциалом скоростей*.

Примером потенциального течения может служить течение жидкости вдоль параллельных прямых линий с постоянной скоростью. Можно показать, что *всякое течение идеальной жидкости, возник-*

шее из состояния покоя под действием консервативных сил, является потенциальным.

3. Примером вихревого движения может служить плоское течение жидкости, когда частицы последней вращаются по концентрическим окружностям с одной и той же угловой скоростью ω (рис. 269). Циркуляция скорости по окружности радиуса r в этом случае равна $\Gamma = 2\pi r v = 2\pi r^2 \omega$, ее отношение к площади контура πr^2 будет $\Gamma/(\pi r^2) = 2\omega$, т. е. не зависит от радиуса r. Если угловая скорость вращения зависит от радиуса r, то вместо отношения $\Gamma/(\pi r^2)$ берут его предел при $r \to 0$. Ясно, что этот предел равен удвоенному значению угловой скорости, с которой вращаются частицы жидкости вблизи оси O. Этот предел называется *вихрем* или *ротором* скорости $\mathbf{v}$, точнее, проекцией ротора на направление, перпендикулярное к плоскости контура. Вообще, для произвольного движения ротор скорости $\mathbf{v}$ определяется своими проекциями на произвольное направление следующим образом. Берется произвольный бесконечно малый контур с площадью ΔS и внешней нормалью $\mathbf{n}$. Проекция вектора $\operatorname{rot} \mathbf{v}$ на направление нормали $\mathbf{n}$ определяется соотношением

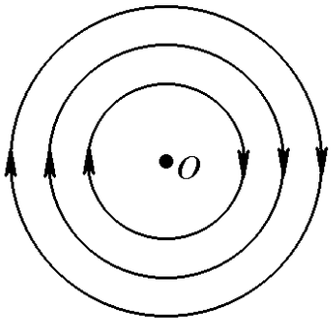


Рис. 269

$$\operatorname{rot}_n \mathbf{v} = \lim_{\Delta S \to 0} \frac{\Gamma}{\Delta S}, \tag{102.3}$$

где Γ — циркуляция вектора $\mathbf{v}$ вдоль рассматриваемого контура.

4. В качестве второго примера рассмотрим плоское течение жидкости параллельно оси X, когда скорость потока меняется в поперечном направлении по линейному закону $v_x = ay$ (рис. 270). Чтобы убедиться в вихревом характере течения, возьмем прямоугольный контур $ABCD$ со сторонами, параллельными координатным осям. Циркуляция скорости по этому контуру будет

$$\Gamma = (x_2 - x_1)\,(v_1 - v_2) = -a(x_2 - x_1)\,(y_2 - y_1).$$

Ее отношение к площади контура $\Delta S = (x_2 - x_1)\,(y_2 - y_1)$, или ротор скорости $\mathbf{v}$ будет

$$\operatorname{rot}_z \mathbf{v} = -a,$$

или

$$\operatorname{rot}_z \mathbf{v} = -\frac{\partial v_x}{\partial y}. \tag{102.4}$$

Если v_x меняется с координатой y не по линейному закону, а произвольно, то формула (102.4) остается верной, однако $\operatorname{rot}_z \mathbf{v}$ становится функцией координаты y.

Заметим еще, что в разбираемом примере скорость $\mathbf{v}$ можно представить в виде векторной суммы двух векторов $\mathbf{v}_1$ и $\mathbf{v}_2$ с компонентами

$$v_{1x} = \frac{v_x}{2} = \frac{a}{2}\,y, \quad v_{2x} = \frac{v_x}{2} = \frac{a}{2}\,y,$$

$$v_{1y} = -\frac{a}{2}\,x, \quad v_{2y} = \frac{a}{2}\,x.$$

Вектор $\mathbf{v}_1$ представляется векторным произведением

$$\mathbf{v}_1 = -\frac{a}{2}[\mathbf{kr}] = \frac{a}{2}\,y\mathbf{i} - \frac{a}{2}\,x\mathbf{j}.$$

Поэтому движение со скоростью $\mathbf{v}_1$ может быть интерпретировано как вращение вокруг оси Z с угловой скоростью $\boldsymbol{\omega} = -\frac{a}{2}\,\mathbf{k}$. Компоненты же вектора $\mathbf{v}_2$ могут быть получены из потенциала скорости $\varphi = \frac{a}{2}\,xy$ по формулам

$$v_{2x} = \frac{\partial \varphi}{\partial x}, \quad v_{2y} = \frac{\partial \varphi}{\partial y}.$$

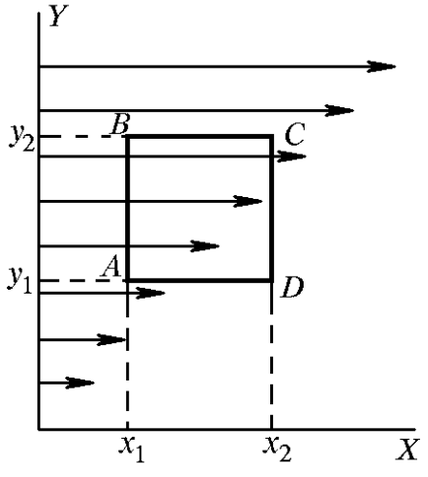


Рис. 270

Значит, движение со скоростью $\mathbf{v}_2$ является потенциальным. Можно в общем виде показать, что произвольное движение жидкости можно разложить на *вращение* и *потенциальное течение*, причем угловая скорость вращения и ее направление в пространстве могут непрерывно меняться от точки к точке.

Тангенциальный разрыв может рассматриваться как пример вихревого течения. В вихревом характере движения в этом случае можно убедиться совершенно так же, как при разборе последнего примера. Распадаясь, тангенциальный разрыв переходит в вихревое турбулентное движение.

§ 103. ПОГРАНИЧНЫЙ СЛОЙ И ЯВЛЕНИЕ ОТРЫВА

1. При больших числах Рейнольдса силы вязкости вдали от поверхности обтекаемого тела не играют существенной роли. Здесь они малы по сравнению с силами, обусловленными разностями давлений. Ими можно пренебречь и считать жидкость идеальной. Не так, однако, обстоит дело вблизи поверхности обтекаемого тела. Силы вязкого трения вызывают *прилипание жидкости* к поверхности обтекаемого тела, т. е. удерживают частицы жидкости в состоянии покоя, несмотря на наличие градиента давления в направлении по-

тока жидкости. Отсюда следует, что вблизи поверхности тела силы вязкого трения *того же порядка*, что и силы разности давлений. Чтобы это было так, скорость жидкости должна очень быстро нарастать при удалении от поверхности тела. Это быстрое нарастание происходит в тонком приповерхностном слое жидкости, называемом *пограничным слоем*. Теория пограничного слоя была создана в основном Л. Прандтлем. Дадим качественное представление о некоторых выводах этой теории.

2. Толщина пограничного слоя δ относится к числу не вполне четко определенных понятий, так как граница слоя со стороны жидкости не является резко очерченной. Толщина слоя зависит не только от свойств жидкости, но и от формы поверхности обтекаемого тела. Она не остается постоянной на поверхности тела, а возрастает в направлении потока от передней части тела к задней. Поэтому о точном выражении для δ говорить не приходится. Речь может идти только об оценке. Толщину пограничного слоя легко оценить, если заметить, что в нем силы вязкости и силы, обусловленные разностями давлений, по порядку величины одинаковы. Оценим сначала силу вязкого трения $f_{\text{тр}}$, действующую на единицу объема жидкости в пограничном слое. Градиент скорости жидкости поперек течения в пограничном слое порядка v/δ. Вязкая сила, действующая на площадку S пограничного слоя, будет $\sim \eta S v/\delta$, а сила, действующая на единицу объема,

$$f_{\text{тр}} \sim \frac{\eta S v/\delta}{S\delta} = \eta \frac{v}{\delta^2}.$$

Оценим теперь силу разности давлений $f_{\text{дав}}$, также отнесенную к единице объема жидкости. Она равна $f_{\text{дав}} = -\operatorname{grad} P$ (см. § 90). Изменения давления поперек пограничного слоя малы, да и вообще не играют роли в рассматриваемом вопросе, — нас интересует только градиент давления *в направлении потока.* Его можно оценить, рассматривая внешний поток жидкости, т. е. поток вне пограничного слоя. К этому потоку применимо уравнение Бернулли $P = P_0 - \frac{1}{2}\rho v^2$, из которого следует $\operatorname{grad} P = -(\rho/2)\operatorname{grad} v^2$. Значит, по порядку величины сила $f_{\text{дав}} \sim \rho v^2/l$, где l — характерный линейный размер обтекаемого тела. Приравнивая обе силы $f_{\text{тр}}$ и $f_{\text{дав}}$, получаем после выполнения элементарных арифметических действий

$$\delta \sim \sqrt{\frac{\eta l}{\rho v}}, \tag{103.1}$$

или

$$\delta \sim \frac{l}{\sqrt{\mathrm{Re}}}. \tag{103.2}$$

Например, для шара диаметром $D = 10$ см в потоке воздуха, движущегося со скоростью $v = 30$ м/с, число Рейнольдса равно

разность давлений должна существовать и в пограничном слое, так как разность давлений между границами слоя пренебрежимо мала (в противоположном случае частицы жидкости в пограничном слое имели бы ускорения, перпендикулярные к поверхности тела). Во внешнем потоке, обтекающем переднюю часть тела, давление падает в направлении движения жидкости. Следовательно, то же самое будет и в пограничном слое. Сила разности давлений направлена вдоль по течению. Поэтому не только во внешнем потоке, но и в пограничном слое скорости частиц жидкости увеличиваются, что позволяет им продолжать движение по поверхности тела, несмотря на действие сил трения. Не то происходит в потоке, обтекающем заднюю часть тела. Здесь давление возрастает в направлении потока. Движение замедляется как во внешнем потоке, так и в пограничном слое. А так как в пограничном слое частицы движутся медленнее, чем во внешнем потоке, то при достаточном замедлении последнего они могут остановится и даже начать движение в обратную сторону. В результате около поверхности обтекаемого тела возникнет *возвратное движение* жидкости, несмотря на то, что внешний поток продолжает по-прежнему двигаться вперед. Новые массы жидкости, подтекающие к месту возникновения возвратного течения, также сначала останавливаются, а затем начинают двигаться назад. (При недостаточно сильном замедлении внешнего потока возвратное движение пограничного слоя может и не возникнуть.) Количество заторможенной жидкости между поверхностью тела и внешним потоком быстро увеличивается, возвратное движение распространяется все шире и шире и, наконец, совершенно оттесняет внешний поток от поверхности тела. Возникает отрыв течения от обтекаемого тела. Получающаяся поверхность разрыва неустойчива и быстро свертывается в вихрь. При этом часть заторможенной жидкости оказывается вовлеченной в область вихря, а самый вихрь уносится течением.

6. Все эти стадии образования вихря хорошо видны на рис. 271 *а, б, в* на котором представлены шесть последовательных фотографий потока воды, обтекающего неподвижный цилиндр *). Для того чтобы линии тока сделать видимыми, поверхность текущей воды обсыпалась порошком алюминия. В первый момент вокруг цилиндра возникало потенциальное течение, линии тока которого, расходившиеся перед цилиндром, вновь смыкались позади него. Дальнейшие фотографии показывают, как меняется последующее течение жидкости. На последних трех фотографиях видно, что за цилиндром образуются два вихря. Сначала один из них, а затем и другой отрываются от цилиндра

*) Фотографии рисунков 271, 272, 273, 279, 280, 281 взяты из книги: Прандтль Л., Титьенс О. Гидро- и аэромеханика. Т. II — М.; Л.: ОНТИ, 1935. На рис. 271 течение направлено слева направо.

и уносятся потоком жидкости. Уносимые вихри сменяются новыми, попеременно возникающими в каждом из двух потоков, отрывающихся сверху и снизу от поверхности обтекаемого тела. Все эти вихри уносятся от тела с одинаковой скоростью. Такая система вихрей называ-

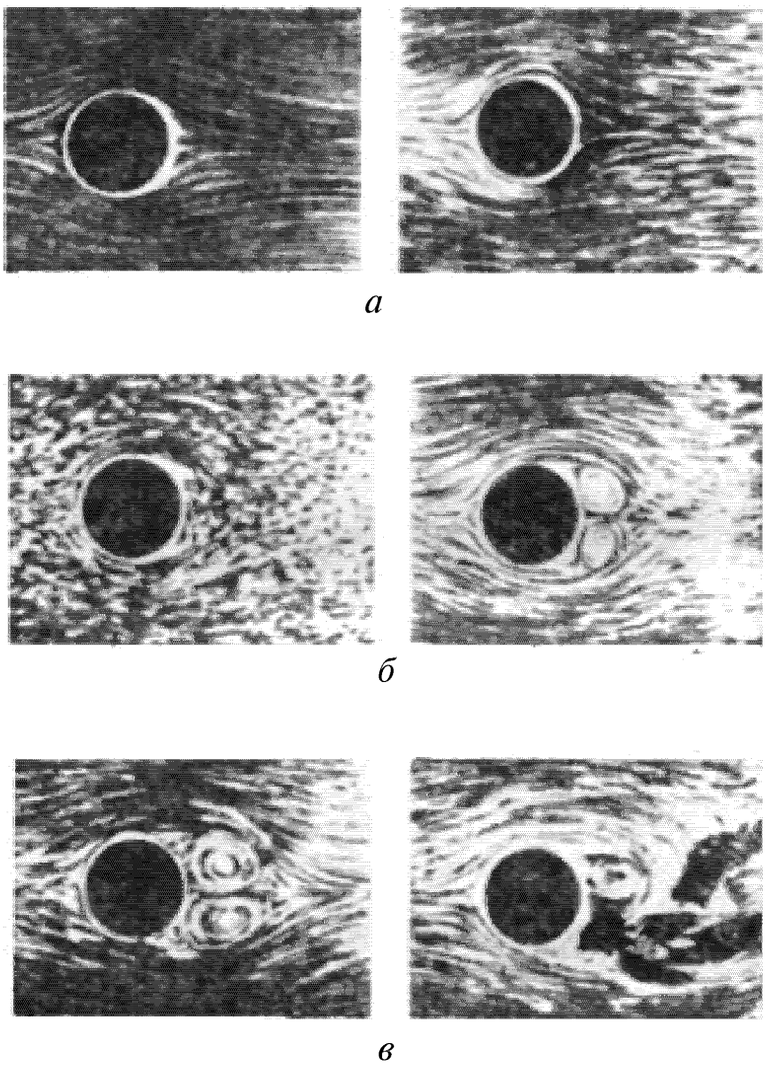


Рис. 271

ется *вихревой дорожкой Кармана* по имени ученого Кармана (1881—1963), теоретически изучившего ее. Она представлена на рис. 272 и 273. Первый рисунок представляет картину течения в системе отсчета, в которой цилиндр неподвижен, а жидкость течет слева направо; второй — в системе отсчета, в которой неподвижна невозмущенная жидкость, а цилиндр движется справа налево. Скорость, с которой уносятся вихри, меньше скорости потока, так как в вихрях собираются как раз те частицы жидкости, которые тормозились при обтекании тела. Поэтому импульс, уносимый потоком жидкости вместе с движущимися в ней вихрями, меньше импульса, который приносит поток, натекающий на тело. Это уменьшение импульса потока жидкости проявляется в возникновении силы лобового сопротивления, действующей на тело в направлении потока.

7. Из изложенного видно, что представлением о пограничном слое можно пользоваться только на передней части тела, простирающейся

до того места, в котором происходит отрыв течения от поверхности тела (это место называется *линией отрыва*). Начиная с этого места за

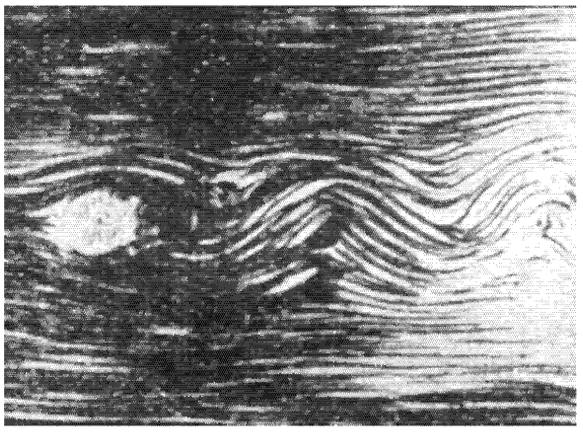


Рис. 272

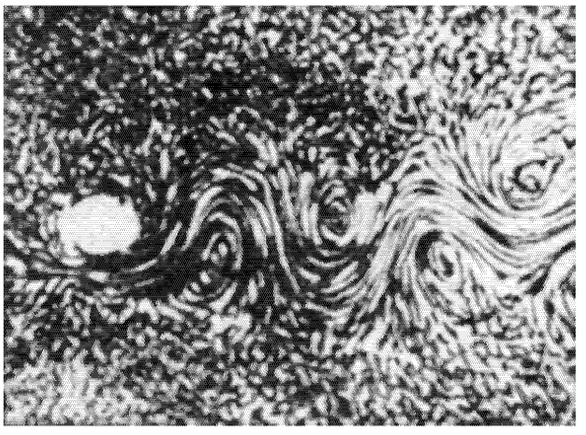


Рис. 273

телом возникает область течения, длина которой обычно намного превосходит характерные размеры самого тела (рис. 274). В эту область попадают частицы из пограничного слоя. Поэтому средняя скорость течения в ней меньше скорости набегающего потока, а само течение вихревое и, как правило, турбулентное. Эта область называется *следом*. Наличием следа и объясняется та часть лобового сопротивления, которая обусловлена разностью давлений на переднюю и заднюю части тела. Чем шире область отрыва, т. е. чем шире след, тем больше при прочих равных условиях лобовое сопротивление. Существованием следа, как мы увидим в следующем параграфе, объясняется и возникновение подъемной силы.

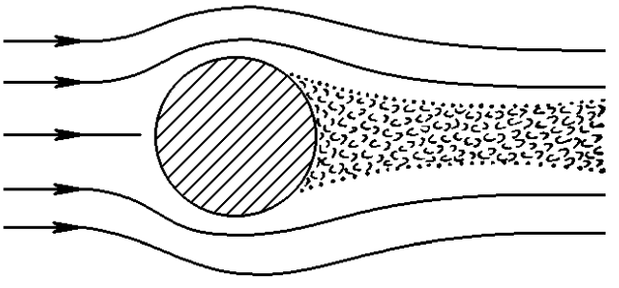


Рис. 274

8. При не очень больших числах Рейнольдса движение в пограничном слое является ламинарным. При увеличении числа Рейнольдса ламинарное течение становится неустойчивым и начинается турбулизация пограничного слоя. Она начинается с задней части пограничного слоя, примыкающей к линии отрыва, и постепенно распространяется на переднюю часть слоя. Таким образом, в передней части пограничный слой ламинарный, затем он переходит в турбулентный, а за линией отрыва образуется след. Для шара турбулизация пограничного слоя начинается при числах Рейнольдса $\sim 3 \cdot 10^5$. Следствием турбулизации пограничного слоя является смещение линии отрыва к задней части тела и связанное с ним сужение следа. В результате коэффициент лобового сопротивления C_x и даже само лобовое сопротивление F_x уменьшаются. Это явление называется *кризисом сопротивления*. Кризис сопртивления может и

не наступить, если с увеличением скорости течения линия отрыва не смещается. Тогда коэффициент сопротивления C_x становится не зависящим от числа Рейнольдса. Так будет, например, при обтекании пластинки с резкими краями, поставленной перпендикулярно к потоку. Здесь линия отрыва определяется чисто геометрическими соображениями и совпадает с краями пластинки.

§ 104. ПОДЪЕМНАЯ СИЛА КРЫЛА САМОЛЕТА

1. С явлением отрыва связано и возникновение подъемной силы. Нас будет интересовать главным образом подъемная сила, действующая на крыло самолета, хотя механизм возникновения подъемной силы в случае тел другой формы сохраняется тем же самым. При полете самолета с постоянной скоростью его ориентация в пространстве остается неизменной. Это указывает на то, что при таком полете моменты всех внешних сил, действующих на самолет, уравновешиваются, а его момент импульса остается неизменным. Для упрощения будем рассматривать отдельное крыло, равномерно движущееся в воздухе и ориентированное перпендикулярно к плоскости рисунка (рис. 275). Длину крыла будем считать бесконечно большой. Такое крыло называется *крылом бесконечного размаха.* Удобно перейти к системе отсчета, связанной с крылом, поместив начало координат в одну из точек крыла, например в его центр масс C. Понятно, что эта система отсчета будет инерциальной.

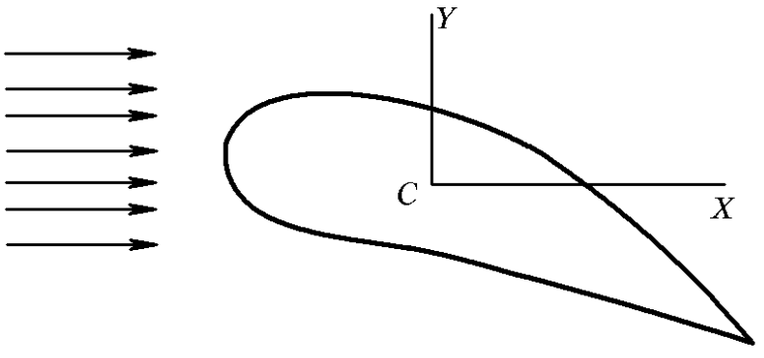


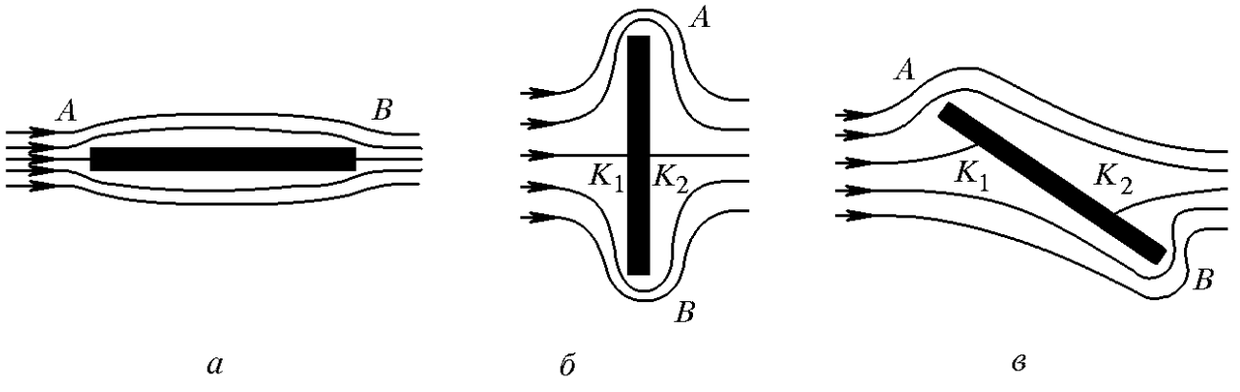
Рис. 275

Таким образом, мы предполагаем, что крыло неподвижно, а течение воздуха плоское. Невозмущенный поток, конечно, будет равномерным. Во избежание недоразумений все моменты импульса, о которых говорится ниже, будет брать относительно точки C. Момент импульса самого крыла равен нулю, и о нем в дальнейшем можно не говорить.

2. Для возникновения подъемной силы необходимо, чтобы крыло было *несимметрично* или *несимметрично расположено* относительно горизонтальной плоскости, в которой оно движется. При движении круглого невращающегося цилиндра, например, никакой подъемной силы возникнуть не может. Поэтому мы предполагаем, что указанной симметрии нет. Напомним теперь, что в пограничном слое скорости частиц воздуха возрастают при удалении от по-

верхности крыла. Благодаря этому движение в пограничном слое вихревое, а потому содержит вращение. Сверху крыла вращение совершается по часовой стрелке, а снизу — против (если поток жидкости натекает слева направо). Допустим, что в результате отрыва какая-то масса воздуха, ранее находившаяся в пограничном слое снизу от крыла, унесена потоком в виде одного или нескольких вихрей. Обладая вращением, эта масса унесет и связанный с ней момент импульса. Но общий момент импульса воздуха не может измениться. Если отрыв пограничного слоя сверху от крыла не произошел, то для сохранения момента импульса воздух во внешнем потоке должен начать вращаться вокруг крыла по часовой стрелке. Иными словами, во внешнем потоке вокруг крыла должна возникнуть циркуляция скорости воздуха по часовой стрелке, накладывающаяся на основной поток. Скорость потока под крылом уменьшится, а над ним — увеличится. К внешнему потоку применимо уравнение Бернулли. Из него следует, что в результате циркуляции давление под крылом возрастет, а над ним — уменьшится. Возникшая разность давлений проявляется в подъемной силе, направленной вверх. Наоборот, если унесенные вихри образовались из частиц пограничного слоя сверху крыла, то возникнет циркуляция против часовой стрелки, а «подъемная» сила будет направлена вниз.

3. Для уяснения явления полезно рассмотреть тонкую пластинку, поставленную на пути потока идеальной жидкости. Если пластинка ориентирована вдоль потока, то критические точки, в которых скорость жидкости обращается в нуль, находятся на краях A и B (рис. 276 *а*). Если пластинка поставлена перпендикулярно к потоку, то обе критические точки смещаются к центру пластинки, а скорость течения достигает максимума на краях пластинки A и B (рис. 276 *б*). Если же пластинка поставлена на-



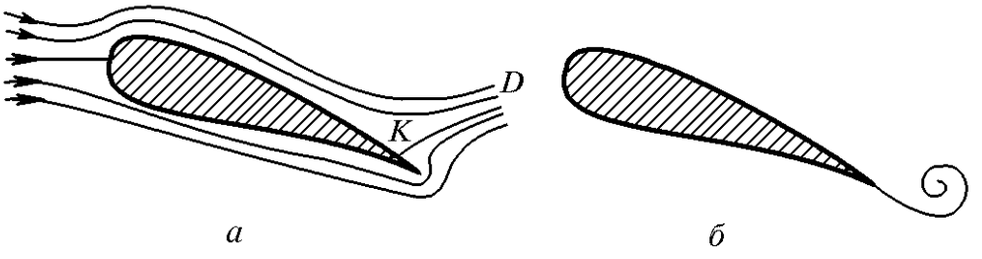
а *б* *в*

Рис. 276

клонно к потоку (рис. 276 *в*), то критические точки K_1 и K_2 занимают промежуточные положения между центром пластинки и ее краями. Скорость течения по-прежнему максимальна на краях пластинки. В окрестности критической точки K_2 она больше снизу, чем сверху, так как нижний поток расположен значительно

ближе к краю B пластинки, чем верхний к краю A. Такая же картина течения образуется в начальный момент и при течении вязкой жидкости.

4. Так и в случае крыла самолета поток воздуха под крылом в начале движения огибает заднюю кромку крыла и встречается



а *б*

Рис. 277

вдоль линии KD с потоком, огибающим крыло сверху. Здесь образуется поверхность раздела, свертывающаяся в дальнейшем в вихрь, причем вращение происходит против часовой стрелки (рис. 277 *а* и *б*). Все это видно на рис. 278, 279, 280 (фотографии), причем первые два рисунка изображают течение в системе

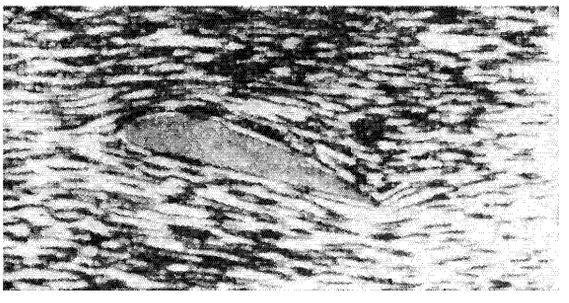


Рис. 278

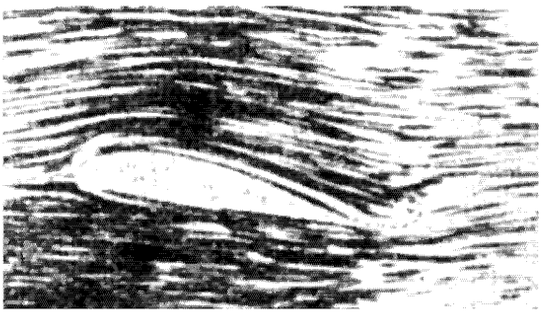


Рис. 279

отсчета, в которой неподвижно крыло, а последний — в системе отсчета, в которой неподвижна невозмущенная жидкость. Вихри уносят момент импульса, а вокруг крыла образуется циркуляция по часовой стрелке. Возрастание скорости течения над крылом и уменьшение ее под крылом приводят к смещению линии отрыва, пока она не достигнет нижней кромки крыла (рис. 281). Если бы

Рис. 280

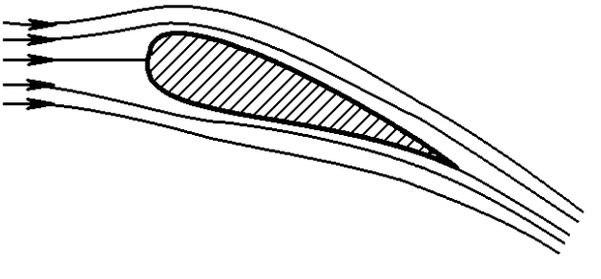


Рис. 281

не было сил вязкости, то дальнейшее образование вихрей, а с ним и циркуляции вокруг крыла прекратились бы. Силы вязкости меняют дело. Благодаря им циркуляция вокруг крыла постепенно затухает. Линия отрыва смещается от кромки крыла вверх, т. е. вновь появляются условия для возникновения вихрей. Появляющиеся вихри вновь усиливают циркуляцию и возвращают линию отрыва к кромке крыла. При постоянной скорости движения самолета описанный процесс носит регулярный характер — вихри периодически отрываются от задней кромки крыла и поддерживают практически постоянную циркуляцию.

5. Зависимость подъемной силы от циркуляции скорости была установлена независимо друг от друга Н. Е. Жуковским и Кутта. Их формула относится к крылу бесконечного размаха и дает значение подъемной силы, отнесенное к единице длины такого крыла. Формула предполагает, что крыло движется равномерно в идеальной жидкости, и вокруг него установилась циркуляция скорости постоянной величины.

Таким образом, в системе отсчета, в которой крыло неподвижно, движение жидкости потенциально, но с циркуляцией. В идеальной жидкости циркуляция практически может быть любой, никак не связанной со скоростью потока, углом атаки и прочими параметрами. Однако вязкость, хотя бы и предельно малая, приводит к однозначной зависимости циркуляции от этих параметров.

При этом сама циркуляция от вязкости практически не зависит. Поэтому формула Жуковского—Кутта дает хорошее приближение для подъемной силы крыла также и в воздухе, обладающем вязкостью.

6. Приведем простейший вывод *формулы Жуковского—Кутта*, из которого с особой отчетливостью выяснится, почему для возникновения подъемной силы существенна циркуляция. Допустим, что поток жидкости простирается во все стороны до бесконечности. Как и раньше, будем предполагать, что невозмущенный поток горизонтален; ось X направлена вдоль потока, а ось Y — вертикально вверх (перпендикулярно к нему).

Пусть крыло K помещено в начале координат (рис. 282). Поместим над крылом и под ним бесконечное множество в точности таких же крыльев, находящихся на равных расстояниях друг от друга. Пусть вокруг каждого крыла возбуждена такая же циркуляция, как и вокруг крыла K. Тогда установившееся течение жидкости будет строго периодично по y. Если расстояние между соседними крыльями очень велико по сравнению с поперечными размерами крыла, то введение добавочных крыльев может исказить течение в непосредственной близости крыла K только пренебрежимо мало. Существенные изменения произойдут лишь вдали от крыла K. Проведем прямолинейный контур $ABCD$, горизонтальные стороны которого проходят посередине между соседними крыльями. Пусть длина его AD бесконечно велика по сравнению с высотой. На боковых сторонах AB и CD скорость v слагается

из горизонтальной скорости $\mathbf{v}_\infty$ невозмущенного потока и вертикальной скорости $\mathbf{v}'$, обусловленной циркуляцией. За положительную цир-

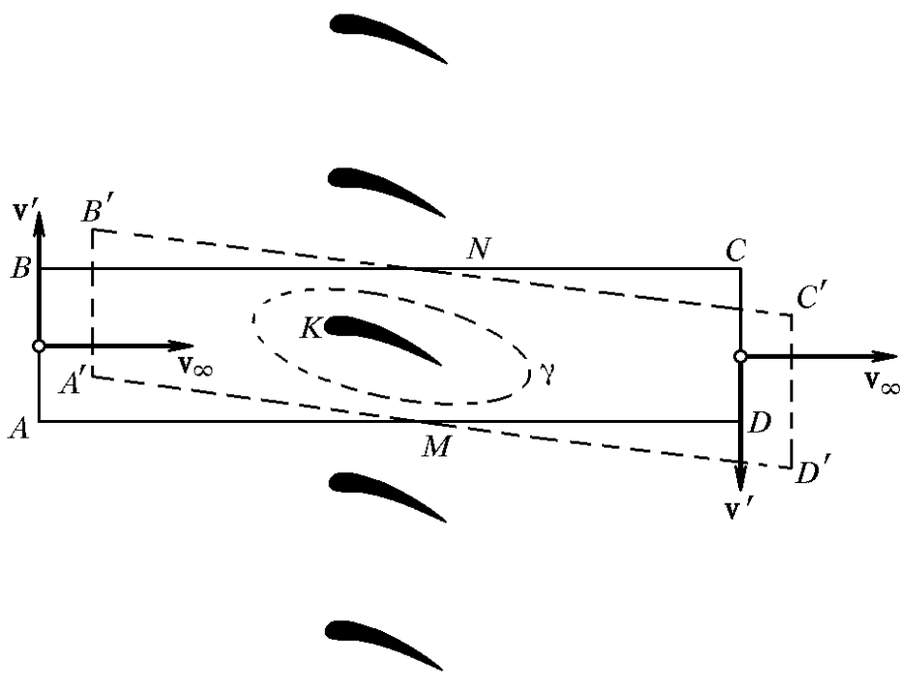


Рис. 282

куляцию примем циркуляцию по часовой стрелке. При такой циркуляции на стороне AB скорость $\mathbf{v}'$ будет направлен вверх (положительна), а на CD — вниз (отрицательна). Рассмотрим жидкость в прямоугольном параллелепипеде с основанием $ABCD$ и единичной высотой, перпендикулярной к плоскости рисунка. Через время dt жидкость, находившаяся в параллелепипеде, переместится в объем $A'B'C'D'$. Рассчитаем приращение импульса ее $d\mathbf{I}$. При стационарном течении это приращение будет равно разности в один и тот же момент времени между импульсом жидкости в новых частях пространства, которые она заняла за время dt, и импульсом в тех частях пространства, из которых она ушла за то же время. Но ввиду полной периодичности картины движения в направлении оси Y импульсы в объемах $AA'M$ и $BB'N$ в точности одинаковы. Одинаковы и импульсы в объемах MDD' и NCC'. Поэтому искомое приращение импульса $d\mathbf{I}$ найдется, если из импульса в объеме $CC'D'D$ вычесть импульс в объеме $AA'B'B$. Каждый из этих объемов равен $lv_\infty\, dt$, где l — длина стороны $AB = CD$; горизонтальные скорости v_∞ одинаковые в обоих объемах, а вертикальные скорости v' отличаются знаками. Поэтому приращение получает только вертикальная составляющая импульса, и это приращение равно

$$dI_y = -2lv_\infty \rho v'\, dt.$$

Но $2lv' = \Gamma$ есть циркуляция скорости v' по контуру $ABCD$, так как стороны AD и BC никакого вклада в циркуляцию не дают. Скорость $\mathbf{v}'$ на этих сторонах одна и та же, и при обходе по контуру $ABCD$ они проходятся в противоположных направлениях. Величина Γ есть в то же время циркуляция по контуру $ABCD$ полной скорости

$\mathbf{v} = \mathbf{v}_\infty + \mathbf{v}'$, так как очевидно, что постоянный член $\mathbf{v}_\infty$ никакого вклада в циркуляцию внести не может. Таким образом,

$$dI_y = -\Gamma \rho v_\infty \, dt.$$

Приращение импульса жидкости равно импульсу внешних сил, действующих на нее. Из них силы давления, действующие на рассматриваемую массу жидкости на поверхности $ABCD$, можно не принимать во внимание, так как равнодействующая всех таких сил давления равна нулю. Остается единственная сила, с которой крыло действует на жидкость. Она равна и противоположна по знаку подъемной силе F_y. Применяя теорему об импульсе силы, получаем

$$F_y = \Gamma \rho v_\infty. \tag{104.1}$$

Из вывода ясно, что под Г следует понимать циркуляцию по контуру $ABCD$. Но для потенциального течения контур циркуляции γ можно провести произвольно. Важно только, чтобы он охватывал крыло K и не охватывал другие крылья. Взяв в качестве γ произвольный контур, будем удалять в бесконечность все остальные крылья, не трогая при этом сам контур γ. Тогда в пределе мы придем к случаю единственного крыла, обтекаемого потоком жидкости. В этом предельном случае результат (104.1) сохраняет силу. Формула (104.1) и есть формула Жуковского—Кутта.

§ 105. ЭФФЕКТ МАГНУСА

1. Если неподвижный круглый цилиндр обтекается равномерным потоком воздуха, перпендикулярным к его оси, то вследствие симметрии возникает только лобовое сопротивление, но никакой подъемной силы не появляется. Если, однако, цилиндр привести во вращение, то появляется подъемная сила, перпендикулярная к направлению внешнего потока, и цилиндр отклоняется в сторону. Это явление называется *эффектом Магнуса* (по имени немецкого физика Генриха Магнуса (1802—1870), ученого, открывшего и исследовавшего это явление экспериментально). Допустим сначала, что цилиндр только вращается с постоянной скоростью, например, по часовой стрелке (рис. 283). Из-за трения приходит в движение и окружающий воздух. Образуется пограничный слой. Движение в пограничном слое вихревое, оно слагается из потенциального движения, на которое накладывается вращение. Ввиду того, что скорость воздуха убывает наружу, вращение в пограничном слое про-

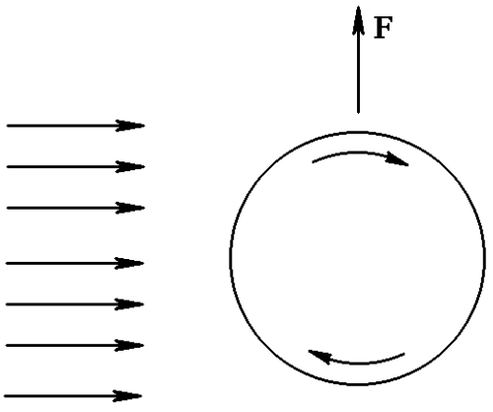
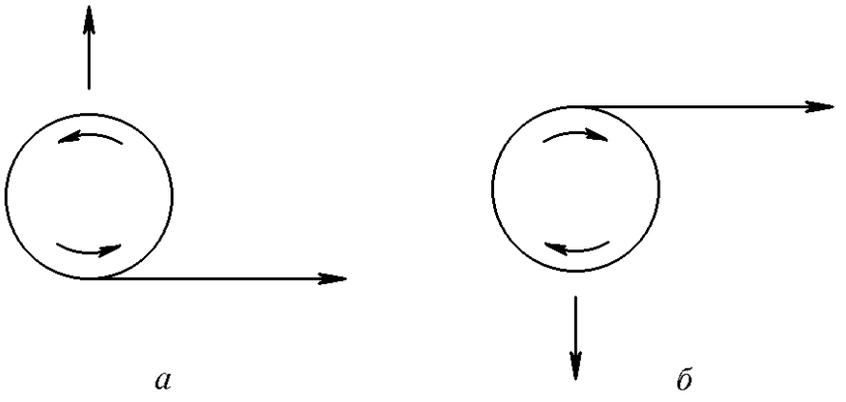


Рис. 283

исходит против часовой стрелки, т. е. противоположно вращению самого цилиндра. При больших числах Рейнольдса ламинарное движение в пограничном слое неустойчиво (см. § 99) и должно переходить в турбулентное. Но и в турбулентном пограничном слое вращение частиц воздуха в основном должно происходить противоположно направлению вращения цилиндра. Допустим теперь, что вращающийся цилиндр обдувается равномерным потоком воздуха слева направо. Сверху цилиндра направление потока совпадает с направлением вращения цилиндра, а снизу — противоположно ему. Частицы в пограничном слое сверху цилиндра ускоряются потоком, что препятствует отрыву пограничного слоя. Наоборот, снизу поток тормозит движение в пограничном слое и способствует его отрыву. Отрывающиеся части пограничного слоя уносятся потоком в виде вихрей, в которых направление вращения происходит против часовой стрелки. Вследствие этого вокруг цилиндра возникает циркуляция скорости в противоположном направлении, т. е. в том же направлении, в каком вращается цилиндр. Вместе с циркуляцией появляется и подъемная сила, направленная вверх. При изменении направления вращения цилиндра на противоположное подъемная сила также меняет направление на противоположное. Разумеется, и в этом случае (для бесконечно длинного цилиндра) величина подъемной силы определяется формулой Жуковского—Кутта (104.1).

2. Эффект Магнуса можно продемонстрировать, поместив прямой круглый цилиндр в вертикальном положении на легкую тележку, стоящую на горизонтальных рельсах. Цилиндр приводится во вращение маленьким электроматорчиком и обдувается потоком воздуха. Если поток воздуха направить перпендикулярно к рельсам, то тележка начинает катиться по ним.



а *б*

Рис. 284

То же происходит, если поток воздуха направлен под углом к рельсам. Можно даже заставить тележку катиться «против ветра» под острым углом. При перемене направления вращения цилиндра тележка катится в противоположную сторону.

Вот другая демонстрация того же эффекта. На легкую картонную катушку наматывается лента, другой конец которой прикрепляется к длинной палке. Катушка кладется в горизонтальном положении на стол. Если быстро дернуть за палку, то катушка начинает вращаться и одновременно приоб-

ретает горизонтальную скорость. Из-за появляющейся подъемной силы катушка взмывает вверх (рис. 284 *а*). Если намотать ленту так, как показано на рис. 284 *б*, то «подъемная сила» изменит направление и будет прижимать катушку к столу.

Бумажный цилиндр, скатившийся с наклонной плоскости, при дальнейшем падении отклоняется назад. Аналогично ведет себя теннисный мяч после «резаного» удара, который сообщает ему вращение. Вот эти явления могут служить иллюстрацией эффекта Магнуса.

3. Флеттнер предложил использовать эффект Магнуса для приведения в движение корабля энергией ветра. Вместо парусов он установил цилиндры (роторы), приводимые в быстрое вращение с помощью моторов. На концах цилиндров помещались выступающие круглые диски (как у катушки) для уменьшения вредного засасывания воздуха в область потока с пониженным давлением. Испытания показали техническую пригодность таких роторных кораблей. Однако в экономическом отношении они оказались менее выгодными, чем обычные моторные суда, а потому не получили распространения.

ПРИЛОЖЕНИЕ

ЕДИНИЦЫ И РАЗМЕРНОСТИ ФИЗИЧЕСКИХ ВЕЛИЧИН В СИ

Определения единиц физических величин для основных (выделенных полужирным шрифтом) и дополнительных единиц СИ. Внесистемные единицы, допустимые к применению наравне с единицами СИ, отмечены кружком

Величина		Единица		
наименование	размерность	наименование	обозначение	связь с основными единицами СИ
Длина	**L**	**метр**	**м**	**Основная единица** Метр представляет собой расстояние, проходимое в вакууме плоской электромагнитной волной за 1/299 792 458 долю секунды
		°астрономическая единица длины	а. е.	1 а. е. = $1{,}49598 \cdot 10^{11}$ м
		°световой год	св. год	1 св. год = $9{,}4605 \cdot 10^{15}$ м
		°парсек	пк	1 пк = $3{,}0857 \cdot 10^{16}$ м
Площадь	L^2	квадратный метр	$м^2$	
		°гектар	га	1 га = 10^4 $м^2$
Объем	L^3	кубический метр	$м^3$	
		°литр	л	1 л = 10^{-3} $м^3$
Плоский угол		радиан	рад	*Дополнительная единица* Радиан равен углу между двумя радиусами окружности, длина дуга между которыми равна радиусу
		°градус	...°	1° = (π/180) рад
		°минута	...′	1′ = (π/10 800) рад
		°секунда	...″	1″ = (π/648 000) рад
Телесный угол	—	стерадиан	ср	*Дополнительная единица* Стерадиан равен телесному углу с вершиной в центре сферы, вырезающему на поверхности сферы площадь, равную площади квадрата со стороной, равной радиусу сферы

Величина		Единица		
наименование	размерность	наименование	обозначение	связь с основными единицами СИ
Время	**T**	**секунда** °минута °час °сутки	**с** мин ч сут	**Основная единица** Секунда равна 9 192 631 770 периодов излучения, соответствующего переходу между двумя сверхтонкими уровнями основного состояния атома цезия-133 1 мин = 60 с 1 ч = 3600 с 1 сут = 86 400 с
Скорость	LT^{-1}	метр в секунду	м/с	
Ускорение	LT^{-2}	метр в секунду в квадрате	$м/с^2$	
Угловая скорость	T^{-1}	радиан в секунду	рад/с	1 рад/с = 1 $с^{-1}$
Угловое ускорение	T^{-2}	радиан на секунду в квадрате	$рад/с^2$	1 $рад/с^2$ = 1 $с^{-2}$
Частота периодического процесса	T^{-1}	герц	Гц	1 Гц = 1 $с^{-1}$
Частота вращения	T^{-1}	секунда в минус первой степени	$с^{-1}$	
Масса	**M**	килограмм °тонна °атомная единица массы	**кг** т а. е. м.	**Основная единица** Килограмм равен массе международного прототипа килограмма 1 т = 10^3 кг 1 а. *е. м.* = $1{,}6605655 \cdot 10^{-27}$ кг
Плотность	$L^{-3}M$	килограмм на кубический метр	$кг/м^3$	

Величина		Единица		
наименование	размерность	наименование	обозначение	связь с основными единицами СИ
Удельный объем	L^3M^{-1}	килограмм на кубический метр	$м^3/кг$	
Массовый расход	MT^{-1}	килограмм в секунду	кг/с	
Объемный расход	L^3T^{-1}	кубический метр в секунду	$м^3/с$	
Сила	LMT^{-2}	ньютон	Н	$1\ Н = 1\ кг \cdot м \cdot с^{-2}$
Давление	$L^{-1}MT^{-2}$	паскаль	Па	$1\ Па = 1\ Н/м^2 = 1\ м^{-1} \cdot кг \cdot с^{-2}$
Жесткость	MT^{-2}	ньютон на метр	Н/м	$1\ Н/м = 1\ кг \cdot с^{-2}$
Напряжение (механическое)	$L^{-1}MT^{-2}$	паскаль	Па	$1\ Па = 1\ Н/м^2 = 1\ м^{-1} \cdot кг \cdot с^{-2}$
Динамическая вязкость	$L^{-1}MT^{-1}$	паскаль-секунда	Па·с	$1\ Па \cdot с = 1\ м^{-1} \cdot кг \cdot с^{-1}$
Кинематическая вязкость	L^2T^{-1}	квадратный метр в секунду	$м^2/с$	
Поверхностное натяжение	MT^{-2}	ньютон на метр	Н/м	$1\ Н/м = 1\ кг \cdot с^{-2}$
Импульс	LMT^{-1}	килограмм-метр в секунду	кг·м/с	
Момент силы	L^2MT^{-2}	ньютон-метр	Н·м	$1\ Н \cdot м = 1\ м^2 \cdot кг \cdot с^{-2}$
Момент импульса	L^2MT^{-1}	килограмм-метр в квадрате в секунду	$кг \cdot м^2/с$	
Момент инерции	L^2M	килограмм-метр в квадрате	$кг \cdot м^2$	
Работа, энергия	L^2MT^{-2}	джоуль	Дж	$1\ Дж = 1\ Н \cdot м = 1\ м^2 \cdot кг \cdot с^{-2}$

Величина		Единица		
наименование	размерность	наименование	обозначение	связь с основными единицами СИ
Мощность, поток энергии	L^2MT^{-3}	ватт	Вт	1 Вт = 1 Дж/с = 1 $м^2 \cdot кг \cdot с^{-3}$
Температура (термодинамическая)	**Θ**	**кельвин** °градус Цельсия	**K** °C	Основная единица Кельвин равен 1/273,16 части термодинамической температуры тройной точки воды $t/°C = T/K - 273{,}15$
Температурный коэффициент	Θ^{-1}	кельвин в минус первой степени	K^{-1}	
Температурный градиент	$L^{-1}\Theta$	кельвин на метр	К/м	
Количество вещества	**N**	**моль**	**моль**	**Основная единица** Моль равен количеству вещества системы, содержащей столько же структурных элементов, сколько содержится атомов в углероде-12 массой 0,012 кг При применении моля структурные элементы должны быть специфированы и могут быть атомами, молекулами, ионами, электронами и другими частицами или специфицированными группами частиц
Молярная масса	MN^{-1}	килограмм на моль	кг/моль	
Молярный объем	L^3N^{-1}	кубический метр на моль	$м^3$/моль	
Количество теплоты (теплота)	L^2MT^{-2}	джоуль	Дж	1 Дж = 1 Н·м = 1 $м^2 \cdot кг \cdot с^{-2}$
Удельная теплота	L^2T^{-2}	джоуль на килограмм	Дж/кг	1Дж/кг = 1$м^2 \cdot с^{-2}$

Величина		Единица		
наименование	размерность	наименование	обозначение	связь с основными единицами СИ
Молярная теплота	$L^2MT^{-2}N^{-1}$	джоуль на моль	Дж/моль	1 Дж/моль = 1 $м^2 \cdot кг \cdot с^{-2} \cdot моль^{-1}$
Теплоемкость, энтропия	$L^2MT^{-2}\Theta^{-1}$	джоуль на кельвин	Дж/К	1 Дж/К = 1 $м^2 \cdot кг \cdot с^{-2} \cdot К^{-1}$
Удельная теплоемкость, удельная энтропия	$L^2T^{-2}\Theta^{-1}$	джоуль на килограмм-кельвин	Дж/(кг·К)	1 Дж/(кг·К) = 1·$м^2 \cdot с^{-2} \cdot К^{-1}$
Молярная теплоемкость, молярная энтропия	$L^2MT^{-2}\Theta^{-1}N^{-1}$	джоуль на моль-кельвин	Дж/(моль·К)	1 Дж/(моль·К) = 1 $м^2 \cdot кг \cdot с^{-2} \cdot К^{-1} \cdot моль^{-1}$
Тепловой поток	L^2MT^{-3}	ватт	Вт	1 Вт = 1 Дж/с = 1 $м^2 \cdot кг \cdot с^{-3}$
Плотность теплового потока	MT^{-3}	ватт на квадратный метр	$Вт/м^2$	1 $Вт/м^2 = кг \cdot с^{-3}$
Теплопроводность	$LMT^{-3}\Theta^{-1}$	ватт на метр-кельвин	Вт/(м·К)	1 Вт/(м·К) = 1 м·кг·$с^{-3} \cdot К^{-1}$
Коэффициент теплопередачи	$MT^{-3}\Theta^{-1}$	ватт на квадратный метр-кельвин	$Вт/(м^2 \cdot К)$	1 $Вт/(м^2 \cdot К) = 1$ кг·$с^{-3} \cdot К^{-1}$
Концентрация (плотность числа частиц)	L^{-3}	метр в минус третьей степени	$м^{-3}$	
Молярная концентрация	$L^{-3}N$	моль на кубический метр	$моль/м^3$	
Коэффициент диффузии	L^2T^{-1}	квадратный метр на секунду	$м^2/с$	

Величина		Единица		
наименование	размерность	наименование	обозначение	связь с основными единицами СИ
Сила электрического тока	**I**	**ампер**	**А**	**Основная единица** Ампер равен силе неизменяющегося тока, который при прохождении по двум параллельным проводникам бесконечной длины и ничтожно малой площади кругового поперечного сечения, расположенным в вакууме на расстоянии 1 м один от другого, вызвал бы на каждом участке проводника длиной 1 м силу взаимодействия, равную $2 \cdot 10^{-7}$ Н
Плотность электрического тока	$L^{-2}I$	ампер на квадратный метр	$А/м^2$	
Количество электричества (электрический заряд)	TI	кулон	Кл	1 Кл = 1 с·А
Поверхностная плотность электрического заряда	$L^{-2}TI$	кулон на квадратный метр	$Кл/м^2$	1 $Кл/м^2 = 1\ м^{-2}$с·А
Пространственная плотность электрического заряда	$L^{-3}TI$	кулон на кубический метр	$Кл/м^3$	1 $Кл/м^3 = 1\ м^{-3}$·с·А

Величина		Единица		
наименование	размерность	наименование	обозначение	связь с основными единицами СИ
Электрическое напряжение, электрический потенциал, разность электрических потенциалов, электродвижущая сила	$L^2MT^{-3}I^{-1}$	вольт	В	$1\ В = 1\ Вт/А = 1\ м^2 \cdot кг \cdot с^{-3} \cdot А^{-1}$
Напряженность электрического поля	$LMT^{-3}I^{-1}$	вольт на метр	В/м	$1\ В/м = 1\ Вт/(А \cdot м) = 1\ м \cdot кг \cdot с^{-3} \cdot А^{-1}$
Электрическое сопротивление	$L^2MT^{-3}I^{-2}$	ом	Ом	$1\ Ом = 1\ В/А = 1 \cdot м^2 \cdot кг \cdot с^{-3} \cdot А^{-2}$
Удельное электрическое сопротивление	$L^3MT^{-3}I^{-2}$	ом·метр	Ом·м	$1\ Ом \cdot м = 1\ м^3 \cdot кг \cdot с^{-3} \cdot А^{-2}$
Электрическая проводимость	$L^{-2}M^{-1}T^3I^2$	сименс	См	$1\ См = 1\ Ом^{-1} = 1\ м^{-2} \cdot кг^{-1} \cdot с^3 \cdot А^2$
Удельная электрическая проводимость	$L^{-3}M^{-1}T^3I^2$	сименс на метр	См/м	$1\ См/м = 1\ Ом^{-1} \cdot м^{-1} = 1\ м^{-3} \cdot кг^{-1} \cdot с^3 \cdot А^2$
Электрическая емкость	$L^{-2}M^{-1}T^4I^2$	фарад	Ф	$1\ Ф = 1\ Кл/В = 1\ м^{-2} \cdot кг^{-1} \cdot с^4 \cdot А^2$
Электрическая постоянная, абсолютная диэлектрическая проницаемость	$L^{-3}M^{-1}T^4I^2$	фарад на метр	Ф/м	$1\ Ф/м = 1\ м^{-3} \cdot кг^{-1} \cdot с^4 \cdot А^2$

Величина		Единица		
наименование	размерность	наименование	обозначение	связь с основными единицами СИ
Поток электрического смещения	TI	кулон	Кл	1 Кл = 1 с·А
Электрическое смещение	$L^{-2}TI$	кулон на квадратный метр	Кл/м^2	1 Кл/м^2 = 1 м$^{-2}$·с·А
Магнитный поток (поток магнитной индукции)	$L^2MT^{-2}I^{-1}$	вебер	Вб	1 Вб = 1 В·с = 1 Тл·м^2 = 1 м^2·кг·с$^{-2}$·А$^{-1}$
Магнитная индукция (плотность магнитного потока)	$MT^{-2}I^{-1}$	тесла	Тл	1 Тл = 1 В·с/м^2 = 1 Вб/м^2 = 1 кг·с$^{-2}$·А$^{-1}$
Индуктивность	$L^2MT^{-2}I^{-2}$	генри	Гн	1 Гн = 1 м^2·кг·с$^{-2}$·А$^{-2}$
Магнитная постоянная, абсолютная магнитная проницаемость	$LMT^{-2}I^{-2}$	генри на метр	Гн/м	1 Гн/м = 1 м·кг·с$^{-2}$·А$^{-2}$
Напряженность магнитного поля	$L^{-1}I$	ампер на метр	А/м	
Энергия излучения	L^2MT^{-2}	джоуль	Дж	1 Дж = 1 м^2·кг·с$^{-2}$
Мощность излучения (поток излучения)	L^2MT^{-3}	ватт	Вт	1 Вт = 1 Дж/с = 1 м^2·кг·с$^{-3}$

Величина		Единица		
наименование	размерность	наименование	обозначение	связь с основными единицами СИ
Интенсивность излучения (плотность потока излучения)	MT^{-3}	ватт на квадратный метр	$Вт/м^2$	$1\ Вт/м^2 = 1\ кг \cdot с^{-3}$
Поток частиц	T^{-1}	секунда в минус первой степени	$с^{-1}$	
Плотность потока частиц	$L^{-2}T^{-1}$	секунда в минус первой степени на метр в минус второй степени	$с^{-1} \cdot м^{-2}$	
Сила света	**J**	**кандела**	**кд**	**Основная единица** Кандела равна силе света в заданном направлении источника, испускающего монохроматическое излучение частотой $540 \cdot 10^{12}$ Гц, энергетическая сила света которого в этом направлении составляет 1/683 Вт/ср
Световой поток	J	люмен	лм	$1\ лм = 1\ кд \cdot ср$
Световая энергия	TJ	люмен-секунда	$лм \cdot с$	$1\ лм \cdot с = 1\ с \cdot кд \cdot ср$
Светимость	$L^{-2}J$	люмен на квадратный метр	$лм/м^2$	$1\ лм/м^2 = 1\ м^{-2} \cdot кд \cdot ср$
Освещенность	$L^{-2}J$	люкс	лк	$1\ лк = 1\ лм/м^2 = 1\ м^{-2} \cdot кд \cdot ср$
Яркость	$L^{-2}J$	кандела на квадратный метр	$кд/м^2$	
Оптическая сила	L^{-1}	°диоптрия	дптр	$1\ дптр = 1\ м^{-1}$

Величина		Единица		
наименование	размерность	наименование	обозначение	связь с основными единицами СИ
Энергетическая сила (сила излучения)	L^2MT^{-3}	ватт на стерадиан	Вт/ср	1 Вт/ср = 1 $м^2 \cdot кг \cdot с^{-3} \cdot ср^{-1}$
Энергетическая светимость (излучательность)	MT^{-3}	ватт на квадратный метр	$Вт/м^2$	1 $Вт/м^2$ = 1 $кг \cdot с^{-3}$
Энергетическая освещенность (облученность)	MT^{-3}	ватт на квадратный метр	$Вт/м^2$	1 $Вт/м^2$ = 1 $кг \cdot с^{-3}$
Энергетическая яркость (лучистость)	MT^{-3}	ватт на стерадиан-квадратный метр	$Вт/(ср \cdot м^2)$	1 $Вт/(ср \cdot м^2) = 1 \cdot кг \cdot с^{-3} \cdot ср^{-1}$

ИМЕННОЙ УКАЗАТЕЛЬ

Аристотель 69
Архимед 12, 476, 477, 481
Атвуд 201

Бернулли Даниил 490, 493, 496, 499, 508, 521, 523, 524, 526, 531, 537
Бессель 383
Бойль 454, 470
Браге Тихо 321
Брагинский 397

Вентури 493
Верн Жюль 296

Гаген 507
Гайзенберг 48
Галилей 12, 89—103, 227, 372, 383
Гамильтон 170, 238
Гаусс 22
Гельмгольц 330
Гук 79, 216, 405, 410—413, 421, 422
Гюйгенс 12, 192, 193, 198, 222—224, 231, 263

Даламбер 520, 522
Дезорм 494
Дикке 395, 397

Евклид 21, 22

Жолли 325
Жуковский 186—192, 295, 364

Кавендиш 324
Карман 534
Кеплер 12, 321, 322, 324, 332, 333, 343, 352, 355, 467
Кёниг 137, 206
Клаузиус 148
Клеман 494
Коперник 71, 72, 342, 371, 380
Кориолис 362, 369, 381, 382, 400
Крылов 68
Кулон 83, 107
Кутта 539, 541

Лавуазье 104
Лаплас 417, 455
Лебедев 93
Лейбниц 49
Ле Шателье 291
Ломоносов 104
Лоренц 98, 102, 142

Магнус 541, 542
Максвелл 269
Мариотт 454, 470
Мах 401, 513
Мещерский 120
Мёссбауэр 403

Ньютон 12—19, 49, 68, 69, 76, 78, 79, 80, 84—87, 96, 103, 112, 133, 172, 173, 185, 211, 213, 219, 324, 326, 333, 345, 353, 356, 369, 385, 389, 392, 393, 454, 455, 457, 491

Обербек 202

Панов 397
Паскаль 468
Паули 157
Пито 495
Пифагор 339
Планк 48, 462
Прандтль 496, 531—533
Пуазейль 506—510
Пуансо 313, 317
Пуассон 414, 422, 447

Резерфорд 342
Рейнольдс 513, 514, 517, 519, 520, 525, 526

Саузернс 395
Сперри 304
Стокс 526, 527
Струхаль 513

Тейлор 466
Титьенс 533
Томсон Вильям 392
Торричелли 496

Флетттнер 543
Фруд 513—516
Фуко 298, 300—305, 380, 383, 384

Циолковский 121, 122, 135

Эйлер 258—260, 475, 480
Эйнштейн 12, 13, 28—31, 103, 326, 397—401
Этвеш 393, 394

Юнг 410, 411, 412, 422, 453—455, 465, 466

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

Автопилот 299
Аддитивность массы 104
Адиабатический инвариант 234
— коэффициент 414
— модуль 414
— процесс 234
Акселерометр 83
Амплитуда колебаний 78
— прилива 385

Барометрическая формула 485
Безразмерные комбинации 463
Бинормаль 43

Ватт 130
Вектор 53, 55
— аксиальный 57, 62
— площади 61
— полярный 62
Векторное произведение 63
Величины основные (первичные) 456
— производные (вторичные) 456
Вес тела 373
Взаимные векторы 65
Взаимодействие прикосновением 92
Вириал сил 148
Вихревая дорожка Кармана 534
Вихрь 533
Водоизмещение 479
Водомер 493
Воздушная подушка 109
Возможные перемещения 174
Волчок 278
— китайский 294
— опрокидывание 294
Время 25
— единое 28
— местное 27
Выпрямляющий момент 478
Высота однородной атмосферы 485
Вязкость 499
— динамическая 508
— кинематическая 508

Гармонический осциллятор 217
Гармоническое колебание 215
Герполодия 317
Гигантские шаги 208
Гирогоризонт 299
Гидродинамика 469
Гидродинамическое подобие 513
Гидростатика 473
Гидростатический парадокс 481
Гироскоп 278
—, вершина 281
—, геометрическая ось 278
—, опрокидывание 301
—, ось фигуры 278
—, приближенная теория 284
— свободный 281
— симметричный 278
—, точка опоры 278
—, точная теория 305
—, уравновешенный (астатический) 279, 298
Гироскопические явления 278
Гироскопический компас 278, 299, 301, 303
Главная нормаль 42
Главные оси 313
Год звездный 45
Год тропический 26, 45
Годограф 39
Гравитационная постоянная 323, 326
Гравитационное смещение спектральных линий 401
Гравитационный заряд 392
Градиент 170, 171, 474

Движение 12
— абсолютное 366
— быстрое 13
— винтовое 252
— вихревое 527
— возвратное жидкости 533
— инфинитное 147, 334, 335
— медленное 13
— относительное 357
— переносное 357
— по инерции 69
— равномерное 36
— равноускоренное 37
— свободное 69
— ультрарелятивистское 134
— финитное 147
Действие 84
— на расстоянии 90, 327
Деформации малые 405
— неоднородные 422
— однородные 422
— пластические (остаточные)— 404
— упругие 404
Джоуль (единица работы) 130
Дина 81
Динамика 68
Динамический (скоростной) напор 495

Длина движущегося стержня 31

Жуковского скамья 186

Закон Архимеда 477
— всемирного тяготения 323
— Гука 79, 405, 410
— Кеплера второй 321, 343
— — первый 321, 343
— — третий 321, 343
— Ньютона второй 68, 77
— — первый 68, 69
— — третий 68, 84
— Паскаля 468, 469
— площадей 181
— подобия течений 513
— Рейнольдса 518
— сложения скоростей нерелятивистский 99
— — — релятивистский 135
— сохранения веса 104
— — вещества 104
— — импульса 75, 86
— — массы 104
— — массы—энергии 105
— — момента импульса 178
— — энергии 144, 156
— эквивалентности инертной и гравитационной масс 392
Законы трения 105
Замедление нейтронов 164
Замкнутая система 73
Зонд 495

Идеальная жидкость 472
Идеально твердое тело 66, 242
— упругое тело 405
Изгиб 426
Изображающая точка 306
Изолированная система 74
Изотермическая атмосфера 485
Изотермический коэффициент 414
— модуль 414
Изотропия пространства 211
Изохронность колебания 217
Импульс 48, 59, 75
— вращательный 186
— поля 93
— силы 112, 114
— системы материальных точек 112
— тела 104
Инвариант 57
Инвариантность уравнений 57
Инверсия 20
Инертность 73
Искусственная вертикаль 299
— тяжесть 375
Искусственный горизонт 299

Капельно-жидкие среды 469
Карданов подвес 278, 279
Касательные силы внутреннего трения 500
Квазистатический процесс 412
Килограмм 74
Кинематика 32
Классический подход 15
Ковариантность уравнений 56
Количество материи 68
Компоненты вектора 55
Конус герполодии 317
— полодии 317
Космическая скорость вторая 122, 348
— — первая 122, 347
— — третья 122, 348, 349
Коэффициент внутреннего трения 501
— жесткости 79
— Пуассона 414
— сжимаемости 469
— сопротивления трубы 520
— трения 108, 110
— упругости 79
Кризис сопротивления 535
Критерий подобия Рейнольдса 514
— — Фруда 514
Критическая длина 432
— скорость 518
— точка 494
Кручение 422

Линия отрыва 524
— тока 488
— центров 158
Лобовое сопротивление 521, 525

Макроскопические тела 13
Малая вода 385
Малые возмущения 437
Масса 68, 73
— гравитационная 391
— инертная 73, 392
— переменная 119
— покоя 75
— приведенная 117
— присоединенная 522
— релятивистская 75
— тяжелая 392
Материальная точка 33
Машина Атвуда 201
Маятник баллистический 153

Маятник гироскопический 287
— — , приведенная длина 221
— конический 309
— математический 221
— физический 220
— — , взаимные точки 222
— — , приведенная длина 221
— — , сопряженные точки 222
— — , точка подвеса 220
— — , центр качания 222
— Фуко 380
— циклоидальный 222
Мгновенная ось вращения 246
Мгновенное вращение 246
Метагалактика 22, 123
Метацентр 478
Метацентрическая высота 478
Метод последовательных приближений 378
— принципов 12
Метр 27
Механика 12
— квантовая 15
— Ньютона 15
— релятивистская 14
— системы 34
— точки 34
Механическое подобие 512
Мировой эфир 91
Модуль всестороннего сжатия 418
— кручения 226, 422
— одностороннего растяжения 419
— сдвига 421
— Юнга 410
Моль 454
Момент импульса относительно оси 184
— — — точки 177
— инерции относительно оси 185
— — — точки 194
— — поперечного сечения 427
— — — — корабля 479
— силы относительно оси 184
— — — точки 176, 177

Набла-оператор 170
Направление отвеса 373
Напряжение 406
Напряжение нормальное 406
— тангенциальное 406
Натяжение 409
Начальная скорость 37
— фаза 215
Начальные условия 94
Невесомость 375
Независимость действия сил 83
Неизменная плоскость 316
Нейтральная линия 426
Нейтральное сечение 426
Нейтрино 157
Несжимаемая жидкость 470
Ньютон (единица силы) 81

Область застоя 109, 524
Обобщенные координаты 66
— скорости 66
Обобщенный закон Галилея 371
Обратная задача механики 369
Объемная плотность силы 475
— — упругой энергии 413, 416, 418, 421, 422
Одновременность 31
Однорельсовая железная дорога 304
Однородность времени 211
— пространства 211
Односвязная область 528
Одностороннее растяжение 418
— сжатие 418
Оператор Гамильтона 170
Основное уравнение гидродинамики идеальной жидкости 475
— — гидростатики 455
Ось изгиба 426
Отвесное направление 373
Отклонение падающих тел от направления отвеса 377
Отклоняющая сила 307
Отлив 385
Относительное сжатие 410
— — , поперечное 413
— удлинение 410
Отражение в начале координат 20

Парадокс Даламбера 522
Параметрические колебания 237
Периметрическое движение гироскопа 296
Период колебания 216
Перманентные оси вращения 314
Плечо силы 183
Плоское движение 252
— течение 529
Плотность истинная 52
— линейная 451
— средняя 51
Поверхность уровня 170
Пограничный слой 531
Подвес бифилярный 224
— трифилярный 226
Подъемная сила 521, 525
Поле 91

Поле гравитационное 399
— скоростей 497
Полевое взаимодействие 92
Полная вода 385
Полное давление 495
Полный напор 495
Полодия 317
Полюс 315
Постоянная Кеплера 322
— Планка 48
Потенциал скоростей 529
Потенциальная кривая 147
— яма 147
Потенциальное движение 527
— течение с циркуляцией 529
Потенциальный барьер 147
Правило буравчика 20
— параллелограма 53
— размерности 464
— Фуко 300
Предел упругости 404
Предельная нагрузка 432
Преобразование Галилея 99
— Лоренца 99
Прецессия быстрая 309
— вынужденная 285
— медленная 309
— псевдорегулярная 290
— регулярная 289
— свободная 282
Прикладной час 386
Прилив 385
Приливообразующий потенциал 387
Приливы большие (сизигийные) 390
— квадратурные 390
— малые 390
Принцип Ле Шателье 291
— Маха 401
— неопределенностей 48
— относительности 102
— — Галилея 100
— — Эйнштейна 103
— суперпозиции гравитационных полей 323
— — малых возмущений 441
— — — деформаций 412
— — скоростей 441
— — смещений 441
— — упругих напряжений 441
— эквивалентности гравитационных сил и сил инерции 399
Прицельное расстояние 342
Простое колебание 215
Противодействие 84
Прямая задача механики 369
Псевдовектор 62
Псевдоскаляр 62
Пульверизатор 493
Пульсары 210

Работа 129
— на конечном перемещении 129
— элементарная 129
Радиус инерции 264
— качения 264
Размерность 457
Разрывные течения 520
Расход жидкости 507
Реакция связей 80
Ротор вектора 529

Сальто 187
Свободные оси вращения 314
Связи 65
— идеальные 174
Сдвиг 420
Секунда 25
Сила 68, 69, 76
— инерции 356
— — Кориолиса 365
— — переносная 365
— — поступательная 365
— — центробежная 365
— Лоренца 142
— приливообразующая 387
— равнодействующая 82
— реактивная 120
— результирующая 180
— сопротивления среды 109
— центральная 138
Силовой центр 139
Силы активные 84
— близкодействия 90
— внешние 85
— внутренние 85
— всемирного тяготения 323
— вязкости 500
— гироскопические 142, 294
— гравитационные 323
— диссипативные 142
— консервативные 141
— массовые 407, 472
— неконсервативные 142
— объемные 407, 472
— пассивные 85
— поверхностные 472
— трения 106
— электромагнитные 70
— ядерные 70
Система единиц LMT 458

Система единиц LMT1 458
— — Международная (СИ) 81
— — МКСА 458
— — СГС 81, 458
— отсчета гелиоцентрическая 71
— — земная 71
— — инерциальная 71, 400
— — Коперника 71
— — лабораторная 159
— — неподвижная 356
— — пространственная 19
— — пространственно-временна́я 26
— — центра масс 159
Скаляр 53, 57
— истинный 62
Скалярное произведение 57, 63, 64
Скоростная точка 39, 40
Скоростной (динамический) напор 495
Скорость абсолютная 357
— газовой струи 120
— звука 454
— истинная 35
— — , вектор 39
— круговая 337
— мгновенная 36
— — , вектор 39
— начальная 37
— относительная 358
— параболическая 337
— переносная 357
— распространения крутильных колебаний 450
— — поперечных возмущений 448, 451
— — — — в натянутом шнуре 451
— — продольных возмущений 448
— секториальная 181
— средняя 39
— угловая 38
— — , вектор 248
Слабые взаимодействия 70
След 535
Сложение векторов математическое 57
— — физическое 57
— вращений 249
— сил 81
Смешанное произведение 64
Сплюснутость Земли 483
Статическая теория приливов 389
Статически неопределенные системы 244
Степени свободы 65
Степени свободы твердого тела 66
Сутки звездные 25
Сутки солнечные 25

Тангенциальные разрывы 520
Твердое тело 242
— — активное 84
— — пассивное 84
Тела аморфные 472
— анизотропные 405
— изотропные 405
Тензор 311
— инерции 311
— упругих напряжений 408
Теорема вириала 148
— Гюйгенса 222
— Гюйгенса—Штейнера 193
— Кёнига 137
— Кориолиса 362
— о движении центра масс 115, 116
— Эйлера 258
Теория подобия 459
— размерности 458
Тепловое равновесие 485
Течение ламинарное 507, 516
— турбулентное 508, 516
Точка поворота 147
Трение внешнее 106
— внутреннее 106, 500
— жидкое 106
— качения 106
— покоя 107
— скольжения 106
— сухое 106
— сцепления 107
Триангуляция 21
Трубка Вентури 493
— Пито 495
— Прандтля 496
— тока 488

Угловой поворот 254
Угол атаки 525
— сдвига 421
Удар абсолютно неупругий 151
— — упругий 157
— высокий 272
— нецентральный 161
— низкий 272
— нормальный 272
— с накатом 272
— с оттяжкой 272
Ударные волны 437
Универсальная газовая постоянная 454
Упругие постоянные 404
Уравнение Бернулли 490
Уравнение движения 77, 357

Уравнение Клапейрона 484
— Мещерского 120
— моментов 178
— — относительно движущегося начала 201
— состояния 471
— Эйлера 475
Ускорение 36
— , вектор 39
— абсолютное 358
— касательное 44
— Кориолиса 362
— нормальное 44
— относительное 358
— переносное 358
— свободного падения 373
— угловое 39
— — , вектор 248
— центростремительное 40, 363
Ускорители на встречных пучках 153
Устойчивость равновесия 176

Фаза 215
Физически бесконечно малые величины 51
Формула Жуковского—Кутта 541
— Ньютона для скорости звука 454
— Пуазейля 507
— размерности 458, 460
— Стокса 526
— Торричелли 496
— Циолковского нерелятивистская 121
— — релятивистская 121
Фотонная ракета 115

Центр инерции 115
— карданова подвеса 279
— качания 222
— масс 115, 116
— пловучести 478
— сил 139
— тяжести 116
— удара 270
Циклическая частота 215
Циркуляция скорости 527

Часы 25
Число Маха 513
— Рейнольдса 513
— — критическое 519
Число Струхаля 513
— Фруда 513

Эллипсоид инерции 322
Энергия внутренняя 156
— гравитационная 329
— кинетическая 131, 133
— покоя 133
— полная 133, 145
— пороговая 167
Энергия потенциальная 143
— релятивистская 133
— упругая 412
Эрг 130
Эффект Доплера 402
— Магнуса 541

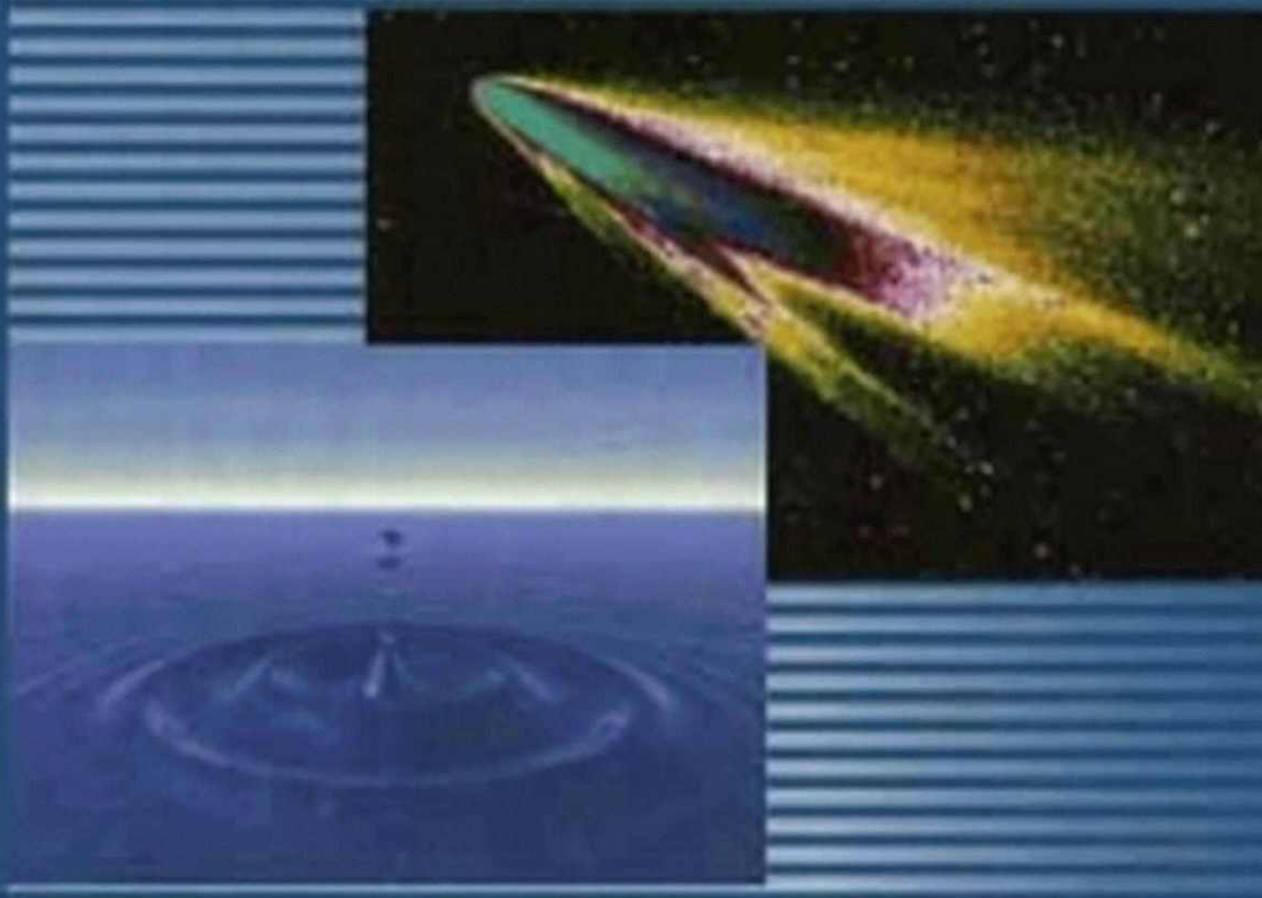
Явление заноса 111
— застоя 108

ОБЩИЙ КУРС ФИЗИКИ

Д.В. Сивухин

ТЕРМОДИНАМИКА И МОЛЕКУЛЯРНАЯ ФИЗИКА

ТОМ II



УДК 530.1 (075.8)
С 34
ББК 22.2

Сивухин Д. В. **Общий курс физики:** Учеб. пособие: Для вузов. В 5 т. Т. II. **Термодинамика и молекулярная физика.** — 5-е изд., испр. — М.: ФИЗМАТЛИТ, 2005. — 544 с. — ISBN 5-9221-0601-5.

Второй том курса физики, широко известного у нас в стране и за рубежом. Книга написана на основе лекций, которые в течение ряда лет читались автором студентам Московского физико-технического института. Основное внимание уделено выяснению физического смысла и содержания основных законов и понятий термодинамики и молекулярной физики, установлению границ применимости этих законов, развитию у студентов навыков физического мышления и умения ставить и решать конкретные задачи.

Четвертое издание второго тома вышло в 1990 г.

Для студентов физических и математических факультетов университетов, физико-технических и инженерно-физических институтов, а также вузов, где физика является основной дисциплиной.

ISBN 5-9221-0601-5

© ФИЗМАТЛИТ, 1990, 2003, 2005

ОГЛАВЛЕНИЕ

Предисловие к первому изданию 7

Введение

Глава I. Температура

1. Температура и термодинамическое равновесие 13
2. Термоскоп и температурные точки 17
3. Эмпирические температурные шкалы 19
4. Идеально-газовая шкала температур 21
5. Виды термометров 26
6. Международная практическая температурная шкала 31
7. Законы идеальных газов 33
8. Уравнение состояния и его следствия для бесконечно малых процессов 35
9. Макроскопические параметры 38

Глава II. Первое начало термодинамики

10. Введение 43
11. Квазистатические процессы 44
12. Макроскопическая работа 45
13. Первое начало термодинамики для системы в адиабатической оболочке 49
14. Внутренняя энергия 52
15. Количество теплоты. Математическая формулировка первого начала термодинамики 57
16. Когда можно пользоваться представлением о количестве теплоты, содержащемся в теле 60
17. Закон Гесса 64
18. Теплоемкость 65
19. Внутренняя энергия идеального газа. Закон Джоуля 67
20. Уравнение Роберта Майера 71
21. Адиабатический процесс. Уравнение Пуассона 73
22. Определение C_P/C_V методом Клемана и Дезорма 76
23. Скорость звука в газах 77
24. Замечания относительно экспериментальных методов определения C_P и C_V для газов 79
25. Уравнение Бернулли 80
26. Скорость истечения газа из отверстия 83

Глава III. Второе начало термодинамики

27. Общие замечания о первом и втором началах термодинамики . . 85
28. Различные формулировки основного постулата, выражающего второе начало термодинамики 86
29. Обратимые и необратимые процессы 91
30. Цикл Карно и теорема Карно 94
31. Термодинамическая шкала температур 96

32. Тождественность термодинамической шкалы температур со шкалой идеально-газового термометра 101
33. Приведение шкалы газового термометра к термодинамической шкале 102
34. Примеры на применение теоремы Карно 104
35. Разность между теплоемкостями C_P и C_V 108
36. Принципиальный способ градуировки термометра в термодинамической шкале 110
37. Неравенство Клаузиуса (для частного случая) 111
38. Неравенство Клаузиуса в общем виде 115
39. Принцип динамического отопления 120
40. Равенство Клаузиуса. Энтропия 122
41. Закон возрастания энтропии 127
42. Обобщение понятия энтропии на неравновесные состояния 129
43. Возрастание энтропии при диффузии газов. Парадокс Гиббса . . 133
44. Различные понимания второго начала термодинамики 135
45. Термодинамические функции 136
46. Термодинамическая теория эффекта Джоуля–Томсона 140
47. Общие замечания о методе термодинамических функций. Примеры 144
48. Максимальная работа и свободная энергия 147
49. Электродвижущая сила гальванического элемента 149
50. Общие критерии термодинамической устойчивости 152
51. Принцип Ле-Шателье–Брауна и устойчивость термодинамического равновесия 153

Глава IV. **Теплопроводность**

52. Уравнение теплопроводности 162
53. Простейшие стационарные задачи на теплопроводность 167
54. Нестационарные задачи. Теорема единственности 169
55. Принцип суперпозиции температур. Температурные волны 174
56. Задача об остывании полупространства 178
57. Внешняя теплопередача 180

Глава V. **Простейшие вопросы молекулярно-кинетической теории вещества**

58. Введение 183
59. Давление газа с точки зрения молекулярно-кинетической теории 186
60. Скорости теплового движения газовых молекул 191
61. Давление фотонного газа 193
62. Молекулярно-кинетический смысл температуры. Равномерное распределение кинетической энергии теплового движения по поступательным степеням свободы 194
63. Равномерное распределение кинетической энергии по степеням свободы 199
64. Броуновское движение 206
65. Вращательное броуновское движение 211
66. Классическая теория теплоемкости идеальных газов 212
67. Адиабатическое нагревание и охлаждение газа с точки зрения молекулярно-кинетической теории 216

68. Классическая теория теплоемкости твердых тел (кристаллов) . . 220
69. Недостаточность классической теории теплоемкостей. Понятие о квантовой теории (качественное рассмотрение) 221

Глава VI. **Статистические распределения**
70. Элементарные сведения из теории вероятностей 228
71. Распределение скоростей молекулы газа. Постановка задачи . . . 239
72. Закон распределения скоростей Максвелла 245
73. Распределение молекул по абсолютным значениям скоростей. Средние скорости молекул 250
74. Другое доказательство закона распределения скоростей Максвелла. Принцип детального равновесия 253
75. Среднее число молекул, сталкивающихся со стенкой сосуда 259
76. Опытная проверка закона распределения скоростей Максвелла . 265
77. Закон распределения Больцмана 268
78. Работы Перрена по определению постоянной Авогадро 275
79. Распределение Больцмана и атмосферы планет 278
80. Энтропия и вероятность 285
81. Флуктуации 292
82. Метод наиболее вероятного распределения в статистике Больцмана 299
83. Статистики Ферми–Дирака и Бозе–Эйнштейна 306
84. Теорема Нернста 313
85. Квантовая теория теплоемкостей Эйнштейна 317

Глава VII. **Явления переноса в газах**
86. Средняя длина свободного пробега 322
87. Эффективное сечение 328
88. Ослабление пучка молекул в газе 332
89. Вязкость и теплопроводность газов 334
90. Самодиффузия в газах 343
91. Связь диффузии с подвижностью частицы 345
92. Концентрационная диффузия в газах 346
93. Броуновское движение как процесс диффузии 348
94. Термическая диффузия в газах 350
95. Явления в разреженных газах 352
96. Молекулярное течение ультраразреженного газа через прямолинейную трубу 359

Глава VIII. **Реальные газы**
97. Молекулярные силы и отступления от законов идеальных газов 366
98. Уравнение Ван-дер-Ваальса 369
99. Другой метод введения поправки на силы притяжения между молекулами. Уравнение Дитеричи 374
100. Изотермы газа Ван-дер-Ваальса 377
101. Изотермы реального газа. Правило Максвелла. Непрерывность газообразного и жидкого состояний вещества 381
102. Свойства вещества в критическом состоянии. Определение критических параметров 387

103. Внутренняя энергия газа Ван-дер-Ваальса 393
104. Эффект Джоуля–Томсона для газа Ван-дер-Ваальса 395
105. Методы получения низких температур и сжижения газов 401

Глава IX. **Поверхностное натяжение**
106. Поверхностное натяжение и некоторые явления, с ним связанные 407
107. Термодинамика поверхностного натяжения 413
108. Краевые углы. Смачивание и несмачивание 416
109. Разность давлений по разные стороны изогнутой поверхности жидкости. Формула Лапласа 419
110. Капиллярно-гравитационные волны малой амплитуды 430

Глава X. **Фазовые равновесия и фазовые превращения**
111. Фазы и фазовые превращения 434
112. Условие равновесия фаз химически однородного вещества. 437
113. Уравнение Клапейрона–Клаузиуса. Испарение и конденсация. Плавление и кристаллизация 440
114. Зависимость давления насыщенного пара от температуры 445
115. Теплоемкость насыщенного пара 449
116. Тройные точки. Диаграммы состояния 452
117. Кипение и перегревание жидкости 455
118. Зависимость давления насыщенного пара от кривизны поверхности жидкости 458
119. Метастабильные состояния 462
120. Фазовые превращения второго рода 466
121. Конвективная устойчивость жидкостей и газов 470

Глава XI. **Растворы**
122. Общие сведения 476
123. Растворимость тел 477
124. Осмос и осмотическое давление 482
125. Закон Рауля 485
126. Повышение точки кипения и понижение точки замерзания раствора 486
127. Правило фаз 489
128. Диаграммы состояния бинарных смесей 492

Глава XII. **Симметрия и строение кристаллов**
129. Симметрия тел 498
130. Кристаллические решетки 502
131. Кристаллические системы 507
132. Пространственные группы и кристаллические классы кристаллов 512
133. Миллеровские индексы и индексы направлений 516
134. Решетки химических элементов и соединений 519
135. Дефекты в кристаллах 523

Именной указатель 529
Предметный указатель 531

Приложение

Предисловие к первому изданию

Второй том предлагаемого учебника курса физики, так же как и первый, написан на основе лекций, читавшихся автором на протяжении многих лет (начиная с весеннего семестра 1957 г.) для студентов первого курса Московского физико-технического института. Поэтому все сказанное в предисловии к первому тому относится и ко второму. По сравненению с лекционным курсом книга, естественно, охватывает более широкий круг вопросов. При выборе материала и способа изложения автор стремился к тому, чтобы все изложенное не выходило за пределы того, что способен усвоить студент первого курса. Однако автор надеется, что его книга может оказаться полезной и для студентов более старших курсов, а также для всех лиц, изучающих и преподающих физику. Некоторые вопросы молекулярной физики, в особенности относящиеся к физике твердого тела, не включены в книгу, так как студенты первого курса еще не подготовлены для их изучения. Эти вопросы предполагается изложить в последующих томах курса.

При выборе вопросов автор не стремился к энциклопедичности изложения. Цель обучения физике заключается не в том, чтобы дать обучающемуся «все», а в том, чтобы научить его главному — умению самостоятельно ставить и решать физические вопросы. В соответствии с этим разбор каждого вопроса, включенного в курс, ведется подробно, чтобы все принципиальное и существенное не прошло мимо внимания изучающего.

Лекционные демонстрации по термодинамике и молекулярной физике осуществлялись лекционными ассистентами Л. Д. Кудряшевой, В. А. Кузнецовой, М. И. Маклаковым и Г. Н. Фрейбергом. Идеи многих задач, включенных в курс, принадлежат преподавателям физики Московского физико-технического института. Перечислить их всех затруднительно.

Второй том этого курса издавался на ротапринте Московского физико-технического института в двух частях в 1972–1973 гг. Организация ротапринтного издания является заслугой Н. И. Петеримовой. По сравнению с ротапринтным изданием настоящее издания несколько исправлено и дополнено главами о растворах, симметрии и строении кристаллов.

Рукопись второго тома была частично просмотрена академиками В. Л. Гинзбургом и М. А. Леонтовичем, профессорами Э. И. Рашбой и Ю. И. Шиманским и доцентом И. Ф. Классен. Профессор И. А. Яковлев взял на себя труд рецензирования рукописи. Рукопись подверглась также внимательному рецензированию на кафедре экспериментальной физики Киевского государственного университета, возглавляемой профессором И. С. Горбанем. Всем этим лицам, а также профессорам А. З. Голику и А. М. Федорченко, с которыми консультировался И. С. Горбань при рецензировании рукописи, автор выражает глубокую благодарность. Их критические замечания немало способствовали улучшению рукописи.

ВВЕДЕНИЕ

1. *Термодинамика* и *молекулярная физика*, которым посвящен настоящий том нашего курса, изучают один и тот же круг явлений, а именно *макроскопические* процессы в телах, т. е. такие явления, которые связаны с колоссальным количеством содержащихся в телах атомов и молекул. Но эти разделы физики, взаимно дополняя друг друга, отличаются различным подходом к изучаемым явлениям.

Термодинамика, или общая теория теплоты, является *аксиоматической* наукой. Она не вводит никаких специальных гипотез и конкретных представлений о строении вещества и физической природе теплоты. Ее выводы основаны на общих *принципах* или *началах*, являющихся обобщением опытных фактов. Она рассматривает теплоту как род какого-то внутреннего движения, но не пытается конкретизировать, что это за движение.

Молекулярная физика, напротив, исходит из представления об *атомно-молекулярном* строении вещества и рассматривает теплоту как беспорядочное движение атомов и молекул. Молекулярная физика в широком смысле слова изучает не только макроскопические явления. Она рассматривает также свойства и строение отдельных молекул и атомов. Но эти вопросы мы здесь затрагивать не будем. Они будут рассмотрены в другом разделе, а именно в атомной физике. Молекулярную физику часто называют также *молекулярно-кинетической теорией* строения вещества.

В XIX веке, когда существование атомов и молекул ставилось под сомнение, гипотетические методы молекулярно-кинетической теории не находили сочувствия среди тех физиков, которые отрицательно относились ко всяким гипотезам и основанным на них теоретическим построениям. В этих условиях строгое разграничение между термодинамикой и молекулярно-кинетической теорией было оправдано; надо было отделить достоверные факты от гипотез, хотя бы и в высшей степени правдоподобных. Но XX век принес окончательные неопровержимые доказательства реальности атомов и молекул. Молекулярно-кинетическая теория в основном утратила гипотетический характер, который был присущ ей в начальный период своего развития. Гипотетический элемент в молекулярно-кинетической теории сохранился лишь постольку, поскольку ей приходится пользоваться упрощенными идеализированными молекулярными моделями, которые не полностью, а лишь частично передают свойства реальных тел. Применять такие модели необходимо либо из-за недостаточности наших знаний молекулярной структуры тел, либо для схематизации и упрощения явлений, без которых теоретическое изучение их было бы вообще невозможно. Поэтому отпала необходимость в том резком разграничении между термодинамикой и молекулярно-кинетической теорией, которое так строго проводилось на ранней стадии развития этих разделов физики. Наш курс мы начнем с аксиоматической термодинамики, но

при ее изложении с самого начала будем привлекать и молекулярные представления.

Термодинамика является одной из важнейших частей физики. Ее выводы достоверны в той же мере, в какой достоверны аксиомы, на которых она построена. Эти выводы используются во всех разделах макроскопической физики: гидродинамике, теории упругости, аэродинамике, учении об электрических и магнитных явлениях, оптике и пр. Пограничные дисциплины — физическая химия и химическая физика — в значительной своей части занимаются приложениями термодинамики к химическим явлениям.

2. Термодинамика возникла в первой половине XIX века как теоретическая основа начавшей развиваться в то время теплотехники. Ее первоначальная задача сводилась к изучению закономерностей превращения теплоты в механическую работу в тепловых двигателях и исследованию условий, при которых такое превращение является наиболее оптимальным. Именно такую цель преследовал французский инженер и физик Сади Карно (1796–1832) в сочинении «О движущей силе огня и о машинах, способных развивать эту силу» (1824 г.), в котором впервые были заложены начатки термодинамики, хотя и сохранились старые ошибочные воззрения на теплоту как на какое-то невесомое вещество «теплород», которое не может быть ни создано, ни уничтожено. В дальнейшем термодинамика далеко вышла за пределы указанной технической задачи. Центр тяжести переместился в сторону изучения физических вопросов. Основным содержанием современной *физической термодинамики* является изучение закономерностей *тепловой формы движения материи* и связанных с ней физических явлений. Приложения к тепловым двигателям, холодильным установкам и прочим вопросам теплотехники выделились в самостоятельный раздел, называемый *технической термодинамикой*. В нашем курсе вопросы технической термодинамики практически будут привлекаться лишь для иллюстрации общих физических законов.

3. Тепловая форма движения материи — это *хаотическое движение* атомов и молекул макроскопических тел. Ее специфичность связана с необычной колоссальностью чисел молекул и атомов во всяком макроскопическом теле. Так, в одном кубическом сантиметре воздуха при нормальных условиях содержится около $2{,}7 \cdot 10^{19}$ молекул. При тепловом движении молекулы сталкиваются между собой и со стенками сосуда, в который заключена система. Столкновения сопровождаются резкими изменениями модуля и направления скоростей молекул. В результате в системе возникает вполне беспорядочное движение, в котором с равными вероятностями представлены все направления скоростей молекул, а сами скорости меняются в широких пределах от очень малых до очень больших значений.

Чтобы получить предварительное представление о характере движения молекул газа, приведем некоторые результаты кинетической теории газов.

Средняя скорость теплового движения газовых молекул весьма велика. Для молекул воздуха она составляет при комнатной температуре почти 500 м/с, возрастая с повышением температуры. Столкновения между молекулами газа происходят чрезвычайно часто. Например, молекула воздуха при нормальной плотности успевает в среднем пройти всего около 10^{-5} см от одного столкновения до следующего. Зная среднюю скорость молекулы, нетрудно подсчитать, что при нормальных температуре и плотности молекула воздуха за одну секунду испытывает до 5 миллиардов столкновений, причем число столкновений возрастает с увеличением температуры и плотности газа. Еще чаще сталкиваются молекулы внутри жидкостей, так как они распределены в пространстве значительно более тесно, чем молекулы газа. Помимо поступательного движения совершаются беспорядочные вращения молекул, а также внутренние колебания атомов, из которых они состоят. Все это создает картину чрезвычайно хаотического состояния, в котором находится совокупность громадного числа молекул газов, а также жидких и твердых тел. Такова природа теплоты с точки зрения молекулярно-кинетической теории строения вещества.

О тепловом движении можно говорить только в тех случаях, когда рассматриваемая физическая система является *макроскопической*. Не имеет смысла говорить о тепловом движении, когда система состоит из одного или небольшого числа атомов.

4. Термодинамика, как правило, изучает только *термодинамически равновесные состояния* тел и *медленные процессы*, которые могут рассматриваться как практически равновесные состояния, непрерывно следующий друг за другом [1]. (Понятие термодинамического равновесия дается в § 1.) Она изучает также *общие закономерности перехода* систем в состояния термодинамического равновесия. Круг задач молекулярно-кинетической теории значительно шире. Она изучает не только термодинамически равновесные состояния тел, но и *процессы в телах*, идущие с *конечными скоростями*. Та часть молекулярно-кинетической теории, которая изучает свойства вещества в состоянии равновесия, называется *статической термодинамикой*, или *статической механикой*. Та же часть, в которой изучаются процессы в телах, идущие с конечными скоростями, называется *физической кинетикой*. Аксиоматическая термодинамика называется также *феноменологической*, или *формальной*. Достоинством термодинамики является то, что ее выводы характеризуются большой общностью, так как они обычно получаются без использования упрощенных моделей, без чего не может обойтись молекулярно-кинетическая теория. Однако последняя, по крайней мере в принципе, позволяет решать и такие вопросы, теоретическое рассмотрение которых невозможно методами одной только аксиоматической термодинамики. Сюда относятся, например, выводы

[1] В начале 30-х годов возникла и стала развиваться *термодинамика неравновесных процессов*. Однако этот раздел физики мы здесь рассматривать не будем.

термического и калорического уравнений состояния вещества. Знание таких уравнений необходимо, чтобы придать общим выводам термодинамики законченный конкретный характер. Аксиоматическая термодинамика заимствует эти уравнения из опыта. Кроме того, опыты, поставленные в связи с различными проблемами молекулярной физики, показали, что принципы аксиоматической термодинамики не в такой степени незыблемы и универсальны, как считали ее основоположники. Как и для большинства законов физики область их применимости ограничена. Так, аксиоматическая термодинамика оставляет в стороне самопроизвольные нарушения состояний термодинамического равновесия, или *флуктуации*, которые проявляются тем отчетливее, чем меньше размеры системы. Статическая термодинамика охватывает и этот круг явлений, устанавливая тем самым границы применимости формальной термодинамики.

5. Мы изучаем молекулярную физику после классической механики. В этом заключается известная научно-педагогическая трудность. Молекулярная физика должна основываться на законах, которым подчиняются атомы и молекулы. Это законы *квантовой механики*, которые будут изучаться нами позднее. Без знания этих законов полное и строгое изложение современной молекулярной физики невозможно. Тем не менее, мы излагаем молекулярную физику после классической механики, и вот почему. Широкий круг макроскопических явлений обусловлен не столько деталями строения атомов и характером управляющих ими законов, сколько *необычайно большим числом* самих атомов в макроскопических системах. При изучении такого рода явлений знание квантовой механики не всегда обязательно. Правда, молекулярная физика, построенная на основе классической механики объясняет не весь диапазон экспериментальных фактов. Квантовая природа атомов и молекул рано или поздно даст о себе знать, например, в вопросах о теплоемкости тел или в явлениях вблизи абсолютного нуля температуры. Но в этих случаях, по крайней мере для понимания самого основного и существенного, достаточной небольших предварительных сведений из квантовой физики, которые можно сообщить в ходе изложения. Систематическое же изложение квантовой механики, хотя бы и в элементарной форме, непосредственно после классической механики педагогически не оправдано. Начинающий должен ознакомиться с широким кругом экспериментальных фактов, которые одни только могут по-настоящему убедить его в недостаточности классических и необходимости введения квантовых представлений.

6. Изложения феноменологической термодинамики необходимо предпослать следующее замечание. Физики XVIII и отчасти первой половины XIX века рассматривали теплоту как особое невесомое вещество, содержащееся в телах. Оно не может быть ни создано, ни уничтожено. Это гипотетическое вещество называлось *теплородом*. Нагревание тел объяснялось увеличением, а охлаждение — уменьшением содержащегося внутри них теплорода. Теория теплорода несостоятельна. Она не может объяснить простейшие явления, например

нагревание тел при трении. Нет необходимости рассматривать эту теорию. Мы упомянули о ней только потому, что терминология, употребляющаяся в учении о теплоте, исторически сложилась под влиянием теории теплорода. Основным понятием теории теплорода было *количество теплоты*. С точки зрения этой теории оно само собой было понятно и не нуждалось в определении. Этим понятием пользуются и теперь, хотя оно неудачно, так как с ним ассоциируется неправильное представление о физической природе теплоты. Терминология всегда более живуча, чем физические представления, сменяющие друг друга. Физикам сплошь и рядом приходится пользоваться не вполне рациональной исторически сложившейся терминологией. Особо большой беды в этом нет, если только каждый термин понимать в смысле данного ему точного определения и не связывать с ним никаких представлений, которые не вытекают из этого определения. В учении о теплоте нет ничего более опасного, чем отнести количество теплоты к числу первоначальных понятий, не требующих определения. Термином «количество теплоты» можно пользоваться только при условии, что ему дано ясное и четкое определение, никак не связанное с представлениями теории теплорода. То же относится к таким понятиям, как *теплоемкость*, *скрытая теплота* и пр., которые наука получила в наследство также от теории теплорода.

Глава I

ТЕМПЕРАТУРА

§ 1. Температура и термодинамическое равновесие

1. Понятие *температуры* вводится для характеристики различной *степени нагретости тел*. Представление о температуре, как и представление о силе, вошло в науку через посредство наших чувственных восприятий. Наши ощущения позволяют различать качественные градации нагретости: *теплый*, *холодный*, *горячий* и пр. Однако количественная мера степени нагретости, пригодная для науки, не может быть установлена с помощью чувственных восприятий. Ощущения субъективны. В зависимости от состояния руки одно и то же тело на ощупь может казаться либо теплым, либо холодным. Погрузим, например, одну руку в горячую, а другую в холодную воду и подержим их в течение некоторого времени. Затем обе руки опустим в воду с комнатной температурой. Тогда первая рука почувствует холод, а вторая тепло. Все испытывали неприятное ощущение холода при медленном вхождении в холодную воду во время купания, быстро исчезающее и сменяющееся чувством бодрости и удовольствия после того, как в результате полного погружения в воду тело купающегося немного охладится. Чувственная оценка температуры сильно зависит от *теплопроводности* тела. Например, при оценке на ощупь металлические предметы в комнате кажутся холоднее деревянных. Если же их нагреть до одной и той же температуры, превышающей температуру руки, то соотношение будет обратным. Чувственная оценка температуры применима только в весьма узком температурном интервале. Она не годится в случае очень горячих и очень холодных тел. Ничего хорошего не получится при попытке определить на ощупь степень нагретости расплавленного железа или жидкого воздуха.

2. В основу количественного определения температуры и построения точной температурной шкалы должны быть положены объективные физические явления и факты, свободные от субъективизма чувственных восприятий. К понятию температуры можно подходить с различных точек зрения. В феноменологическом учении о теплоте температура вводится через понятие *теплового*, или *термического*, *равновесия*. Более общим является понятие *термодинамического равновесия*. Как то, так и другое понятия трудно поддаются логическому определению. К ним приходят в результате рассмотрения конкретных примеров и последующего обобщения.

Если два тела, температуры которых при оценке с помощью наших органов чувств сильно отличаются друг от друга (например, раскаленный металл и холодную воду), привести в соприкосновение друг с

другом, то опыт показывает, что одно тело будет нагреваться, а другое охлаждаться, пока в системе не прекратятся всякие макроскопические изменения. (Мы предполагаем, что соприкасающиеся тела химически не реагируют друг с другом.) Тогда, применяя терминологию, заимствованную из механики, говорят, что эти два тела находятся в *термодинамическом равновесии* друг с другом и имеют *одинаковые температуры*. Термодинамическое равновесие, как показывает опыт, в конце концов наступает не только в случае соприкосновения двух, но и в случае соприкосновения скольких угодно тел. Если соприкасающиеся тела химически реагируют друг с другом, то в результате химических реакций может происходить дополнительное нагревание или охлаждение. Но после прекращения химических реакций в конце концов устанавливается термодинамическое равновесие, при котором никаких макроскопических процессов уже не происходит.

3. Отмеченные факты допускают обобщение. Назовем *изолированной*, или *замкнутой*, *системой* систему тел, которые не могут обмениваться энергией с окружающими телами. Такая система есть идеализация и в действительности точно никогда не реализуется. Даже в тех случаях, когда вблизи соприкасающихся тел никаких других тел не имеется, тела все же не образуют замкнутую систему, так как они непрерывно испускают лучистую энергию и поглощают излучение, исходящее от других удаленных тел. Однако искусственными приемами можно создать такие условия, когда обмен энергий системы с другими телами станет пренебрежимо малым. Этого можно достигнуть путем заключения системы в твердую *теплонепроводящую*, или *адиабатическую*, оболочку, т. е. такую оболочку, что состояние системы, помещенной внутри нее, не меняется при нагревании или охлаждении тел, находящихся вне оболочки (подробнее см. § 13). Адиабатическая оболочка является физической абстракцией. Таких оболочек в действительности не существует, но можно создать оболочки, по своим свойствам приближающиеся к адиабатическим. Наиболее совершенными адиабатическими оболочками в современной физике и технике являются стенки *сосудов Дьюара*, или *термосов*. Это стеклянные или металлические баллоны с двойными стенками, между которыми создан высокий вакуум. Они хорошо предохраняют помещаемые в них тела от теплового воздействия внешних тел. Идеализируя подобные приборы, мы приходим в пределе к представлению об идеальной адиабатической оболочке. Тело, заключенное в адиабатическую оболочку с твердыми неподвижными стенками, при условии, что внешние силовые поля поддерживаются постоянными, полностью защищено от внешних воздействий, т. е. является изолированной системой.

Обобщение, о котором говорилось выше, заключается в следующем. *Каково бы ни было начальное состояние тел изолированной системы, в ней в конце концов установится термодинамическое равновесие, в котором прекратятся все макроскопические процессы.* Это положение играет важную роль в термодинамике и принимается в ней за постулат, иногда называемый *общим началом термодинамики*.

4. Приведем еще несколько примеров на установление термодинамического равновесия. Допустим, что жесткая теплонепроводящая оболочка разделена также теплонепроводящей перегородкой на две части. В одной из частей находится жидкость, в другой создан вакуум. Быстро удалим перегородку. Жидкость закипит. В пространстве, ограниченном оболочкой, возникает сложное движение жидкости и ее пара. Но в конце концов оно, а также дальнейшее преобразование жидкости прекратятся. Получится либо только один пар (если жидкости вначале было мало), либо система, состоящая из жидкости и насыщенного пара. В обоих случаях конечное состояние термодинамически равновесно. Это не есть состояние абсолютного покоя, в котором прекращаются все без исключения процессы. Рассматриваемое с молекулярной точки зрения, оно характеризуется непрерывным и интенсивным обменом молекулами между жидкостью и паром. Это значит, что непрерывно идет процесс преобразования жидкости и обратный ему процесс конденсации пара в жидкость. Однако в состоянии термодинамического равновесия эти два процесса как бы компенсируют друг друга: среднее число испаряющихся молекул равно среднему числу молекул, возвращающихся обратно из пара в жидкость. Термодинамическое равновесие, таким образом, может быть охарактеризовано как *динамическое равновесие*, когда весьма интенсивно идут процессы молекулярного масштаба, но все макроскопические процессы прекращаются. Это относится ко всякому термодинамическому равновесию, рассмотренному в приведенном примере.

Если в стакан с водой бросить кусок сахара, то начальное состояние системы будет термодинамически неравновесным — сахар начнет растворяться в жидкости. Однако по прошествии некоторого времени, когда процесс растворения прекратится, возникнет термодинамически равновесное состояние, в котором получится либо однородный раствор, либо неоднородная система, состоящая из куска сахара и окружающего его насыщенного раствора. В последнем случае динамический характер равновесного состояния проявляется в том, что процесс растворения сахара, если его рассматривать с молекулярной точки зрения, никогда не прекращается. Однако в состоянии равновесия он компенсируется обратным процессом кристаллизации сахара из раствора.

5. Самопроизвольный процесс перехода системы в состояние термодинамического равновесия называется *релаксацией*, а время, затрачиваемое на такой переход, — *временем релаксации*. Время релаксации относится к числу нечетко определенных понятий. При его измерении никогда не дожидаются момента, когда наступает полное динамическое равновесие. Вместо этого измеряют время, по истечении которого система переходит в такое неравновесное состояние, которое с требуемой точностью может быть принято за равновесное. Это время и принимают за время релаксации. Таким образом, во всех случаях речь идет не о точном определении времени релаксации, а о его приближенных оценках.

6. Термодинамическое равновесие предполагает, что тела, приводимые в контакт, находятся также в *механическом* и *химическом* равновесии друг с другом. Это означает, в частности, что давления в обоих телах одни и те же и тела при контакте химически не реагируют между собой. Если это не так, то тела можно отделить одно от другого абсолютно жесткой теплопроводящей оболочкой, химически не реагирующей ни с одним из рассматриваемых тел. Примером теплопроводящей оболочки может служить тонкая фольга из химически нейтрального металла. Такая оболочка, ввиду ее тонкости, не влияет существенно на физическое состояние тел. Однако она не препятствует обмену энергиями между контактирующими телами. Поэтому при контакте тел через разделяющую их абсолютно жесткую теплопроводящую перегородку в конце концов должно наступить равновесие, при котором обмен энергией между телами прекращается. При этом в отсутствие перегородки механического и химического равновесия между телами может и не быть. В этом случае говорят, что тела находятся в *тепловом*, или *термическом*, *равновесии* между собой, или имеют *одинаковые температуры*.

Таким образом, по определению, два тела находятся в тепловом равновесии друг с другом, или имеют одинаковые температуры, если при приведении их в тепловой контакт через разделяющую абсолютно жесткую теплопроводящую оболочку равновесие не нарушается. (Оболочка не нужна, если тела находятся в механическом и химическом равновесии друг с другом.) Если же теплового равновесия при контакте не получается и для его достижения требуется время, то говорят, что температуры тел до контакта были разные. Это определение освобождает понятие температуры от субъективизма, свойственного температуре, вводимой с помощью наших ощущений. Подчеркнем, что температура относится к величинам, которые зависят только от *внутреннего состояния тела*.

Во всем дальнейшем подразумевается, что приводимые в контакт тела, если они химически реагируют друг с другом, заключены в тонкие теплопроводящие оболочки, препятствующие этим химическим реакциям.

7. Температура есть одна из *макроскопических характеристик* внутреннего состояния тел. Это понятие не имеет смысла для систем, состоящих из одного или небольшого числа атомов и молекул. Хотя оно строго применимо только для систем, находящихся в термодинамическом равновесии, однако им постоянно пользуются также и в тех случаях, когда полного термодинамического равновесия еще нет. Говорят, например, о неравномерно нагретых телах, разные точки которых имеют разные температуры. Это возможно потому, что время релаксации уменьшается с уменьшение размеров системы. Мысленно разобьем неравновесную систему на достаточно малые макроскопические части. Ввиду малости времени релаксации таких частей каждая из них быстро придет практически в состояние термодинамического

равновесия. Это значит, что если такую малую часть мгновенно заключить в твердую адиабатическую оболочку, то ее состояние практически окажется равновесным и не будет меняться с течением времени. Поэтому и можно говорить о температурах таких малых частей. Система в целом при этом не находится в термодинамическом равновесии — происходит медленный процесс выравнивая температур ее различных частей. Но могут быть и такие сильно неравновесные состояния, что разделение системы на малые макроскопические части, практически являющиеся равновесными, невозможно. К таким состояниям понятие температуры неприменимо.

§ 2. Термоскоп и температурные точки

1. Для суждения об одинаковости или различии температур двух тел A и B нет необходимости обязательно приводить их в тепловой контакт друг с другом. Можно воспользоваться для этой цели третьим телом C, приводимым последовательно в контакт с телами A и B. Этот способ имеет то преимущество, что он позволяет сравнивать температуры и в том случае, когда тела A и B реагируют друг с другом химически, не используя при этом теплопроводящих перегородок. В основе его лежит следующий опытный факт.

Если тело C находится в тепловом равновесии с телами A и B, то тела A и B, приведенные в контакт друг с другом, также будут находиться в тепловом равновесии. Иными словами, если температура C равна температурам тел A и B, то тела A и B имеют одну и ту же температуру, равную по определению температуре тела C.

Допустим теперь, что тело C настолько мало, что при приведении в контакт с телами A и B оно не меняет заметно температуры этих тел, хотя изменения температуры самого тела C и могут быть значительными. Такое тело C может служить «пробным телом», с помощью которого можно констатировать одинаковость или различие температур тел A и B. Приведем тело C в контакт с телом A и подождем, пока не наступит тепловое равновесие между ними. Тело C примет температуру тела A, тогда как температура последнего практически остается неизменной. Затем приведем тело C в контакт с телом B. Если окажется, что при этом температура тела C не изменилась, то тела A и B имеют одну и ту же температуру; в противном случае их температуры разные.

Достаточно малое тело C, служащее для констатации одинаковости или различия температур двух или нескольких тел, называется *термоскопом*. Малость тела C существенна. В противном случае термоскоп заметно искажал бы температуру испытываемого тела.

2. О постоянстве или изменении температуры термоскопа можно судить по изменению различных величин, характеризующих его физические свойства. Опыт показывает, что практически все физические свойства тел изменяются с изменением температуры. Так, при

нагревании большинство тел расширяется, т. е. увеличивается их объем. Электрическое сопротивление металлов возрастает с повышением температуры, а полупроводников — убывает. Если две проволоки из разнородных металлов спаять своими концами и включить гальванометр, как показано на рис. 1, то прибор не обнаружит электрического тока, если все места соединений разнородных металлов имеют одну и ту же температуру. Если же нагреть или охладить один спай, то в цепи возникнет электрический ток, называемый *термоэлектрическим током*. Совокупность разнородных проволок, спаянных, как указано выше, составляет так называемую *термоэлектрическую пару* или, короче, *термопару*.

Все подобные явления могут быть использованы для построения термоскопа. Примером может служить воздушный термоскоп, представленный на рис. 2 в двух различных вариантах. При соприкос-

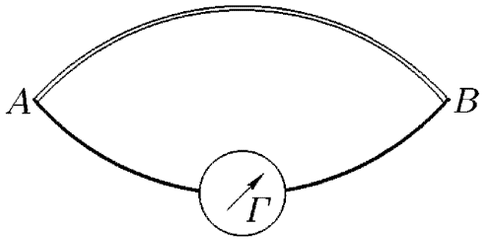


Рис. 1

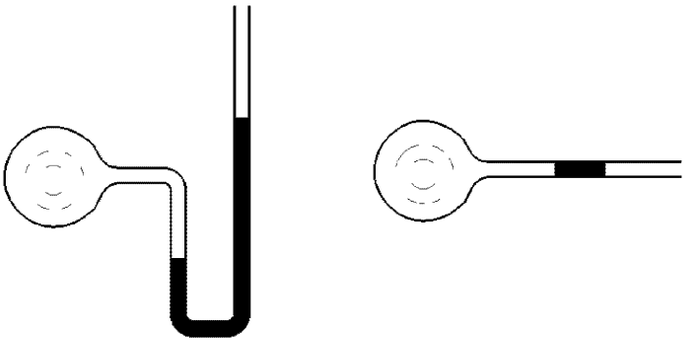


Рис. 2

новении шарика термоскопа с исследуемым телом меняется объем содержащегося внутри него воздуха. Изменение объема воздуха констатируется с помощью жидкостного манометра или по перемещению столбика жидкости в трубке, соединенной с шариком.

3. С помощью термоскопа можно установить ряд постоянных *температурных точек*, т. е. неизменных, хорошо воспроизводимых температур. Они используются при построении *температурных шкал*. Отметим наиболее важные из таких точек.

Если твердое тело и жидкость, состоящие из одного и того же вещества, находятся в контакте, то, в зависимости от температуры, твердое тело будет плавиться или, наоборот, жидкость затвердевать. При этом предполагается, что давление в системе поддерживается постоянным. Лишь при вполне определенной температуре, как показывает опыт, оба процесса — плавление и затвердевание — взаимно компенсируют друг друга. В этом случае массы жидкой и твердой фаз остаются неизменными. Тогда говорят, что жидкость и твердое тело находятся или сосуществуют в *фазовом равновесии*. Температура, при которой твердая и жидкая фазы одного и того же вещества сосуществуют в фазовом равновесии при нормальном атмосферном давлении (101325 Па), называется *нормальной точкой плавления* (НТП) рассматриваемого вещества.

Аналогичные определения вводятся для процессов *кипения* и *возгонки* (возгонкой называется превращение твердой фазы непосредственно в газообразную, минуя промежуточную — жидкую — фазу). Температура, при которой жидкость сосуществует в фазовом равновесии со своим паром при нормальном атмосферном давлении, называется *нормальной точкой кипения* (НТК). Аналогично определяется *нормальная точка возгонки* (НТВ). Примером вещества, которое при нормальном давлении не плавится, а возгоняется, может служить твердая углекислота (см. § 104, п. 4).

Наконец, могут сосуществовать в фазовом равновесии три фазы одного и того же вещества — твердая, жидкая и газообразная. Однако для этого требуется не только вполне определенная температура, но и вполне определенное давление. Для воды, например, давление должно составлять 4,58 мм рт. ст. (611 Па). Температура, при которой сосуществуют в фазовом равновесии твердая, жидкая и газообразная фазы одного и того же вещества, называется *тройной точкой* (ТТ) этого вещества.

§ 3. Эмпирические температурные шкалы

1. Мы описали лишь способ, позволяющий убедиться в равенстве или различии температур двух тел. Но мы еще не определили температуру *количественно*. Решение этого вопроса сводится к установлению *температурной шкалы*, т. е. к принятию системы правил, с помощью которых каждая температура может быть охарактеризована определенным числом. Пока мы не располагаем количественным определением температуры, в выборе этих правил имеется широкий произвол. Необходимо только, чтобы соответствие между температурами и числами было взаимно однозначным.

Температурная шкала может быть установлена с помощью любого термоскопа. Проградуированный по определенному правилу термоскоп превращается в *термометр*, т. е. прибор, предназначенный для измерения температур. Важными требованиями, предъявляемыми к термометру, являются его *чувствительность*, *точность измерений* и их *воспроизводимость*. Другим полезным свойством термометра является *быстрота перехода* его в состояние теплового равновесия с телом, температура которого подлежит измерению.

Основной частью большинства термометров является *термометрическое тело*, приводимое в тепловой контакт с телом, температуру которого надо измерить. Физическая величина, служащая индикатором температуры, называется *термометрической величиной*. Так, в жидкостных термометрах термометрическим телом служит жидкость (например, ртуть) в резервуаре термометра, а термической величиной — ее объем. В термометрах сопротивления термометрическими телами являются металлические проволоки или полупроводники, термометрическими величинами — их электрические сопротивления. В

термопарных термометрах термометрическим телом является термопара, термометрической величиной — электродвижущая сила, возникающая при наличии разности температур между спаями термопары.

2. Возьмем произвольный термометр и обозначим буквой a термометрическую величину (объем жидкости, электрическое сопротивление, электродвижущую силу и т. п.). При нагревании величина a должна меняться *монотонно*. Иначе между a и температурой T не будет взаимно однозначного соответствия. Так, объем определенной массы воды не годится в качестве термометрической величины, так как при нагревании он проходит через минимум (около 4 °C). По самому смыслу величины a между ней и температурой T имеется функциональная связь:

$$T = f(a).$$

Однако вид функции $f(a)$ определить нельзя, пока не выбрана та или иная температурная шкала. Более того, само введение температурной шкалы принципиально сводится к выбору определенной функции $f(a)$. Надо условиться в качестве $f(a)$ взять какую-либо произвольную *монотонную функцию* величины a. Таким образом, с помощью одного и того же термометра можно построить бесконечное множество температурных шкал, отличающихся друг от друга выбором функции $f(a)$. Заранее нет никаких принципиальных соображений, за исключением требований простоты, которые заставили бы предпочесть какую-либо одну из этих шкал другим. Простейший способ выбора функции состоит в том, чтобы в качестве $f(a)$ взять *линейную однородную функцию*, т. е. положить

$$T = Aa. \tag{3.1}$$

Постоянную A можно выбрать произвольно. Выбором этой постоянной однозначно определится и единица температуры — *градус*. Однако в действительности поступают иначе. А именно, постоянную A вычисляют, приписывая какой-либо температурной точке определенную температуру или двум температурным точкам — определенную разность температур. Такие температурные точки называются *реперными*. До 1954 г. температурная шкала строилась по *двум реперным точкам* — нормальной точке кипения воды $T_к$ и нормальной точке плавления льда $T_{пл}$. (Под последней понимают равновесную температуру чистого льда и насыщенной воздухом воды.) Принималось, по определению, что разность температур этих точек $T_к - T_{пл}$ равна 100°. После этого постоянная A вычислялась по очевидной формуле

$$A = \frac{T_к - T_{пл}}{a_к - a_{пл}} = \frac{100}{a_к - a_{пл}}, \tag{3.2}$$

где $a_к$ и $a_{пл}$ — значения термометрической величины a в соответствующих реперных точках.

Экспериментальные исследования показали, однако, что тройная точка воды $T_{тр}$ обладает лучшей воспроизводимостью, чем нормальные точки плавления льда и кипения воды. Поэтому было принято

международное соглашение строить температурную шкалу по *одной реперной точке* — тройной точке воды. В так называемой *абсолютной термодинамической шкале температур*, или *шкале Кельвина*, принимается по определению, что температура этой точки равна точно 273,16 К. Постоянная A при таком построении температурной шкалы должна вычисляться по формуле

$$A = \frac{T_{\text{тр}}}{a_{\text{тр}}} = \frac{273{,}16}{a_{\text{тр}}}. \tag{3.3}$$

Такой выбор числового значения температуры тройной точки воды сделан для того, чтобы промежуток между нормальными точками плавления льда и кипения воды с максимально возможной точностью составлял 100 К, если пользоваться газовым термометром с идеальным газом. Тем самым устанавливается преемственность шкалы Кельвина с ранее применявшейся шкалой с двумя реперными точками. Измерения показали, что температуры нормальных точек плавления льда и кипения воды в описанной шкале равны приближенно 273,15 и 373,15 К соответственно.

3. Описанный способ построения температурной шкалы, равно как и значение температуры T, зависят от того, какая величина a принята в качестве индикатора температуры. Температура T, отсчитываемая по какому-либо термометру, называется *эмпирической температурой*, а соответствующая шкала температур — *эмпирической температурной шкалой*. Таким образом, в зависимости от выбора термометрической величины a, можно осуществить бесконечное множество эмпирических температурных шкал. Они будут совпадать между собой, вообще говоря, только в основных реперных точках, принятых при построении температурной шкалы (т. е. в тройной точке воды, когда шкала строится по одной реперной точке, или в нормальных точках плавления льда и кипения воды, если шкала строится по двум реперным точкам). Во всех остальных точках, как правило, шкалы совпадать не будут. Это означает, что различные термометры, например ртутный и спиртовой или термометры сопротивления с железной и медной проволокой, при измерении температуры одного и того же тела будут давать, как правило, разные показатели. Для устранения возникшей неоднозначности можно было бы условно принять какой-либо определенный термометр за основной и по нему градуировать все остальные термометры. Но какой термометр следует принять за основной?

§ 4. Идеально-газовая шкала температур

1. При предварительном решении поставленного вопроса можно воспользоваться тем обстоятельством, что разреженные газы, достаточно близкие по свойствам к так называемым *идеальным газам* (см. § 7), с большой точностью подчиняются закону Бойля–Мариотта: *произведение объема V данной массы газа на его давление P зависит*

только от температуры. Это произведение можно принять за термометрическую величину a, а самый газ — за термометрическое тело. Таким путем приходят к идеально-газовой шкале температур. Идеально-газовая температура — это температура, отсчитываемая по газовому термометру, наполненному сильно разреженным газом. Температуру по идеально-газовой шкале будем обозначать большой буквой T. При этом формула (3.1) переходит в формулу

$$PV = CT, \tag{4.1}$$

где C — постоянная, зависящая только от массы и химической природы газа.

Преимущество идеально-газовой шкалы температур перед всеми прочими эмпирическими температурными шкалами состоит в том, что, как показывает опыт, температура T, определенная по формуле (4.1), очень слабо зависит от химической природа газа, которым наполнен резервуар газового термометра. Показания различных газовых термометров при измерении температуры одного и того же тела очень мало отличаются друг от друга. Поэтому, впредь до более совершенного определения (см. § 31), мы будем понимать под температурой T величину, измеряемую с помощью идеально-газового термометра.

Практически газовый термометр можно реализовать двумя способами. В одном способе объем газа V поддерживается постоянным, индикатором температуры служит давление P. Во втором способе поддерживается постоянное давление газа, измерение температуры сводится к измерению объема V. Принципиально оба способа равноправны. Но первый способ более удобен, а поэтому он и применяется на практике.

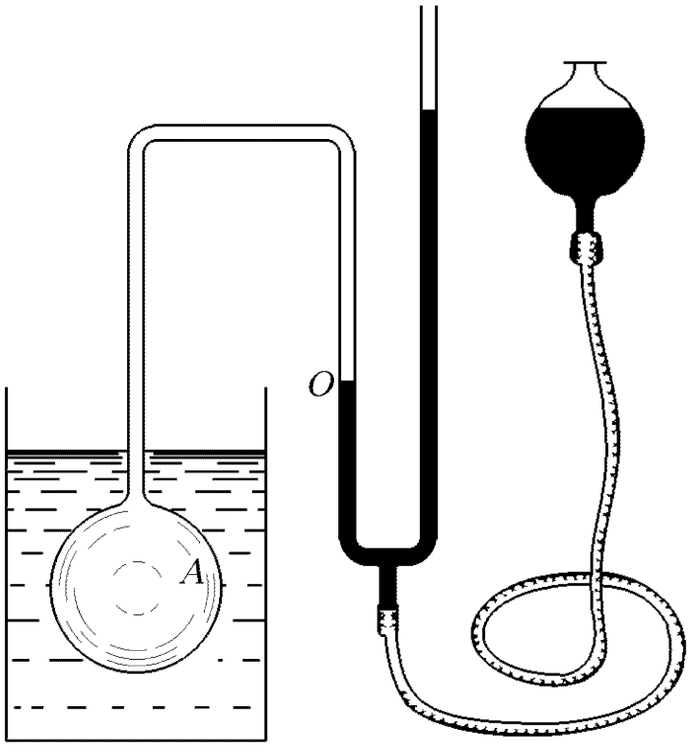


Рис. 3

2. Принцип устройства газового термометра можно понять из схематического рис. 3. Уровень ртути в левом колене манометра при измерениях доводится до постоянной отметки O, чтобы обеспечить постоянство объема газа в баллоне A. В действительности газовый термометр представляет собой весьма сложное устройство, а измерение температуры с помощью этого термометра — нелегкую экспериментальную процедуру, требующую большой тщательности. Не останавливаясь на этих вопросах, ограничимся общим замечанием, что

газовый термометр представляет собой стеклянный, кварцевый или металлический баллон неизменного объема, заполненный реальным газом и соединенный с манометром при помощи капилляра. В качестве термометрического газа раньше употреблялся водород. В настоящее время применяются гелий и азот (последний в области высоких температур, где гелий не годится, так как он проходит сквозь стенки баллона). Давление газа измеряется манометром, находящимся при стандартной температуре. По давлению газа вычисляется измеряемая температура с учетом поправок на изменение объема баллона, неидеальность газа и пр.

3. При $T = 0$ уравнение (4.1) дает $PV = 0$. На этом основании заключают, что при такой температуре $P = 0$, так как объем V в нуль обращаться не может. Далее, говорят, что это есть наиболее низкая из всех возможных температур, называемая *абсолютным нулем температуры*. Хотя это заключение и правильно, однако приведенная аргументация основана на экстраполяции и не имеет доказательной силы. По мере приближения к абсолютному нулю наблюдаются все более и более заметные отступления от законов идеальных газов. Все газы начинают конденсировать еще до достижения абсолютного нуля. Строгое доказательство существования абсолютного нуля температуры основано на *втором начале термодинамики* (см. § 31).

Классическая физика полагала, что при абсолютном нуле атомно-молекулярное движение прекращается. В действительности это не так. Предположение о полном покое атомов и молекул противоречит *принципу неопределенности Гейзенберга*. При абсолютном нуле остается еще весьма интенсивное движение частиц, из которых состоит тело. Ему соответствует минимальная энергия, которая от тела уже не может быть отнята (если остаются постоянными объем и прочие внешние параметры, определяющие состояние тела). Но это движение уже не является тепловым. Соответствующая ему энергия называется *нулевой энергией*. Наличием нулевой энергии объясняются многие явления. Так, при температуре, близкой к абсолютному нулю, энергия жидкого гелия столь велика (в три раза больше, чем теплота испарения), что кристалл гелия, находящийся под давлением собственных насыщенных паров, становится нестабильным и переходит в жидкое состояние. Затвердевание гелия возможно лишь при повышенном давлении. Под влиянием давления атомы гелия сближаются, и силы молекулярного притяжения возрастают. Это обстоятельство и делает возможным переход гелия в твердое состояние.

Абсолютный нуль может быть определен как такая температура, при которой в теле прекращается тепловое движение и остается только движение частиц, связанное с нулевой энергией. Абсолютный нуль является самой низкой и «холодной» из всех температур. Температура, отсчитываемая от абсолютного нуля, называется *абсолютной температурой*. Сначала единица температуры в идеально-газовой абсолютной шкале называлась *градусом Кельвина* в честь английского

физика лорда Кельвина (1824–1907), предложившего термодинамическую шкалу температур, тождественную с идеально-газовой шкалой. Затем такую единицу стали называть просто *кельвин* (сокращенно К). Эта единица однозначно определяется требованием, чтобы температура тройной точки воды была равна 273,16 К.

4. Мы еще не дали точных определений понятиям «более холодный» и «более горячий». Полагаться на чувственные восприятия, приводящие к такого рода представлениям, нельзя по причинам, о которых было сказано выше. Точное определение использует понятие энергии. Тело A называется *более горячим*, или *более нагретым*, чем тело B, если при приведении этих тел в тепловой контакт энергия переходит от тела A к телу B; тело же B в этом случае называется *более холодным*, или *менее нагретым*.

5. Поскольку обе величины P и V для идеальных газов не могут быть отрицательными, абсолютная температура T, определяемая соотношением (4.1), может быть либо только положительной, либо только отрицательной. Первый случай реализуется, когда в качестве постоянной C взята положительная величина, второй — эта постоянная отрицательна. Так как условились температуру тройной точки воды считать положительной, то при таком соглашении второй случай отпадает. (Можно было бы поступать наоборот, тогда все абсолютные температуры были бы отрицательными.) Таким образом, абсолютная температура, определенная с помощью идеально-газового термометра, может быть *только положительной*. Более нагретым телам соответствуют более высокие, а менее нагретым — более низкие температуры.

6. Существование абсолютного нуля, как температуры, ниже которой тело охладить нельзя, конечно, не зависит ни от каких определений. Однако то обстоятельство, что такой температуре приписывается значение нуль градусов, является всего лишь произвольным соглашением. Можно было бы температуру определить так, что абсолютному нулю соответствовало любое, наперед выбранное число градусов. Действительно, температурой можно было бы назвать не обязательно величину T, определяемую соотношением (4.1), а любую функцию этой величины. Например, за температуру можно было бы принять логарифм T. Тогда температура абсолютного нуля выражалась бы числом $-\infty$, а температурная шкала заполняла бы весь бесконечный интервал от $-\infty$ до $+\infty$. Однако нет оснований отступать от традиционного определения.

7. Наряду с абсолютной температурной шкалой в физике употребляется также *шкала Цельсия* (1701–1744), строящаяся известным способом по двум основным реперным точкам — нормальной точке плавления льда и нормальной точке кипения воды. Шкала Цельсия отличается от абсолютной шкалы положением нуля. Если t — температура в градусах Цельсия, а T — абсолютная температура в кельвинах, то приближенно

$$t = T - 273{,}15\,^{\circ}\mathrm{C}. \tag{4.2}$$

Пользуясь этим выражением, уравнение (4.1), выражающее закон Бойля–Мариотта, можно записать в виде

$$PV = P_0V_0(1 + \alpha t), \tag{4.3}$$

где P_0 и V_0 — давление и объем газа при температуре $t = 0\,^\circ\mathrm{C}$, а α — постоянный коэффициент, числовое значение которого приближенно равно

$$\alpha = \frac{1}{273{,}15}\ \mathrm{K}^{-1}.$$

Этому коэффициенту можно дать двоякое физическое толкование. Если давление газа поддерживать постоянным, то формула (4.3) переходит в $V = V_0(1 + \alpha t)$ и показывает, что α есть *температурный коэффициент объемного расширения газа*. Наоборот, если поддерживать постоянным объем газа, то $P = P_0(1 + \beta t)$. Отсюда видно, что β есть приращение давления газа $P - P_0$, отнесенное к единице приращения температуры $t - t_0$, когда в качестве исходной взята температура $t_0 = 0\,^\circ\mathrm{C}$. Так определяемая величина называется *температурным коэффициентом давления газа*. Мы видим, что для идеальных газов коэффициент объемного расширения и термический коэффициент давления совпадают.

Формулы (4.3) и (4.1), устанавливающие линейную связь между произведением PV и температурой, не выражают какого-либо физического закона, а являются лишь следствиями выбранного способа построения температурной шкалы. Объективный физический закон состоит в том, что для всех идеальных газов коэффициент α один и тот же. Этот закон называется *законом Гей-Люссака* (1778–1850). Наличием такого закона, как уже отмечалось выше, и объясняется предпочтение, отдаваемое идеально-газовой шкале температур перед всеми прочими эмпирическими температурными шкалами.

8. Идеально-газовая шкала температур все же может быть признана вполне удовлетворительной. Идеальных газов в строгом смысле слова не существует. В поведении реальных газов наблюдаются отступления от законов Бойля–Мариотта и Гей-Люссака, и притом не одинаковые для различных газов. Поэтому газовые термометры, наполненные различными газами, дают хотя и близкие, но все же не вполне совпадающие показания. Правда, расхождения между ними стремятся к нулю по мере увеличения степени разрежения газов. Однако это имеет место не для всех температур. При температурах порядка тысячи и нескольких тысяч градусов многоатомные газы начинают диссоциировать, т. е. их молекулы распадаются на атомы. При еще более высоких температурах атомы ионизуются, т. е. распадаются на электроны и положительно заряженные ионы. Благодаря этому газы перестают подчиняться закону Бойля–Мариотта, даже если их степень разрежения сколь угодно велика. Это показывает, что газовый термометр не пригоден для установления температурной шкалы в области очень высоких температур. Газовые термометры не пригодны и для измерения очень низких температур — при понижении температуры

все реальные газы в конце концов конденсируются. Самая низкая температура, которая может быть измерена с помощью газового термометра, составляет примерно 1 К при условии, что термометр наполнен гелием при низком давлении.

Рациональная температурная шкала не должна зависеть от индивидуальных свойств термометрического вещества. Она должна строиться так, чтобы в принципе можно было измерять любые температуры, как бы высоки или низки они ни были. Этим требованиям удовлетворяет *абсолютная термодинамическая шкала температур*, построенная на основе *второго начала термодинамики*. Ее называют также *шкалой Кельвина*. Кельвин впервые предложил эту шкалу в 1848 г. С ней мы подробно познакомимся в § 31.

Абсолютная термодинамическая шкала является *основной температурной шкалой* в физике. Там, где пригоден газовый термометр, эта шкала практически не отличается от идеально-газовой шкалы температур. Поэтому в интервале от 4 до 1338 К (точка затвердевания золота) абсолютная термодинамическая шкала практически осуществляется с помощью газового термометра.

ЗАДАЧА

Газовый термометр постоянного объема наполнен неидеальным газом, уравнение состояния которого имеет вид $P + \pi(V) = \varphi(V)T$, где $\pi(V)$ и $\varphi(V)$ — произвольные функции объема, а T — температура по шкале идеально-газового термометра. Термометр градуируется обычным способом по двум реперным точкам. Доказать, что его показания будут совпадать с показаниями такого же термометра, наполненного идеальным газом, если только температуры реперных точек в обоих случаях выбраны одинаковыми.

§ 5. Виды термометров

1. Среди различных термометров вне конкуренции по чувствительности, точности измерений и воспроизводимости стоит *газовый термометр*, в котором индикатором температуры является *давление газа* при постоянном объеме резервуара термометра. Шкала газового термометра почти точно совпадает с абсолютной термодинамической шкалой Кельвина, что является большим достоинством термометра. Однако устройство газового термометра и работа с ним довольно сложны — термометры имеют большие размеры, громоздки и медленно приходят в состояние теплового равновесия. По этим причинам для технических целей газовые термометры применяются редко. Они довольно редко используются и в физических исследованиях, где с их помощью иногда измеряются низкие температуры. Основное назначение газовых термометров состоит в том, что по ним градуируются другие термометры. В этой связи газовые термометры называются *первичными*, а градуированные по ним термометры, применяющиеся при практических измерениях — *вторичными*. Не каждый вторичный

термометр должен градуироваться непременно по газовому термометру. Газовые термометры используются для градуировки главным образом в бюро стандартов, метрологических институтах и в некоторых исследовательских лабораториях. Обычно же для градуировки термометров применяются достаточно точные вторичные термометры, уже проградуированные по газовому.

Из вторичных термометров наибольшее значение имеют *жидкостные* (главным образом ртутные) *термометры*, *термометры сопротивления* и *термоэлементы*.

2. Устройство жидкостных термометров общеизвестно, и нет необходимости подробно останавливаться на этом вопросе. Жидкостные термометры не отличаются большой точностью. По этой и ряду других причин они не всегда пригодны для физических исследований. Область их применения ограничена со стороны низких температур свойствами жидкостей, а со стороны высоких температур — свойствами стекла: при понижении температуры все используемые жидкости замерзают, при высоких температурах стекло размягчается. Все же жидкостные термометры могут применяться в довольно широкой области температур — примерно от -200 до $+600\,^\circ$C. Наиболее употребительными жидкостями, применяемыми в жидкостных термометрах, являются: 1) пентан (от -200 до $+20\,^\circ$C); 2) этиловый спирт (от -110 до $+50\,^\circ$C); 3) толуол (от -70 до $+100\,^\circ$C); 4) ртуть (от $-38{,}86$ до $600\,^\circ$C). Самыми распространенными являются ртутные термометры. Температура кипения ртути при нормальном атмосферном давлении равна $357\,^\circ$C. При повышении давления она увеличивается (при давлении в 4 атм до $450\,^\circ$C, а при давлении в 30 атм — до $500\,^\circ$C). Поэтому для измерения температур выше $357\,^\circ$C необходимо, чтобы ртуть в капилляре термометра находилась под давлением выше атмосферного. С этой целью капилляр заполняют газом. Давление газа может доходить до 70 атм. Чтобы выдержать такое давление, стенки капилляра должны быть толстыми. Шкала наносится на стенках самого капилляра путем травления. Другие жидкости, применяемые в термометрах, отличаются от ртути прежде всего тем, что они смачивают стенки капилляра. В таких термометрах для предотвращения разрыва столбика жидкости капилляр должен наполняться газом всегда, а не только для измерения высоких температур.

3. *В термометрах сопротивления* термометрическим телом является проволока обычно из чистого металла, а иногда сплава; термометрической величиной служит ее *электрическое сопротивление*. Электрическое сопротивление металлов возрастает с повышением температуры. Отношение увеличения сопротивления при нагревании на один градус к сопротивлению при $0\,^\circ$C называется *температурным коэффициентом сопротивления*. Для большинства чистых металлов (вблизи комнатной температуры) этот коэффициент равен примерно $0{,}4\,\%$. Линейная зависимость сопротивления металлов от температуры соблюдается только приближенно. Отступления от нее особенно отчетливо проявляются при низких температурах. Поэтому термометры

сопротивления имеют нелинейную шкалу и нуждаются в градуировке по газовому термометру.

Наиболее употребительными металлами для термометров сопротивления являются чистые платина и медь. *Платиновый термометр* применяется в интервале температур примерно от 10 до 1100 °С, *медный* — от температур жидкого водорода приблизительно до 120 °С. Точность термометров сопротивления в различных температурных интервалах разная. Для платинового термометра (в интервале температур от 20 до 70 К) она составляет около 0,01 К, а при измерении малых разностей температур — 0,001 К. Точность измерения температур медным термометром примерно на порядок меньше. В термометрах сопротивления применяются и другие материалы, например никель (ниже точки Кюри, т. е. 358 °С), а также сплав железа с никелем (около 70 % никеля и 30 % железа) — ниже 0 °С до 600 °С. Для измерения низких температур (от 3,4 до 273 К) может использоваться термометр сопротивления из чистого индия.

Конструкция термометра сопротивления может быть весьма разнообразной в зависимости от его предназначения. Платиновый термометр обычно изготовляется из тонкой платиновой проволоки (диаметром 0,05–0,1 мм с сопротивлением при комнатной температуре 10–100 Ом), закрепляемой на изоляционном каркасе из слюды (в технических термометрах), фарфора или кварца. Концы проволоки соединяются медными проводами с измерительной схемой сопротивления, например с мостиком Уитстона (1802–1875).

4. Особым видом термометров сопротивления являются *полупроводниковые термометры*, называемые также *термосопротивлениями*, или *термисторами*. В них термометрическим телом служит полупроводник, например уголь или кристаллик германия. Угольные и германиевые термометры сопротивления применяются при температурах ниже 20 К. Сопротивление полупроводников убывает с повышением температуры. Их удельное сопротивление в десятки и сотни раз больше, чем у металлов. Температурный коэффициент сопротивления также примерно в 10 раз больше. Благодаря этому полупроводниковые термометры сопротивления при большой чувствительности могут иметь исключительно малые размеры. Такими термометрами можно надежно измерять изменения температуры в тысячные доли градуса.

5. Многие металлы и сплавы вблизи абсолютного нуля переходят из *нормального* в *сверхпроводящее* состояние, в котором их электрическое сопротивление обращается в нуль. Этот переход, однако, происходит настолько круто, что делает невозможным использование соответствующих материалов для устройства термометров сопротивления. Некоторое распространение в низкотемпературной термометрии получил термометр сопротивления из фосфористой бронзы, содержащей в качестве примеси несколько сотых процента свинца. Здесь переход из нормального в сверхпроходящее состояние сильно растянут. Сопротивление падает почти линейно в температурном интервале от 7 до 1 К.

6. О принципе устройства термопарных термометров уже говорилось выше. Термопарный термометр состоит из двух проволок A и B, изготовленных из разнородных материалов и сваренных своими концами (см. рис. 1). Если спаи имеют разные температуры, то в цепи термопары возникает электрический ток. Один из спаев, называемый *спаем сравнения*, поддерживается при постоянной температуре, например, погружается в тающий лед. Эта температура называется *опорной температурой*. Другой спай, называемый *измерительным*, приводится в тепловой контакт с телом, температуру которого надо измерить. В цепь термопары включается *милливольтметр* для измерения возникающей ЭДС. По величине ЭДС судят о температуре рассматриваемого тела. Градуировка может быть выполнена, например, с помощью газового термометра. При малой разности температур между спаями термо-ЭДС пропорциональна этой разности.

Принципиальная схема включения термопары показана на рис. 4. Разнородные проволоки термопары A и B изображены жирной и двойными линиями, медные провода C, идущие к милливольтметру, — тонкими линиями. Любой из спаев *1* или *2* может быть использован в качестве измерительного или спая сравнения. Необходимо, чтобы спаи проволок A и B с концами медных проводов C, а также места соединения последних с милливольтметром имели одинаковые температуры, так как в противном случае в цепи может появиться дополнительный термоэлектрический ток, искажающий показания приборов. На рис. 5

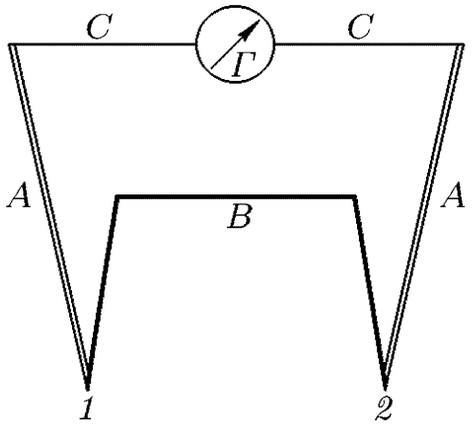


Рис. 4

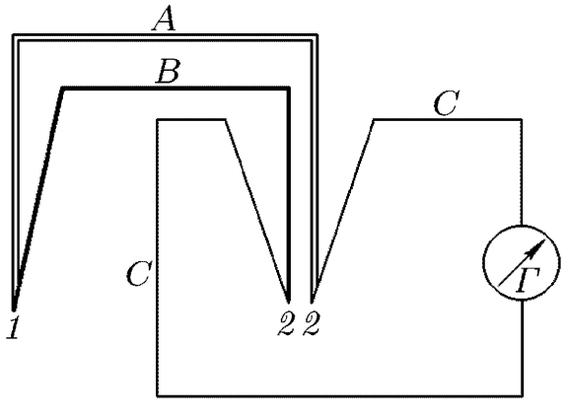


Рис. 5

приведена более совершенная схема включения термопарного термометра. Спай сравнения *2,2* в сущности состоит из двух спаев: металла A с медью и металла B с медью. Эти два спая поддерживаются при любой желаемой постоянной температуре.

Термопары изготовляются как из благородных, так и неблагородных металлов. К последним относятся:

1) медьконстантановая термопара (применяется примерно от -200 до $+350\,^\circ$C);

2) железоконстантановая термопара (от 0 примерно до $750\,^\circ$C);

3) хромельалюмелевая термопара (от -200 до $1100\,^\circ$C);

4) хромельконстантановая термопара (от 20 K до 1000 °C). (Сплав хромель содержит 90 % Ni и 10 % Cr; алюмель — около 94 % Ni, 3 % Mn, 2 % Al и 1 % Si).

К термопарам из благородных металлов относятся:

1) платинородиевая термопара (до 1400–1600 °C);

2) платиноиридиевая термопара (до 1500 °C);

3) иридийродиевая термопара (до 2200 °C).

Особенно важное значение имеет платинородиевая термопара. В этой термопаре один из проводников изготавливается из чистой платины, а другой — из сплава 90 % платины и 10 % родия. Применяются также термопары, оба проводника которых изготовлены из указанного сплава, но с различным содержанием платины и родия. Для измерения очень высоких температур (до 2600–3000 °C) применяется вольфраморениевая термопара.

При соответствующем выборе материала термопары могут обеспечить измерения в широком диапазоне температур (начиная приблизительно от 4 и почти от 3000 K) с высокой точностью (в некоторых случаях до ±0,01 °C) и высокой чувствительностью (до 100 мкВ/К для металлических и 1–5 мкВ/К для полупроводниковых термопар). Они являются идеальными приборами для измерения относительно небольшой разности температур, которые в отдельных случаях могут быть найдены с точностью до ±0,001 °C. Среди прочих достоинств термопары надо отметить простоту изготовления, малую теплоемкость и малое время установления температурного равновесия. Недостатком термопары является малая величина термо-ЭДС при низких температурах. Термо-ЭДС уменьшается в понижении температуры и обращается в нуль при абсолютном нуле. Вместе с ней уменьшается и чувствительность термопарного термометра. Так, например, чувствительность наиболее часто применяемой в низкотемпературной области медьконстантановой термопары составляет при комнатной температуре около 40, при 90 K — около 17, а при 20 K — всего 5 мкВ/К.

7. Измерение очень высоких температур — тысячи градусов и выше — наталкивается на ту очевидную трудность, что не существует тугоплавких термометров, выдерживающих такие температуры. При достаточно высоких температурах все тела плавятся. О температуре тел в этих случаях судят по испускаемому ими излучению. Роль термометрического тела играет само излучающее тело, а термометрической величины — *интенсивность испускаемого ими излучения*. Приборы, работающие на этом принципе, называются *пирометрами*. Их устройство основано, таким образом, на *законах излучения* нагретых тел. Эти законы будут изложены в отделе оптики нашего курса (т. IV).

8. Измерение очень низких температур (ниже 1 K) также наталкивается на большие трудности. В этих случаях тепловой контакт термометра с охлаждаемым телом длительное время не приводит к установлению теплового равновесия между ними. Кроме того, многие термометрические величины, употребляемые при измерении обычных

температур, становятся непригодными в области очень низких температур: давление газа становится неизмеримо малым, сопротивление перестает зависеть от температуры и т. п. Поэтому здесь о температуре охлаждаемого тела судят по изменению физических свойств самого тела, например магнитных. Однако на этом пути возникают серьезные и еще не совсем преодоленные трудности, связанные с согласованием измеряемой таким образом температуры с термодинамической шкалой температур.

§ 6. Международная практическая температурная шкала

Экспериментальные трудности измерения температуры по абсолютной термодинамической шкале привели к необходимости введения *Международной практической температурной шкалы* (МПТШ). При этом преследовалась цель создания такой температурной шкалы, которую легко и быстро можно было бы использовать для калибровки научных и технических приборов и которая в то же время воспроизводила бы термодинамическую шкалу с наибольшей возможной точностью, допускаемой современной техникой измерений. Единицами температуры в этой шкале являются кельвин и градус Цельсия в зависимости от выбора начала отсчета температуры. Шкала несколько раз уточнялась. Последний раз это было сделано в 1968 г.

Международная практическая температурная шкала 1968 г. (МПТШ-68) основана на двенадцати хорошо воспроизводимых температурных точках, которым приписаны определенные значения температуры (*первичные реперные точки*). Температуры первичных реперных точек приведены в табл. 1. Между первичными реперными точками температурная шкала устанавливается с помощью *интерполяционных формул*, дающих соотношение между температурой и показаниями стандартных термометров (платиновый термометр, платинородиевая термопара, оптический пирометр), градуированных по этим точкам. Вся область температур, охватываемая МПТШ-68, делится на ряд интервалов, в каждом из которых рекомендуются свои методы воспроизведения температур и свои интерполяционные формулы.

В интервале от 13,81 K (тройная точка водорода) до 630,74 °C (точка затвердевания сурьмы) стандартным прибором является платиновый термометр сопротивления. Весь этот интервал разбивается на пять более мелких интервалов, в каждом из которых для платинового термометра применяются свои интерполяционные формулы. (Мы не приводили их, так как они занимают довольно много места.)

Выше температуры 630,74 °C до 1064 °C (температура затвердевания золота) Международная практическая температурная шкала устанавливается с помощью платинородиевой термопары Pt 10 Rh/Pt. Если один спай термопары находится при температуре 0 °C, то температуры t другого спая связана с электродвижущей силой $\mathscr{E}$ термопары

Таблица 1

Реперные точки МПТШ-68		Температура	
		K	°C
Опорная	Тройная точка H_2O	273,16	0,01
Первичные	Тройная точка кипения H_2 [1])	13,81	−259,34
	Точка кипения H_2 [1]) при 25/76 атм	17,042	−256,108
	Точка кипения H_2 [1])	20,28	−252,87
	Тройная точка O_2	54,361	−218,789
	Тройная точка Ar	83,798	−189,352
	Точка кипения O_2	90,188	−182,962
	Точка кипения H_2O	373,15	100
	Точки затвердевания		
	Sn	505,1181	231,9681
	Zn	692,73	419,58
	Ag	1235,08	961,93
	Au	1337,58	1064,43
Вторичные	Точка кипения ^4He	4,215	−268,935
	Тройная точка N_2	63,148	−210,002
	Точка кипения N_2	77,348	−195,802
	Точка сублимации CO_2	194,674	−78,476
	Точка затвердевания Hg	234,288	−38,862
	Точка замерзания H_2O	273,15	0
	Тройная точка феноксибензина	300,02	26,87
	Точки затвердевания		
	In	429,784	156,634
	Bi	544,592	271,442
	Cd	594,258	321,108
	Pb	600,652	327,502
	Точки кипения		
	Hg	629,81	356,66
	S	717,824	444,674
	Точки затвердевания эвтектики Cu–Al	821,38	548,23
	Sb	903,89	630,74
	Al	933,52	660,37
	Cu	1357,6	1084,5
	Ni	1728	1455
	Co	1767	1494
	Pd	1827	1554
	Pt	2045	1772
	Rh	2236	1963
	Ir	2720	2447
	W	3660	3387

[1]) Равновесная смесь орто- и параводорода.

Примечание. За исключением температур тройных точек и $T = 17{,}042$ K приведенные в этой таблице значения относятся к давлению в одну стандартную атмосферу.

соотношением

$$\mathscr{E} = a + bt + ct^2,$$

где постоянные a, b, c вычисляются из измерений ЭДС в точках затвердевания сурьмы, серебра и золота.

В реперных точках Международная практическая температурная шкала согласуется очень трудно с термодинамической шкалой. В промежутках между ними расхождения с термодинамической шкалой настолько малы, что ими в большинстве практических измерений можно пренебречь.

Выше температуры затвердевания золота применяется оптический пирометр, проградуированный с помощью законов теплового излучения тел.

Для установления температурной шкалы в диапазоне 0,8–5 К, как правило, используют измерения давления насыщенных паров гелия. Ниже 0,8 К температура обычно определяется путем измерений либо давления паров редкого изотопа ^{3}He (до 0,4 К), либо магнитной восприимчивости некоторых парамагнитных солей (до 10^{-3} К).

§ 7. Законы идеальных газов

1. *Идеальными называются газы, строго подчиняющиеся законам Бойля–Мариотта и Гей-Люссака.* Эти законы выражаются уравнениями (4.1) и (4.3) при дополнительном условии, что постоянная α одинакова для всех идеальных газов. Все прочие газы называются *реальными*.

Уравнение (4.1), связывающее температуру, давление и объем идеального газа в состоянии теплового равновесия, называется его *уравнением состояния*. Входящая в это уравнение постоянная C пропорциональна массе и зависит от химической природы газа. Введем для измерения количества вещества новую единицу — *моль*, масса которой различна для различных веществ. Молем какого-либо вещества называется количество этого вещества, содержащее столько молекул, сколько атомов содержится в 12 г изотопа углерода ^{12}C. Тысяча молей называется *киломолем*. Таким образом, по определению, в одном моле различных веществ содержится одно и то же число молекул. В кинетической теории газов будет показано, что для идеальных газов справедлив *закон Авогадро*. Согласно этому закону *в одинаковых объемах идеальных газов при одинаковых температурах и давлении содержится одно и то же число молекул*. Отсюда следует, что при одинаковых температурах произведение PV для одного моля различных идеальных газов имеет одно и то же значение. На этом основано измерение моля в граммах, а также молекулярных масс газов. Найдено, что моль кислорода содержит приблизительно 32 г, моль азота — 28 г, моль водорода — 2,016 г и т. д. Эти числа в то же время дают значения *молекулярных масс* соответствующих веществ.

До 1961 г. определение моля и молекулярных масс основывалось на *кислородной шкале*. Существовало две кислородных шкалы: *физическая* и *химическая*. В физической шкале основному изотопу кислорода ^{16}O приписывалась атомная масса точно 16. В химической шкале принималось, что средняя атомная масса естественной смеси изотопов кислорода равна точно 16. Однако при масс-спектрометрических измерениях более удобна *углеродная шкала*, так как углерод дает гораздо больше различных соединений, чем кислород. Углеродная шкала и была принята в 1961 г. Международным союзом по чистой и прикладной химии. В этой шкале изотопу углерода ^{12}C приписывается атомная масса точно 12. Углеродная шкала очень мало отличается от физической кислородной шкалы.

Для одного моля или одного киломоля идеального газа уравнение состояния запишется в виде

$$pV = RT, \tag{7.1}$$

где постоянная R, ввиду закона Авогадро, имеет одно и то же значение для всех газов. На это было указано Д. И. Менделеевым (1834–1907). Уравнение (7.1) обычно называют *уравнением Клапейрона* (1799–1864) или *Клапейрона–Менделеева*, а постоянную R — *универсальной газовой постоянной*. Ее числовое значение можно найти, измерив объем V моля идеального газа при нормальном давлении ($P = 1{,}01325$ Па $= 1{,}01325 \cdot 10^6$ дин/см^2) и при температуре 0 °C ($T = 273{,}15$ K). По современным данным $V = (22413{,}6 \pm 0{,}9)$ см^3/моль. Используя это значение, нетрудно получить $R = (8{,}31434 \pm 0{,}00035) \cdot 10^7$ эрг/(моль · K) $= (8{,}31434 \pm 0{,}00035)$ Дж/(моль · K) $= 0{,}0821$ л · атм/(моль · K).

Если газ содержит ν молей, то уравнение Клапейрона–Менделеева примет вид

$$PV = \nu RT. \tag{7.2}$$

2. Найдем теперь уравнение состояния для смеси идеальных газов. Для этого воспользуемся эмпирически установленным *законом Дальтона* (1766–1844). Пусть в различных сосудах одинакового объема V заключены различные идеальные газы, поддерживаемые при одной и той же температуре T. Обозначим давления этих газов через $P_1, P_2, P_3, \ldots$ Какое давление P установится, если все газы смешать в том же объеме и поддерживать прежнюю температуру смеси T? Закон Дальтона утверждает, что

$$P = P_1 + P_2 + P_3 + \ldots \tag{7.3}$$

Давления $P_1, P_2, P_3, \ldots$ называются *парциальными давлениями* газов, входящих в смесь. Таким образом, по закону Дальтона *давление смеси идеальных газов равно сумме парциальных давлений этих газов*. Пусть ν_i — число молей i-го газа. Тогда $P_i V = \nu_i RT$. Поэтому, умножая обе части соотношения (7.3) на V, получим

$$PV = \nu RT,$$

где ν — общее число молей в смеси, т. е. $\nu = \nu_1 + \nu_2 + \nu_3 + \ldots$ Уравнение состояния смеси идеальных газов, таким образом, имеет такой же вид, как и уравнение состояния химически однородного идеального газа.

Поэтому на основании уравнения состояния идеального газа нельзя решить, имеем ли мы дело с химически однородным газом или с механической смесью таких газов.

§ 8. Уравнение состояния и его следствия для бесконечно малых процессов

1. Опыт показывает, что в состоянии термодинамического равновесия объем V, давление P и температура T находятся в функциональной зависимости не только для идеальных газов, но и для реальных газов, а также для любых физически однородных и изотропных тел. Эту функциональную зависимость можно выразить уравнением

$$f(P, V, T) = 0. \tag{8.1}$$

Вид функции $f(P, V, T) = 0$ различен для различных тел. Соотношение (8.1) называется *уравнением состояния тела*. Для идеальных газов уравнением состояния является уравнение Клапейрона–Менделеева (7.1). Реальные газы лишь приблизительно следуют уравнению Клапейрона–Менделеева. К нему необходимо ввести поправки, простейшие из которых будут рассмотрены в гл. VII.

Уравнение состояния принадлежит к числу важнейших характеристик макроскопических свойств физически однородных тел. Как уже отмечалось во введении, его нельзя вывести теоретически из общих принципов термодинамики. Термодинамика заимствует уравнения состояния либо из опыта, либо из статистической физики, где они могут быть выведены теоретически.

2. Ввиду наличия уравнения состояния изменения величин P, V, T при термодинамическом равновесии не независимы, а связаны определенным соотношением. Если изменения состояния бесконечно малы, то это соотношение может быть установлено без знания конкретного вида функции $f(P, V, T)$. С этой целью разрешим уравнение (8.1) относительно одного из переменных, например V, т. е. представим объем V в виде функции остальных двух переменных P и T: $V = V(P, T)$. Если поддерживать температуру постоянной, а давление изменить на бесконечно малую величину dP, то объем V получит также бесконечно малое приращение, определяемое выражением

$$d_1 V = \left(\frac{\partial V}{\partial P}\right)_T dP.$$

Значок T у производной $(\partial V/\partial P)_T$ указывает на то, что при дифференцировании V и P температура T должна оставаться постоянной. Производные, получаемые дифференцированием какой-либо функции двух или нескольких аргументов по одному из них в предположении, что все остальные аргументы остаются постоянными, называются в математике *частными производными*. Таким образом, $(\partial V/\partial P)_T$ есть частная производная объема по давлению при постоянной температуре. Пусть теперь давление P поддерживается постоянным, а

температура T получает бесконечно малое приращение dT. Тогда соответствующее приращение объема V представится выражением

$$d_2V = \left(\frac{\partial V}{\partial P}\right)_P dT.$$

Если, наконец, изменяются и давление P и температура T, то с точностью до бесконечно малых высшего порядка приращение объема представится суммой $dV = d_1V + d_2V$, или

$$dV = \left(\frac{\partial V}{\partial P}\right)_T dP + \left(\frac{\partial V}{\partial T}\right)_P dT. \tag{8.2}$$

Это соотношение и решает поставленную задачу.

Соотношение (8.2) справедливо при любых бесконечно малых приращениях dP и dT. Приращения dP и dT могут поэтому рассматриваться как независимые переменные. Но формула (8.2) останется в силе и в том случае, когда на изменения P и T наложено какое-либо ограничение. Допустим, например, что тело участвует в процессе, при котором давление P является определенной функцией температуры T. Тогда приращения dP и dT перестанут быть независимыми. Например, для процесса при постоянном объеме $dV = 0$, и соотношение (8.2) переходит в

$$\left(\frac{\partial V}{\partial P}\right)_T dP + \left(\frac{\partial V}{\partial T}\right)_P dT = 0.$$

Если разрешить это уравнение относительно dP/dT, то полученная таким образом величина даст частную производную $(\partial P/\partial T)_V$, так как dP и dT означают приращения давления и температуры при постоянном объеме. Таким образом,

$$\left(\frac{\partial P}{\partial T}\right)_V = -\frac{(\partial V/\partial T)_P}{(\partial V/\partial P)_T}.$$

Ввиду очевидного соотношения

$$\left(\frac{\partial V}{\partial P}\right)_T = \frac{1}{(\partial P/\partial V)_T},$$

последнее тождество может быть записано в виде

$$\left(\frac{\partial P}{\partial T}\right)_V = -\left(\frac{\partial P}{\partial V}\right)_T \left(\frac{\partial V}{\partial T}\right)_P, \tag{8.3}$$

а также в виде

$$\left(\frac{\partial P}{\partial V}\right)_T \left(\frac{\partial V}{\partial T}\right)_P \left(\frac{\partial T}{\partial P}\right)_V = -1. \tag{8.4}$$

Примечательно, что для справедливости этих тождеств совершенно не существенно, какой физический смысл имеют величины P, V и T. Они являются чисто математическими тождествами, выражающими в дифференциальной форме наличие функциональной связи между величинами P, V и T. Каковы бы ни были величины x, y, z, связанные функциональной зависимостью $f(x, y, z) = 0$, между их частными

производными существует соотношение

$$\left(\frac{\partial x}{\partial y}\right)_z = -\left(\frac{\partial x}{\partial z}\right)_y \left(\frac{\partial z}{\partial y}\right)_x, \tag{8.5}$$

или

$$\left(\frac{\partial x}{\partial y}\right)_z \left(\frac{\partial y}{\partial z}\right)_x \left(\frac{\partial z}{\partial x}\right)_y = -1. \tag{8.6}$$

Эти математические тождества широко используются в термодинамике.

3. Не лишне предостеречь читателя от искушения сократить выражение в левой части (8.4) на ∂P, ∂V и ∂T. В результате такого «сокращения» получился бы неправильный результат $+1$ вместо правильного -1. Сокращение нельзя производить потому, что величины ∂P, ∂V и ∂T в числителе имеют иной смысл, чем аналогичные величины в знаменателе. Например, ∂V в знаменателе есть приращение объема, испытываемое им при увеличении давления на ∂P, когда температура T остается постоянной. Величина же ∂V в числителе означает приращение того же объема при повышении температуры на ∂T при условии, что постоянным остается давление. Таким образом, речь идет о совершенно различных приращениях объема, ввиду чего и нельзя сокращать на ∂V. Но формула (8.4) показывает, что формально можно производить такое «сокращение» при условии, что оно сопровождается изменением знака рассматриваемого выражения. Это дает удобное правило для запоминания тождества (8.4).

Соотношение (8.3) напоминает правило дифференцирования функции от функции. Оно отличается от этого правила знаком минус в правой части. Это по-прежнему объясняется тем, что величины ∂V в знаменателе и числителе соотношения (8.3) имеют разный смысл, указанный выше.

4. Тождество (8.3) или (8.4) позволяет установить связь между температурным коэффициентом объемного расширения, температурным коэффициентом давления и модулем объемного сжатия физически однородного и изотропного вещества.

Температурным коэффициентом объемного расширения α называется отношение приращения объема тела при нагревании на 1 К к его объему V_0 при 0 °C при условии, что давление P поддерживается постоянным. Таким образом, по определению можно написать

$$\alpha = \frac{V_{T+1} - V_T}{V_0} \quad (P = \text{const}).$$

Объем, как правило, относительно медленно изменяется с температурой. Поэтому приведенное выражение практически не отличается от теоретически более совершенного выражения

$$\alpha = \frac{1}{V_0}\left(\frac{\partial V}{\partial T}\right)_P. \tag{8.7}$$

Температурным коэффициентом давления β называется отношение увеличения давления тела при нагревании на 1 К к давлению P_0 при 0 °С при условии, что объем тела поддерживается постоянным. Переходя снова от отношения конечных приращений к производным, можем написать

$$\beta = \frac{1}{P_0}\left(\frac{\partial P}{\partial T}\right)_V. \tag{8.8}$$

Наконец, *модулем* (точнее, *изотермическим модулем*) *объемного*, или *всестороннего*, *сжатия вещества* K называется отношение бесконечно малого приращения давления к вызванному им относительному сжатию вещества при постоянной температуре:

$$K = \partial P : \left(-\frac{\partial V}{V}\right)_T = -V\left(\frac{\partial P}{\partial V}\right)_T. \tag{8.9}$$

Ввиду тождества (8.4) между величинами α, β и K должно существовать соотношение

$$\frac{V_0}{P_0 V}\,\frac{\alpha K}{\beta} = 1. \tag{8.10}$$

Опыт подтверждает справедливость такого соотношения, а с ним и наличие функциональной связи между P, V и T.

Возьмем для примера ртуть при 0 °С и атмосферном давлении. Коэффициент расширения

$$\alpha = 1{,}8 \cdot 10^{-4}\ \text{К}^{-1},$$

изотермический модуль объемного сжатия

$$K = 2{,}56 \cdot 10^5\ \text{атм}.$$

Следовательно,

$$\beta = \alpha K / P_0 = 46\ \text{К}^{-1}.$$

Отсюда следует, что для сохранения постоянным объема ртути при нагревании от 0 до 1 °С требуется увеличение давления приблизительно на 46 атм.

§ 9. Макроскопические параметры

1. В классической механике мгновенное состояние механической системы определяется *координатами* и *скоростями* частиц, из которых состоит система. В молекулярной физике буквальное применение такого способа описания состояний физических систем сводилось бы к определению в каждый момент времени координат и скоростей всех молекул и атомов, а также электронов, атомных ядер и прочих частиц, из которых построены тела. Состояние, описанное столь детально, называется *динамическим состоянием* или *микросостоянием*. Квантовая механика дает иной способ описания микросостояний, на котором здесь нет необходимости останавливаться. Важно заметить

только, что подобное детальное описание состояний макроскопических систем, ввиду колоссальности числа частиц в них, не только невозможно осуществить фактически, но оно само по себе не представляет никакого интереса. Понятие микросостояния в классическом или квантовом смысле полезно лишь постольку, поскольку оно может быть связано с *макроскопическими свойствами вещества* и может служить для определения последних. В термодинамике равновесные состояния макроскопических систем описываются несравненно более грубо — с помощью небольшого числа различных *макроскопических параметров*. К ним относятся например, давление, плотность, температура, концентрация, объем системы, напряженность электрического и магнитного полей и т. д. Состояние, описанное с помощью макроскопических параметров, называется *макроскопическим состоянием* или *макросостоянием*. Именно в этом смысле понятие состояния употребляется в термодинамике. Термодинамически равновесное состояние газа, например, в отсутствие внешних силовых полей полностью определяется его массой, химической природой, давлением и температурой. Объем V, в котором заключен газ, не является независимым параметром, а может быть вычислен из уравнения состояний. Нельзя указать все макроскопические параметры, определяющие состояние системы, не конкретизируя последнюю и внешние условия, в которых она находится.

2. Чтобы выяснить смысл макроскопических параметров с молекулярной точки зрения, рассмотрим в качестве примера плотность газообразного, жидкого или твердого тела. Выделим мысленно в пространстве малую неизменную область с объемом V. Пусть M — масса содержащегося в ней вещества. Плотностью вещества внутри объема V называется отношение $\rho = M/V$. Ввиду теплового движения число молекул или атомов в объеме V, а с ним и масса M непрерывно и беспорядочно меняются с течением времени. Это ведет к беспорядочному изменению и плотности ρ. Такие беспорядочные изменения плотности или других физических величин называются *флуктуациями*. Обозначим через $\rho_1, \rho_2, \ldots, \rho_n$ значения величины ρ в равноотстающие моменты времени $t_1, t_2, \ldots, t_n$. Их среднее арифметическое по определению равно $(1/n)(\rho_1 + \rho_2 + \ldots + \rho_n)$. Опыт показывает, что при неизменных внешних условиях эта величина приближается к определенному пределу $\overline{\rho}$, когда число n, а также общее время наблюдения $t_n - t_1$ становятся достаточно большими. Величина $\overline{\rho}$ при макроскопическом описании и принимается за плотность тела. Если в объеме V содержится одна или небольшое число молекул, то отклонения мгновенных значений величины ρ от ее среднего значения $\overline{\rho}$ очень велики — они сравнимы с самой плотностью $\overline{\rho}$. При увеличении объема V отклонения становятся все меньше и меньше. Для больших объемов, содержащих огромное количество молекул, флуктуации плотности мало заметны. Феноменологическая термодинамика от флуктуаций отвлекается.

Подобно плотности ведут себя и другие макроскопические параметры, например давление газа. Давление газа на стенку сосуда есть результат ударов о стенку молекул, беспорядочно движущихся с тепловыми скоростями. Возьмем малый участок стенки с площадью S. Пусть F означает мгновенное значение силы, с которой действуют на этот участок ударяющие молекулы. Сила, отнесенная к единице площади, равна $P = F/S$. Величина P, как и плотность, беспорядочно флуктуирует во времени. В термодинамике имеют дело не с мгновенными, а со средними значениями величины P за большие промежутки времени — очень большие по сравнению с длительностью одного удара, с длительностью промежутка времени между двумя последовательными ударами или другими характерными временами. Эту среднюю величину $\overline{P}$ при макроскопическом рассмотрении и принимают за давление газа. То обстоятельство, что давление газа воспринимается нашими грубыми измерительными приборами как сила, непрерывная во времени и непрерывно распределенная по площади, на которую она действует, объясняется колоссальностью числа молекул, бомбардирующих эту площадь, а также исключительной малостью этих частиц. При повышении чувствительности приборов регистрируются и самопроизвольные беспорядочные колебания давления — тепловые флуктуации.

Из этих примеров видно, что макроскопические термодинамические параметры имеют смысл *средних значений* (за большой промежуток времени) каких-то функций, характеризующих динамическое состояние системы. Только с подобными средними и имеет дело феноменологическая термодинамика. Их мы будем обозначать через $\rho, P, V, \ldots$, т. е. будем опускать черту или любой другой знак усреднения.

3. Макроскопические параметры, определяющие состояние системы и ее отношение к окружающим телам, разделяются на *внутренние* и *внешние*. Внутренние параметры определяют *внутреннее состояние системы*. Внешними параметрами характеризуются *внешние тела* и *силовые поля*, воздействующие на систему. Приведем примеры внутренних и внешних параметров.

Пусть газ заключен в сосуде с твердыми стенками. Объем сосуда определяется положением внешних тел — стенок. Это внешний параметр. Давление, оказываемое газом на стенку сосуда, зависит от скоростей теплового движения его молекул. Оно является внутренним параметром. Если газ двух- или многоатомный, то при его нагревании молекулы диссоциируют, т. е. распадаются на атомы или группы атомов. При дальнейшем нагревании атомы газа ионизируются — расщепляются на заряженные ионы или электроны. Отношение числа диссоциировавшихся молекул к общему их числу называется *степенью диссоциации* молекул газа. Аналогично определяется *степень ионизации* газа, как отношение числа ионизовавшихся атомов к общему числу

их. Степень диссоциации и степень ионизации газа — внутренние параметры. В электрическом поле газ поляризуется, а в магнитном — намагничивается, возникают электрический и магнитный моменты газа. Эти величины являются внутренними параметрами. Напряженности же внешних электрического и магнитного полей, в которых помещен газ, являются параметрами внешними.

Допустим теперь, что газ заключен в цилиндре под поршнем, который может свободно перемещаться. Пусть на поршне лежит груз веса Q. Если площадь поршня равна S, то груз Q оказывает давление $P = Q/S$ на каждую единицу площади поршня. Давление P в таком смысле будет внешним параметром, так как оно определяется весом внешнего тела — груза Q. Объем газа V определяется положением подвижной стенки сосуда — поршня. Однако теперь объем V становится внутренним параметром, так как положение поршня с грузом зависит от внутреннего давления, оказываемого газом на поршень.

В состоянии термодинамического равновесия каждый внутренний параметр является функцией внешних параметров и температуры системы. Это утверждение является результатом обобщения опытных фактов. Примером может служить уравнение состояния. Внутренний параметр — давление газа P — здесь однозначно определяется температурой газа и внешним параметром — объемом сосуда V, в котором заключен газ. Сформулированное общее положение является одним из важнейших постулатов аксиоматической термодинамики. Уравнение, выражающее функциональную связь между внутренними и внешними параметрами системы в состоянии термодинамического равновесия, называется *обобщенным уравнением состояния системы.*

4. Следует заметить, что даже в состоянии равновесия внутренние параметры системы определяются внешними параметрами и температурой окружающей среды *не всегда однозначно*. Так, твердый брусок на твердой наклонной плоскости в однородном поле тяжести может удерживаться силами сухого трения на различных высотах. Тело из пластического материала, деформированное внешними силами, после удаления последних принимает различные равновесные формы, в зависимости от того, каковы были эти силы. Ферромагнитное твердое тело из-за гистеризиса может быть намагничено по-разному в одном и том же магнитном поле, в зависимости от того, как создавалось это поле. То же относится к поляризации сегнетоэлектриков в электрическом поле (см. т. III, § 39,74). Во всех этих и аналогичных случаях одним и тем же значениям внешних параметров и температуры окружающей среды соответствует множество равновесных состояний системы, отличающихся различными значениями внутренних параметров.

Возникает вопрос, не существует ли среди этого множества равновесных состояний состояние, *наиболее устойчивое* по отношению к малым возмущениям внешних условий, в которое в конце концов самопроизвольно переходят все остальные менее устойчивые равновесные состояния. Если бы это было так, то такое единственное самое устойчивое состояние было бы целесообразно принять за состояние *истинного*

термодинамического равновесия, и все прочие равновесные состояния считать состояниями *заторможенного равновесия*. Такая гипотеза не противоречит наблюдаемым фактам. Более того, она лежит в основе статистической физики, важнейшей задачей которой является нахождение значений различных внутренних параметров именно в состоянии термодинамического равновесия.

В свете изложенной гипотезы окружающие нас предметы находятся в состоянии не истинного, а заторможенного равновесия. Последнее воспринимается нами как истинное термодинамическое равновесие только потому, что релаксационные процессы, приближающие системы к такому равновесию, идут *чрезвычайно медленно* и потому не замечаются нами. Соответствующие им времена релаксации могут быть порядка возраста Земли и даже значительно больше.

В дальнейшем, если не сделано специальной оговорки, рассматриваются системы, равновесные состояния которых при заданных внешних условиях единственны.

Глава II
ПЕРВОЕ НАЧАЛО ТЕРМОДИНАМИКИ

§ 10. Введение

Первое начало термодинамики выражает принцип сохранения энергии для тех макроскопических явлений, в которых одним из существенных параметров, определяющих состояние тел, является температура. Его открытие относится к 40-м годам XIX века, когда было установлено, что теплота — не вещество, а какое-то *внутреннее движение* тела.

В механике энергия складывается из кинетической энергии макроскопического движения и потенциальной энергии макроскопических тел во внешних силовых полях. В механике доказывается, что для изолированной системы полная механическая энергия сохраняется, т. е. ее количество остается неизменным. Но это справедливо не всегда, а только в тех случаях, когда все действующие в системе силы *консервативны*. При наличии диссипативных сил — сил трения — механическая энергия замкнутой системы уменьшается. Однако опыты показали, что работа диссипативных сил всегда сопровождается выделением теплоты. Оказалось, что принцип сохранения энергии остается справедливым и при наличии диссипативных сил, если только расширить понятие энергии введением новой формы ее, а именно *внутренней энергии*, называемой также (не совсем удачно) *тепловой энергией*. Это было сделано Майером (1814–1878), Джоулем (1818–1889), Гельмгольцем (1821–1894) и другими учеными, с именами которых связывают открытие принципа сохранения энергии в его общефизическом смысле.

С атомистической точки зрения нарушения механического закона сохранения энергии объясняются тем, что макроскопическая механика учитывает не все движения и силовые взаимодействия. Из ее поля зрения ускользают *внутренние невидимые движения отдельных атомов и молекул*, а также силы взаимодействия между ними. Диссипативные силы годятся только для описания макроскопических движений. В мире атомов и молекул никаких диссипативных сил не существует, все силы консервативные и гироскопические. Всякое тело с атомистической точки зрения является консервативной системой колоссального числа частиц — атомов и молекул. Для такой системы механический закон сохранения энергии справедлив, но только при непременном условии, что к энергии макроскопического движения добавляется *энергия беспорядочного атомно-молекулярного движения*. Это последнее движение и есть теплота. Однако последовательная

феноменологическая термодинамика должна вводить расширенное понятие механической энергии, не обращаясь к атомистическим представлениям, руководствуясь одними только общими принципами, являющимися обобщениями опытных фактов. Ближайшей нашей целью и является введение этого понятия.

§ 11. Квазистатические процессы

1. В теоретических методах термодинамики широко используются так называемые *квазистатические*, или *квазиравновесные*, процессы, т. е. идеализированные процессы, состоящие из непрерывно следующих друг за другом состояний равновесия.

В качестве примера возьмем цилиндр с поршнем, который может свободно перемещаться в нем без трения. Пусть в цилиндре находится газ и поршень удерживается в положении равновесия лежащим на нем грузом Q (рис. 6). Если снять часть груза или добавить новый, то равновесие нарушится. Возникнет довольно сложное движение газа и колебания поршня с грузом. Но в конце концов все эти движения затухнут, и рассматриваемая система вновь придет в состояние равновесия. Подобные неравновесные процессы очень сложны. Квазистатические процессы несравненно более просты. Чтобы заставить газ в цилиндре расширяться или сжиматься квазистатически, допустим, что разгрузка или нагрузка поршня идет бесконечно малыми порциями. Для наглядности можно вообразить, что поршень нагружен мелким песком. Снимем или добавим одну песчинку. От этого равновесие нарушится бесконечно мало. Когда оно восстановится, снимем или добавим вторую песчинку. Повторив эту операцию много раз, можно в конце концов снять весь груз или изменить его на любую конечную величину. Вместе с грузом изменится на конечную величину и объем газа. Этот процесс состоит из последовательности бесконечно малых процессов, каждый из которых лишь ничтожно мало нарушает состояние равновесия. В пределе, когда масса каждой песчинки стремится к нулю, а полное число песчинок — к бесконечности, получится бесконечно медленный процесс, состоящий из последовательности равновесных состояний. Такой предельный процесс называется *равновесным*. Логически равновесный процесс надо отличать от квазиравновесного, или квазистатического, процесса. Однако надо иметь в виду, что строго равновесный процесс никогда не реализуется в природе. Поэтому в большинстве рассуждений квазиравновесный процесс называют просто равновесным, опуская приставку «квази». Это будем делать и мы, когда такое сокращение не вызывает недоразумений.

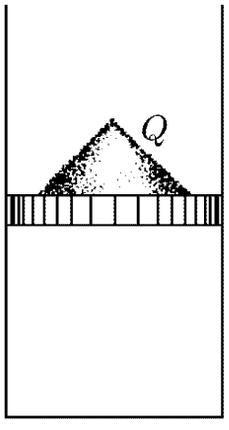


Рис. 6

2. Значение квазистатических процессов состоит в том, что они сильно упрощают термодинамические исследования. Это объясняется

тем, что для мгновенного описания состояния системы, совершающей квазистатический процесс, требуется *столько же параметров, сколько и для макроскопического описания равновесного состояния*. В случае газа таких параметров два, например объем и температура. Для более сложных систем число параметров может быть другим, но, если процесс квазистатический, оно, как правило, невелико. Напротив, для описания состояния системы, совершающей какой-либо сложный неквазистатический процесс, например турбулентное движение жидкости или газа, требуется, вообще говоря, *бесконечное множество даже макроскопических параметров*.

Квазистатические процессы в строгом смысле этого слова никогда не реализуются в природе. Они являются абстракциями. Но к ним можно подойти сколь угодно близко. Очень многие реальные процессы, идущие с конечными скоростями, часто могут считаться приблизительно квазистатическими. Таковы, например, процессы расширения газов в цилиндрах тепловых двигателей или компрессоров. Образование сгущений или разрежений воздуха в звуковой волне также может рассматриваться как приблизительно квазистатический процесс. Конечно, не всякий медленный процесс является равновесным. Примером может служить процесс теплопроводности.

3. В термодинамике часто встречаются следующие квазистатические процессы: 1) *изохорный процесс* — процесс, происходящий при постоянном объеме ($V = \text{const}$); 2) *изобарный процесс* — процесс, в котором давление остается постоянным ($P = \text{const}$); 3) *изотермический процесс* — процесс, происходящий при постоянной температуре ($T = \text{const}$). Как и все квазистатические процессы, указанные процессы можно графически изобразить непрерывными линиями (см. следующий параграф). Соответствующие кривые называются *изохорой* ($V = \text{const}$), *изобарой* ($P = \text{const}$) и *изотермой* ($T = \text{const}$).

§ 12. Макроскопическая работа

1. Рассмотрим снова газ в цилиндре с поршнем (рис. 7). Вычислим бесконечно малую или *элементарную работу* δA, совершаемую газом при бесконечно малом квазистатическом расширении, в котором его объем увеличивается на dV. Сила давления газа на поршень равна $F = PS$, где S — площадь поршня. Если поршень переместится на расстояние dx, то газ совершит работу $\delta A = F\,dx = PS\,dx$ или

$$\delta A = P\,dV, \tag{12.1}$$

так как приращение объема равно $dV = S\,dx$.

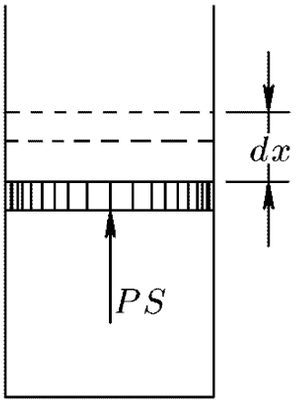


Рис. 7

Выражение (12.1) справедливо и в общем случае квазистатического изменения объема любого тела, находящегося под постоянным внешним давлением. Допустим, например, что газ заключен в

мягкую эластичную оболочку и эта оболочка квазистатически расширяется (рис. 8). Работа, совершаемая газом при перемещении элемента площади dS оболочки на расстояние dn вдоль нормали, равна $P\,dS\,dn$, или $P\,dV$, где $dV = dS\,dn$ — элементарный объем, заштрихованный на рис. 8. Чтобы найти элементарную работу δA при перемещении всех элементов оболочки, надо выражение $P\,dV$ проинтегрировать по объему слоя между двумя последовательными бесконечно близкими положениями оболочки. Так как давление P одно и то же по всей оболочке, то его можно вынести из-под знака интеграла. Таким путем получится $\delta A = P\int dV = P\Delta V$, где ΔV — объем вышеуказанного слоя, равный приращению объема газа в рассматриваемом процессе. Введя для него прежнее обозначение dV, мы снова придем к формуле (12.1). Для применимости вывода несущественно, что в оболочке помещен газ. Вывод справедлив для любого вещества, находящегося под постоянным давлением. Несущественно также наличие оболочки. Ее роль может играть поверхность тела.

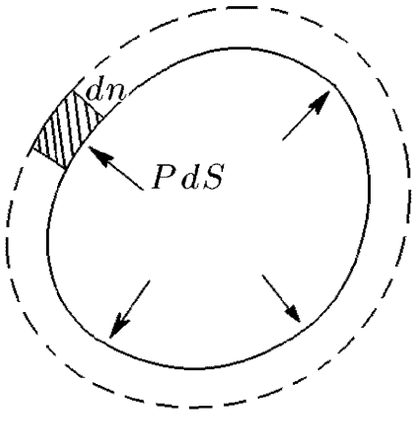


Рис. 8

В случае квазистатических процессов внутреннее давление P газа в пределе всегда равно внешнему давлению на поршень $P_{\text{внеш}}$. Только тогда внутреннее состояние газа может быть охарактеризовано двумя параметрами P и V, и только тогда процесс может быть равновесным и идти бесконечно медленно. В противном случае возникает ускоренное макроскопическое движение поршня и частей газа с конечными скоростями и для описания внутреннего состояния газа потребуется бесконечное множество параметров. Если процесс неквазистатический, но внешнее давление по всей поверхности системы одно и то же, то работа внешних сил представится выражением

$$\delta A_{\text{внеш}} = -P_{\text{внеш}}\,dV. \tag{12.2}$$

Для квазистатических процессов $P_{\text{внеш}} = P$, а потому $\delta A_{\text{внеш}} = -\delta A$. (В общем случае работа тела A над телом B не равна работе тела B над телом A.) В случае неравновесных процессов, происходящих с ускорением, это, вообще говоря, несправедливо и для работы системы нельзя написать никакого простого выражения. Таким образом, формула (12.1) относится только к случаю квазистатических процессов. В этом параграфе предполагается, что все процессы — квазистатические.

2. Чтобы от элементарной работы δA перейти к работе для конечного процесса, надо вычислить интеграл

$$A = \int dV. \tag{12.3}$$

Однако такое вычисление возможно только тогда, когда давление является определенной функцией объема V. Между тем, согласно

уравнению состояния, P зависит не только от V, но и от T. Меняя в ходе процесса различным образом температуру системы, можно квазистатически перевести систему из начального состояния в конечное бесчисленным множеством способов. Каждому из этих способов соответствует своя функция $P = P(V)$ и свое значение интеграла в формуле (12.3). Таким образом, *работа A не определяется заданием начального и конечного состояний системы. Ее величина зависит также от способа или пути перехода системы из начального состояния в конечное*. Про величины такого рода говорят, что они *не являются функциями состояния*. Напротив, величины, имеющие вполне определенные значения в каждом состоянии системы, называются *функциями состояния*. Такова, например, температура системы в состоянии термодинамического равновесия.

3. Разобраться в существе дела проще всего с помощью графического метода. Он использует то обстоятельство, что равновесное состояние физически однородного и изотропного тела полностью определяется заданием двух параметров, например V и P. Температура T может быть найдена по ним из уравнения состояния $T = T(V, P)$. Состояние тела задается точкой на координатной плоскости, причем по горизонтальной оси откладывается объем V, а по вертикальной — давление P. Такая плоскость для краткости называется плоскостью VP. Когда система совершает квазистатический процесс, точка, изображающая ее состояние, описывает на плоскости VP непрерывную линию. Таким образом, квазистатические процессы изображаются непрерывными кривыми. Вместо переменных V, P можно пользоваться переменными T, V или T, P. Однако для графического представления работы наиболее удобны переменные V, P. Неравновесные состояния и неравновесные процессы нельзя изображать точками и кривыми на плоскости, так как для задания неравновесного состояния двух параметров недостаточно. Неравновесные состояния характеризуются, вообще говоря, бесконечным множеством параметров.

Пусть система квазистатически переходит из состояния M в состояние N вдоль кривой $M1N$ (рис. 9). Эта кривая определяет давление P как вполне определенную функцию объема V. После этого работа системы A определится однозначно. Она численно равна площади «криволинейной трапеции» $M_1M1NN_1M_1$. Если систему заставить переходить из того же начального в то же конечное состояние вдоль другой кривой $M2N$, то соответствующая работа A_1 изобразится другой площадью $M_1M2NN_1M_1$. Вообще говоря, $A_1 \neq A$.

Вычислим, например, работу, совершаемую одним молем идеального газа при изотермическом расширении (т. е. при таком расширении, когда температура газа поддерживается постоянной). Для наглядности представим себе газ, заключенный в цилиндр с поршнем, на котором находится груз. Будем бесконечно медленно и непрерывно уменьшать нагрузку на поршень и в то же время подогревать газ, чтобы обеспечить постоянство его температуры во время расширения. Графически процесс расширения газа изобразится на плоскости VP

гиперболой $PV = RT = \text{const}$ (рис. 10). Работа, совершаемая газом,

$$A = \int_{V_1}^{V_2} P\,dV = RT \int_{V_1}^{V_2} \frac{dV}{V} = RT \ln \frac{V_2}{V_1}. \tag{12.4}$$

Можно перевести газ из начального состояния *1* в конечное состояние *2* бесчисленным множеством других способов. Например, можно, сохраняя давление газа постоянным (т. е. не изменяя нагрузку на поршень), нагреть газ, доведя его объем до значения $V = V_2$. Этот

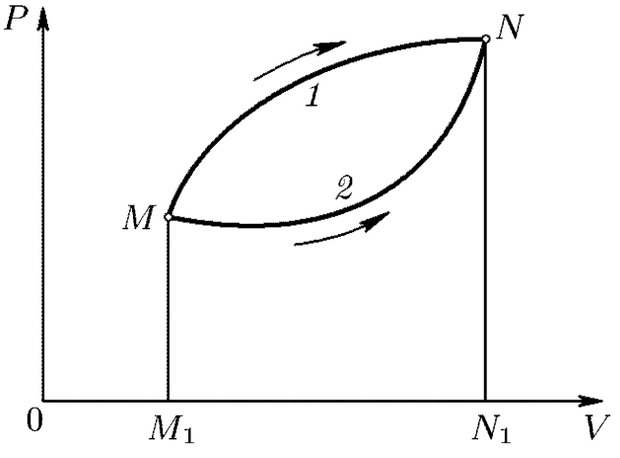


Рис. 9

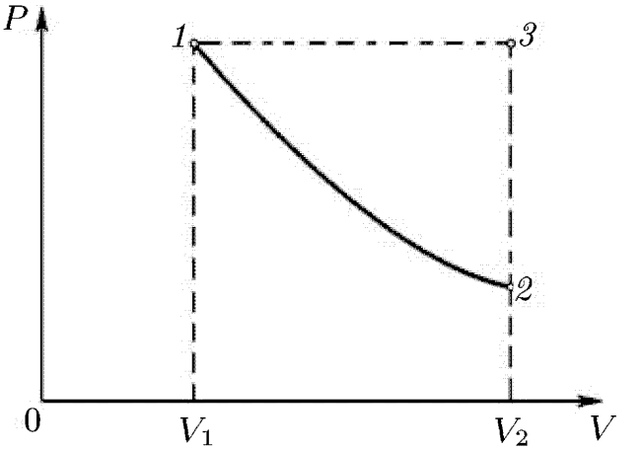


Рис. 10

процесс на рис. 10 изображен горизонтальной прямой *1–3*; при этом газ совершит работу $A_1 = P_1(V_2 - V_1)$. Затем, закрепив неподвижно поршень и охлаждая газ, можно довести его давление до значения P_2. Этот процесс происходит без совершения работы; на рис. 10 он изображен вертикальной прямой *3–2*. В результате система перейдет в то же конечное состояние *2*, совершив работу $A_1 = P_1(V_2 - V_1) > A$. Приведенный пример наглядно показывает, что работа зависит не только от начального и конечного состояний, но и от способа (или пути) переходя системы из одного состояния в другое. Значит, работа не есть функция состояния.

4. Если в результате изменений система вернулась в исходное состояние, то говорят, что она совершила *круговой процесс*, или *цикл*. Такой процесс, если он квазистатический, на диаграмме VP изображается *замкнутой кривой* (рис. 11). Работа, совершенная системой в круговом процессе, численно равна площади цикла, заштрихованной на рис. 11. При этом если точка, изображающая состояние системы, описывает цикл по часовой стрелке, то работа системы положительна. Если же цикл проходится в направлении против часовой стрелки, то она отрицательна.

5. Мы нашли выражение для элементарной работы и выяснили свойства этой величины на примере газа или изотропного однородного тела, находящегося под постоянным внешним давлением. Внутреннее состояние таких систем определяется двумя параметрами, например P

и V. Поэтому их можно называть простыми системами или *системами с двумя степенями свободы*. Могут быть системы со многими степенями свободы, внутреннее состояние их определяется температурой T и какими-то внешними макроскопическими параметрами $a_1, a_2, \ldots, a_n$.

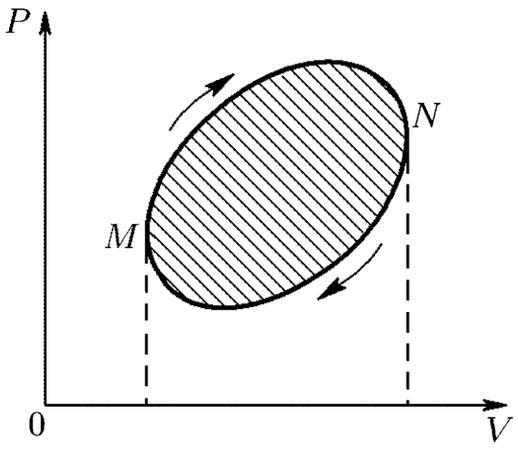


Рис. 11

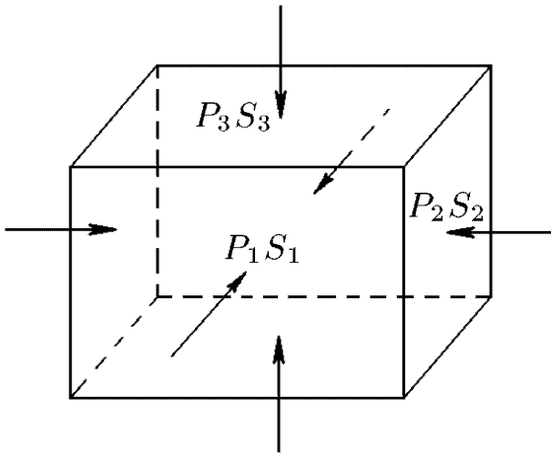


Рис. 12

В этом случае работа по-прежнему зависит от пути перехода, однако вместо формулы (12.1) следует писать

$$\delta A = A_1\, da_1 + A_2\, da_2 + \ldots + A_n\, da_n, \tag{12.5}$$

где для квазистатического процесса используются $A_1, A_2, \ldots$ — функции параметров $a_1, a_2, \ldots$ и температуры T, называемые *обобщенными силами*. Рассмотрим, например, прямоугольный параллелепипед из однородного изотропного вещества (рис. 12). Если на его грани действуют нормальные давления P_1, P_2, P_3, то элементарная работа, совершаемая системой, представится выражением

$$\delta A = P_1 S_1\, dx_1 + P_2 S_2\, dx_2 + P_3 S_3\, dx_3,$$

где dx_1, dx_2, dx_3 — удлинения ребер параллелепипеда, а S_1, S_2, S_3 — площади соответствующих граней. Обобщенные силы определяются выражениями

$$A_1 = P_1 S_1, \quad A_2 = P_2 S_2, \quad A_3 = P_3 S_3.$$

§ 13. Первое начало термодинамики для системы в адиабатической оболочке

1. Пусть термодинамическая система заключена в какую-то оболочку, отделяющую ее от других тел. Различные части оболочки могут перемещаться. Примером может служить цилиндр с поршнем, в котором находится газ. Изменять состояние системы внутри оболочки можно различными способами. Один из них состоит в *механическом перемещении частей оболочки* или, вообще, в *изменении внешних параметров*, определяющих наряду с температурой внутреннее макроскопическое состояние системы. Этот способ, как правило, сопровождается производством механической работы. Работа внешних сил, связанная с перемещением оболочки или с изменением внешних

параметров, называется *макроскопической работой*, производимой над системой. Эту величину мы обозначаем буквой $A_{\text{внеш}}$ в отличие от работы, производимой самой системой, которая обозначается через A. В предыдущем параграфе приводились формулы для вычисления работы A или $A_{\text{внеш}}$ при квазистатических процессах. В этом случае всегда $A = -A_{\text{внеш}}$. Теперь рассматривается более общий случай, когда процесс не обязательно квазистатический. Для такого процесса, вообще говоря, $A \neq A_{\text{внеш}}$.

Не надо думать, что производство работы обязательно связано с изменением объема системы. Так, в общеизвестных опытах Джоуля по определению механического эквивалента теплоты лопасти мешалки являются составными частями оболочки, в которую заключена система, например вода. При вращении мешалки над системой совершается механическая работа. Она проявляется в нагревании воды в калориметре.

Приведем другой пример, в котором внутреннее состояние системы меняется в результате производства механической работы над ней. Возьмем толстостенную трубку из плексигласа, в которую входит плотно подогнанный поршень. К нижней части поршня прикрепим кусочек пироксилиновой ваты. При быстром вдвигании поршня воздух в трубке нагревается настолько сильно, что пироксилин воспламеняется (пневматическое, или воздушное, огниво).

Но состояние тел в оболочке можно изменять и без механического перемещения ее стенок. Так, воду в калориметре Джоуля или воздух в трубке предыдущего опыта можно нагреть на газовой горелке до одной и той же температуры, до которой они были доведены ранее путем механического перемещения стенок оболочки. В результате эти тела окажутся в тех же самых конечных состояниях. Состояние тел в оболочке можно также менять, воздействуя на них извне различными силовыми полями, например электрическим и магнитным. Вообще, состояние системы в оболочке можно менять путем изменения внешних параметров и температуры окружающих тел.

2. В § 1 для частных случаев было введено понятие адиабатической оболочки и приведены примеры оболочек, являющихся приблизительно адиабатическими. Дадим теперь общее определение адиабатической оболочки. Оболочка называется *адиабатической*, когда состояние заключенной в ней системы остается неизменным при любых изменениях температуры окружающих тел, если только значения внешних параметров поддерживаются постоянными. Таким образом, изменить состояние системы в адиабатической оболочке можно только путем изменения внешних параметров. Система, заключенная в адиабатическую оболочку, называется *адиабатически изолированной*.

3. Основное положение, с помощью которого в термодинамике можно расширить понятие энергии, состоит в следующем. *Если система тел адиабатически изолирована, то работа внешних сил над этой системой зависит только от ее начального и конечного состояний, но совсем не зависит от способа или пути, каким осуществляется*

переход системы из начального состояния в конечное. Это не противоречит высказанному в § 12 утверждению о зависимости работы от пути перехода. Дело в том, что сейчас мы рассматриваем не все возможные переходы, а только такие, при которых сохраняется адиабатическая изоляция системы.

Сформулированное положение мы примем за постулат, выражающий содержание первого начала термодинамики. Его справедливость устанавливается опытом. Прямыми опытами, подтверждающими этот постулат, могут служить, например, классические опыты Джоуля по определению *механического эквивалента теплоты*. Они общеизвестны, и нет надобности останавливаться на их описании. Необходимо, однако, четко сформулировать, что эти опыты доказывают. Стенки калориметра в опытах Джоуля с хорошим приближением являются адиабатической оболочкой, в которой заключена вода или другая жидкость. Состояние жидкости в калориметре, если она находится в покое, определяется двумя параметрами, например давлением и температурой. Давление в опытах Джоуля оставалось постоянным. Единственным переменным параметром, полностью определяющим состояние покоящейся жидкости, была температура. Состояние жидкости в калориметре можно менять различными способами. Можно применять мешалки и перегородки в калориметре различной формы, изготовлять их из различных материалов, менять их число и расположение, короче, можно как угодно варьировать устройство калориметра. Можно вращать мешалку быстро или медленно, равномерно или неравномерно. Жидкость в промежуточных состояниях может совершать спокойное ламинарное движение, или ее движение может быть бурным и турбулентным. Все это отражается на ходе и скорости процесса. Но опыты Джоуля доказали, что механическая работа, которую необходимо затратить для перевода системы из одного строго фиксированного состояния в другое, также строго фиксированное состояние, не зависит от способа перевода системы из начального состояния в конечное. К этому прежде всего и сводятся результаты опытов Джоуля. Другой результат опытов Джоуля состоит в установлении числового значения механического эквивалента теплоты.

Но справедливость постулата первого начала термодинамики доказывается не только прямыми опытами типа опытов Джоуля. Постулат позволяет, не вдаваясь в рассмотрение механизма явлений, получать многочисленные следствия и количественные соотношения. В этом его громадное познавательное значение. Подтверждаемые опытом такие следствия и соотношения дают несравненно более точное и надежное доказательство самого постулата, чем прямые опыты.

4. Сделаем еще одно существенное замечание, на которое мы будем опираться в следующем параграфе. При вращении мешалки вода в калориметре Джоуля с адиабатическими стенками *всегда нагревается, но никогда не охлаждается*. Поэтому вращением мешалки невозможно адиабатически вернуть воду из конечного в прежнее начальное состояние. В дальнейшем на основе второго начала термодинамики

будет показано, что это невозможно сделать никакими способами, если только вода остается адиабатически изолированной. То же относится к любой термодинамической системе. Если адиабатически изолированная система переходит из состояния *1* в состояние *2*, то обратный адиабатический переход в состояние *1* может оказаться невозможным. Таким образом, не всегда возможно адиабатически перевести систему из какого-либо состояния в другое, произвольно заданное состояние. Однако, каковы бы ни были состояния *1* и *2*, опыт показывает, что *всегда возможен* один из двух адиабатических переходов: либо из состояния *1* в состояние *2*, либо обратный переход из состояния *2* в состояние *1* [1]. Однако одних квазистатических процессов для осуществления таких переходов недостаточно, для этого требуются также неравновесные адиабатические процессы. Рассмотрим, например, два состояния воды в калориметре Джоуля: состояние *1* при температуре 20 °C и состояние *2* при температуре 30 °C. Поскольку давление в калориметре поддерживается постоянным, состояние воды полностью фиксируется заданным значением температуры. Нельзя адиабатически перевести воду из более нагретого состояния *2* в менее нагретое состояние *1*. Но обратный адиабатический переход *1* в *2* возможен и фактически осуществляется в опытах Джоуля.

§ 14. Внутренняя энергия

1. Внутренняя энергия является одной из важнейших величин, характеризующих систему в состоянии термодинамического равновесия. Поэтому мы будем здесь предполагать, что начальное и конечное состояния системы равновесны. Однако процессы, переводящие систему из одного состояния в другое, а следовательно, и промежуточные состояния, вообще говоря, могут быть и неравновесными.

Внутренней энергией U системы называется функция состояния, приращение которой во всяком процессе, совершаемом системой в адиабатической оболочке, равно работе внешних сил над системой при переходе ее из начального равновесного состояния в конечное, также равновесное состояние. Возможность введения такой функции состояния основана на том, что работа над системой в адиабатической оболочке зависит только от начального и конечного состояний системы, но не зависит от способа перехода. Начальное состояние мы будем обозначать индексом 1, конечное — индексом 2. Таким образом, если система помещена в адиабатическую оболочку, то

$$U_2 - U_1 = A_{12}^{\text{внеш}}, \tag{14.1}$$

[1] Забегая вперед, укажем, что второе начало термодинамики вводит функцию состояния системы — *энтропию*, с помощью которой можно сформулировать общий критерий возможности или невозможности того или иного процесса. Согласно этому критерию в адиабатически изолированной системе возможны только такие процессы, в которых энтропия либо возрастает, либо остается неизменной.

где U_1 и U_2 — внутренние энергии системы в состояниях *1* и *2*, а $A_{12}^{\text{внеш}}$ — работа, производимая над системой внешними силами при переходе ее из состояния *1* в состояние *2 по любому пути*. Разумеется, слово «приращение» надо понимать не в арифметическом, а в алгебраическом смысле. Внутренняя энергия системы может как возрастать, так и убывать, а работа $A_{12}^{\text{внеш}}$ может быть и положительной, и отрицательной.

Может случиться, что адиабатический переход из состояния *1* в состояние *2* невозможен. Тогда определение (14.1) теряет смысл. Но в таких случаях, как подчеркивалось в предыдущем параграфе, возможен обратный адиабатический переход из состояния *2* в состояние *1*. Им и надо воспользоваться для определения изменения внутренней энергии, т. е. вместо формулы (14.1) следует писать

$$U_1 - U_2 = A_{21}^{\text{внеш}}. \tag{14.2}$$

Таким образом, используя либо адиабатический переход $1 \to 2$, либо обратный адиабатический переход $2 \to 1$, всегда можно определить разность внутренних энергий в любых равновесных состояниях *1* и *2*.

2. Внутренняя энергия определена не однозначно, а с точностью до произвольной *аддитивной постоянной*. Такая неоднозначность может отразиться на реальном содержании физических выводов. Реальный смысл имеют не сами энергии, а их *разности* в различных состояниях, значения которых от выбора произвольной постоянной не зависят. Одно из состояний, безразлично какое, можно принять за нулевое и условиться считать, что внутренняя энергия системы в этом состоянии равна нулю. Тогда внутренняя энергия в любом другом состоянии определится уже однозначно. Таким образом, можно дать следующее определение внутренней энергии. *Внутренней энергией системы в каком-либо (равновесном) состоянии называется работа, которую должны совершить внешние силы, чтобы любым возможным адиабатическим путем перевести систему из нулевого состояния в рассматриваемое*. Если адиабатический процесс в таком направлении невозможен, то надо воспользоваться обратным процессом и заменить приведенное определение следующим. *Внутренней энергией системы в каком-либо (равновесном) состоянии называется взятая с противоположным знаком работа, которую должны произвести внешние силы, чтобы любым адиабатическим путем перевести систему из рассматриваемого состояния в нулевое*.

3. Если за нулевое принять другое состояние, то значения внутренней энергии во всех остальных состояниях изменятся на одну и ту же постоянную. Действительно, пусть *0*, *0′* и *1* — три произвольные состояния термодинамической системы, из которых *0* и *0′* принимаются за нулевые. Внутренние энергии системы в состоянии *1* относительно нулевых состояний *0* и *0′* будем обозначать соответственно U_1 и U_1'. Так как всегда возможен в том или ином направлении адиабатический переход из одного произвольного состояния в другое произвольное состояние, то могут представиться только четыре различных случая,

изображенные на рис. 13. Стрелки означают направления, в которых возможны адиабатические переходы. В случае *а* для определения энергии U_1' изберем адиабатический переход из $0'$ в 1 через промежуточное состояние 0, т. е. переход $0'01$. По определению внутренней энергии

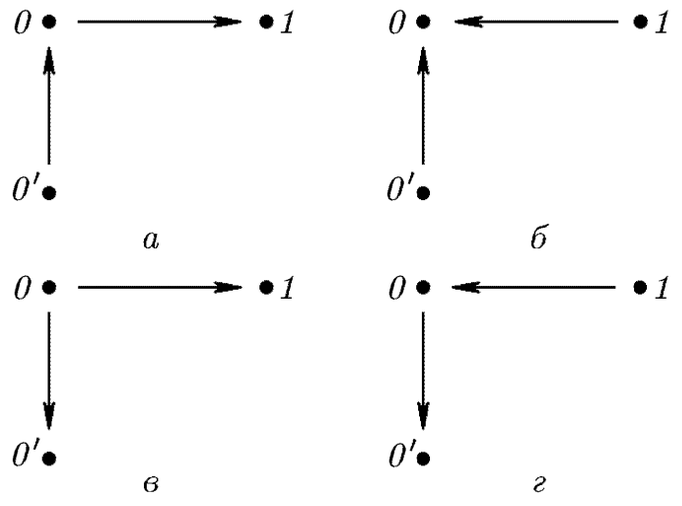


Рис. 13

$$U_1' = A_{0'0} + A_{01},$$

$$U_1 = A_{01}, \quad U_{00'} = A_{0'0}.$$

Поэтому

$$U_1' - U_1 = U_{00'}. \qquad (14.3)$$

То же самое соотношение справедливо и во всех остальных случаях. Действительно, в случае *г* применимо уже рассмотренное рассуждение, но для обратного адиабатического перехода $100'$. В случае *б* могут быть две возможности, в зависимости от того, какой из двух адиабатических переходов $0' \to 1$ или $1 \to 0'$ возможен. Достаточно рассмотреть одну из этих возможностей, например первую (рис. 14). Совершив адиабатический переход из $0'$ в 0 через промежуточное состояние 1, можно написать

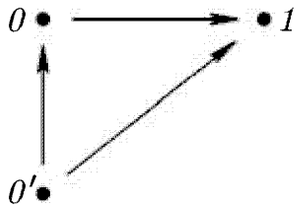


Рис. 14

$$U_{00'} = A_{0'1} + A_{10} = U_1' - U_1,$$

т. е. мы снова приходим к соотношению (14.3). Оно показывает, что разность $U_1' - U_1$ не зависит от выбора состояния 1, а только от состояний 0 и $0'$. Тем самым наше утверждение доказано. Случай *в* совершенно аналогичен случаю *б*.

4. Для квазистатических процессов $A^{\text{внеш}} = -A$ (газ в цилиндре). В этом случае вместо (14.1) можно написать

$$U_1 - U_2 = A, \qquad (14.4)$$

т. е. работа системы при адиабатических процессах совершается за счет убыли внутренней энергии.

5. Отметим одну трудность, которая встречается при описанном способе введения понятия внутренней энергии. При очень высоких температурах все вещества переходят в состояние полностью или частично ионизованного газа, называемого плазмой. Не существует материалов, из которых можно было бы изготовить оболочку для удержания вещества в таком состоянии [1].

[1] Удержание плазмы может быть существенно только силовыми полями. В звездах оно осуществляется полями тяготения. В земных условиях проблема удержания горячей плазмы возникла в связи с попыткой осуществления

Таким образом, не всегда возможно даже приближенно осуществить адиабатическую изоляцию системы. В этих условиях понятие адиабатической оболочки превращается в абстракцию, лишенную реального содержания. Становится бессодержательным и определение внутренней энергии с помощью адиабатической оболочки. Существует, по-видимому, единственный путь, каким может быть преодолена эта трудность. Он дается на основе представлений молекулярно-кинетической теории. В этом случае внутренняя энергия тела определяется как сумма кинетической энергии внутренних движений составляющих его частиц (молекул, атомов, электронов и т. п.) и потенциальной энергии силовых полей, с помощью которых осуществляется взаимодействие между этими частицами. Однако если температуры не очень высоки, определение внутренней энергии, неиспользующее представление об адиабатической оболочке, вполне возможно и более соответствует духу аксиоматической термодинамики.

6. Следует подчеркнуть, что, говоря о внутренней энергии, отвлекаются от макроскопических движений системы и от воздействий на нее внешних силовых полей. Внутренняя энергия зависит только от параметров, характеризующих внутреннее состояние тела, но не может явно зависеть от его макроскопического движения и от внешних сил, действующих на тело. Понятие внутренней энергии, строго говоря, относится к телам, находящимся в состоянии термодинамического равновесия. На неравновесные состояния оно может быть распространено в тех случаях, когда рассматриваемая система тел может быть разделена на достаточно малые макроскопические части, каждая из которых в рассматриваемый момент времени практически находится в равновесном состоянии. В такой системе различные малые макроскопические части могут двигаться с различными скоростями, находиться под разными давлениями и иметь разные температуры. В этих случаях полная энергия системы складывается из: 1) кинетической энергии движения ее макроскопических частей; 2) потенциальной энергии системы в поле внешних сил и 3) внутренней энергии.

Внутренняя энергия в свою очередь состоит из суммы внутренних энергий макроскопических подсистем, на которые мысленно можно разделить всю систему, а также энергии взаимодействия этих подсистем. Взаимодействие между подсистемами осуществляется посредством *молекулярных сил*, действие которых простирается на малые расстояния порядка 10^{-8}–10^{-7} см. Ввиду этого энергия взаимодействия макроскопических подсистем пропорциональна площади поверхности, вдоль которой они соприкасаются между собой. Она называется поэтому *поверхностной энергией*. В большинстве случаев

управляемой термоядерной реакции. Плазму пытаются удерживать с помощью сильных магнитных полей, воздействующих на заряженные частицы, из которых она состоит. Но таким путем не может быть достигнута полная адиабатическая изоляция системы, хотя бы уже потому, что плазма, нагретая до весьма высоких температур, интенсивно теряет энергию на электромагнитное излучение. Кроме того, силовые поля, удерживающие плазму, оказывают влияние на ее состояние и, следовательно, меняют запас энергии в ней.

поверхностной энергией пренебрегают, так как ею обладают только молекулы тонкого пограничного слоя, объем которого пренебрежимо мал по сравнению с объемами самих подсистем. В таком приближении внутренняя энергия обладает свойством *аддитивности*, т. е. внутренняя энергия системы равна сумме внутренних энергий подсистем, из которых она состоит. Но пренебречь поверхностной энергией можно не всегда. Например, этого нельзя делать при рассмотрении явлений *поверхностного натяжения*, так как сами эти явления обусловлены именно наличием поверхностной энергии.

Если адиабатическая оболочка, в которую заключена система, жесткая, то при всех процессах работа системы равна нулю. При отсутствии внешних силовых полей такая система является изолированной или замкнутой. Ее внутренняя энергия при любых изменениях сохраняется постоянной.

7. Приведенное выше термодинамическое определение внутренней энергии указывает принципиальный способ измерения этой величины. Для этого тело надо адиабатически изолировать и измерить работу, производимую над ним внешними силами, когда тело переходит по любому пути из нулевого состояния в рассматриваемое или обратно. Таким путем можно найти внутреннюю энергию тела как функцию макроскопических параметров, характеризующих его состояние.

В результате тело будет «энергетически проградуировано», т. е. каждому состоянию его будет поставлено в соответствие вполне определенное значение внутренней энергии. Такое тело может быть использовано в качестве калориметра для измерения внутренней энергии других тел. Всякий калориметр, в сущности, представляет собой тело, помещенное в жесткую адиабатическую оболочку и проградуированное по внутренней энергии. Такое тело называют *калориметрическим телом*. В ту же оболочку могут помещаться другие тела, внутреннюю энергию которых требуется измерить. Калориметрическим телом может служить, например, определенная масса воды или другой жидкости. Внутренняя энергия жидкости практически зависит только от температуры, так как давление обычно поддерживается постоянным или меняется мало. Малые же изменения давления, ввиду малой сжимаемости жидкости, практически не сказываются на ее внутренней энергии. Измерение внутренней энергии тела в калориметре сводится к более удобной манипуляции — измерению температуры. Если температура тела, погруженного в калориметр, отличается от температуры калориметра, то начнется процесс выравнивания температур. Внутренние энергии тела и калориметра начнут меняться, пока в системе не установится общая температура. Внутренняя энергия, потерянная телом, будет равна внутренней энергии, полученной калориметром, так как тело и калориметр образуют замкнутую систему, энергия которой меняться не может. Таким образом, по изменению внутренней энергии калориметра можно судить об изменении внутренней энергии

рассматриваемого тела. С изложенной здесь точки зрения классические работы Джоуля по определению механического эквивалента теплоты сводятся к градуировке водяного калориметра в механических единицах энергии.

§ 15. Количество теплоты. Математическая формулировка первого начала термодинамики

1. Если система помещена в адиабатическую оболочку, то единственным способом изменить ее внутреннюю энергию является производство над ней макроскопической работы, что достигается путем изменения внешних параметров. (К числу внешних параметров относятся и силовые поля, в которые помещена система.) Однако если адиабатической изоляции нет, то изменение внутренней энергии возможно и без производства макроскопической работы. Так, при соприкосновении горячего и холодного тел внутренняя энергия переходит от горячего тела к холодному, хотя при этом макроскопическая работа и не совершается.

Процесс обмена внутренними энергиями соприкасающихся тел, не сопровождающийся производством макроскопической работы, называется теплообменом. Энергия, переданная телу окружающей средой в результате теплообмена, называется количеством теплоты, или просто теплотой, полученной телом в таком процессе.

Изменение внутренней энергии тела во время теплообмена, в сущности, обусловлено также работой каких-то внешних сил. Но это не есть макроскопическая работа, связанная с изменением внешних макроскопических параметров. Она является *микроскопической работой*, т. е. складывается из работ, производимых молекулярными силами, с которыми на молекулы и атомы тела действуют молекулы и атомы окружающей среды. Так, при приведении тела в контакт с горячим газом передача энергии от газа к телу осуществляется посредством столкновений молекул газа с молекулами тела.

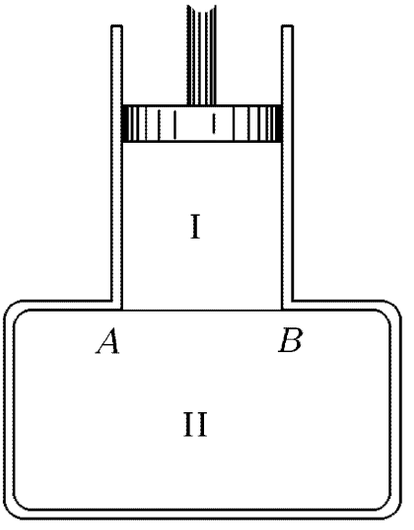


Рис. 15

2. Сформулируем математически начало термодинамики с учетом теплообмена. Пусть интересующая нас термодинамическая система I (рис. 15) находится в тепловом контакте с какой-то системой II. Вся система I + II заключена в адиабатическую оболочку, однако граница AB между системами является теплопроводящей. При этих условиях система I+II не может обмениваться теплотой с окружающей средой, однако теплообмен между системами I и II может происходить. Допустим далее, что оболочка, в которую заключена система II, жесткая, так что никакой работы система II производить не может.

Система I, напротив, может совершать работу над окружающей средой. На схематическом рис. 15 система I изображена в виде цилиндра с подвижным теплонепроводящим поршнем. Стенки цилиндра — адиабатические, а дно AB может проводить тепло. Адиабатические стенки на рисунке изображены двойными линиями, проводящая перегородка AB — сплошной тонкой линией.

Пусть система I + II перешла из произвольного состояния *1* в другое состояние *2*, в результате чего системой совершена работа A_{12} над внешними телами. Эту работу совершила только система I. Так как составная система I + II адиабатически изолирована, то

$$A_{12} = (U_1 + U_1') - (U_2 + U_2'),$$

где U — внутренняя энергия системы I, а U' — системы II. Поскольку нас интересует поведение только системы I, перепишем это соотношение так:

$$A_{12} = U_1 - U_2 + (U_1' - U_2').$$

Убыль внутренней энергии системы II есть, по определению, количество теплоты, полученное системой I в рассматриваемом процессе. Обозначим эту величину через Q. Тогда по определению

$$Q = U_1' - U_2' = -\Delta U', \tag{15.1}$$

и предыдущее соотношение примет вид

$$Q = U_2 - U_1 + A_{12}. \tag{15.2}$$

Это уравнение и дает математическую формулировку первого начала термодинамики. Оно утверждает, что *теплота Q, полученная системой, идет на приращение ее внутренней энергии $\Delta U = U_2 - U_1$ и на производство внешней работы.*

3. Для уяснения понятия количества теплоты полезно обратить внимание на следующее обстоятельство. Согласно определению о величине Q *следует судить по изменению состояния не самой системы* I, *а системы* II, *с которой она обменивается теплотой.* Иначе соотношение (15.2) было бы не физическим законом, а простым определением понятия теплоты Q. Если же по определению величину Q находить по изменению внутренней энергии системы II, то соотношение (15.2) становится физическим утверждением, допускающим экспериментальную проверку. Действительно, все три величины: $U_2 - U_1$, A_1 и Q, входящие в соотношение (15.2), могут быть измерены независимо, а потому можно экспериментально проверить, удовлетворяют они соотношению (15.2) или не удовлетворяют. Исторически понятие о теплоте как о количественной величине возникло из калориметрических понятий. В этих измерениях о количестве теплоты, сообщенной телу, судят по изменению температуры калориметра, с которым тело обменивается теплотой. Все это согласуется с определением величины Q, даваемым соотношением (15.1): система II играет роль калориметра.

4. Соотношение (15.2) остается справедливым и в общем случае, т. е. при любых изменениях состояния системы I. Однако выражение (15.1) подлежит обобщению. Допустим, например, что составная система I + II заключена в адиабатическую оболочку, но оболочка, окружающая систему II, не является жесткой. В этих условиях система II может совершать работу. Если $A_{\text{полн}}$ — полная работа составной системы I + II, то

$$A_{\text{полн}} = U_1 - U_2 + (U_1' - U_2').$$

Полная работа слагается из работы A_{12} системы I и работы A_{12}' системы II: $A_{\text{полн}} = A_{12} + A_{12}'$. Поэтому предыдущее соотношение попрежнему можно записать в виде (15.2), если величину Q определить выражением

$$Q = U_1' - U_2' - A_{12}'. \tag{15.3}$$

5. Для квазистатических процессов, при которых термодинамические параметры испытывают бесконечно малые, или элементарные, изменения, уравнение (15.2) принимает вид

$$\delta Q = dU + \delta A, \tag{15.4}$$

а в случае работы против давления P

$$\delta Q = dU + P\,dV. \tag{15.5}$$

Если процесс круговой, т. е. в результате него система возвращается в исходное состояние, то $U_2 = U_1$ и, следовательно, $Q = A$. В круговом процессе все количество теплоты, полученной системой, идет на производство внешней работы.

Если $U_1 = U_2$ и $Q = 0$, то $A = 0$. *Это значит, что невозможен процесс, единственным результатом которого является производство работы без каких бы то ни было изменений в других телах*. Механизм, в котором осуществляется такой процесс, называется *перпетуум мобиле* (*вечным двигателем*). Таким образом, из первого начала термодинамики следует невозможность перпетуум мобиле. Нескончаемые попытки построить такой двигатель, с настойчивостью предпринимавшиеся в предыдущие века, да иногда встречающиеся и теперь (конечно, среди неграмотных в физике людей), неизменно оканчивались неудачей. Это привело к тому, что невозможность перпетуум мобиле была возведена в принцип, по своему содержанию эквивалентный принципу сохранения энергии.

6. Ясно, что единицы работы и энергии могут служить также единицами количества теплоты, хотя исторически было не так. Пока не была установлена кинетическая природа теплоты, для этих величин пользовались различными единицами. Использование одних и тех же единиц удобно не только в теоретических исследованиях, но и на практике, так как при этом во всех соотношениях выпадают числовые множители, являющиеся коэффициентами перевода количества теплоты из тепловых единиц в механические или обратно. В системе СИ единицей количества теплоты является *джоуль*, в системе СГС —

эрг. Однако во времена теории теплорода для измерения количества теплоты была введена особая единица — *калория*. *Малая калория или грамм-калория есть количество теплоты, которое надо сообщить одному грамму чистой воды, чтобы повысить ее температуру на один градус Цельсия.* Тысяча грамм-калорий составляет *большую калорию* или *килокалорию*. Для большей определенности различали нулевую калорию, 15-градусную, 20-градусную и др. Например, 20-градусная калория определялась как количество теплоты, потребное для нагревания 1 г чистой воды от 19,5 до 20,5 °С. В настоящее время надобность в особой единице для измерения количества теплоты отпала. Калория не входит в число единиц СИ или системы СГС. Она является *внесистемной единицей*. Однако, ввиду привычности и наглядности этой единицы, имеет смысл в отдельных случаях пользоваться ею. Часто применяется *международная килокалория*, содержащая по определению 4,1868 кДж. Тысячную долю этой величины мы будем называть просто *калорией*. Полезно заметить, что универсальная газовая постоянная, выраженная в калориях,

$$R = 1{,}9858 \text{ кал}/(\text{моль} \cdot \text{К}) \approx 2 \text{ кал}/(\text{моль} \cdot \text{К}).$$

§ 16. Когда можно пользоваться представлением о количестве теплоты, содержащемся в теле

1. *Количество теплоты, полученное телом, не является функцией состояния*. Это непосредственно видно из уравнения (15.2). Действительно, разность $U_2 - U_1$ зависит только от начального и конечного состояний системы, когда работа A_{12} зависит еще от пути перехода. Поэтому и количество теплоты $Q = U_2 - U_1 + A_{12}$ также зависит от пути перехода, т. е. от способа, каким система была приведена в рассматриваемое состояние. Это и значит, что Q не есть функция состояния системы. Перевести систему в рассматриваемое состояние можно бесчисленным множеством способов. Во всех этих способах системе сообщается одна и та же внутренняя энергия, но разделение этой величины на работу и сообщенную теплоту в разных способах будет разным. Если задано состояние системы, но не указан процесс, каким это состояние было достигнуто, то ничего нельзя сказать о количестве теплоты, запасенном системой при переходе в рассматриваемое состояние. В этом смысле нельзя говорить о количестве теплоты, содержащемся или запасенном в теле. Но всегда можно говорить о запасе внутренней энергии, так как он не зависит от способа приведения системы в рассматриваемое состояние. О количестве же теплоты, сообщенному телу, имеет смысл говорить лишь в том случае, когда *указан процесс*, каким система переходила в рассматриваемое состояние.

Следующая аналогия может разъяснить суть дела. Допустим, что озеро пополняется водой за счет дождя и за счет воды втекающей в него реки. Имеет смысл говорить о количестве воды, которое попадает в озеро за сутки из реки, а также о количестве воды, выпадающей в него за то же время в

виде дождя. Но лишено смысла утверждение, что в озере, рассматриваемом безотносительно к процессам его наполнения, содержится столько-то кубических метров дождевой и столько-то кубических метров речной воды. В озере есть единая вода, и она могла попасть туда различными способами. Так и о внутренней энергии, содержащейся в теле, не имеет смысла говорить, что она состоит из стольких-то единиц работы и стольких-то единиц теплоты. Безотносительно к процессам, какими тело получало внутреннюю энергию, ее разделение на работу и теплоту лишено всякого содержания.

2. Очень распространен термин «тепловая энергия». Его ни в коем случае не следует понимать в смысле количества теплоты, запасенного телом в процессе теплообмена. Под тепловой энергией следует понимать внутреннюю энергию тела. Поэтому нет надобности в особом термине «тепловая энергия». Но при правильном понимании его можно и сохранить, чтобы напомнить о связи строго научного понятия внутренней энергии с субъективными ощущениями тепла и холода, из которых оно возникло. Кроме того, термином «тепловая энергия» с особой выразительностью подчеркивается, что речь идет об энергии, связанной с внутренним хаотическим движением атомов и молекул. Тело, обладающее запасом тепловой (внутренней) энергии, в термодинамике называется *тепловым резервуаром*.

Про тепловой резервуар говорят также, что он содержит определенный *запас теплоты*. Но в этом случае под тепловой понимают внутреннюю энергию, а не количество теплоты, полученное телом в процессе теплообмена. Все эти несовершенства терминологии, как и самый термин «количество теплоты», наука получила в наследство от теории теплорода. Последняя рассматривала теплоту как гипотетическую невесомую жидкость, содержащуюся в телах, которая не может быть создана или уничтожена. Говорили о законе сохранения теплорода, аналогичном закону сохранения обыкновенного вещества. С этой точки зрения естественно было говорить о запасе теплоты или теплорода в теле безотносительно к способу, каким тело было приведено в рассматриваемое состояние. Эти воззрения, не представляющие сейчас никакого интереса, довольно долго удерживались в науке. Причина этого в том, что они подтверждались калориметрическими измерениями и некоторыми другими явлениями природы. И сейчас в калориметрии часто рассуждают так, как если бы был справедлив «закон сохранения количества теплоты». Так поступают, например, при вычислении температуры смеси двух жидкостей или теплоемкости тел при калориметрических измерениях. Так же поступают и в математической теории теплопроводности. Основы последней были заложены французским математиком Фурье (1768–1830) в первой четверти XIX века и полностью сохранили свое значение в настоящее время. Математическая теория теплопроводности (в твердых и жидких телах) исходит из предположения, что при всех процессах тепло не возникает и не уничтожается, а может только перемещаться из одних мест пространства в другие. Такое представление, конечно, противоречит механической теории теплоты. Однако при выполнении некоторых условий явления протекают так, как если бы это представление было

верно. Выяснение этих условий представляет не только исторический интерес, но и необходимо по существу.

3. Представление о количестве теплоты как об определенной величине, содержащейся в теле, и о сохранении ее возникло на основе таких опытов, в которых на изменения системы тел наложены определенные ограничения. Допустим, например, что система заключена в адиабатическую оболочку. Рассечем ее мысленно на две подсистемы B и C, соприкасающиеся друг с другом вдоль некоторой поверхности. Каждая из этих подсистем уже не будет адиабатически изолирована. Если процесс, происходящий в оболочке, является квазистатическим, то на основании первого начала термодинамики для этих подсистем можно написать

$$Q^B = U_2^B - U_1^B + A^B, \quad Q^C = U_2^C - U_1^C + A^C,$$

где Q^B и Q^C — теплоты, полученные подсистемами B и C, а A^B и A^C — произведенные ими работы. Предполагая, что внутренняя энергия всей системы равна сумме внутренних энергий подсистем $U = U^B + U^C$, и складывая предыдущие равенства, получим

$$Q^B + Q^C = (U_2 - U_1) + (A^B + A^C).$$

Работа A^B, совершенная подсистемой B, состоит из двух частей: работы над внешними телами и работы над подсистемой C. Аналогичное утверждение справедливо и для подсистемы C. Ввиду равенства действия и противодействия работа подсистемы B над подсистемой C равна и противоположна по знаку работе подсистемы C над подсистемой B (так как процесс квазистатический). Поэтому из суммы $A^B + A^C$ эти работы выпадают, и мы получаем $A^B + A^C = A$, где A — работа всей системы над внешними телами. Итак,

$$Q^B + Q^C = U_2 - U_1 + A.$$

Но для адиабатически изолированной системы $U_1 - U_2 = A$, а потому

$$Q^B + A^C = 0.$$

Количество теплоты, полученное подсистемой B, равно и противоположно по знаку количеству теплоты, полученному подсистемой C. Это значит, что квазистатические процессы в адиабатически изолированной системе происходят так, как если бы выполнялся «закон сохранения количества теплоты».

Полученный результат без труда распространяется на более общий случай. Допустим, что термодинамическая система, заключенная в адиабатическую оболочку, состоит из n не перемешивающихся подсистем, химически не реагирующих друг с другом и имеющих разные температуры. Между подсистемами может происходить только теплообмен. Если Q_i — тепло, полученное квазистатически i-подсистемой, то

$$\sum_{i=1}^{n} Q_i = 0.$$

4. Рассмотрим, далее, явления, происходящие в системе при постоянном объеме или постоянном давлении. Именно такие явления привели к представлению о «законе сохранения количества теплоты», так как все калориметрические опыты производились либо при постоянном объеме, либо при постоянном давлении. То же самое относится к явлениям теплопроводности. Если объем постоянен, то работа A_{12} равна нулю, и формула (15.2) переходит в

$$Q = U_2 - U_1 = \Delta U \quad (V = \text{const}). \tag{16.1}$$

Если же постоянно давление, то $A_1 = P(V_2 - V_1) = \Delta(PV)$, и мы получаем

$$Q = \Delta U + \Delta(PV).$$

Введем новую величину

$$I = U + PV. \tag{16.2}$$

Она, очевидно, есть функция состояния, так как величины U, P, V сами являются функциями состояния. Тогда

$$Q = I_2 - I_1 = \Delta I \quad (P = \text{const}). \tag{16.3}$$

Функция I играет важную роль в термодинамике. Она называется *энтальпией* или *тепловой функцией*. Иногда ее называют также *теплосодержанием*. Но этот последний термин неудачен, и мы им пользоваться не будем. *Энтальпию можно определить как функцию состояния, приращение которой при изобарическом процессе дает теплоту, полученную системой.*

Таким образом, если объем системы остается постоянным, то количество теплоты Q равно приращению внутренней энергии системы. Если же постоянно давление, то оно выражается приращением энтальпии. В обоих случаях Q не зависит от пути перехода, а только от начального и конечного состояний системы. Поэтому на основании опытов при постоянном объеме или при постоянном давлении и могло сложиться представление о какой-то величине Q, содержащейся в теле и не зависящей от способа приведения его из нулевого состояния в рассматриваемое. Величина Q имеет различный смысл в зависимости от того, что остается постоянным: объем или давление. В первом случае под Q следует понимать внутреннюю энергию, во втором — энтальпию. Но в ранних опытах это различие ускользало от наблюдений, так как опыты производились с твердыми и жидкими телами, для которых оно незначительно благодаря малости коэффициентов теплового расширения твердых и жидких тел. В обоих случаях имеет место сохранение Q, но оно сводится к закону сохранения энергии.

5. В связи с тем, что теплота и работа не являются функциями состояния, для бесконечно малых количеств теплоты и работы мы употребляем обозначения δQ и δA, но не пользуемся обозначениями dQ и dA. Этим мы хотим подчеркнуть, что величины δQ и δA не могут рассматриваться как полные дифференциалы, т. е. не всегда могут быть представлены как бесконечно малые приращения каких-то

функций состояния. Такое представление возможно лишь в частных случаях, например, для δQ в случае процессов при постоянном объеме или давлении. Количества δQ и δA являются какими-то бесконечно малыми величинами, но не приращениями определенных функций. Напротив, бесконечно малые изменения внутренней энергии, энтальпии, давления, объема, температуры и пр. мы обозначаем символами $dU, dI, dP, dV, dT, \ldots$ Этим мы хотим указать, что такие величины являются *полными дифференциалами*, т. е. во всех без исключения случаях могут рассматриваться как бесконечно малые приращения функций состояния $U, I, P, V, T, \ldots$ Если забыть, что величины типа δQ и δA в общем случае не являются приращениями функций состояния, и обращаться с ними как с полными дифференциалами, то можно прийти к грубым ошибкам. Один пример такого рода будет приведен в § 47.

§ 17. Закон Гесса

1. Применим уравнения (16.1) и (16.3) к химическим реакциям. В термохимии *тепловым эффектом реакции называется количество теплоты, выделяющееся в этой реакции*. Тепловой эффект реакции называется также *теплотой образования*, если в результате реакции получается какое-либо определенное химическое соединение. Реакция называется *экзотермической*, если она сопровождается выделением теплоты, и *эндотермической*, если она идет с поглощением теплоты. В первом случае тепловой эффект считается положительным, во втором — отрицательным. Обозначая тепловой эффект реакции буквой W, можно написать $W = -Q$. Если реакция идет при постоянном объеме, то в силу уравнения (16.1)

$$W_V = U_1 - U_2 = -\Delta U. \tag{17.1}$$

Индекс V указывает на постоянство объема системы во время реакции. Аналогично, при постоянном давлении из уравнения (16.3) получаем

$$W_P = I_1 - I_2 = -\Delta I. \tag{17.2}$$

Из уравнений (17.1) и (17.2) непосредственно следует, что *тепловой эффект реакции зависит лишь от природы и физического состояния исходных веществ и конечных продуктов, но не зависит от промежуточных стадий реакции*. Это положение было установлено эмпирически в 1840 г. русским академиком Гессом (1802–1850) еще до открытия первого начала термодинамики, следствием которого оно является. Оно называется *правилом* или *законом Гесса*. Ясно, что правило Гесса относится либо к реакциям, идущим при постоянном давлении, либо к реакциям при постоянном объеме.

2. Если начальные и конечные продукты реакции все находятся в твердом или жидком состояниях, то обе величины W_V и W_P практически не отличаются одна от другой. Это связано с тем, что объем системы практически остается неизменным, а потому работа, производимая системой против внешнего давления, пренебрежимо мала. Не так обстоит дело в тех случаях, когда среди исходных или конечных продуктов реакции встречаются газообразные. Тогда работа системы против внешних сил сравнима с изменением внутренней энергии во время реакции, и тепловые эффекты W_V и W_P становятся существенно разными. В этих случаях тепловой эффект обычно рассматривают при постоянном давлении. При заданной температуре тепловой эффект реакции W_P очень слабо зависит, а в случае идеальных газов

совсем не зависит от внешнего давления (разумеется, поддерживаемого постоянным). Если тепловой эффект определен при 25 °C и 760 мм рт. ст., то он называется *стандартным*. Значение теплового эффекта указывается в виде слагаемого в правой части уравнения соответствующей химической реакции. Так, уравнение

$$H_2 + \frac{1}{2}\,O_2 = (H_2O)_{ж} + 287 \text{ кДж}$$

означает, что при сгорании одного моля газообразного водорода (при 25 °C и 760 мм рт. ст.) с образованием одного моля жидкой воды выделяется 287 кДж теплоты. Химические символы H_2, O_2 и H_2O в уравнениях такого типа указывают не только химический состав и количество соответствующих веществ, но и обозначают их энтальпии (или внутренние энергии, когда реакция идет при постоянном объеме). Предыдущее уравнение означает, например, что сумма энтальпий газообразных водорода и кислорода больше энтальпии образовавшейся жидкой воды на 287 килоджоулей. Для большей определенности принято заключать химический символ твердого вещества в квадратные, жидкого — в круглые и газообразные — в фигурные скобки. Так, $[H_2O]$, (H_2O) и $\{H_2O\}$ означают энтальпии соответственно моля льда, жидкой воды и водяного пара. В этих обозначениях предыдущее уравнение запишется так:

$$\{H_2\} + \frac{1}{2}\,\{O_2\} = (H_2O) + 287 \text{ кДж}.$$

3. С уравнениями такого типа можно обращаться как с обычными алгебраическими уравнениями. Они очень удобны при термохимических расчетах. Ввиду закона Гесса они позволяют, например, рассчитывать тепловые эффекты таких реакций, которые трудно или невозможно осуществить в чистом виде. Так, теплота образования газообразной окиси углерода из твердого углерода и кислорода не может быть непосредственно измерена по той простой причине, что углерод никогда не сгорает целиком в окись углерода, а всегда с образованием некоторого количества двуокиси. Для ее определения измеряют, во-первых, теплоту полного сгорания твердого углерода и получают

$$[C] + \{O_2\} = \{CO_2\} + 394 \text{ кДж}.$$

Во-вторых, измеряют теплоту сгорания окиси углерода в двуокись:

$$\{CO\} + \frac{1}{2}\,\{O_2\} = \{CO_2\} + 283 \text{ кДж}.$$

Вычитая второе уравнение из первого, находим искомую теплоту образования окиси углерода:

$$[C] + \frac{1}{2}\,\{O_2\} = \{CO\} + 111 \text{ кДж}.$$

§ 18. Теплоемкость

1. *Теплоемкостью тела C называется отношение бесконечно малого количества теплоты δQ, полученного телом, к соответствующему приращению dT его температуры*:

$$C = \frac{\delta Q}{dT}. \tag{18.1}$$

Когда масса тела равна единице, теплоемкость называют *удельной* — ее мы будем обозначать малой буквой c. Более удобны *молярная теплоемкость*. Так называют теплоемкость одного моля вещества. Эту величину мы будем обозначать буквой C.

Приращение температуры dT не определяет еще полностью того бесконечно близкого состояния, в которое переходит система из заданного состояния. Рассмотрим, например, физически однородное тело, состояние которого полностью определяется двумя параметрами, в качестве которых можно взять объем и температуру. Пусть исходное состояние изображается точкой $M(V,T)$ (рис. 16). Проведем прямую RS, параллельную оси объемов и отстоящую от точки M на dT. Все точки этой прямой изображают состояния с одной и той же температурой $T+dT$, но с различными объемами. Система из состояния M может перейти в различные бесконечно близкие состояния $R, N, S, \ldots$, лежащие на этой прямой. Всем этим переходам соответствует одно и то же повышение температуры, но, вообще говоря, различные количества теплоты δQ. Будут разными и теплоемкости системы при таких переходах. Поэтому теплоемкость есть характеристика не одного какого-либо состояния системы, а двух бесконечно близких состояний ее, из которых одно является начальным, а другое конечным. Вместо двух бесконечно близких состояний можно задать одно из них и направление пути перехода системы в бесконечно близкое состояние. Таким образом, теплоемкость не есть функция состояния тела, а является характеристикой *бесконечно малого процесса*, совершаемого телом.

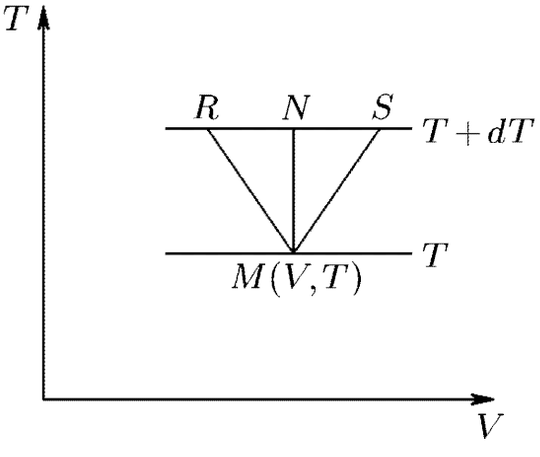


Рис. 16

2. Придадим этим рассуждениям количественную форму. На основании формул (18.1) и (15.5) можно написать

$$C = \frac{dU + P\,dV}{dT}.$$

Так как

$$dU = \left(\frac{\partial U}{\partial T}\right)_V dT + \left(\frac{\partial U}{\partial V}\right)_T dV,$$

то

$$C = \left(\frac{\partial U}{\partial T}\right)_V + \left[\left(\frac{\partial U}{\partial V}\right)_T + P\right]\frac{dV}{dT}. \tag{18.2}$$

Объем V зависит не только от температуры T, но и от давления P. В зависимости от того, как меняется давление, отношение dV/dT может принять любое значение. Чтобы придать выражению (18.2) однозначный смысл, надо фиксировать значение этого отношения. Иными словами, надо указать в плоскости VT направление пути,

по которому система переходит в бесконечно близкое состояние. Так как это направление может быть любым, то теплоемкость C, вообще говоря, может принимать любые значения от $-\infty$ до $+\infty$. В частности, для изотермического процесса $C = \pm\infty$, так как в этом случае $dT = 0$, $\delta Q \neq 0$. Для адиабатического процесса $\delta Q = 0$, $C = 0$.

Особое значение имеют теплоемкости при постоянном объеме и постоянном давлении, обозначаемые символами C_V и C_P. Если объем остается постоянным, то $dV = 0$, и, следовательно,

$$C_V = \left(\frac{\partial U}{\partial T}\right)_V. \tag{18.3}$$

Если же постоянно давление, то отношение dV/dT переходит в частную производную $(\partial V/\partial T)_P$. В этом случае формула (18.2) дает

$$C_P = \left(\frac{\partial U}{\partial T}\right)_V + \left[\left(\frac{\partial U}{\partial V}\right)_T + P\right]\left(\frac{\partial V}{\partial T}\right)_P. \tag{18.4}$$

Для разности теплоемкости $C_P - C_V$ получаем

$$C_P - C_V = \left[\left(\frac{\partial U}{\partial V}\right)_T + P\right]\left(\frac{\partial V}{\partial T}\right)_P. \tag{18.5}$$

Можно написать выражение для C_P, которым часто пользуются. Если процесс протекает при постоянном давлении, то на основании определения энтальпии $\delta Q = dI$. Поэтому

$$C_P = (\partial I/\partial T)_P. \tag{18.6}$$

3. Формулы (18.3) и (18.6) позволяют установить зависимость теплового эффекта реакции от температуры. С этой целью продифференцируем уравнения (17.1) и (17.2) по температуре и используем формулы (18.3) и (18.6). В результате придем к соотношениям

$$dW_V/dT = (C_V)_1 - (C_V)_2, \tag{18.7}$$

$$dW_P/dT = (C_P)_1 - (C_P)_2. \tag{18.8}$$

Здесь буквой C с соответствующими индексами обозначены теплоемкости всей системы до и после реакции.

§ 19. Внутренняя энергия идеального газа. Закон Джоуля

1. Чтобы из общих термодинамических соотношений, установленных в предыдущих параграфах, можно было получить конкретные результаты, надо знать, во-первых, уравнение состояния

$$f(P, V, T) = 0. \tag{19.1}$$

Во-вторых, надо знать внутреннюю энергию тела как функцию параметров, определяющих его состояние, например

$$U = U(V, T). \tag{19.2}$$

Зависимость типа (19.2) называется *калорическим* (или *калориметрическим*) *уравнением состояния*, в отличие от зависимости типа (19.1), называемой *термическим уравнением состояния*. Оба эти уравнения не могут быть получены теоретически методами формальной термодинамики. Формальная термодинамика заимствует их из опыта.

2. Рассмотрим прежде всего приложения первого начала термодинамики к идеальным газам. Термическим уравнением состояния таких газов является уравнение Клапейрона–Менделеева. Для одного моля газа оно имеет вид

$$PV = RT.$$

Чтобы получить калорическое уравнение состояния, исследуем сначала, как зависит внутренняя энергия U от объема газа V.

Первый опыт, позволяющий в принципе дать приближенный ответ на этот вопрос, был поставлен Гей-Люссаком, хотя сам Гей-Люссак не уяснил его значение и не сделал из него надлежащих выводов. Два медных сосуда A и B одинаковых объемов (рис. 17) были соединены трубкой с краном C. Сосуд A был наполнен воздухом, сосуд B — откачан. При открытии крана C воздух из A устремлялся в B. Гей-Люссак наблюдал, что температура воздуха в A несколько понижалась, а в B — повышалась. Такое изменение температуры объясняется тем, что воздух в A при расширении совершал работу и на это затрачивал часть своей внутренней энергии. При достижении теплового равновесия между сосудами A и B в них устанавливалась одна и та же температура, равная первоначальной температуре воздуха в сосуде A. Какой вывод следует сделать из результата опыта? Весь воздух был заключен в жесткую оболочку, состоящую из стенок сосудов A и B и соединительной трубки. Внешняя работа не производилась. Тепло из окружающей среды, если и подводилось, то за время опыта было пренебрежимо мало. Поэтому внутренняя энергия воздуха в системе измениться не могла. Опыт показал, что температура газа не изменилась, тогда как объем его удвоился. Отсюда можно сделать вывод, что при неизменной температуре внутренняя энергия газа не зависит от его объема. Опыт в измененном виде был повторен Джоулем. В одном из сосудов находился воздух под давлением 22 атм, другой сосуд был откачан. Оба сосуда погружались в воду, перемешиваемую во время опыта, чтобы температура ее во всех частях была одной и той же. При открывании крана C воздух перетекал из одного сосуда в другой. Однако никакого изменения температуры окружающей воды не наблюдалось.

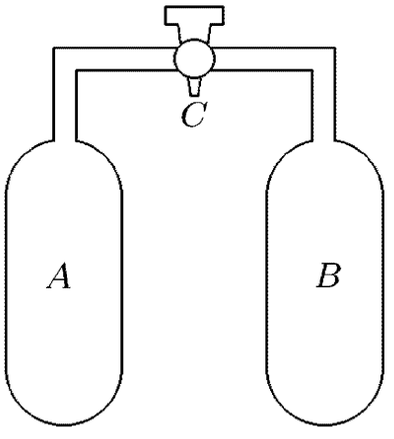


Рис. 17

3. Опыты Гей-Люссака и Джоуля не обладали достаточной чувствительностью. Причина малой точности этих опытов состоит, главным образом, в том, что теплоемкость воздуха была мала по сравнению с теплоемкостью сосудов и воды в калориметре. Вследствие этого малые изменения температуры было трудно обнаружить. Безупречные экспериментальные исследования были выполнены совместно Джоулем и В. Томсоном в течение десятилетия 1852–1862 гг. Эти классические исследования позволили ответить не только на вопрос о зависимости внутренней энергии газа от объема. Их значение много шире. Было открыто важное физическое явление, получившее название *эффекта Джоуля–Томсона.*

В опытах Джоуля и Томсона бралась цилиндрическая трубка, окруженная теплоизолирующим материалом. В середине трубки между двумя металлическими сетками MN и $M'N'$ (рис. 18) помещалась пробка из плотной ваты или очесов шелка. Исследуемый газ под действием разности давлений медленно протекал через пробку. Благодаря наличию пробки течение газа было спокойным, т. е. не возникало никаких турбулентных движений. Кроме того, пробка обеспечивала медленность течения, при котором кинетической энергией газа как величиной, пропорциональной квадрату скорости, можно было полностью пренебречь. В этих условиях в каждый момент времени газ по обе стороны пробки находился в термодинамически равновесных состояниях. Наличие тепловой защиты делало процесс течения адиабатическим. Давления газа по разные стороны пробки P_1 и P_2 поддерживались постоянными. Пробка и газ во время процесса обменивались между собой теплотой. Но когда процесс становился стационарным, теплообмен прекращался, физическое состояние пробки, а потому и ее внутренняя энергия оставались неизменными. При стационарном течении по одну сторону пробки устанавливалась постоянная температура газа T_1, а по другую — постоянная температура T_2. Эти температуры и измерялись в опыте. Стационарное течение газа через пробку называется *процессом Джоуля–Томсона*, а изменение температуры газа при таком течении — *эффектом Джоуля–Томсона.* Несмотря на медленность течения, этот процесс неквазистатичен, так как в системе в целом нет термодинамического равновесия.

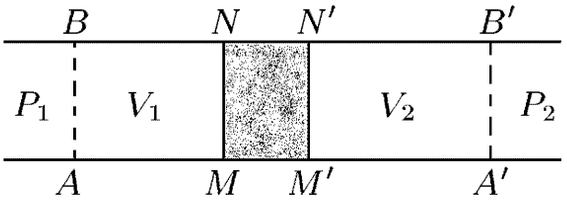


Рис. 18

4. Выделим мысленно по левую сторону пробки объем газа V_1, занимающий пространство $ABNM$. После прохождения через пробку выделенная порция газа займет пространство $M'N'B'A'$ с объемом V_2. Применим к ней первое начало термодинамики. Граница AB переходит в положение MN. При этом над газом производится работа $P_1 \cdot S \cdot AM = P_1 V_1$ (S — площадь поперечного сечения трубки). Граница же $M'N'$ переходит в положение $A'B'$, и газ производит внешнюю

работу $P_2 \cdot S \cdot M'A' = P_2V_2$. Полная работа, совершаемая газом, равна $A = P_2V_2 - P_1V_1$. Теплоты газ не получал, так как стенки трубки — адиабатические. Физическое состояние пробки и ее внутреняя энергия остались неизменными. Поэтому, обозначая внутреннюю энергию выделенной массы газа через U, можно написать

$$U_2 - U_1 + A = 0,$$

или

$$U_1 + P_1V_1 = U_2 + P_2V_2.$$

По определению энтальпии $I = U + PV$. Поэтому последнее равенство означает, что в *процессе Джоуля–Томсона энтальпия I газа не меняется*:

$$I_1 = I_2. \tag{19.3}$$

5. Следствия из этого соотношения, являющегося основным в теории эффекта Джоуля–Томсона, могут быть выведены только с помощью второго начала термодинамики. Они будут рассмотрены в § 46 и 104. Сейчас же мы ограничимся более скромной целью. Используя результаты опытов Джоуля–Томсона, найдем зависимость внутренней энергии идеального газа от его объема.

На опыте измерялись температуры газа по обе стороны пробки при стационарном течении его. Джоуль и Томсон нашли, что температура всех исследуемых газов немного понижалась, за исключением водорода, для которого она несколько повышалась. Разность температур $T_2 - T_1$ была тем меньше, чем лучше выполнялось уравнение Клапейрона–Менделеева $PV = RT$, т. е. чем ближе газ следовал законам идеальных газов. Отсюда можно заключить, что для идеальных газов $T_1 = T_2$. А так как по закону Бойля–Мариотта $P_1V_1 = P_2V_2$, то уравнение (19.2) дает $U_1 = U_2$ или

$$U(T, V_1) = U(T, V_2). \tag{19.4}$$

Таким образом, *внутренняя энергия идеального газа при неизменной температуре не зависит от его объема*. Иными словами, *для идеального газа U есть функция только температуры*. Этот опытный факт называется *законом Джоуля*. В дальнейшем (см. § 34) будет показано, что с помощью второго начала термодинамики закон Джоуля можно вывести теоретически из уравнения Клапейрона. Однако вид функции

$$U = U(T) \tag{19.5}$$

не может быть установлен методами формальной термодинамики. Эта функция может быть только выражена через теплоемкость C_V идеального газа. В самом деле, для любого тела величина C_V определяется выражением (18.3). А так как для идеального газа U от V не зависит, то в этом случае

$$C_V = dU/dT. \tag{19.6}$$

Отсюда следует, что сама *теплоемкость C_V идеального газа не зависит от объема, а является функцией только температуры*. Поэтому для идеального газа

$$U = \int_V (T)\, dT. \tag{19.7}$$

Опыт показывает, что во многих случаях C_V в широких температурных интервалах остается почти постоянной. Это имеет место для таких газов, как водород, гелий, аргон, неон, азот, кислород и пр., начиная с температур порядка 100 К до температур порядка 1000 К. Если совсем пренебречь зависимостью C_V от температуры, то вместо (19.7) можно написать более простую формулу

$$U = C_V T. \tag{19.8}$$

Как показывает опыт и молекулярно-кинетическая теория, для одноатомных газов $C_V \approx 3$ кал/(моль · К), для двухатомных $C_V \approx 5$ кал/(моль · К), для многоатомных $C_V \approx 6$ кал/(моль · К).

ЗАДАЧИ

1. Доказать, что энтальпия идеального газа не зависит от давления, а является функцией только его температуры.

2. Если начальные и конечные продукты реакции являются идеальными газами, то тепловой эффект реакции W_V не зависит от объемов газов, а W_P — от их давлений до и после реакции. Обе эти величины зависят только от температур газов до и после реакции. Доказать.

3. Показать, что внутренняя энергия воздуха в помещении не зависит от температуры, если давление наружного воздуха остается постоянным. Воздух считать идеальным газом.

Указание. Представить внутреннюю энергию в виде $U = C_V(PV/R)$.

§ 20. Уравнение Роберта Майера

1. Применим формулу (18.5) к идеальному газу. По закону Джоуля $(\partial U/\partial V)_T = 0$, из уравнения Клапейрона–Менделеева следует $(\partial V/\partial T)_P = R/P$. Поэтому указанная формула дает

$$C_P - C_V = R. \tag{20.1}$$

Это важное соотношение называется *уравнением Роберта Майера.*

Приведем еще один вывод уравнения (20.1). Пусть один моль идеального газа находится в цилиндре с поршнем. Закрепив поршень, повысим температуру газа на dT. Так как объем газа остается постоянным, то количество теплоты, необходимое для такого нагревания, $\delta_V Q = C_V dT$. А так как при этом не производится работа, то это количество теплоты равно приращению внутренней энергии газа:

$$C_V\, dT = dU. \tag{20.2}$$

Произведем теперь с тем же газом другой опыт. Пусть начальное состояние (T, V) будет тем же самым, что и в предыдущем опыте,

но поршень не закреплен, а может свободно перемещаться под постоянным внешним давлением P. По определению теплоемкости C_P для повышения температуры газа на dT потребуется теплота $\delta_P Q = C_P\,dT$. При этом газом будет совершена работа $\delta A = P\,dV$. Так как давление постоянно, то эту величину можно записать в виде $\delta A = d(PV) = d(RT) = R\,dT$. А так как внутренняя энергия газа зависит только от температуры, то она изменится на столько же, на сколько и в предыдущем опыте. Таким образом, во втором опыте

$$C_P dT = dU + R\,dT.$$

Подставив вместо dU выражение (20.2), мы снова придем к формуле (20.1).

Последний вывод особенно ясно показывает, что различие между C_P и C_V в случае идеального газа обусловлено только работой, которую совершает газ при расширении против постоянного внешнего давления. Важно подчеркнуть, что вывод использует закон Джоуля о независимости внутренней энергии идеального газа от занимаемого им объема.

2. Измерив теплоемкости C_P и C_V газа, можно вычислить *механический эквивалент теплоты*. Для этого можно воспользоваться уравнением Майера (20.1). Измеряя количество теплоты в калориях, можно на опыте найти разность $C_P - C_V$ в тепловых единицах. С другой стороны, газовую постоянную R можно измерить в механических единицах. Если пользоваться 20-градусной калорией, то измерения дают

$$C_P - C_V = 1{,}986 \text{ кал}/(\text{моль}\cdot\text{К}), \quad R = 8{,}314\cdot 10^7 \text{ эрг}/(\text{моль}\cdot\text{К}).$$

Приравнивая эти две величины, получаем

$$1 \text{ кал} = 4{,}18\cdot 10^7 \text{ эрг} = 4{,}18 \text{ Дж}.$$

Именно таким путем в 1842 г. механический эквивалент теплоты был вычислен теоретически немецким врачом Робертом Майером — одним из основоположников механической теории теплоты и первого начала термодинамики. Правда, найденное им значение механического эквивалента было заметно меньше истинного, что объясняется неточными значениями теплоемкостей воздуха C_P и C_V, которыми он располагал. Впервые достаточно точное значение механического эквивалента теплоты было определено в классических опытах Джоуля, начатых в 1847 г. Эти опыты общеизвестны, и нет необходимости останавливаться на их описании. Последующие измерения уточнили значение C_P и C_V для газов. Пользуясь этими значениями, было показано, что метод Майера, метод Джоуля и другие прямые методы, аналогичные методу Джоуля, приводят к одинаковым значениям механического эквивалента теплоты. Это было бы не так, если бы энергия идеального газа, помимо его температуры, зависела еще от объема. Поэтому отмеченное совпадение может служить одним из экспериментальных подтверждений закона Джоуля о независимости внутренней энергии идеального газа от его объема.

§ 21. Адиабатический процесс. Уравнение Пуассона

1. *Процесс, происходящий без подвода и отвода теплоты, называется адиабатическим.* Рассмотрим, как связаны между собой параметры, определяющие состояние идеального газа, когда газ совершает квазистатический адиабатический процесс. Полагая в уравнении (15.5) $\delta Q = 0$, $dU = C_V dT$, получим

$$C_V dT + P\,dV = 0.$$

Из уравнения Клапейрона–Менделеева

$$dT = \frac{d(PV)}{R} = \frac{P\,dV + V\,dP}{R} = \frac{P\,dV + V\,dP}{C_P - C_V}.$$

Исключая dT, получим

$$C_P P\,dV + C_V V\,dP = 0.$$

Введем обозначение

$$\gamma = C_P/C_V. \tag{21.1}$$

Тогда

$$\gamma P\,dV + V\,dP = 0. \tag{21.2}$$

Таково дифференциальное уравнение квазистатического адиабатического процесса для идеального газа. Теплоемкости C_P и C_V для идеальных газов могут зависеть от температуры. Но во многих случаях они остаются практически постоянными в широких температурных интервалах. Если это так, то постоянно также и их отношение γ. Тогда уравнение (21.2) легко интегрируется и дает

$$PV^\gamma = \text{const}. \tag{21.3}$$

Это уравнение называется *уравнением Пуассона* (1781–1840). Оно является уравнением *адиабаты*, т. е. кривой, графически изображающей квазистатический адиабатический процесс. Величина γ называется *адиабатической постоянной*. Поскольку $PV = RT$, уравнение адиабаты можно записать еще в двух видах:

$$TV^{\gamma-1} = \text{const}, \tag{21.4}$$

$$P^{\gamma-1}/T^\gamma = \text{const}. \tag{21.5}$$

Так как $\gamma > 1$, то из (21.4) следует, что при адиабатическом сжатии газ нагревается, а при адиабатическом расширении — охлаждается. На этом основано явление *пневматического огнива* (см. § 13). Это явление находит применение в дизелях, где воспламенение горючей смеси осуществляется путем адиабатического сжатия. Нагревание газа при адиабатическом сжатии объясняется тем, что во время сжатия над газом производится работа, которая идет на увеличение его внутренней энергии. А так как внутренняя энергия идеального газа зависит только от температуры, то это увеличение внутренней энергии проявляется в повышении его температуры. Аналогично объясняется и охлаждение газа при адиабатическом расширении.

2. Уравнения адиабаты (21.3)–(21.5) относятся только к квазистатическому адиабатическому процессу. Для неквазистатических адиабатических процессов эти уравнения неприменимы. Рассмотрим, например, цилиндр с адиабатическими стенками, разделенный на две равные половины адиабатической перегородкой. Пусть газ вначале занимал одну из этих половин. Если внезапно убрать перегородку, то произойдет адиабатический процесс расширения газа в пустоту. Этот процесс не квазистатический. Сначала возникает резко неравновесное состояние, сопровождающееся весьма бурными и сложными макроскопическими движениями газа. Затем эти макроскопические движения затухнут из-за внутреннего трения, их кинетическая энергия перейдет во внутреннюю энергию. В конце концов установится равновесное состояние, в котором газ будет занимать весь объем цилиндра при постоянной плотности и температуре. В ходе процесса газ не совершил никакой работы, тепло к нему не подводилось, а потому внутренняя энергия газа осталась без изменения. Отсюда на основании закона Джоуля можно заключить, что в конечном состоянии температура газа будет такой же, как в начале процесса. Было бы ошибочным применять к начальному и конечному состояниям газа уравнение адиабаты, например (21.4). Если это сделать, то мы пришли бы к ошибочному выводу, что в описанном адиабатическом процессе газ должен охлаждаться.

Разумеется, если отступления от неравновесности невелики, то можно пользоваться уравнением адиабаты и для не вполне равновесных процессов. Такие условия выполняются, например, в опытах Клемана и Дезорма (см. § 22) по определению адиабатической постоянной газа γ, а также в обычных звуковых волнах, распространяющихся в газах.

ЗАДАЧИ

1. Процесс, происходящий при постоянной теплоемкости, называется *политропическим*, а кривая, являющаяся его графическим изображением, — *политропой*. Найти уравнение политропы для идеального газа, если молярная теплоемкость его в политропическом процессе равна C.

Ответ. $TV^{n-1} = \text{const}$ или $PV^n = \text{const}$; $n = (C - C_P)/(C - C_V)$. Постоянная n называется *показателем политропы*.

2. При каких значениях показателя политропы идеальный газ нагревается при сжатии, а при каких охлаждается?

Ответ. Нагревается при $n > 1$, охлаждается при $n < 1$.

3. Нагревается или охлаждается идеальный газ, если он расширяется при постоянном давлении?

4. Вычислить работу одного моля идеального газа в политропическом процессе, если объем газа изменяется от начального значения V_1 до конечного V_2.

Ответ. $A = \dfrac{P_1 V_1^n}{1-n}\,(V_2^{1-n} - V_1^{1-n}) = \dfrac{P_2 V_2^n}{1-n}\,(V_2^{1-n} - V_1^{1-n})$.

5. Путем предельного перехода $n \to 1$ получить из предыдущей формулы выражение для работы идеального газа при изотермическом процессе.

6. На диаграмме PV (рис. 19) через производную точку проведена изотерма TT и адиабата SS для идеального газа, теплоемкость C_V которого не зависит от температуры. Показать, что политропе, проходящей через ту же точку и лежащей в заштрихованной области, соответствует отрицательная теплоемкость, а политропе в незаштрихованной области — положительная.

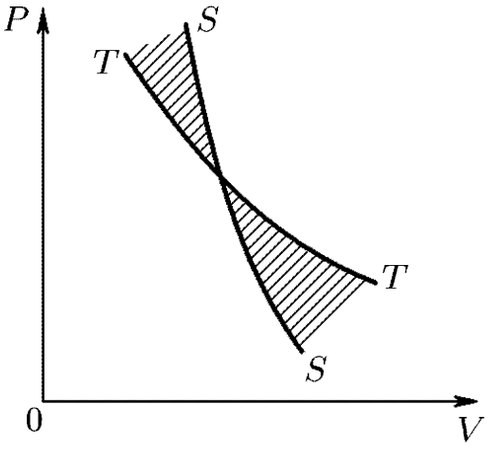


Рис. 19

7. Идеальный газ находится в эластичной адиабатической оболочке под давлением P_1, имея температуру T_1. Определить температуру газа T_2, которая установится после того, как внешнее давление на газ скачкообразно изменится до P_2. Сравнить изменение температуры в этом процессе с изменением, которое получилось бы, если бы адиабатический процесс проходил квазистатически.

Р е ш е н и е. При переходе из начального состояния (объем V_1, температура T_1) в конечное (объем V_2, температура T_2) внешнее давление совершает над газом работу $A_{\text{внеш}} = P_2(V_1 - V_2)$, которая идет на приращение внутренней энергии $U_2 - U_1 = C_V(T_2 - T_1)$. Применяя уравнение Клапейрона–Менделеева $PV = RT$, а также соотношение $C_P - C_V = R$, после несложных преобразований получим

$$T_2 = \Big(1 + \frac{\gamma - 1}{\gamma}\,\frac{P_2 - P_1}{P_1}\Big)T_1.$$

При квазистатическом адиабатическом процессе, как следует из (21.5),

$$T_2^{\text{квст}} = T_1\Big(\frac{P_2}{P_1}\Big)^{(\gamma-1)/\gamma}.$$

В первом случае T_2 меняется линейно, а во втором по степенному закону с изменением P_2, причем в бесконечно малой окрестности точки P_1 оба изменения идут одинаково быстро. А так как $0 < (\gamma - 1)/\gamma < 1$, то всегда $T_2^{\text{квст}} < T_2$. Значит, повышение температуры при внезапном адиабатическом расширении меньше, чем при квазистатическом адиабатическом процессе.

8. В длинной вертикальной цилиндрической трубке, закрытой с нижнего конца, может ходить без трения поршень, масса M которого велика по сравнению с массой газа, заключенного внутри трубки. В положении равновесия расстояние между поршнем и дном трубки равно l_0. Определить период малых колебаний, которые возникнут при отклонении поршня из положения равновесия, в предположении, что они являются изотермическими, а газ идеальным. Площадь поперечного сечения трубки равна S, нормальное атмосферное давление P_0. Рассмотреть предельный случай, когда $P_0 = 0$.

О т в е т. $T = 2\pi\sqrt{\dfrac{Ml_0}{Mg + P_0S}}$.

В предельном случае, когда $P_0 = 0$, $T = 2\pi\sqrt{l_0/g}$, т. е. период колебаний совпадает с периодом математического маятника длины l_0.

9. Решить предыдущую задачу в предположении, что колебания — адиабатические. Будет ли сказываться на результате зависимость адиабатической постоянной газа γ от температуры?

О т в е т. $T = 2\pi\sqrt{\dfrac{1}{\gamma}\,\dfrac{Ml_0}{Mg + P_0S}}$.

Формула верна и в том случае, когда γ зависит от температуры, так как для ее получения используется уравнение адиабаты в дифференциальной форме. В предельном случае, когда $P_0 = 0$, $T = 2\pi\sqrt{\dfrac{1}{\gamma}\dfrac{l_0}{g}}$.

§ 22. Определение C_P/C_V методом Клемана и Дезорма

Клеман (1779–1842) и Дезорм (1777–1862) в 1819 г. предположили и осуществили следующий метод измерения отношения теплоемкостей $\gamma = C_P/C_V$ для газов. Стеклянный баллон вместимостью в несколько литров (рис. 20) наполняется исследуемым газом при атмосферном давлении. С помощью насоса в баллон дополнительно накачивается небольшая порция того же газа, затем кран K_1 закрывается. Спустя короткое время температура газа в баллоне сравняется с температурой окружающего воздуха. После этого водяным манометром измеряют давление газа в баллоне. Обозначим это давление через P_1, а температуру газа через T_0. Затем на короткое время открывают кран K_2. При открытом кране K_2 часть газа выйдет из баллона, и его давление P_0 сравняется с атмосферным. При этом газ, оставшийся в баллоне, адиабатически расширится, совершив работу против давления окружающего воздуха. Вследствие этого его температура понизится до некоторого значения T. Во все время этого кратковременного процесса кран K_2 открыт. Затем кран K_2 быстро закрывается, и газ начинает медленно нагреваться, пока его температура не сравняется с температурой T_0 окружающего воздуха. Пусть давление газа в этот момент равно P_2. По измеренным давлениям P_1, P_2, P_0 можно вычислить отношение теплоемкостей $\gamma = C_P/C_V$.

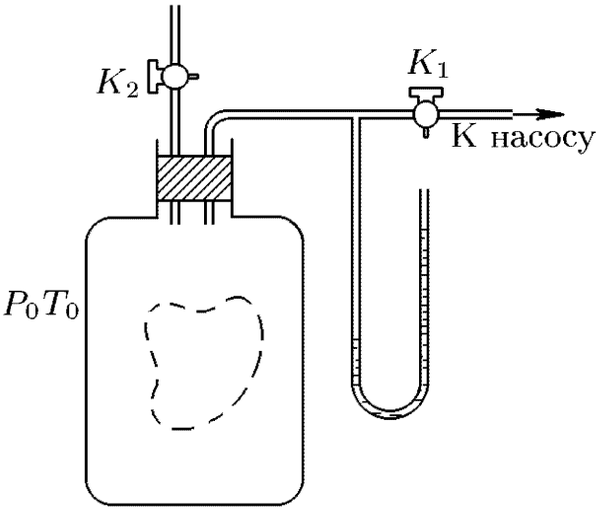


Рис. 20

Для этого мысленно выделим внутри нашего баллона произвольную порцию газа, ограниченную замкнутой поверхностью. Эта поверхность на рис. 20 изображена штриховой линией. Она играет роль оболочки, в которую заключена рассматриваемая порция газа. В различных процессах газ, заключенный в эту «оболочку», будет расширяться и сжиматься, совершая работу против давления окружающего газа и обмениваясь с ним теплотой. Поскольку кинетическая энергия возникающего макроскопического движения невелика, эти процессы могут рассматриваться как квазистатические. В моменты отсчета

давления параметры, характеризующие состояние газа внутри «оболочки», имеют следующие значения:

состояние 1: $P_1\, T_0\, V_1$,

состояние 2: $P_0\, T\, V_2$,

состояние 3: $P_2\, T_0\, V_2$.

Разности давлений $P_1 - P_0$ и $P_2 - P_1$ в сотни и тысячи раз меньше атмосферного давления P_0, а потому для упрощения вычислений с этими разностями можно обращаться как с бесконечно малыми дифференциалами. То же относится и к соответствующим изменениям объема выделенной порции газа. Переход газа из состояния 1 в состояние 2 совершается адиабатически, а потому соответствующие изменения давления и объема связаны уравнением адиабаты (21.2). Полагая в нем $dV = V_2 - V_1$, $dP = P_0 - P_1$, можно написать

$$\gamma P(V_2 - V_1) + V(P_0 - P_1) = 0.$$

В состояниях же 1 и 3 температуры газа одинаковы, а потому в этих состояниях произведение PV одно и то же. Следовательно, соответствующие изменения давления и объема связаны соотношением $P\,dV + V\,dP = 0$, или

$$P(V_2 - V_1) + V(P_2 - P_1) = 0.$$

Из этого соотношения совместно с предыдущим получаем

$$\gamma = \frac{P_1 - P_0}{P_1 - P_2}. \tag{22.1}$$

В эту формулу входит отношение давлений, а потому безразлично, в каких единицах измерять изменения давления. Проще всего разности давлений измерять в миллиметрах водяного столба с помощью манометра, как это показано на рис. 20.

ЗАДАЧА

Предполагая, что адиабатическая постоянная γ не зависит от температуры, обобщить формулу (22.1), не вводя предположения о малости разности давлений $P_1 - P_0$ и $P_1 - P_2$.

Ответ. $\gamma = \dfrac{\ln(P_1/P_0)}{\ln(P_1/P_2)}$.

При выводе формулы (22.1) использовалось уравнение адиабаты в дифференциальной форме (21.2), а потому формула справедлива не только в том случае, когда отношение γ постоянно, но и в тех случаях, когда оно меняется при изменении температуры. По этой причине мы и отдали предпочтение уравнению адиабаты в дифференциальной форме и не пользовались при выводе уравнением в интегральной форме (21.3), предполагающей постоянство отношения γ.

§ 23. Скорость звука в газах

1. В механике (см. т. I, § 85) выводится следующая формула для скорости распространения звука в газах:

$$c = \sqrt{dP/d\rho}, \tag{23.1}$$

где ρ — плотность газа. Но давление P зависит не только от ρ, а также и от температуры T. Поэтому надо указать, в каком смысле понимается производная $dP/d\rho$. Ньютон считал, что давление связано с плотностью законом Бойля–Мариотта: $P/\rho = \text{const}$. Это соответствует предположению, что разности температур между сгущениями и разряжениями воздуха в звуковой волне мгновенно выравниваются, так что распространение звука есть изотермический процесс. Если верно это предположение, то под $dP/d\rho$ следует понимать частную производную $(\partial P/\partial\rho)_T$. Тогда формула (23.1) перейдет в *формулу Ньютона*

$$c_{\mathrm{N}} = \sqrt{\frac{P}{\rho}} = \sqrt{\frac{RT}{\mu}}, \tag{23.2}$$

где μ — молярная масса газа, а индекс N указывает, что скорость звука вычислена по формуле Ньютона. Полагая для воздуха $\mu = 28{,}8$; $T = 273$ K, получаем по формуле (23.2): $c_{\mathrm{N}} = 280$ м/с, тогда как опыт дает $c = 330$ м/с.

2. Расхождение было устранено Лапласом (1749–1827). Он указал, что колебания плотности и связанные с ними колебания температуры в звуковой волне происходят настолько быстро, а теплопроводность воздуха настолько мала, что для таких процессов теплообмен не играет никакой роли. Разности температур между сгущениями и разрежениями воздуха в звуковой волне не успевают выравниваться, так что распространение звука можно считать адиабатическим процессом. В таком случае надо пользоваться не уравнением изотермы, а уравнением адиабаты (21.2). Если в это уравнение вместо объема V ввести плотность $\rho \sim 1/V$, то оно перейдет в уравнение

$$\gamma P\, d\rho - \rho\, dP = 0, \tag{23.3}$$

откуда для адиабатического процесса

$$\frac{dP}{d\rho} = \left(\frac{\partial P}{\partial \rho}\right)_{\text{ад}} = \gamma\,\frac{P}{\rho} = \gamma\left(\frac{\partial P}{\partial \rho}\right)_T. \tag{23.4}$$

Поэтому вместо формулы Ньютона (23.2) получается *формула Лапласа*

$$c_{\mathrm{L}} = \sqrt{\gamma\,\frac{P}{\rho}} = c_{\mathrm{N}}\sqrt{\gamma}. \tag{23.5}$$

Она дает для скорости звука величину в $\sqrt{\gamma}$ раз большую, чем формула Ньютона. Измерения γ для воздуха привели к результату $\gamma = 1{,}4$. Поэтому согласно формуле Лапласа при $T = 273$ K скорость звука в воздухе должна быть

$$c = 280\sqrt{1{,}4} = 380 \text{ м/с},$$

что находится в хорошем согласии с опытом.

3. На формулах Ньютона и Лапласа основан второй удобный метод экспериментального измерения отношения теплоемкостей γ, превосходящий по точности метод Клемана и Дезорма. Экспериментально измеряется скорость звука c в исследуемом газе. Значение γ вычисляется

по формуле

$$\gamma = (c/c_{\mathrm{N}})^2, \tag{23.6}$$

где c_{N} — так называемая *ньютонова скорость звука*, т. е. величина, определяемая формулой (23.2). Величина же, определяемая формулой (23.5), называется *лапласовой скоростью звука*.

ЗАДАЧА

Показать, что соотношение (23.5) между скоростью звука, рассчитанной по формуле Лапласа (адиабатической) и Ньютона (изотермической), справедливо для любого физически однородного изотропного вещества.

Решение. Для адиабатического процесса $du + P\,dv = 0$, где u и v — удельные значения внутренней энергии и объема вещества. Взяв за независимые переменные P и v, получаем

$$\Big(\frac{\partial u}{\partial P}\Big)_v dP + \Big(\frac{\partial u}{\partial v}\Big)_P dv = -P\,dv,$$

а затем

$$\Big(\frac{\partial P}{\partial v}\Big)_{\text{ад}} = -\frac{(\partial u/\partial v)_P + P}{(\partial u/\partial P)_v}.$$

Если объем v поддерживается постоянным, то остается только одна независимая переменная, от которой зависят все остальные величины. Таковой является либо P, либо T. Рассматривая $u(P)$ как сложную функцию $u[T(P)]$ и дифференцируя, получим

$$\Big(\frac{\partial u}{\partial P}\Big)_v = \Big(\frac{\partial u}{\partial T}\Big)_v \Big(\frac{\partial T}{\partial P}\Big)_v$$

или, используя соотношения (8.4) и (18.3), запишем

$$\Big(\frac{\partial u}{\partial P}\Big)_v = -c_v \Big(\frac{\partial T}{\partial v}\Big)_P \Big(\frac{\partial v}{\partial P}\Big)_T.$$

Поступая аналогично, находим

$$\Big(\frac{\partial u}{\partial v}\Big)_P + P = \Big(\frac{\partial u}{\partial T}\Big)_P \Big(\frac{\partial T}{\partial v}\Big)_P + P =$$
$$= \Big[\Big(\frac{\partial u}{\partial T}\Big)_P + P\Big(\frac{\partial v}{\partial T}\Big)_P\Big]\Big(\frac{\partial T}{\partial v}\Big)_P = c_P \Big(\frac{\partial T}{\partial v}\Big)_P.$$

После соответствующей подстановки получим

$$\Big(\frac{\partial P}{\partial v}\Big)_{\text{ад}} = \frac{c_P}{c_v}\Big(\frac{\partial P}{\partial v}\Big)_T. \tag{23.7}$$

Соотношения (23.4) и (23.5) являются следствиями формулы (23.7), если ввести плотность $\rho = 1/v$.

§ 24. Замечания относительно экспериментальных методов определения C_P и C_V для газов

Для идеальных газов значением величины $\gamma = C_P/C_V$ однозначно определяются их теплоемкости C_P и C_V, так как в случае идеальных газов эти теплоемкости связаны дополнительным соотношением

$$C_P - C_V = R. \tag{24.1}$$

Разрешая эти уравнения относительно C_P и C_V, находим

$$C_V = \frac{R}{\gamma - 1}, \quad C_P = \frac{\gamma R}{\gamma - 1}. \tag{24.2}$$

Детальное описание методов измерения теплоемкостей газов C_P и C_V не входит в задачи нашего курса. Поэтому ограничимся только одним замечанием. Непосредственное измерение на опыте C_V затруднительно, так как при постоянном V масса газа, а следовательно, и его теплоемкость всегда малы по сравнению со значением соответствующих величин для калориметра. На опыте удобнее измерять теплоемкость C_P и адиабатическую постоянную γ, а теплоемкость C_V вычислять по формуле $C_V = C_P/\gamma$.

Для измерения C_P исследуемый газ, нагретый до определенной температуры, заставляют протекать через спиральную металлическую трубку (змеевик), опущенную в воду калориметра (рис. 21). На одном конце змеевика поддерживаются постоянными давление P_1 и температура T_1 газа. На выходе змеевика поддерживается постоянным давление P_2 и измеряется температура газа T_2. Температура T_2 ниже T_1, так как при прохождении через змеевик газ отдает тепло воде калориметра. Обычно при прохождении через змеевик газ успевает охладиться и принять температуру воды в калориметре. Через змеевик можно пропустить большую массу газа и заметно нагреть воду в калориметре. Благодаря этому отпадает отмеченная выше трудность, встречающаяся при прямых измерениях теплоемкости C_V. По повышению температуры воды в калориметре можно определить количество теплоты, полученное калориметром. Обозначим это количество теплоты через Q. Эта же величина, взятая с противоположным знаком, дает количество теплоты, полученное газом, прошедшим через змеевик. Для упрощения расчета предположим, что через змеевик прошел один моль газа. Считая газ идеальным, вычислим работу A, совершенную им. Она равна $A = P_2V_2 - P_1V_1 = R(T_2 - T_1)$. Приращение внутренней энергии газа $U_2 - U_1 = C_V(T_2 - T_1)$. Подставляя эти величины в уравнение — $Q = U_2 - U_1 + A$, получим

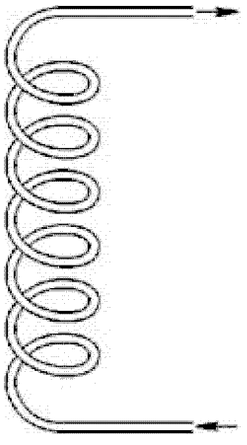


Рис. 21

$$Q = (C_V + R)(T_1 - T_2),$$

или на основании уравнения Майера (24.1)

$$Q = C_P(T_1 - T_2).$$

Отсюда легко вычислить искомую теплоемкость C_P.

§ 25. Уравнение Бернулли

1. Уравнение Бернулли было выведено в § 94 первого тома нашего курса. Однако там мы могли рассмотреть движение только несжимаемых жидкостей. Исследование движений сжимаемых жидкостей

и газов существенно опирается на законы термодинамики. Поэтому мы дополним материал первого тома термодинамическими соображениями.

Уравнение Бернулли относится к ламинарному стационарному течению идеальной жидкости. Жидкость понимается здесь в обобщенном смысле — газ считается частным случаем сжимаемой жидкости. Идеальность жидкости понимается в *гидродинамическом смысле*. Это значит, что, каково бы ни было движение жидкости, в ней никогда не возникают тангенциальные силы вязкости; взаимодействие между соприкасающимися элементами жидкости осуществляется исключительно с помощью нормальных сил давления. Кроме того, мы совершенно пренебрежем теплообменом между различными частями жидкости, считая его малым. По отношению к любой движущейся части жидкости окружающая жидкость играет роль адиабатической оболочки. Наше исследование относится поэтому к *адиабатическому ламинарному течению идеальной сжимаемой жидкости.*

2. Движущаяся жидкость, конечно, не является равновесной термодинамической системой. Однако если скорость макроскопического движения жидкости не очень быстро меняется в пространстве и во времени, то жидкость можно мысленно разбить на достаточно малые макроскопические части, каждая из которых, как целое, движется с определенной макроскопической скоростью v и внутреннее состояние которой может быть охарактеризовано теми же параметрами, что и в состоянии термодинамического равновесия, — температурой, давлением и плотностью. Эти параметры связаны между собой уравнением состояния $f(T, R, \rho) = 0$. Кроме того, между ними существует дополнительная связь, выражающая адиабатичность течения. В случае идеального газа, например, эта связь выражается соотношением (21.2) или при постоянном γ — соотношением $P = \text{const} \cdot \rho^\gamma$. В других случаях условие адиабатичности течения не может быть записано в столь простой форме. Но во всех случаях при адиабатическом течении, ввиду наличия уравнения состояния, из трех параметров T, P, ρ независимым остается только один, например плотность.

3. Уравнение Бернулли утверждает, что при стационарном ламинарном течении идеальной жидкости величина $\varepsilon + P/\rho$ остается постоянной вдоль линии тока:

$$\varepsilon + \frac{P}{\rho} = \text{const}. \tag{25.1}$$

Нет необходимости повторять вывод этого уравнения, так как в первом томе оно было получено без использования предположения о постоянстве плотности ρ. Единственное, что нужно сделать здесь, — это раскрыть смысл полной энергии ε, учитывая при этом сжимаемость жидкости. Величина ε есть полная энергия единицы массы жидкости. Она слагается из трех частей: кинетической энергии $v^2/2$ макроскопического движения, потенциальной энергии φ во внешнем силовом поле и внутренней энергии u. Если внешним полем является однородное

поле тяжести, то $\varphi = gh$, где h — высота, отсчитываемая от некоторого произвольного уровня. В этом случае уравнение (25.1) принимает вид

$$u + \frac{P}{\rho} + gh + \frac{v^2}{2} = \text{const}, \tag{25.2}$$

т. е. величина, стоящая слева, постоянна вдоль линии тока.

Величина $1/\rho$ равна удельному объему жидкости, а потому $u + P/\rho$ есть *удельная энтальпия*, т. е. энтальпия единицы массы жидкости. Обозначая ее буквой i, можно записать уравнение Бернулли в форме

$$i + gh + \frac{v^2}{2} = \text{const}. \tag{25.3}$$

Если течение происходит в горизонтальном направлении, то gh остается постоянной. В этих случаях

$$i + \frac{v^2}{2} = \text{const}. \tag{25.4}$$

При больших скоростях v соотношением (25.4) можно пользоваться и тогда, когда течение не горизонтально, так как в этих случаях изменениями потенциальной энергии gh с высотой можно пренебречь. Иными словами, можно полностью отвлечься от наличия силы тяжести. Именно с такими случаями мы и будем иметь дело в дальнейшем.

При медленных течениях можно пренебречь кинетической энергией. Тогда

$$i = \text{const}, \tag{25.5}$$

т. е. энтальпия вдоль линии тока остается постоянной. Этот результат был получен также при рассмотрении опыта Джоуля–Томсона.

4. Технический эффект Джоуля–Томсона может быть осуществлен без использования пробки. Газ, находящийся под высоким давлением (порядка сотен атмосфер), заставляют перетекать в пространство с низким давлением (порядка атмосферного) через вентиль или узкое отверстие. Такой процесс называется *дросселированием газа*. Он аналогичен течению газа по широкой трубе, в которой имеется очень узкое отверстие, за которым труба неограниченно расширяется. В этом случае к начальному и конечному состояниям газа также применимо соотношение (25.5). Действительно, применим уравнение Бернулли (25.4) к линии тока, начало и конец которой находятся перед и за узким отверстием, через которое протекает газ. Выберем эти точки в широких участках трубы, где скорость течения очень мала. Тогда в уравнении (25.4) кинетической энергией можно полностью пренебречь, и мы снова приходим к соотношению (25.5). Таким образом, процесс Джоуля–Томсона, независимо от того, осуществляется ли он продавливанием газа через пористую пробку или путем дросселирования через вентиль, может быть охарактеризован как такой процесс, при котором энтальпия газа в начальном и конечном состояниях одна и та же.

§ 26. Скорость истечения газа из отверстия

1. Вычислим скорость истечения сжатого газа из баллона через малое отверстие или сопло (рис. 22). Считая течение ламинарным и установившимся, возьмем произвольную линию тока, один конец которой (*2*) находится снаружи баллона вблизи отверстия, а другой (*1*) — внутри баллона, где скорость газа v_1 пренебрежимо мала. Применим уравнение Бернулли (25.4) к точкам *1* и *2* линии тока; тогда получим

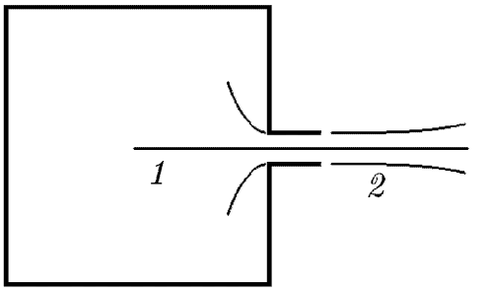


Рис. 22

$$i_1 + \frac{v_1^2}{2} = i_2 + \frac{v_2^2}{2}.$$

Величиной v_1^2 можно пренебречь. Опустим индекс у скорости v_2, в результате чего получаем

$$v = \sqrt{2(i_1 - i_2)}. \tag{26.1}$$

Эта формула применима как для идеальных, так и для реальных газов. Допустим теперь, что газ идеальный и что зависимостью теплоемкости C_V от температуры можно пренебречь. Тогда

$$i = u + \frac{P}{\rho} = \frac{C_V T}{\mu} + \frac{RT}{\mu},$$

или на основании (20.1)

$$i = \frac{C_P T}{\mu}. \tag{26.2}$$

Следовательно,

$$v = \sqrt{\frac{2}{\mu} C_P (T_1 - T_2)}. \tag{26.3}$$

В таком виде, однако, эта формула непригодна для вычислений, так как не известна температура T_2 струи газа при ее выходе из отверстия. Известны давление P_1 и температура T_1 газа в баллоне, а также наружное давление P_2. Температуру T_2 можно найти из уравнения адиабаты

$$\frac{P_1^{\gamma-1}}{T_1^{\gamma}} = \frac{P_2^{\gamma-1}}{T_2^{\gamma}}.$$

Оно дает

$$T_2 = T_1 \left(\frac{P_2}{P_1}\right)^{(\gamma-1)/\gamma}.$$

После подстановки этого значения в формулу (26.3) получаем

$$v = \sqrt{\frac{2}{\mu} C_P T_1 \left[1 - \left(\frac{P_2}{P_1}\right)^{(\gamma-1)/\gamma}\right]}. \tag{26.4}$$

Максимальная скорость достигается при истечении в вакуум. Она равна

$$v_{\text{вак}} = \sqrt{\frac{2}{\mu} C_P T}. \tag{26.5}$$

(Индекс 1 у температуры T мы опустили.) Подставляя вместо C_P ее значение из формулы (24.2), получим

$$v_{\text{вак}} = \sqrt{\frac{2}{\mu} \frac{\gamma}{\gamma - 1} RT} = \sqrt{\frac{2}{\gamma - 1} c}, \tag{26.6}$$

где c — скорость звука. Для молекулярного водорода при температуре $T = 1000$ K эта формула дает

$$v_{\text{вак}} = \sqrt{\frac{1{,}4}{0{,}4} \cdot 8{,}314 \cdot 10^3 \cdot 10^3} = 5400 \text{ м/с}.$$

2. Получение больших скоростей истечения газов является одной из важнейших проблем ракетной техники. Формула (26.6) показывает, что скорость истечения пропорциональна квадратному корню из абсолютной температуры и обратно пропорциональна квадратному корню из молекулярной массы газа. Поэтому в ракетной технике выгодно применять горючее с малой молекулярной массой, обладающее высокой калорийностью (чтобы температура T была возможно выше).

Сравнение формулы (26.5) с формулой (26.3) показывает, что при истечении в вакуум $T_2 = 0$, т. е. газ охлаждается до абсолютного нуля. Не следует придавать этому выводу большого значения. Он получен в предположении ламинарности течения, тогда как реальное истечение газа в вакуум всегда турбулентное. Кроме того, использована незаконная экстраполяция — газ считается идеальным вплоть до абсолютного нуля, а его теплоемкости C_P и C_V при истечении сохраняют постоянные значения, не зависящие от температуры.

ЗАДАЧА

Тело (например, космический корабль) движется в идеальном газе со скоростью v. В какой точке тела температура газа будет максимальной? Определить эту температуру, если температура окружающего газа равна T.

Р е ш е н и е. Перейдем в систему отсчета, в которой тело покоится. Считая течение газа в этой системе стационарным, применим к нему уравнение Бернулли (25.4). В рассматриваемом случае оно имеет вид $c_P T + v^2/2 = \text{const}$. Температура будет максимальна в точке, где $v = 0$, т. е. в критической точке. Так в гидродинамике называют точку на поверхности тела, в которой скорость натекающей жидкости обращается в нуль. В критической точке $i = c_P T_{\text{макс}}$, а потому

$$T_{\text{макс}} = T\left(1 + \frac{v^2}{2T c_P}\right),$$

или

$$T_{\text{макс}} = T\left[1 + \frac{M^2(\gamma - 1)}{2}\right],$$

где $M = v/c$ — число Маха (c — скорость звука).

Глава III

ВТОРОЕ НАЧАЛО ТЕРМОДИНАМИКИ

§ 27. Общие замечания о первом и втором началах термодинамики

1. Первое начало термодинамики не дает никаких указаний относительно направления, в котором могут происходить процессы в природе. Для изолированной системы, например, первое начало требует только, чтобы при всех процессах энергия системы оставалась постоянной. Если *1* и *2* — два состояния такой системы, то первое начало ничего не может сказать, будет ли система переходить из состояния *1* в состояние *2*, или из состояния *2* в состояние *1*. Вообще, на основании первого начала нельзя выяснить, будут ли в изолированной системе происходить какие-либо процессы.

Пусть адиабатически изолированная система состоит из двух тел, взаимодействующих между собой, но не взаимодействующих с другими телами. Тогда, как было показано в § 16, теплообмен между ними подчиняется условию $Q_1 = -Q_2$. Теплота Q_1, полученная одним телом, равна теплоте $-Q_2$, отданной другим телом. В каком направлении будет переходить теплота — на этот вопрос первое начало термодинамики ответить не может. Первому началу не противоречил бы, например, процесс, в котором теплота самопроизвольно переходит от тела менее нагретого к телу более нагретому. Вопрос о количественной мере температуры чужд первому началу термодинамики. Это проявляется в том, что первое начало не привело ни к какой рациональной шкале температур.

Второе начало термодинамики, наоборот, позволяет судить о *направлении процессов*, которые могут происходить в действительности. Но этим значение второго начала не исчерпывается. Второе начало позволяет вполне удовлетворительно решить вопрос о количественной мере температуры и построить рациональную температурную шкалу, не зависящую от произвола выбора термодинамического тела и устройства термометра. Оно, совместно с первым началом, позволяет также установить множество точных количественных соотношений между различными макроскопическими параметрами тел в состоянии термодинамического равновесия. Все такие точные соотношения получили общее называние *термодинамических соотношений*.

2. Основоположником второго начала термодинамики считается французский инженер и физик Сади Карно. В своем сочинении «О движущей силе огня и о машинах, способных развивать эту силу», вышедшем в 1824 г. (т. е. значительно ранее открытия первого начала Майером, Джоулем и Гельмгольцем), Карно исследовал условия

превращения теплоты в работу. Однако тогда Карно стоял на точке зрения теории теплорода (позднее он от нее отказался), а поэтому ему не удалось дать ясную и четкую формулировку второго начала термодинамики. Это было сделано только в 1850–1851 гг. независимо друг от друга немецким физиком Рудольфом Клаузиусом и шотландским физиком Вильямом Томсоном (лордом Кельвином). Они сформулировали основной постулат, выражающий второе начало термодинамики, и вывели из него главнейшие следствия.

§ 28. Различные формулировки основного постулата, выражающего второе начало термодинамики

1. Чтобы перейти к формулировке постулата второго начала термодинамики, следуя историческому ходу идей, рассмотрим схематически работу тепловой машины. В цилиндре машины (рис. 23) помещается

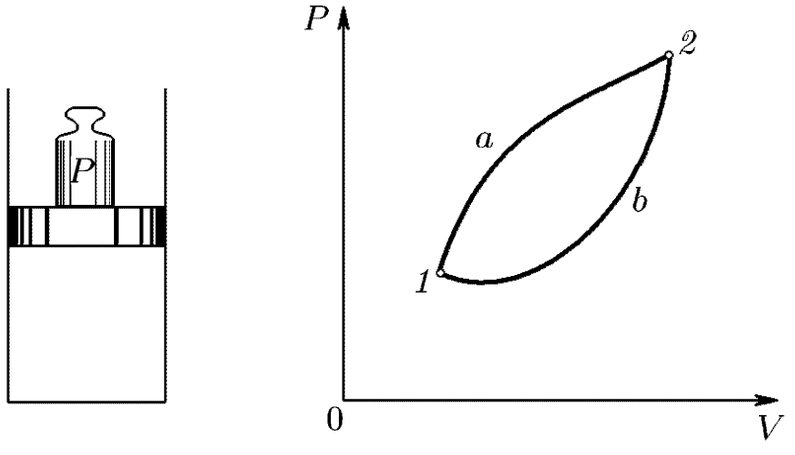


Рис. 23

газ или какое-либо другое вещество, называемое *рабочим телом*. Для определенности будем считать, что рабочим телом является газ. Пусть на диаграмме VP начальное состояние рабочего тела изображается точкой *1*. Приведем дно цилиндра в тепловой контакт с *нагревателем*, т. е. телом, температура которого выше температуры газа в цилиндре. Газ будет нагреваться и расширяться — этот процесс изображен кривой *1a2*. Рабочее вещество получит от нагревателя теплоту Q_1 и совершит положительную работу A_1. По первому началу

$$Q_1 = U_2 - U_1 + A_1. \tag{28.1}$$

Теперь надо вернуть поршень в исходное положение, т. е. сжать газ. Это надо сделать так, чтобы работа A_2, затраченная на сжатие, была меньше A_1. С этой целью приведем дно цилиндра в тепловой контакт с *холодильником*, т. е. телом, температура которого ниже температуры газа в цилиндре, и сожмем газ по пути *2b1*. В результате газ вернется в исходное состояние *1*. При этом он отдаст холодильнику количество

теплоты Q_2. По первому началу

$$-Q_2 = U_1 - U_2 - A_2. \tag{28.2}$$

Отсюда в комбинации с (28.1)

$$Q_1 - Q_2 = A_1 - A_2. \tag{28.3}$$

Таким образом, тепловая машина совершила круговой процесс, в результате которого нагреватель отдал количество теплоты Q_1, холодильник получил количество теплоты Q_2, $Q = Q_1 - Q_2$ пошло на производство работы $A_1 - A_2$. Отношение

$$\eta = \frac{A}{Q_1} = \frac{Q_1 - Q_2}{Q_1} \tag{28.4}$$

называется *коэффициентом*, или *экономическим коэффициентом полезного действия* (КПД) тепловой машины.

2. Возникает вопрос, нельзя ли построить периодически действующую тепловую машину без холодильника, т. е. добиться того, чтобы $Q_2 = 0$ и, следовательно, $\eta = 1$? Такая машина могла бы превращать в работу всю теплоту, заимствованную от одного теплового резервуара. Возможность ее построения не противоречит закону сохранения энергии. По своему практическому значению она почти не уступала бы перпетуум мобиле, так как с ее помощью можно было бы производить работу за счет практически неисчерпаемых запасов внутренней энергии, содержащихся в водах океанов и морей, воздушной атмосфере и недрах Земли. Такую машину Вильгельм Оствальд (1853–1932) называл *перпетуум мобиле второго рода* в отличие от *перпетуум мобиле первого рода*, т. е. вечного двигателя, производящего работу из ничего, возможность которого отрицается законом сохранения энергии.

Но уже Карно понял, что такая машина принципиально невозможна. Работу тепловых двигателей он сравнивал с работой двигателей водяных. Производство работы в последних двигателях связано с падением воды с более высокого на более низкий уровень. Так и возможность производства работы тепловыми двигателями обусловлена по Карно переходом теплоты от тела более нагретого к телу менее нагретому. Исходя из этой аналогии, Карно вывел ряд правильных положений, с которыми мы ознакомимся в дальнейшем. Ошибка Карно состояла лишь в том, что он вместе со своими современниками считал, что во всех процессах теплота не может создаваться и уничтожаться.

Опытные факты говорят против возможности построения перпетуум мобиле второго рода. Поэтому невозможность построения такого перпетуум мобиле была возведена в постулат. Он называется *постулатом второго начала термодинамики* и является обобщением опытных фактов. Доказательством этого постулата является согласие всех вытекающих из него следствий с опытом. До сих пор, применяя этот постулат к макроскопическим системам, размеры которых не очень малы, физика нигде не натолкнулась на противоречия. Поэтому

постулат второго начала термодинамики покоится на надежной экспериментальной основе. Приведем три точных формулировки постулата.

3. Вильям Томсон (получивший позднее за научные заслуги титул лорда Кельвина) в 1851 г. дал такую формулировку постулата второго начала термодинамики: «*Невозможен круговой процесс, единственным результатом которого было бы производство работы за счет охлаждения теплового резервуара*».

Напомним, что под тепловым резервуаром понимают тело или систему тел, находящуюся в состоянии термодинамического равновесия и обладающую запасом внутренней энергии. Но тепловой резервуар сам макроскопической работы не совершает, а может только передавать внутреннюю энергию другому телу или системе тел. Если последняя система производит работу за счет внутренней энергии теплового резервуара, то она называется в термодинамике *рабочим телом*. Таким образом, согласно Томсону, *невозможен круговой процесс, единственным результатом которого было бы производство работы за счет уменьшения внутренней энергии теплового резервуара.*

4. Можно конкретизировать, в чем должно выражаться производство внешней работы, и таким путем получить множество разнообразных формулировок основного постулата. Одна из них принадлежит Планку (1858–1947). Она состоит в следующем: «*Невозможно построить периодически действующую машину, единственным результатом которой было бы поднятие груза за счет охлаждения теплового резервуара*».

Указание на периодичность действия машины в формулировке Планка существенно. Точно так же в формулировке Томсона существенно, что процесс должен быть круговым. Действительно, возможен процесс (но не круговой), единственным результатом которого было бы поднятие груза за счет внутренней энергии, заимствованной от теплового резервуара. Планк приводит следующий пример. Допустим, что в цилиндре с поршнем (см. рис. 7) находится идеальный газ. На поршне лежит груз P. Приведем дно цилиндра в тепловой контакт с достаточно большим тепловым резервуаром, температура которого превышает температуру газа на бесконечно малую величину. Затем будем бесконечно малыми порциями разгружать поршень. Тогда газ будет изотермически расширяться, производя работу A по поднятию груза. По первому началу

$$Q = U_2 - U_1 + A.$$

Так как внутренняя энергия идеального газа U зависит только от температуры, которая в описанном процессе не меняется, то $U_2 - U_1 = 0$ и, следовательно,

$$Q = A.$$

Таким образом, теплота Q, заимствованная у теплового резервуара, полностью перешла в работу поднятия груза. Это не противоречит постулату второго начала термодинамики, так как описанный процесс не круговой, т. е. машина не является периодически действующей. Вот

если бы удалось каким угодно способом, оставляя груз в поднятом положении, сжать груз, приведя его в начальное состояние, и вернуть поршень в исходное положение таким образом, чтобы во всех остальных телах (за исключением теплового резервуара, часть внутренней энергии которого пошла на производство работы) *не произошло никаких изменений*, то тогда получилось бы противоречие с постулатом второго начала термодинамики, ибо постулат второго начала термодинамики утверждает, что сделать это невозможно никакими способами.

Формулировка Планка отличается от формулировки Томсона лишь по форме. Для удобства условимся называть *процессом Томсона–Планка* воображаемый круговой процесс, единственным результатом которого является производство работы за счет охлаждения теплового резервуара. Тогда постулат сводится к утверждению, что *процесс Томсона–Планка невозможен.*

5. Клаузиус (1822–1888) в 1850 г. дал существенно иную формулировку основного постулата. Он выдвинул следующее положение: «Теплота не может самопроизвольно переходить от тела менее нагретого к телу более нагретому». Под теплотой здесь надо понимать внутреннюю энергию тела. Постулат Клаузиуса никоим образом не сводится к утверждению, что при непосредственном тепловом контакте двух тел теплота всегда переходит от тела более нагретого к телу менее нагретому. Это утверждение вообще не составляет содержания физического закона, а является просто определением того, какое из двух тел условились называть более, а какое менее нагретым (см. § 4, п. 4). Передачу теплоты (точнее, внутренней энергии) можно осуществить не только тепловым контактом, но и бесчисленным множеством других способов. Например, все тела излучают и поглощают видимые или невидимые лучи (электромагнитные волны). Излучение одного тела можно с помощью линзы или сферического зеркала сконцентрировать на другом теле и таким путем передать ему тепло. Однако не всякая передача возможна. Содержание постулата Клаузиуса как раз и состоит в том, что невозможно *каким бы то ни было способом* забрать теплоту от тела менее нагретого, целиком передать его телу более нагретому и притом так, чтобы в природе больше не произошло никаких изменений. Любой воображаемый процесс, в котором осуществляется такая передача тепла, называется *процессом Клаузиуса.* Таким образом, постулат утверждает, что *процесс Клаузиуса невозможен.*

Но постулат Клаузиуса не утверждает, что передача тепла от тела менее нагретого к телу более нагретому вообще невозможна. Она невозможна при условии, что во всех остальных телах *никаких изменений не должно произойти.* В этом смысле слова «*самопроизвольно*», употребленного при формулировке второго начала термодинамики. Если же допустить другие процессы, то передача теплоты от тела менее нагретого к телу более нагретому становится невозможной. Такие процессы называются *компенсирующими процессами* или, короче, *компенсациями.* Так, в холодильных машинах теплота, заимствованная от менее нагретого тела, передается более нагретому телу. Это не

противоречит постулату Клаузиуса, так как такой переход происходит здесь не самопроизвольно, а сопровождается работой электрического мотора. Электрический холодильник перестает действовать, если выключить питающий его ток. Простейшая тепловая машина, о которой говорилось в начале этого параграфа (см. рис. 23), может работать как *холодильная машина*. Для этого расширение рабочего вещества следует производить по кривой $1b2$, а сжатие — по кривой $2a1$, лежащей выше. Совершая расширение $1b2$, машина будет заимствовать от холодильника теплоту Q_2; при сжатии по кривой $2a1$ она передаст нагревателю теплоту $Q_1 > Q_2$. При этом над машиной будет произведена положительная работа $A' = Q_1 - Q_2$. Производство этой работы и является здесь компенсирующим процессом.

6. Из невозможности процесса Томсона–Планка следует невозможность процесса Клаузиуса.

Для доказательства предположим противное, т. е. что процесс Клаузиуса возможен. Взяв простейшую тепловую машину, произведем круговой процесс, в результате которого машина отнимет от нагревателя теплоту Q_1, передаст холодильнику теплоту Q_2 и совершит положительную работу $A = Q_1 - Q_2$. Затем с помощью процесса Клаузиуса теплоту Q_2 вернем от холодильника к нагревателю. Тогда получится круговой процесс, единственным результатом которого является производство работы A за счет эквивалентного ей количества теплоты $Q_1 - Q_2$, отнятого от нагревателя; никаких других изменений в природе не произойдет. Но это есть процесс Томсона–Планка, а он по предположению невозможен. Получившееся противоречие и доказывает наше утверждение.

7. Обратно, из невозможности процесса Клаузиуса вытекает невозможность и процесса Томсона–Планка. Для доказательства предположим противное, т. е. что процесс Томсона–Планка возможен. Тогда, пользуясь этим круговым процессом, отнимем от менее нагретого тела теплоту Q и за счет этой теплоты произведем механическую работу, например, подняв груз. Затем используем энергию поднятого груза для нагревания, например путем трения, более нагретого тела. В результате Q перейдет от менее нагретого тела к телу более нагретому, и никаких других изменений не произойдет. Но это есть процесс Клаузиуса, а он невозможен. Получившееся противоречие и доказывает высказанное утверждение. При доказательстве мы использовали не только постулат Клаузиуса, но воспользовались также утверждением, что потенциальная энергия поднятого груза может быть целиком превращена в теплоту. Это утверждение является следствием повседневных наблюдений, которые показывают, что при столкновении падающего груза с препятствием он в конце концов останавливается. Потенциальная энергия груза пропадает, зато появляется теплота. По первому принципу термодинамики количество этой теплоты точно равно потерянной потенциальной энергии груза.

Таким образом, *постулаты Клаузиуса и Томсона–Планка эквивалентны*.

§ 29. Обратимые и необратимые процессы

1. Если в результате какого-либо процесса система переходит из состояния A в другое состояние B и если возможно вернуть ее хотя бы одним способом в исходное состояние A и притом так, чтобы во всех остальных телах не произошло никаких изменения, то этот процесс называется *обратимым*. Если же это сделать невозможно, то процесс называется *необратимым*. Примером необратимого процесса может служить переход теплоты от более нагретого тела к телу менее нагретому при тепловом контакте этих тел. Необратимость такого процесса непосредственно следует из постулата Клаузиуса. Необратимым является процесс получения теплоты путем трения. Его необратимость является непосредственным следствием постулата Томсона–Планка.

Если систему из конечного состояния B можно вернуть в исходное состояние A *безразлично каким способом*, не требуя, чтобы она обязательно проходила через ту же последовательность состояний, что и в прямом процессе $A \to B$, то такой процесс называют *обратимым в широком смысле слова*. Если же возможен обратный процесс $B \to \to A$, переводящий систему в исходное состояние A через ту же последовательность состояний, через которую прошла система в прямом процессе $A \to B$, то процесс $A \to B$ называется *обратимым в узком смысле слова*. Всякий процесс, обратимый в узком смысле, очевидно, обратим и в широком смысле слова.

2. Ясно, что *процесс, обратимый в узком смысле, необходимо должен быть квазистатическим*. Действительно, неквазистатический процесс идет с какой-то скоростью, отличной от нуля. Если же скорость отлична от нуля, то она должна входить в число внутренних параметров, определяющих состояние системы. Но тогда состояния системы в прямом и обратном процессах не могут быть тождественными — они всегда будут отличаться направлениями своих скоростей. Это различие исчезнет только в пределе, когда процесс идет квазистатически — бесконечно медленно.

Обратно, *если внутренние параметры системы в состоянии равновесия определяются внешними условиями однозначно, то всякий квазистатический процесс будет обратимым в узком смысле*. Оговорка об однозначности существенна, так как в приводимом ниже доказательстве используется следующее положение. Если внешними воздействиями система выводится из положения равновесия, то по устранении этого воздействия она в конце концов возвращается в прежнее равновесное состояние. Если же однозначности нет, то может произойти переход и в другие равновесные состояние.

Приведем теперь само доказательство, хотя после сделанных разъяснений почти самоочевидно. Квазистатический процесс есть бесконечно медленный процесс, состоящий из последовательности состояний равновесия, точнее, состояний, бесконечно мало отличающихся от равновесных. Если взять какое-либо равновесное состояние, то по самому определению равновесия в отсутствие внешних воздействий

оно будет сохраняться неограниченно долго. Чтобы начался процесс, надо с помощью внешних воздействий нарушить равновесие, т. е. менять внешние параметры и температуру окружающей среды. Для квазистатичности процесса необходимо, чтобы эти изменения совершались настолько медленно, чтобы в каждый момент времени система находилась либо в равновесном состоянии, либо в состоянии, как угодно мало отличающемся от равновесного. В пределе получится идеализированный процесс, идущий с бесконечно малой скоростью и состоящий из последовательных состояний равновесия. С помощью таких процессов можно перевести систему из начального состояния A в конечное состояние B, отстоящее от начального как угодно далеко; для этого требуется только достаточно большое время. Конкретно, существенные изменения состояния системы должны совершаться за времена очень большие по сравнению с временами релаксации. Если изменить знаки бесконечно малых приращений внешних параметров и температуры на противоположные, то система снова вернется в исходное состояние A, проходя в обратном порядке через состояния, бесконечно мало отличающиеся от состояний, через которые она проходила ранее. В пределе, когда прямой и обратный процессы сделаются строго равновесными, исчезнет и это бесконечно малое различие. При этом в результате прямого и обратного процесса в окружающих телах не произойдет никаких изменений, поскольку внешние параметры и температура окружающей среды вернутся в точности к своим исходным значениям. Таким образом, при оговоренной выше однозначности квазистатический процесс не только обратим вообще, но обратим в узком смысле слова. Это утверждение постоянно используется в термодинамике. В частности, при том же условии *всякий квазистатический круговой процесс может происходить как в прямом, так и в обратном направлении*. Ниже, когда квазистатичность и обратимость являются синонимами, употребляется термин «квазистатический».

3. В качестве примера, иллюстрирующего приведенные рассуждения, рассмотрим адиабатически изолированную систему — газ в цилиндре с поршнем, который может в нем свободно перемещаться. Внешнее давление P можно осуществить, положив на поршень груз. Для того чтобы груз можно было увеличивать или уменьшать малыми порциями, предположим, что поршень нагружен мелким песком. Пусть газ адиабатически расширяется, переходя из начального равновесного состояния M в конечное равновесное состояние N (рис. 24). Этот процесс можно осуществить, снимая с поршня песчинку за песчинкой. Снимем сначала одну песчинку. Внешнее давление уменьшится, а газ расширится. Это расширение очень мало, и его трудно заметить.

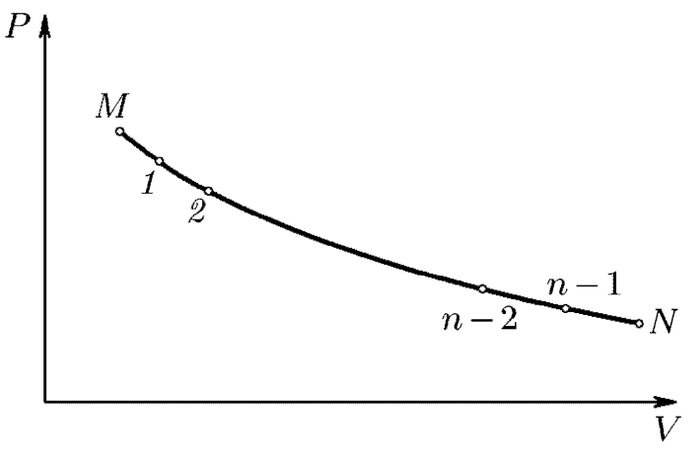


Рис. 24

Но по существу оно представляет собой неравновесный процесс, сопровождающийся весьма сложными макроскопическими движениями газа. Однако в конце концов газ придет в состояние равновесия, и это состояние изобразится на графике точкой *1*. Сняв вторую песчинку, заставим газ совершить второй неравновесный процесс, переводящий его в равновесное состояние *2*. Повторив эту операцию n раз, переведем газ в конце концов в равновесное состояние N, пройдя при этом через конечное состояние $(n-1)$ равновесных состояний *1*,*2*, ..., $(n-1)$. Каждое из этих состояний получается из предыдущего путем малого, но неравновесного процесса, так что процесс $M \to N$ в целом является неравновесным.

Попытаемся теперь вернуть газ в исходное состояние M, последовательно нагружая поршень по одной песчинке. При этом мы пройдем через ту же конечную последовательность равновесных состояний $(n-1), (n-2), \ldots, M$, что и в прямом процессе. Однако промежуточные малые неравновесные процессы будут уже иными. Например, в обратном процессе $N \to (n-1)$ газ сжимается при несколько большем давлении (число песчинок на поршне больше на одну), чем в прямом процессе $(n-1) \to N$. Поэтому при обратном процессе над газом надо совершить несколько большую работу, чем в прямом процессе. Процесс $M \to N$, состоящий из конечного числа равновесных состояний, в целом является необратимым.

Допустим теперь, что вес песчинки неограниченно уменьшается, а общее число песчинок неограниченно растет, однако так, что общий вес песка остается неизменным. Тогда в пределе неравновесный процесс $M \to N$ перейдет в квазистатический процесс и изобразится непрерывной линией MN. Той же линией, но проходимой в противоположном направлении, изобразится и обратный процесс. Работа, совершаемая газом как в прямом, так и в обратном процессе, численно одна и та же и изображается площадью криволинейной трапеции, ограниченной сверху кривой MN. Для приведения газа в исходное состояние должна быть затрачена такая же работа, какую совершил сам газ при расширении. Ясно поэтому, что квазистатическое расширение газа, которое мы рассмотрели, есть процесс обратимый и притом в узком смысле слова.

ЗАДАЧИ

1. Моль идеального газа с постоянной теплоемкостью C_V заключен в цилиндр с адиабатическими стенками и поршнем, который может перемещаться в цилиндре без трения. Поршень находится под постоянным внешним давлением P_1. В некоторый момент времени внешнее давление скачкообразно уменьшают или увеличивают до P_2. (Этого можно достигнуть, снимая часть груза с поршня или добавляя новый груз.) В результате газ адиабатически изменяет свой объем. Вычислить температуру и объем газа после того как установится термодинамическое равновесие.

Р е ш е н и е. Теплота, полученная газом при адиабатическом расширении или сжатии, равна нулю. Работа, совершаемая газом, $A = P_2 \Delta V$, поэтому $\Delta U + P_2 \Delta V = 0$. Так как $U = C_V T$, то отсюда находим

$$C_V(T_2 - T_1) + P_2(V_2 - V_1) = 0, \quad \text{или} \quad C_V(T_2 - T_1) + RT_2 = P_2 V_1. \quad (29.1)$$

Следовательно,

$$T_2 = \frac{C_V T_1 + P_2 V_2}{C_P}, \quad V_2 = \frac{RT_2}{P_2}. \quad (29.2)$$

2. В предыдущей задаче после того как установилось состояние равновесия, давление газа снова меняют скачкообразно до первоначального значе-

ния P_1. Вычислить окончательную температуру T_3 и окончательный объем газа V_3, когда он опять придет в состояние термодинамического равновесия.

Р е ш е н и е. Используя решение предыдущей задачи, находим

$$T_3 = \frac{C_V T_2 + P_1 V_2}{C_P}, \quad V_3 = \frac{RT_3}{P_1}. \tag{29.3}$$

С помощью уравнений Клапейрона–Менделеева $PV = RT$ и соотношения Майера $C_P - C_V = R$ выражение для T_3 нетрудно преобразовать к виду

$$T_3 = T_1 + \frac{C_V}{C_P^2} \frac{V_1 (P_2 - P_1)^2}{P_2}. \tag{29.4}$$

Отсюда видно, что в результате обоих адиабатических процессов температура, а с ней и объем газа всегда возрастают. Если давление меняется бесконечно мало, то из (29.4) следует, что температура и объем меняются на бесконечно малые величины *второго порядка*. В первом порядке они остаются неизменными. Отсюда следует, что если адиабатически расширять газ, последовательно снимая с поршня бесконечно малые грузы, а затем снова положить эти грузы на поршень в обратном порядке, то температура и объем газа в конечном состоянии будут отличаться от их значений в исходном состоянии бесконечно мало. В пределе, когда массы последовательно снимаемых грузов стремятся к нулю, а их число к бесконечности, газ совершит конечный процесс, пройдя при сжатии в обратном порядке через ту же последовательность равновесных состояний, через которые он проходил при расширении.

§ 30. Цикл Карно и теорема Карно

1. Из различных круговых процессов особое значение в термодинамике имеет *круговой процесс*, или *процесс Карно*. Это квазистатический процесс, в котором систему можно приводить в тепловой контакт с двумя тепловыми резервуарами, имеющими постоянные температуры T_1 и T_2. В дальнейшем предполагается, что $T_1 > T_2$. Тепловой резервуар с более высокой температурой T_1 называется *нагревателем*, а с более низкой температурой T_2 — *холодильником*. Цикл Карно заключается в следующем. Сначала система, имея температуру T_1, приводится в тепловой контакт с нагревателем. Затем, бесконечно медленно уменьшая внешнее давление, ее заставляют квазистатически расширяться по изотерме *12* (рис. 25). При этом она заимствует теплоту Q_1 от нагревателя и производит работу A_{12} против внешнего давления. После этого систему адиабатически изолируют и заставляют квазистатически расширяться по адиабате *23*, пока ее температура не достигнет температуры холодильника T_2. При адиабатическом расширении

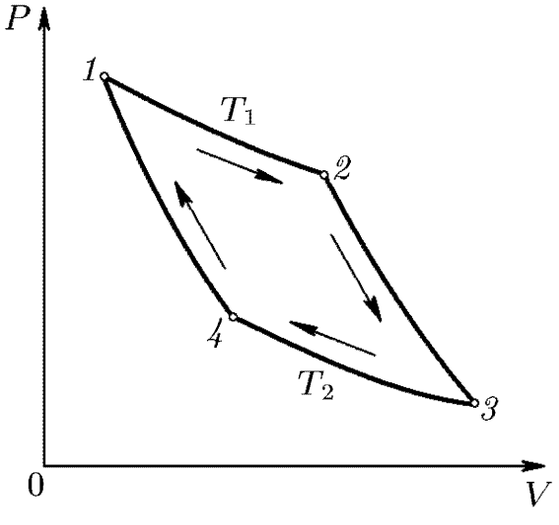


Рис. 25

система также совершает некоторую работу A_{23} против внешнего давления. В состоянии *3* систему приводят в тепловой контакт с холодильником и непрерывным увеличением давления изотермически сжимают ее до некоторого состояния *4*. При этом над системой производится работа (т. е. сама система совершает отрицательную работу A_{34}), и она отдает холодильнику некоторое количество теплоты Q_2. Состояние *4* выбирается так, чтобы можно было квазистатическим сжатием по адиабате *14* вернуть систему в исходное состояние *1*. Для этого надо, разумеется, над системой совершить работу (т. е. сама система должна произвести отрицательную работу A_{41}). В результате кругового процесса Карно внутренняя энергия системы не изменится, а поэтому произведенная ею работа $A = A_{12} + A_{23} + A_{34} + A_{41} = Q_1 - Q_2$. Коэффициент полезного действия η цикла Карно определяется соотношением (28.4), из которого следует

$$Q_2 = (1 - \eta)Q_1. \tag{30.1}$$

2. Докажем знаменитую теорему Карно: *коэффициент полезного действия тепловой машины, работающей по циклу Карно, зависит только от температур T_1 и T_2 нагревателя и холодильника, но не зависит от устройства машины, а также от вида используемого рабочего вещества.*

Для доказательства рассмотрим две машины Карно, имеющие общий нагреватель при температуре T_1 и общий холодильник при температуре T_2. Пусть КПД первой машины равен η, а второй η'. Допустим, что $\eta > \eta'$, и покажем, что это допущение приводит к противоречию с постулатом второго начала термодинамики. Цикл Карно — квазистатический, а потому он может совершаться как в прямом, так и в обратном направлении, т. е. производить работу. Пусть в результате m циклов она отберет от нагревателя теплоту Q_1, передаст холодильнику теплоту Q_2 и произведет работу $A = Q_1 - Q_2$, например, поднимет груз. Остановим после этого первую машину и используем потенциальную энергию поднятого груза, чтобы привести в действие вторую машину в обратном направлении. Вторая машина Карно будет, следовательно, работать как холодильная машина. Пусть в результате m' циклов она заберет теплоту Q_2' от холодильника и передаст теплоту Q_1' нагревателю; при этом над машиной будет совершена работа $A' = Q_1' - Q_2'$. Результат действия m циклов первой и m' циклов второй машины представится схемой

нагреватель отдал теплоту $Q_1 - Q_1'$,
холодильник отдал теплоту $Q_2' - Q_2$,
машина совершила работу

$$A - A' = (Q_1 - Q_2) - (Q_1' - Q_2') = \eta Q_1 - \eta' Q_1'.$$

Дальнейший ход доказательства зависит от того, какой постулат второго начала термодинамики положить в основу: постулат Томсона–Планка или постулат Клаузиуса. Если исходить из постулата Томсона–Планка, то доказательство следует вести так. Выберем целые числа m и m' так, чтобы $Q_1 - Q_1' = 0$. Это всегда можно достигнуть.

Действительно, $Q_1 = mq_1$, $Q_1 = m'q_1'$, где q_1 — количество теплоты, полученное первой машиной от нагревателя в результате одного цикла, а q_1' — количество теплоты, отданное тому же нагревателю второй машиной также в результате одного цикла. Если q_1 и q_1' соизмеримы, то всегда можно подобрать целые числа m и m' так, чтобы $mq_1 - m'q_1' = 0$, т. е. $Q_1 - Q_1' = 0$. Если же эти величины не соизмеримы, то целые числа m и m' можно выбрать настолько большими, чтобы это равенство выполнялось с какой угодно заранее заданной точностью. Поэтому физически всегда возможно выбрать целые числа m и m' так, чтобы $Q_1 - Q_2' = 0$. Тогда результат кругового процесса представится в виде

состояние нагревателя не изменилось,

холодильник отдал теплоту $Q_2' - Q_2 = (\eta - \eta')Q_1 > 0$,

машина совершила работу $\eta Q_1 - \eta' Q_1' = (\eta - \eta')Q_1 > 0$.

Таким образом, единственным результатом кругового процесса будет производство работы $(\eta - \eta')Q_1 > 0$ за счет эквивалентного количества теплоты, заимствованной от холодильника. Это процесс Томсона–Планка, возможность которого противоречит постулату второго начала термодинамики. Поэтому предположение $\eta > \eta'$ неверно — так же неверно предположение $\eta' > \eta$. Чтобы убедиться в этом, надо заставить вторую машину пройти цикл Карно в прямом, а первую — в обратном направлении и повторить наше рассуждение. Следовательно, $\eta = \eta'$, и теорема Карно доказана.

3. Если в основу доказательства положить постулат Клаузиуса, то рассуждение немного изменится. Выберем целые числа m и m' так, чтобы работа, выполненная обеими машинами, равнялась нулю: $\eta Q_1 - \eta' Q_1' = 0$. Тогда результат кругового процесса представится в виде

нагреватель получил теплоту $Q_1' - Q_1 = \dfrac{\eta - \eta'}{\eta'} Q_1 > 0$,

холодильник отдал теплоту $Q_2' - Q_2 = (1 - \eta')Q_1' - (1 - \eta)Q_1 = Q_1' - Q_1 > 0$.

Никаких других изменений не произошло. Единственным результатом процесса получился переход теплоты от тела менее нагретого к телу более нагретому. Это противоречит постулату Клаузиуса, что и доказывает теорему Карно.

§ 31. Термодинамическая шкала температур

1. В 1848 г. Вильям Томсон (лорд Кельвин) указал, что теоремой Карно можно воспользоваться для построения рациональной температурной шкалы, совершенно не зависящей от индивидуальных особенностей термометрического вещества и устройства термометра.

Из теоремы Карно следует, что КПД цикла Карно может зависеть только от температур нагревателя и холодильника. Обозначим буквами t_1 и t_2 эмпирические температуры нагревателя и холодильника, измеренные каким-либо термометром (например, газовым, ртутным,

термометром сопротивления и т. п.). Тогда

$$\eta \equiv \frac{Q_1 - Q_2}{Q_1} = f(t_1, t_2), \tag{31.1}$$

где $f(t_1, t_2)$ — *универсальная функция выбранных эмпирических температур* t_1 *и* t_2. Ее вид совершенно не зависит от устройства машины Карно и от рода используемого рабочего вещества. Этим обстоятельством и воспользовался Вильям Томсон, предложивший применить цикл Карно для построения температурной шкалы.

2. Чтобы построить *термодинамическую шкалу температур*, введем более простую универсальную функцию температур t_1 и t_2:

$$\frac{Q_1}{Q_2} = \varphi(t_1, t_2). \tag{31.2}$$

Эта функция легко выражается через прежнюю функцию $f(t_1, t_2)$. Определим общий вид функции $\varphi(t_1, t_2)$. С этой целью возьмем три тепловых резервуара, температуры которых поддерживаются постоянными. Эмпирические температуры этих резервуаров обозначим через t_1, t_2, t_3 соответственно. Используя их в качестве нагревателей и холодильников, проведем три цикла Карно, изображенные на рис. 26. Предполагается, что t_1, t_2, t_3 — температуры на изотермах *12*, *43*, *65*. Для циклов Карно *1234* и *4356* можно написать

$$\frac{Q_1}{Q_2} = \varphi(t_1, t_2), \quad \frac{Q_2}{Q_3} = \varphi(t_2, t_3).$$

Исключив отсюда теплоту Q_2, получим

$$\frac{Q_1}{Q_3} = \varphi(t_1, t_2)\varphi(t_2, t_3).$$

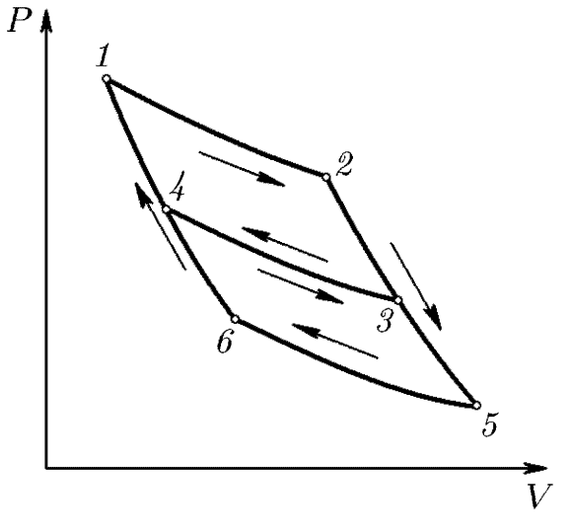


Рис. 26

Но эти два цикла, объединенные вместе по схеме рис. 26, эквивалентны одному циклу Карно *1256*. Это происходит потому, что по изотерме *43* проходим дважды в противоположных направлениях и она может быть исключена из рассмотрения. Следовательно,

$$\frac{Q_1}{Q_3} = \varphi(t_1, t_3).$$

Сравнивая это соотношение с предыдущим, получим

$$\varphi(t_1, t_2)\varphi(t_2, t_3) = \varphi(t_1, t_3), \tag{31.3}$$

откуда

$$\varphi(t_1, t_2) = \frac{\varphi(t_1, t_3)}{\varphi(t_2, t_3)}, \tag{31.4}$$

или

$$\frac{Q_1}{Q_2} = \frac{\varphi(t_1, t_3)}{\varphi(t_2, t_3)}. \tag{31.5}$$

Такое соотношение справедливо при любом значении аргумента t_3. Левая часть его не зависит от значения температуры t_3. Поэтому и отношение в правой части не может меняться с изменением t_3. Можно фиксировать t_3, не меняя значения самого отношения. Но тогда числитель в правой части формулы (31.5) будет функцией одного только аргумента t_1. Обозначим эту функцию через $\Theta(t_1)$. Знаменатель будет такой же функцией, но от аргумента t_2. Итак,

$$\frac{Q_1}{Q_2} = \frac{\Theta(t_1)}{\Theta(t_2)}. \tag{31.6}$$

Таким образом, $\varphi(t_1, t_2)$ есть отношение значений *одной и той же функции* $\Theta(t)$ при $t = t_1$ и $t = t_2$. Так как величина $\Theta(t)$ зависит только от температуры, то она сама может быть принята за меру температуры тела. Величину $\Theta(t)$ и называют *абсолютной термодинамической температурой*. Отношение двух термодинамических температур $\Theta_1 \equiv \Theta(t)$ и $\Theta_2 \equiv \Theta(t_2)$ определяется соотношением

$$\frac{\Theta_1}{\Theta_2} = \frac{Q_1}{Q_2}. \tag{31.7}$$

3. Отношение Θ_1/Θ_2 в принципе может быть найдено экспериментально. Для этого надо измерить теплоты Q_1 и Q_2. Однако значением этого отношения сами температуры Θ_1 и Θ_2 еще не определяются однозначно. Это видно также из того, что функция $\Theta(t) = \varphi(t, t_3)$ зависит от параметра t_3, которому можно придать произвольное значение. Отношение (31.7) не зависит от параметра t_3. Однако сами термодинамические температуры будут иметь разные значения при различном выборе этого параметра. Вместо функции $\Theta(t)$ можно было бы в качестве термодинамической температуры принять, например, величину $\Theta'(t) = \psi(t_3)\Theta(t)$, где $\psi(t_3)$ — произвольная функция. От этого значение отношения (31.7) не изменилось бы. Но, придавая параметру t_3 различные значения, мы получили бы *бесконечное множество температурных шкал*, отличающихся друг от друга масштабами единицы температуры. Чтобы однозначно определить термодинамическую температуру Θ, можно поступить двояко.

Во-первых, можно взять какие-либо *две постоянные температурные точки*, например нормальную точку плавления льда и нормальную точку кипения воды. Обозначим термодинамические температуры этих точек $\Theta_\text{п}$ и $\Theta_\text{к}$, а соответствующие им количества теплоты в цикле Карно — $Q_\text{п}$ и $Q_\text{к}$. Фиксируем далее значение разности $\Theta_\text{к} - \Theta_\text{п}$, например примем, что она равна 100 градусам. Тогда температурный интервал между нормальными точками плавления льда и кипения воды разделится на 100 равных частей, каждая из которых ранее называлась *градусом Кельвина*, а теперь — просто кельвином. Из двух

уравнений

$$\frac{\Theta_к}{\Theta_п} = \frac{Q_к}{Q_п}, \quad \Theta_к - \Theta_п = 100 \tag{31.8}$$

можно в отдельности вычислить $\Theta_п$ и $\Theta_к$. Для этого надо измерить отношение $Q_к/Q_п$. Хотя ни в одном реальном опыте это не делалось, но путем косвенных измерений было найдено

$$\Theta_п \approx 273{,}15 \text{ K}, \quad \Theta_к = 373{,}15 \text{ K}. \tag{31.9}$$

Термодинамическую температуру Θ любого тела можно вычислить, например, по формуле

$$\frac{\Theta}{\Theta_п} = \frac{Q}{Q_п}, \tag{31.10}$$

если предварительно провести цикл Карно между данным телом и тающим льдом и измерить соответствующие количества теплоты Q и $Q_п$. Построенная таким образом температурная шкала называется *абсолютной термодинамической шкалой температур*.

Во-вторых, можно условно приписать какой-либо постоянной температурной точке определенное значение Θ, а затем по формуле типа (31.10) вычислять температуру любого другого тела. За постоянную температурную точку можно, например, принять точку плавления льда при нормальном давлении и условиться, что для этой точки $\Theta_п = 273{,}15$ K. Тогда мы придем к абсолютной термодинамической шкале температур, совпадающей в пределах ошибок измерений со шкалой, построенной первым способом. Температура тройной точки воды, как показали измерения, в этой температурной шкале равна приблизительно 273,16 K.

4. Таким образом, в первом способе при построении абсолютной термодинамической шкалы температур используются *две* постоянные реперные точки, а во втором — *одна*. Теоретически оба способа эквивалентны. Однако практически необходимо считаться с погрешностями, с которыми могут быть воспроизведены реперные точки. Погрешность воспроизведения нормальной точки кипения воды составляет 0,002–0,01 °C, а нормальной точки таяния льда 0,0002–0,001 °C. Между тем, тройная точка воды может быть воспроизведена в специальных приборах с погрешностью не больше 0,0001 °C. Учитывая это, Десятая генеральная конференция по мерам и весам (1954 г.) утвердила построения абсолютной термодинамической шкалы температур *по одной реперной точке*, а именно тройной точке воды, и приписала ей температуру 273,16 K **точно**. Таким образом, в современной термодинамической шкале температур разность между температурами нормальных точек кипения воды и плавления льда равна 100 °C лишь приближенно. Приближенными являются и значения самих температур обеих точек, а именно 273,15 K и 373,15 K. Температура же тройной точки 273,16 K является **точной по определению**.

5. Абсолютная термодинамическая температура не может менять своего знака. А так как абсолютную температуру реперной точки,

положенной в основу построения температурной шкалы, условились считать положительной, то *абсолютная термодинамическая температура не может принимать отрицательных значений.* Докажем это утверждение.

Для доказательства допустим, что существует тело, абсолютная температура Θ_2 которого отрицательна: $\Theta_2 < 0$. Используем это тело в качестве холодильника в тепловой машине Карно. В качестве нагревателя возьмем другое тело, абсолютная температуры Θ_1 которого положительна: $\Theta_1 > 0$ (по крайней мере одно такое тело существует, так как по определению абсолютная температура основной реперной точки положительна). Пусть в процессе Карно нагреватель отдал количество теплоты $Q_1 > 0$. Тогда холодильник получит теплоту $Q_2 = (\Theta_2/\Theta_1)Q_1$. Так как по предположению $\Theta_2/\Theta_1 < 0$, то $Q_2 < 0$. Это значит, что в действительности холодильник не получил, а отдал теплоту $-Q_2 = |Q_2|$. В результате цикла произведена положительная работа $A = Q_1 - Q_2 = Q_1 + |Q_2|$. Будем рассматривать нагреватель и холодильник как один тепловой резервуар. Единственный результат кругового процесса Карно состоит в том, что такой тепловой резервуар отдал теплоту $Q_1 + |Q_2|$, за счет которой произведена эквивалентная работа $A = Q_1 + |Q_2|$. Это — процесс Томсона–Планка, возможность которого противоречит постулату второго начала термодинамики. Поэтому предположение $\Theta_2 < 0$ — неправильное: абсолютная термодинамическая температура не может быть отрицательной. Самая низкая температура, допускаемая постулатом второго начала термодинамики, есть $\Theta = 0$. Эта температура называется *абсолютным нулем температур.* Абсолютный нуль лежит на 273,16 К ниже температуры тройной точки воды. Таким образом, из второго начала термодинамики строго следует существование абсолютного нуля. Конечно, второе начало термодинамики не может ответить на вопрос, достижим или не достижим абсолютный нуль температур. Оно позволяет лишь утверждать, что *охладить тело ниже абсолютного нуля невозможно.*

Что касается приведенного выше рассуждения, то, как уже отмечалось выше, оно доказывает лишь, что абсолютная термодинамическая температура есть *величина одного знака.* Абсолютные температуры двух тел не могут отличаться знаками. Какой знак следует взять — положительный или отрицательный — это вопрос соглашения. Условились температуру основной реперной точки, а с ней и все абсолютные температуры считать положительными. Можно было бы поступить наоборот. Тогда все абсолютные температуры стали бы отрицательными.

6. В квантовой статистической физике вводится обобщение понятия температуры. Некоторые квантовые системы могут находиться в состояниях, которые формально характеризуются как состояния с *отрицательными абсолютными температурами.* Это не противоречит термодинамике, так как последняя определяет температуру лишь для *термодинамически равновесных состояний.* Состояния же с отри-

цательными абсолютными температурами, рассматриваемые в статистической физике, *термодинамически неравновесны*. К ним обычное термодинамическое понятие температуры неприменимо.

§ 32. Тождественность термодинамической шкалы температур со шкалой идеально-газового термометра

Докажем теперь, что *абсолютная термодинамическая шкала температур тождественна с абсолютной шкалой идеально-газового термометра*. (Температуру по шкале такого термометра по-прежнему будем обозначать буквой T.) Для доказательства осуществим цикл Карно, взяв в качестве рабочего тела идеальный газ. Для простоты будем предполагать, что количество газа равно одному молю. Вычислим сначала количество теплоты Q_1, отданное нагревателем на изотерме *12* (см. рис. 25). По первому началу $\delta Q = dU + P\,V$. Так как для идеального газа внутренняя энергия U зависит только от температуры, то на изотерме $dU = 0$, а следовательно,

$$\delta Q = P\,V = RT_1\,\frac{dV}{V}.$$

Интегрируя это выражение, находим

$$Q_1 = RT_1 \ln \frac{V_2}{V_1}.$$

При адиабатическом расширении по пути *23* газ тепла не получает. Поэтому величина Q_1 есть полное количество теплоты, отданное нагревателем за один цикл. Аналогично вычисляется количество теплоты Q_2, полученное холодильником за тот же цикл:

$$Q_2 = RT_2 \ln \frac{V_3}{V_4}.$$

Следовательно,

$$\frac{Q_1}{Q_2} = \frac{T_1}{T_2}\,\frac{\ln(V_2/V_1)}{\ln(V_3/V_4)}.$$

Логарифмический множитель в правой части этого соотношения равен единице. Действительно, если $\gamma = C_P/C_V$ не зависит от температуры, то в этом проще убедиться с помощью уравнения адиабаты в форме $TV^{\gamma-1} = \text{const}$. Применив это уравнение к адиабатам *23* и *14* (см. рис. 25), получим

$$T_1V_2^{\gamma-1} = T_2V_3^{\gamma-1}, \quad T_1V_1^{\gamma-1} = T_2V_4^{\gamma-1}.$$

Почленное деление приводит к соотношению $V_2/V_1 = V_3/V_4$. Этим соотношением наше утверждение доказано. Но приведенное соотношение справедливо и для таких идеальных газов, у которых величина γ зависит от температуры. Для доказательства замечаем, что при

адиабатическом расширении или сжатии

$$\delta Q = C_V\,dT + P\,V = 0, \quad \text{или} \quad C_V\,dT + RT\,\frac{dV}{V} = 0.$$

Отсюда

$$\frac{dV}{V} = -\frac{C_V}{R}\,\frac{dT}{T}.$$

Теплоемкость C_V идеального газа зависит только от температуры. Поэтому при интегрировании последнего уравнения вдоль адиабат *23* и *14* получатся одинаковые результаты:

$$\int\limits_{V_2}^{V_3} \frac{dV}{V} = \int\limits_{V_1}^{V_4} \frac{dV}{V}, \quad \text{или} \quad \ln\frac{V_3}{V_2} = \ln\frac{V_4}{V_1}.$$

Отсюда

$$\frac{V_2}{V_1} = \frac{V_3}{V_4},$$

что и доказывает наше утверждение. Следовательно,

$$\frac{T_1}{T_2} = \frac{Q_1}{Q_2}. \tag{32.1}$$

Сравнивая это соотношение с (31.7), получаем

$$\frac{T_1}{T_2} = \frac{\Theta_1}{\Theta_2}. \tag{32.2}$$

Из этого соотношения следует, что термодинамическая шкала температур станет тождественной с соответствующей температурной шкалой идеального термометра, если в обоих случаях температуре основной реперной точки (или разности температур двух основных реперных точек) приписать *одно и то же значение*. Поскольку так и поступают на практике, тождественность обеих температурных шкал доказана: $T = \Theta$. Поэтому в дальнейшем термодинамическую и идеально-газовую температуру мы будем обозначать одной и той же буквой T. Подчеркнем еще раз, что тождественность обеих температурных шкал имеет место для любых идеальных газов, независимо от того, зависит или не зависит их теплоемкость C_V от температуры.

§ 33. Приведение шкалы газового термометра к термодинамической шкале

1. Машина Карно позволяет лишь принципиально построить температурную шкалу. Для практических измерений температуры она непригодна. На практике всегда измеряется *эмпирическая температура* с помощью каких-либо реальных термометров. Задача заключается в том, чтобы в показания таких термометров ввести *поправки* и таким образом привести эти показания к абсолютной термодинамической шкале. Как показано в предыдущем параграфе, в случае идеально-газового термометра такие поправки

вводить не требуется. Но строго идеальных газов не существует. Реальные газы обнаруживают отступления от идеальных, особенно существенные при очень низких и очень высоких температурах. Рассмотрим, как могут быть найдены поправки к показаниям термометра, использующего в качестве термометрического тела *реальный газ* (лучше всего — гелий).

2. Уравнение состояния всякого вещества можно записать в виде $P = P(T, V)$, или после умножения на V: $PV = VP(T, V) = f(T, V)$. Вместо объема V введем в качестве аргумента плотность $\rho \sim 1/V$. Тогда $PV = F(T, \rho)$. Предполагая, что веществом является какой-то реальный газ, разложим функцию $F(T, \rho)$ по степеням ρ:

$$PV = F(T, 0) + \frac{\partial F(T, 0)}{\partial \rho}\,\rho + \frac{1}{2}\,\frac{\partial^2 F(T, 0)}{\partial \rho^2}\,\rho^2 + \dots$$

Первый член этого разложения $F(T, 0)$ дает значение произведения PV при $\rho \to 0$, т. е. при бесконечном разрежении газа, когда он становится идеальным. Поэтому, если газ взят в количестве одного моля, $F(T, 0) = RT$. Остальные коэффициенты являются функциями одной только температуры. Если вместо плотности снова ввести объем $V \sim 1/\rho$, то разложение представится в виде

$$PV = RT\Big(1 + \frac{B_2}{V} + \frac{B_3}{V^2} + \dots\Big), \tag{33.1}$$

где $B_2, B_3, \dots$ — определенные функции температуры, называемые *вириальными коэффициентами* (B_2 — второй вириальный коэффициент, B_3 — третий и т. д., первый вириальный коэффициент есть единица). Если плотность газа невелика, то можно ограничиться только вторым вириальным коэффициентом, т. е. написать

$$PV = RT\Big(1 + \frac{B}{V}\Big) \tag{33.2}$$

(индекс 2 ради краткости опущен).

3. Возьмем теперь реальный гелиевый термометр. Объем газа в термометре V_0 должен оставаться постоянным. В нормальном гелиевом термометре объем V_0 выбирается таким, чтобы при $t = 0\,^\circ\mathrm{C}$ давление газа составляло 1000 мм рт. ст. Эмпирическая температура определяется по давлению газа в термометре. Обозначим ее буквой τ. По определению эмпирической температуры

$$PV_0 = A\tau, \tag{33.3}$$

где A — постоянная. С другой стороны, из (33.2) находим

$$PV_0 = RT\Big(1 + \frac{B}{V_0}\Big).$$

Путем сравнения получаем

$$T\Big(1 + \frac{B}{V_0}\Big) = \frac{A}{R}\,\tau. \tag{33.4}$$

Существенно, что для измерения постоянных A и R не требуется предварительного построения термодинамической шкалы температур. Для этого достаточно знания числового значения температур T и τ в основной реперной точке, т. е. тройной точке воды. Для воды по определению $\tau_{\text{тр}} = T_{\text{тр}} = 273{,}16$ К. Измерив давление газа $P_{\text{тр}}$ при температуре тройной точки,

найдем по формуле (33.3)

$$A = \frac{P_{\text{тр}} V_0}{\tau_{\text{тр}}}.$$

Постоянную R можно определить, если при $\tau = \tau_{\text{тр}} = T_{\text{тр}}$ измерить произведение PV при бесконечном разрежении газа ($V \to \infty$). По формуле (33.2)

$$R = \frac{(PV)_\infty}{T_{\text{тр}}}.$$

Таким образом, коэффициенты A и R в формуле (33.4) могут считаться известными. Значения функции $B(\tau)$ могут быть также измерены экспериментально для любого значения аргумента τ. Для этого достаточно измерить при температуре τ давления газа P_1 и P_2 для двух значений объема V_1 и V_2. Тогда по формуле (33.2)

$$P_1 V_1 = RT\Big[1 + \frac{B(\tau)}{V_1}\Big], \quad P_2 V_2 = RT\Big[1 + \frac{B(\tau)}{V_2}\Big]$$

и, следовательно,

$$P_2 V_2 - P_1 V_1 = RT B(\tau)\Big(\frac{1}{V_2} - \frac{1}{V_1}\Big).$$

Ввиду того, что разность $\Delta T = T - \tau$ является малой поправкой, в правой части последнего равенства температуру T в рамках принятой точности расчета можно заменить на τ. Тогда

$$B(\tau) = \frac{P_2 V_2 - P_1 V_1}{R\tau(1/V_2 - 1/V_1)}.$$

С той же точностью величину BT в левой части соотношения (33.4) можно заменить на $B\tau$. В результате получится

$$T = \Big(\frac{A}{R} - \frac{B}{V_0}\Big)\tau, \tag{33.5}$$

а для поправки ΔT

$$\Delta T \equiv T - \tau = \Big(\frac{A - R}{R} - \frac{B}{V_0}\Big)\tau. \tag{33.6}$$

Значения поправок ΔT для гелиевого термометра, вычисленные по формуле типа (33.6), приведены в табл. 2 [1].

§ 34. Примеры на применение теоремы Карно

С помощью теоремы Карно можно получить много важных соотношений между физическими величинами, характеризующими систему в состоянии термодинамического равновесия. Для этого надо заставить систему надлежащим образом осуществить цикл Карно и применить к нему теорему Карно. Этот метод называется *методом циклов*. Поясним его на примерах.

[1]) Другие способы градуировки термометра в термодинамической шкале приведены в § 36, 46.

Таблица 2

T, К	ΔT	T, К	ΔT	T, К	ΔT
3	0,045	140	0,011	300	−0,001
10	44	150	10	320	−1
20	40	160	9	350	−1
30	37	170	7	370	0
40	34	180	6	400	1
50	31	190	5	420	3
60	28	200	4	450	5
70	25	210	4	470	8
80	22	220	3	500	0,011
90	20	230	2	520	14
100	18	240	2	550	17
110	16	250	1	570	20
120	15	260	0	620	25
130	13	270	0	670	31
		280	0		

Пример 1. Рассмотрим физически однородное тело, состояние которого характеризуется двумя параметрами, например T и V. Внутренняя энергия такого тела есть однозначная функция тех же параметров: $U = U(T, V)$. Если известно термическое уравнение состояния $f(P, V, T) = 0$, то теорема Карно позволяет в общем виде решить вопрос о зависимости внутренней энергии U от объема. Этот вопрос мы и рассмотрим.

Будем изображать состояние тела точкой на диаграмме V, P. Рассмотрим в плоскости VP семейство изотерм и семейство адиабат. Они разбивают эту плоскость на клетки, имеющие форму криволинейных

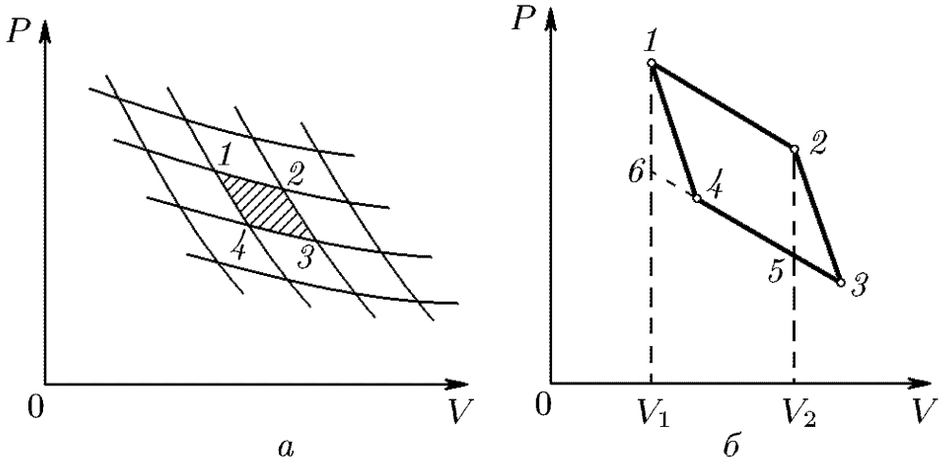


Рис. 27

четырехугольников (рис. 27 *а*). Если изотермы и адиабаты провести достаточно густо, то клетки будут сколь угодно мало отличаться от

параллелограмма. Возьмем один из таких бесконечно малых параллелограммов *1234*, изображенный на рис. 27 *б* в увеличенном масштабе. Цикл *1234* есть цикл Карно. Обозначим абсолютную температуру на изотерме *12* через T_1, а на изотерме *34* — через T_2. Так как эти температуры бесконечно мало отличаются друг от друга, то индексы *1* и *2* будем опускать во всех соотношениях, в которые T_1 и T_2 входят в виде множителей при бесконечно малых величинах. То же относится и к другим величинам, например P_1, P_2, V_1, V_2 и т. п. Работа A, произведенная системой в результате цикла *1234*, численно равна площади параллелограмма *1234*. Чтобы вычислить ее, проведем прямые *16* и *25*, параллельные оси давлений. Ясно, что искомая площадь равна площади параллелограмма *1256*. Высота этого параллелограмма численно равна приращению $V_2 - V_1$ объема при изотермическом процессе *12*. Основание же *61* дает приращение давления при повышении температуры на $T_1 - T_2$, когда объем системы поддерживается постоянным. Оно равно $(\partial P/\partial T)_V(T_1 - T_2)$. Для работы цикла, которая численно равна его площади, получаем

$$A = \left(\frac{\partial P}{\partial T}\right)_V (T_1 - T_2)\,(V_2 - V_1).$$

Вычислим теперь количество теплоты Q_1, отданное нагревателем на изотерме *12*. По первому началу $Q_1 = U_2 - U_1 + P(V_2 - V_1)$.

Так как на изотерме *12* температура постоянна, то

$$U_2 - U_1 = \left(\frac{\partial U}{\partial V}\right)_T (V_2 - V_1),$$

что дает

$$Q_1 = \left[\left(\frac{\partial U}{\partial V}\right)_T + P\right](V_2 - V_1).$$

По теореме Карно

$$\frac{A}{Q_1} = \frac{T_1 - T_2}{T_1}. \tag{34.1}$$

Подставляя сюда значения A и Q, найденные выше, получим

$$\left(\frac{\partial U}{\partial V}\right)_T = T\left(\frac{\partial P}{\partial T}\right)_V - P. \tag{34.2}$$

Эта формула и решает поставленную задачу.

Пример 2. Формулу, аналогичную формуле (34.2), можно вывести и для энтальпии. Энтальпию I будем рассматривать как функцию температуры T и давления P. Преобразуем выражения для количества теплоты Q_1 и работы A. Очевидно,

$$dI = d(U + PV) = (dU + P\,V) + V\,dP,$$

откуда на основании первого начала

$$dI = \delta Q + V\,dP, \quad \text{или} \quad \delta Q = dI - V\,dP. \tag{34.3}$$

Соотношение (34.3) также выражает первое начало термодинамики, но в иной форме. Пользуясь им, для теплоты Q_1 получаем

$$Q_1 = I_2 - I_1 - V(P_2 - P_1).$$

Так как на изотерме *12* температура постоянна, то

$$I_2 - I_1 = \left(\frac{\partial I}{\partial P}\right)_T (P_2 - P_1).$$

Таким образом,

$$Q_1 = \left[\left(\frac{\partial I}{\partial P}\right)_T - V\right](P_2 - P_1).$$

Преобразуем теперь выражение для работы A. С этой целью воспользуемся тождеством (8.3):

$$\left(\frac{\partial P}{\partial T}\right)_V = -\left(\frac{\partial V}{\partial T}\right)_P \left(\frac{\partial P}{\partial V}\right)_V,$$

а также соотношением

$$\left(\frac{\partial P}{\partial V}\right)_T (V_2 - V_1) = P_2 - P_1.$$

Тогда получим

$$A = -\left(\frac{\partial V}{\partial T}\right)_P (T_1 - T_2)(P_2 - P_1).$$

Подставляя Q_1 и A в формулу (34.1), находим искомое соотношение:

$$\left(\frac{\partial I}{\partial P}\right)_T = -T\left(\frac{\partial V}{\partial T}\right)_P + V. \tag{34.4}$$

Применим уравнение (34.2) к идеальному газу. В этом случае $P = RT/V$ и, следовательно,

$$T\left(\frac{\partial P}{\partial T}\right)_V = \frac{RT}{V} = P.$$

После подстановки этих значений в формулу (34.2) получаем

$$\left(\frac{\partial U}{\partial V}\right)_V = 0. \tag{34.5}$$

Отсюда следует, что внутренняя энергия идеального газа не зависит от объема, а является функцией только температуры. Это — закон Джоуля, который использовался нами ранее как эмпирический факт. Мы видим, что он является следствием уравнения Клапейрона и второго начала термодинамики. Далее, из соотношения $C_V = (\partial U/\partial T)_V$ получаем

$$\left(\frac{\partial C_V}{\partial V}\right)_T = \frac{\partial}{\partial T}\frac{\partial U}{\partial V} = 0. \tag{34.6}$$

Значит, *теплоемкость C_V идеального газа не зависит от объема, а может зависеть от температуры.*

Аналогичные соотношения получаются из уравнения (34.4). Именно

$$\left(\frac{\partial I}{\partial P}\right)_T = 0, \tag{34.7}$$

$$\left(\frac{\partial C_P}{\partial P}\right)_T = 0. \tag{34.8}$$

Отсюда видно, что *энтальпия идеального газа и его теплоемкость* C_P *являются функциями одной только температуры.*

Пример 3. Уравнение (34.2) можно обобщить на случай произвольной термодинамической системы, состояние которой определяется заданием каких-то внешних параметров $a_1, a_2, \ldots, a_n$ и температуры T. В этом случае элементарная работа представляется выражением

$$\delta A = A_1\, da_1 + \ldots + A_n da_n,$$

причем величины $A_i(a_1, a_2, \ldots, a_n, T)$ играют роль *обобщенных сил*. Если $(n-1)$ параметров a_j, за исключением одного a_i, поддерживать постоянными, то остается только один свободный параметр a_i. Тогда можно без всяких изменений повторить рассуждения, приведшие нас к формуле (34.2). Роль объема V будет играть параметр a_i, роль давления — обобщенная сила A_i. В результате получится

$$\frac{\partial U}{\partial a_i} = T\left(\frac{\partial A_i}{\partial T}\right)_a - A_i. \tag{34.9}$$

Аналогично можно обобщить формулу (34.4). Если ввести обозначения

$$I = U + A_1 a_1 + A_2 a_2 + \ldots + A_n a_n, \tag{34.10}$$

то получится

$$\frac{\partial I}{\partial A_i} = a_i - T\left(\frac{\partial a_i}{\partial T}\right)_A. \tag{34.11}$$

Формулы (34.9) и (34.11) имеют многочисленные применения.

§ 35. Разность между теплоемкостями C_P и C_V

1. В § 18 из первого начала термодинамики была выведена формула

$$C_P - C_V = \left[\left(\frac{\partial U}{\partial V}\right)_T + P\right]\left(\frac{\partial V}{\partial T}\right)_P. \tag{35.1}$$

Чтобы по этой формуле вычислить $C_P - C_V$, надо знать термическое и калорическое уравнения состояния. Второе начало термодинамики позволяет решить ту же задачу *без использования калорического уравнения состояния*. Оно приводит к формуле (34.2), с помощью которой из соотношения (35.1) можно исключить производную $(\partial U/\partial V)_T$. Это дает

$$C_P - C_V = T\left(\frac{\partial P}{\partial T}\right)_V \left(\frac{\partial V}{\partial T}\right)_P. \tag{35.2}$$

С помощью тождества (8.4) эту формулу можно преобразовать к виду

$$C_P - C_V = -T\left(\frac{\partial V}{\partial T}\right)_P^2 \left(\frac{\partial P}{\partial V}\right)_T = -T\left(\frac{\partial P}{\partial T}\right)_V^2 \left(\frac{\partial V}{\partial P}\right)_T. \qquad (35.3)$$

Зная термическое уравнение состояния, можно по любой из этих формул вычислить разность $C_P - C_V$. В частности, для идеального газа $PV = RT$. В этом случае формулы (35.2) и (35.3) приводят к соотношению Майера $C_P - C_V = R$.

Если ввести температурный коэффициент объемного расширения

$$\alpha = \frac{1}{V_0}\left(\frac{\partial V}{\partial T}\right)_P$$

и изотермический модуль объемного сжатия

$$K = -V\left(\frac{\partial P}{\partial V}\right)_T,$$

то из первой формулы (35.3) получится

$$C_P - C_V = T\alpha^2 K \frac{V_0^2}{V}. \qquad (35.4)$$

Так как $K > 0$, то из этой формулы следует, что для всех веществ $C_P > C_V$. Доказательство этого неравенства при менее ограниченных предположениях будет дано в § 51.

2. Для твердых и жидких тел разницей объемов V_0 и V обычно можно пренебречь. Если требуется получить $c_P - c_v$ удельных теплоемкостей, то молярный объем V следует заменить на удельный объем $v = 1/\rho$, где ρ — плотность вещества. Таким образом, для твердых и жидких тел приближенно

$$c_P - c_v = \frac{K}{\rho} T\alpha^2. \qquad (35.5)$$

Для воды при $t = 0\,^\circ\text{C}$ ($T = 273{,}15$ K) коэффициент $\alpha = -6{,}10 \times \times 10^{-5}$ K^{-1}, $K = 2 \cdot 10^9$ Па, $\rho = 10^3$ кг/м^3 и формула (35.5) дает $c_P - c_v = 2$ Дж/(кг · K) $= 5 \cdot 10^{-4}$ ккал/(кг · K). Столь ничтожная разница удельных теплоемкостей c_P и c_v для воды объясняется малостью температурного коэффициента объемного расширения α. Эта малость в свою очередь обусловлена тем, что коэффициент α при $t = = 4\,^\circ\text{C}$, где вода имеет наибольшую плотность, обращается в нуль.

3. В качестве второго примера вычислим удельную теплоемкость c_v ртути при $t = 0\,^\circ\text{C}$. Воспользуемся следующими экспериментальными данными:

$$T = 273 \text{ K}, \quad c_P = 140 \text{ Дж/(кг} \cdot \text{K)}, \quad \rho = 13{,}6 \cdot 10^3 \text{ кг/м}^3,$$

$$K = 2{,}6 \cdot 10^{10} \text{ Па}, \quad \alpha = 1{,}81 \cdot 10^{-4} \text{ K}^{-1}.$$

После подстановки в формулу (35.5) получится

$$c_P - c_v = 17 \text{ Дж/(кг} \cdot \text{K)},$$

$$c_v = 123 \text{ Дж/(кг} \cdot \text{К)} \approx 0{,}0292 \text{ ккал/(кг} \cdot \text{К)}$$
$$c_P/c_v = 1{,}13.$$

Здесь разница между c_P и c_v — величина, вполне заметная. Это в основном объясняется значительно большим коэффициентом теплового расширения ртути при $t = 0\,^\circ\text{C}$, чем воды при той же температуре.

4. Выясним причину различия между теплоемкостями c_P и c_v для жидких и твердых тел. С этой целью воспользуемся формулой (35.1) и представим разность $c_P - c_v$ в виде

$$c_P - c_v = v\left[\left(\frac{\partial u}{\partial v}\right)_T + P\right] \cdot \frac{1}{v}\left(\frac{\partial v}{\partial T}\right)_P = \frac{\alpha}{\rho}\left[\left(\frac{\partial u}{\partial v}\right)_T + P\right].$$

Отсюда

$$\left(\frac{\partial u}{\partial v}\right)_T + P = \frac{\rho}{\alpha}(c_P - c_v). \tag{35.6}$$

После подстановки числовых данных получаем для воды

$$\left(\frac{\partial u}{\partial v}\right)_T + P \approx -0{,}33 \cdot 10^8 \text{ Па} \approx -0{,}33 \cdot 10^3 \text{ атм},$$

и для ртути

$$\left(\frac{\partial u}{\partial v}\right)_T + P \approx 1{,}28 \cdot 10^9 \text{ Па} \approx 1{,}3 \cdot 10^4 \text{ атм}.$$

Следовательно, при обычных условиях величина $(\partial u/\partial v)_T$ в тысячи и десятки тысяч раз превосходит атмосферное давление. Отсюда вытекает, что для жидких и твердых тел разность $c_P - c_v$ обусловлена главным образом работой, которая идет на изменение внутренней энергии тела при его расширении или сжатии при постоянном давлении. Работа против внешнего давления практически не играет роли. Для газов, как мы видели, положение обратное: здесь разность $c_P - c_v$ обусловлена почти исключительно работой против постоянного внешнего давления P.

Если температурный коэффициент объемного расширения α положителен (как в примере со ртутью), то при тепловом расширении тела надо затратить положительную работу против молекулярных сил. Если коэффициент теплового расширения отрицателен (как в примере с водой), то при нагревании тело сжимается, но при этом работа против молекулярных сил также положительна. Этим и объясняется, почему в обоих случаях разность $c_P - c_v$ положительна.

§ 36. Принципиальный способ градуировки термометра в термодинамической шкале

Машина Карно теоретически может быть использована для градуировки термометра в термодинамической шкале температур. Для той же цели можно воспользоваться любым точным термодинамическим соотношением, в которое, помимо температуры T, входят только экспериментально измеримые величины. Примером может служить соотношение (34.2). Покажем, как оно может быть использовано для указанной цели. Обозначим буквой τ какую-

либо эмпирическую температуру, отсчитываемую по шкале практического термометра. Очевидно, $T = T(\tau)$, и задача состоит в нахождении функции $T(\tau)$. Так как постоянство T влечет за собой постоянство τ и наоборот, то $(\partial U/\partial V)_T = (\partial U/\partial V)_\tau$. Поэтому формулу (34.2) можно переписать в виде

$$\left(\frac{\partial U}{\partial V}\right)_\tau = T\left(\frac{\partial P}{\partial \tau}\right)_V \frac{d\tau}{dT} - P.$$

Отсюда

$$\frac{dT}{T} = \frac{(\partial P/\partial \tau)_V}{(\partial U/\partial V)_\tau + P}\, d\tau,$$

или после интегрирования

$$T = T_0 \exp \int_{\tau_0}^{\tau} \frac{(\partial P/\partial \tau)_V}{(\partial U/\partial V)_\tau + P}\, d\tau. \tag{36.1}$$

Частная производная $(\partial P/\partial \tau)_V$ может быть найдена из уравнения состояния $P = P(\tau, V)$. Величину же $(\partial U/\partial V)_\tau + P$ можно найти, измеряя количество теплоты, которое тело получает при изотермическом квазистатическом расширении или отдает при изотермическом квазистатическом сжатии. В самом деле, поделив обе части равенства $\delta Q = dU + P\,dV$ на dV и предполагая, что температура остается постоянной, получим

$$\left(\frac{\delta Q}{dV}\right)_\tau = \left(\frac{\partial U}{\partial V}\right)_\tau + P.$$

Таким образом, все величины, стоящие под знаком интеграла в формуле (36.1), могут быть экспериментально измерены при всех значениях эмпирической температуры τ. После этого самый интеграл может быть вычислен численно. Тем самым будет найдена термодинамическая температура T как функция эмпирической температуры τ.

§ 37. Неравенство Клаузиуса (для частного случая)

1. Из рассуждений, приведенных в § 30 для доказательства теоремы Карно, можно извлечь другое интересное следствие. Рассмотрим произвольную термодинамическую систему (назовем ее системой I), которая может обмениваться теплом с двумя тепловыми резервуарами R_1 и R_2. Температуры этих резервуаров обозначим через T_1 и T_2 соответственно. Теперь мы не будем различать, какой резервуар является нагревателем, а какой — холодильником. Количество теплоты, отданное тепловым резервуаром (т. е. полученное системой II), условимся считать положительным. В противоположном случае количество теплоты считается отрицательным. Благодаря этому окончательные результаты формулируются симметрично относительно обоих резервуаров.

Пусть система I совершила произвольный круговой процесс — обратимый или необратимый, в котором она получила количество теплоты Q_1 от резервуара R_1 и Q_2 от резервуара R_2. Так как система вернулась в исходное состояние, то полное количество теплоты $Q_1 + Q_2$, полученное ею, будет равно работе, которую она произвела. Возьмем теперь обратимую машину Карно и заставим ее работать

между теми же тепловыми резервуарами R_1 и R_2. Для того чтобы наличие машины Карно никак не отразилось на количествах теплоты Q_1 и Q_2, полученных системой от тепловых резервуаров во время кругового процесса, можно присоединить машину Карно уже *после того*, как круговой процесс в системе I закончился. Если с этого момента теплоизолировать систему I, то тепловые резервуары R_1 и R_2 начнут обмениваться теплом только с машиной Карно. Наличие последней никак не скажется на ходе интересующего нас кругового процесса в системе I, поскольку этот процесс является событием *прошлым* по отношению к моменту присоединения машины Карно. Пусть сама машина Карно совершила круговой процесс, в ходе которого она заимствовала теплоту Q'_1 от резервуара R_1 и теплоту Q'_2 от резервуара R_2. Для дальнейшего существенно, что машина Карно обратима. Ее можно заставить работать и как двигатель, и как холодильник. Кроме того, изотерма *12* (см. рис. 25) в цикле Карно может быть взята сколь угодно короткой. Следовательно, и работа, совершаемая машиной Карно в одном цикле, может быть как угодно малой. Можно получить и сколь угодно большую работу, заставив машину Карно совершить много одинаковых или различных циклов. Таким образом, машина Карно позволяет получать как положительную, так и отрицательную работу *любой наперед заданной величины*. Это обстоятельство делает возможным по произволу распорядиться одной из величин Q'_1 или Q'_2. Всегда можно достигнуть, чтобы одна из них приняла произвольное значение, как положительное, так и отрицательное.

На основании теоремы Карно и определения абсолютной температуры можно написать

$$\frac{Q'_1}{T_1} + \frac{Q'_2}{T_2} = 0. \tag{37.1}$$

Объединим машину Карно и систему I в одну сложную систему. Круговые процессы, последовательно совершенные системой I и машиной Карно, очевидно, также можно объединить в один общий круговой процесс. В этом процессе сложная система

получила от резервуара R_1 теплоту $Q_1 + Q'_1$,
получила от резервуара R_2 теплоту $Q_2 + Q'_2$,
совершила работу $A = Q_1 + Q'_1 + Q_2 + Q'_2$.

Дальнейшие рассуждения зависят от того, какой постулат второго начала термодинамики положить в их основу. Если воспользоваться постулатом Томсона–Планка, то надо рассуждать следующим образом. Подберем Q'_1 так, чтобы $Q'_1 = -Q_1$. Тогда ввиду соотношения (37.1)

$$Q'_2 = T_2 \frac{Q_1}{T_1}.$$

В результате кругового процесса состояние резервуара R_1 не изменится. Тепловой резервуар R_2 отдаст количество теплоты

$$Q_2 + Q'_2 = Q_2 + T_2 \frac{Q_1}{T_1} = T_2\Big(\frac{Q_1}{T_1} + \frac{Q_2}{T_2}\Big).$$

За счет этой теплоты будет произведена эквивалентная работа $A = Q_2 + Q_2'$. Если бы эта работа была положительна, то получился бы процесс Томсона–Планка, что невозможно. Поэтому должно быть $A \leqslant 0$. Так как абсолютная температура T_2 существенно положительна, то это приводит к неравенству

$$\frac{Q_1}{T_1} + \frac{Q_2}{T_2} \leqslant 0. \tag{37.2}$$

Это неравенство является частным случаем более общего *неравенства Клаузиуса*, которое мы рассмотрим в следующем параграфе.

2. Соотношение (37.2) можно, конечно, получить и из постулата Клаузиуса. Для этого выберем Q_1' так, чтобы

$$A = Q_1 + Q_1' + Q_2 + Q_2' = 0, \quad \text{или} \quad Q_1 + Q_1' = -(Q_2 + Q_2').$$

Тогда единственным результатом кругового процесса будет передача теплоты $Q_1 + Q_1'$ от резервуара R_1 к резервуару R_2. Теплоту Q_1' можно найти из условия $A = 0$, если воспользоваться соотношением (37.1). Это дает

$$Q_1 + Q_1' + Q_2 - \frac{T_2}{T_1} Q_1' = 0.$$

Определив отсюда Q_1', находим дальше

$$Q_1 + Q_1' = \frac{T_1 T_2}{T_2 - T_1}\left(\frac{Q_1}{T_1} + \frac{Q_2}{T_2}\right).$$

Если $T_1 > T_2$, то согласно постулату Клаузиуса должно быть $Q_1 + Q_1' \geqslant 0$. Если же $T_1 < T_2$, то должно быть $Q_1 + Q_1' \leqslant 0$. Так как абсолютные температуры существенно положительны, то в обоих случаях мы приходим к неравенству

$$\frac{Q_1}{T_1} + \frac{Q_2}{T_2} \leqslant 0.$$

3. Придадим неравенству Клаузиуса (37.2) другую форму, в которой отчетливее выявляется его связь с техническими проблемами. Вернемся к прежним обозначениям, которыми мы пользовались в § 30. Пусть T_1 — температура нагревателя, а T_2 — холодильника. Количество теплоты Q_2, как и ранее в § 30, будем считать положительным, если холодильник его получает. При таком выборе знаков

$$\frac{Q_1}{T_1} - \frac{Q_2}{T_2} \leqslant 0.$$

Отсюда легко получить

$$\frac{Q_1 - Q_2}{Q_2} \leqslant \frac{T_1 - T_2}{T_1}. \tag{37.3}$$

Слева стоит коэффициент полезного действия тепловой машины, обозначавшийся ранее буквой η. В результате мы доказали следующую (*вторую*) теорему, принадлежащую Карно. *Коэффициент полезного действия всякой тепловой машины не может превосходить коэффициент полезного действия идеальной машины, работающей по циклу*

Карно с теми же самыми температурами нагревателя и холодильника.

Эта теорема позволяет, таким образом, оценить *верхний предел* КПД тепловой машины. Возьмем, например, паровую машину. Пусть максимальная температура пара в котле $t_1 = 150\,^\circ\mathrm{C}$, температура холодильника $t_2 = 20\,^\circ\mathrm{C}$. В термодинамической шкале соответствующие температуры будут $T_1 = 423\ \mathrm{K}$, $T_2 = 293\ \mathrm{K}$. При этом КПД такой машины не может превосходить

$$\frac{T_1 - T_2}{T_1} = \frac{130}{423} \approx 30\,\%.$$

В действительности КПД паровых машин значительно меньше.

Вторая теорема Карно ясно показывает, что энергия должна характеризоваться не только количественно, но и *качественно*. Под качеством энергии мы понимаем ее способность превращаться в другие виды энергии при заданных внешних условиях. Конечно, если на способы превращения не накладывать никаких ограничений, то внутренняя (тепловая) энергия может быть целиком затрачена на производство работы. В этом смысле теплота и работа эквивалентны между собой. Однако если на внешние условия, в которых находится тело, наложить определенные ограничения, то полное превращение тепла в работу может стать невозможным. Например, невозможно полностью превратить теплоту в работу при помощи периодически действующей тепловой машины. Для этого надо было бы располагать холодильником, температура которого равна абсолютному нулю. Поскольку такого холодильника нет, периодически действующая машина может превратить в работу только часть тепловой (внутренней) энергии тела. Чем выше температура тела, тем выше *качество* запасенной в нем тепловой энергии. Всякий естественно идущий необратимый процесс приводит к *обесцениванию энергии* в указанном выше смысле. Обратный процесс, в котором качество энергии повышается, возможен только при наличии другого процесса, в котором энергия обесценивается. Этот другой процесс Клаузиус назвал *компенсирующим процессом* или, короче, *компенсацией*. Например, можно отнять теплоту от холодильника и передать его нагревателю. Но для этого необходим компенсирующий процесс, скажем, производство работы.

ЗАДАЧИ

1. Какую максимальную работу можно получить из системы двух тел, нагретых до разных абсолютных температур T_{10} и T_{20} ($T_{10} > T_{20}$), если тела используются в качестве нагревателя и холодильника в тепловой машине? Теплоемкости тел C_1 и C_2 считать не зависящими от температуры. Найти окончательную температуру T, которую будут иметь тела, когда установится тепловое равновесие между ними.

Р е ш е н и е. Максимальная работа получается тогда, когда тепловая машина работает последовательно повторяющимися бесконечно малыми циклами Карно. Пусть в результате одного из таких циклов первое тело отдало теплоту $\delta Q_1 = -C_1 dT_1$, а второе $\delta Q_2 = -C_2 dT_2$ (T_1 и T_2 означают переменные температуры тел). Произведенная работа $\delta A = \delta Q_1 + \delta Q_2$,

причем

$$\frac{\delta Q_1}{T_1}+\frac{\delta Q_2}{T_2}=0, \quad \text{или} \quad C_1\frac{dT_1}{T_1}+C_2\frac{dT_2}{T_2}=0. \tag{37.4}$$

Интегрируя это соотношение с учетом начальных условий, получим

$$T_1^{C_1}T_2^{C_2}=T_{10}^{C_1}T_{20}^{C_2}. \tag{37.5}$$

Окончательная температура T найдется из условия $T_1=T_2=T$. Оно дает

$$T^{C_1+C_2}=T_{10}^{C_1}T_{20}^{C_2}. \tag{37.6}$$

Максимальная работа, которую может совершить система,

$$A=\int\delta A=-C_1\int\limits_{T_1}^{T}dT-C_2\int\limits_{T_2}^{T}dT=(C_1T_1+C_2T_2)-(C_1+C_2)\,T. \tag{37.7}$$

Она равна убыли внутренней энергии системы.

2. Рассмотреть предельный случай предыдущей задачи, когда теплоемкость холодильника C_2 бесконечно велика. (Нагретое тело, погруженное в бесконечную среду, температура которой T_{20} поддерживается постоянной.)

Р е ш е н и е. Температура T находится из (37.6) предельным переходом $C_2\to\infty$, который дает $T=T_{20}$. Этот результат непосредственно очевиден. Предельный переход в окончательном выражении (37.7) выполнить затруднительно, так как он приводит к неопределенности вида $\infty-\infty$. Удобнее выполнить предельный переход в выражении для элементарной работы $\delta A=\delta Q_1+\delta Q_2$. Выразив δQ_2 через δQ_1 по формуле (37.4) и учтя, что $T_2=T_{20}=\text{const}$, получим

$$\begin{gathered}\delta A=\delta Q_1-\frac{T_{20}}{T_1}\,\delta Q_1=-C_1dT_1+C_1T_{20}\frac{dT_1}{T_1},\\ A=C_1\Big(T_{10}-T_{20}-T_{20}\ln\frac{T_{10}}{T_{20}}\Big).\end{gathered} \tag{37.8}$$

Работа A меньше убыли внутренней энергии нагретого тела $C_1(T_{10}-T_{20})$. Часть внутренней энергии тело передает окружающей среде в виде теплоты (см. § 48).

§ 38. Неравенство Клаузиуса в общем виде

1. Обобщим теперь неравенство (37.2) на случай произвольного числа тепловых резервуаров $R_1, R_2, \ldots, R_n$. Резервуары должны быть достаточно велики (в пределе бесконечно велики), чтобы в ходе теплообмена их температуры $T_1, T_2, \ldots, T_n$ оставались практически постоянными. Пусть какая-либо термодинамическая система (назовем ее по-прежнему системой I) совершила произвольный круговой процесс — обратимый или необратимый, в ходе которого она заимствовала от тепловых резервуаров теплоты $Q_1, Q_2, \ldots, Q_n$ и за их счет произвела эквивалентную работу $Q_1+Q_2+\ldots+Q_n$ (рис. 28).

После того как указанный круговой процесс закончился, теплоизолируем систему I. Возьмем вспомогательный тепловой резервуар R_0 настолько большой, что в процессе теплообмена его температура T_0 практически не меняется. Возьмем, кроме того, n идеальных машин

Карно и включим их между вспомогательным резервуаром R_0 и резервуарами $R_1, R_2, \ldots, R_n$. Таким образом, i-я машина будет совершать цикл Карно между резервуарами R_0 и R_i $(i = 1, 2, \ldots, n)$. Не имеет значения, совершаются ли все эти n циклов Карно одновременно, последовательно или с произвольным наложением друг на друга. Пусть в результате своего цикла i-я машина Карно забирает от резервуара R_0 теплоту Q_0, а от резервуара R_i — теплоту Q'_i. На основании теоремы Карно и определения абсолютной температуры

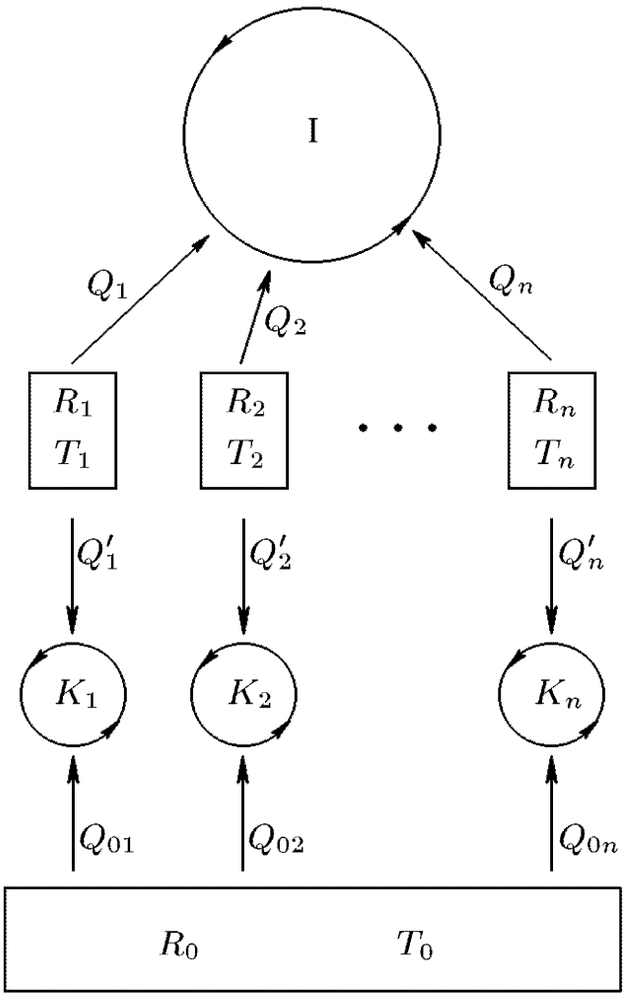


Рис. 28

$$\frac{Q_{0i}}{T_0} + \frac{Q'_i}{T_i} = 0. \tag{38.1}$$

Суммируя по i, находим общее количество теплоты Q_0, отданное вспомогательным резервуаром:

$$Q_0 = \sum Q_{0i} = -T_0 \sum_{i=1}^{n} \frac{Q'_i}{T_i}.$$

Все n циклов Карно можно объединить с круговым процессом, ранее совершаемым системой I, в один сложный круговой процесс. В результате такого процесса

$$\begin{aligned}
&\text{резервуар } R_0 \text{ отдал теплоту } Q_0,\\
&\text{резервуар } R_1 \text{ отдал теплоту } Q_1 + Q'_1,\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&\text{резервуар } R_n \text{ отдал теплоту } Q_n + Q'_n,\\
&\text{совершена работа } A = Q_0 + (Q_1 + Q'_1) + \ldots + (Q_n + Q'_n).
\end{aligned} \tag{38.2}$$

Дальнейшие рассуждения построим на постулате Томсона–Планка. Выберем $Q'_1, Q'_2, \ldots$ так, чтобы $(Q_1 + Q'_1) = \ldots = (Q_n + Q'_n) = 0$. Это всегда возможно, так как тепловые резервуары $R_1, R_2, \ldots, R_n$ предполагаются достаточно большими. В результате все тепловые резервуары вернутся в свои исходные состояния. Вспомогательный резервуар R_0 отдаст

$$Q_0 = T_0 \sum_{i=1}^{n} \frac{Q_i}{T_i}. \tag{38.3}$$

Получился круговой процесс, совершенный системой I и n машинами Карно, в результате которого

резервуар R_0 отдал количество теплоты Q_0,
совершена эквивалентная работа $A = Q_0$.

Больше никаких изменений не произошло. Произведенная работа $A = Q_0$ не может быть положительной, так как в противоположном случае получилось бы противоречие с постулатом Томсона–Планка. Таким образом, должно быть $Q_0 \leqslant 0$, или, учитывая (38.3) и положительность абсолютной температуры T_0,

$$\bigcirc \sum_{i=1}^{n} \frac{Q_i}{T_i} \leqslant 0. \tag{38.4}$$

Это неравенство мы и хотели доказать. Кружок у знака суммы означает, что соотношение (38.4) относится к круговому процессу. Имеется в виду круговой процесс, выполненный самой системой I. Вспомогательные приспособления — машины Карно и тепловой резервуар R_0 — использовались только для доказательства. Их наличие никак не может отразиться на справедливости соотношения (38.4). Это ясно хотя бы уже из того, что вспомогательные приспособления были привлечены только после того, как произошел круговой процесс, к которому относятся величины Q_i и T_i.

2. При доказательстве предполагалось, что каждый из тепловых резервуаров R_i может обмениваться теплотой только с рассматриваемой системой I. Обмен теплотой между самими тепловыми резервуарами, а также между резервуарами и остальными телами не учитывался. Но такого рода теплообмен не играет роли. Неравенство (38.4) останется справедливым и при его наличии. Действительно, всегда можно ввести адиабатические перегородки, исключающие указанный теплообмен и в то же время не меняющие физическое состояние тепловых резервуаров и не влияющие на их теплообмен с рассматриваемой системой.

3. Далее, доказательство предполагало, что тепловые резервуары $R_1, R_2, \ldots, R_n$ достаточно велики. Такое предположение необходимо, чтобы температуры T_i могли считаться постоянными, каковы бы ни были количества теплоты, получаемые или отдаваемые тепловыми резервуарами. Общий случай, когда резервуары конечны, а их температуры произвольно меняются во времени, сводится к разобранному частному случаю. Действительно, пусть температура T_i резервуара R_i меняется во времени. Процесс теплообмена, в результате которого резервуар R_i отдает системе I теплоту Q_i, можно разбить на сколь угодно большое число N бесконечно малых процессов, в которых резервуар R_i отдает бесконечно малые количества теплоты $\delta Q_{i1}, \ldots, \delta Q_{iN}$. В каждом из таких процессов температуру резервуара R_i можно считать постоянной. А это означает, что по отношению к этим процессам резервуар R_i ведет себя как бесконечно большой. Один резервуар R_i с переменной температурой как бы эквивалентен N последовательно включаемым резервуарам с постоянными, но разными температурами; в течение короткого времени первый резервуар отдает системе I тепло-

ту δQ_{i1}, а в остальное время кругового процесса он теплоизолирован; в течение следующего короткого времени второй резервуар отдает теплоту δQ_{i2}, оставаясь теплоизолированным в остальное время, и т. д. Ясно поэтому, что в общем случае неравенство (38.4) следует записать в виде

$$\oint \frac{\delta Q}{T} \leqslant 0. \tag{38.5}$$

Это фундаментальное соотношение называется *неравенством Клаузиуса.*

4. При окончательной формулировке неравенства (38.5) нет необходимости вводить какие-то тепловые резервуары $R_1, R_2, \ldots, R_n$, с которыми система I обменивается теплом. Лучше пользоваться представлением о теплообмене между системой I и *окружающей средой.* Тогда величина T будет означать *температуру этой среды* и может меняться как в пространстве, так и во времени. Надо мысленно представить себе, что окружающая среда разделена на малые области, каждая из которых характеризуется определенной, вообще говоря, переменной температурой. Символ δQ означает бесконечно малое количество теплоты, переданное системе I одной или несколькими из таких областей при температуре T. Кружок у знака интеграла должен напоминать, что неравенство (38.5) относится к круговому процессу, совершенному рассматриваемой системой I.

5. Отметим еще, почему при доказательстве был введен вспомогательный тепловой резервуар R_0. Это было сделано для того, чтобы иметь в распоряжении неограниченный источник внутренней энергии, из которого можно было бы черпать или которому можно было бы передавать ничем не ограниченное количество теплоты. Если бы один из тепловых резервуаров $R_1, R_2, \ldots$, например первый, обладал неограниченным запасом внутренней энергии, то надобности во вспомогательном резервуаре R_0 не было бы. Его функции можно было бы возложить на резервуар R_1. Поскольку в общем случае дело обстоит не так, то и потребовался резервуар R_0. Таким образом, для доказательства существенно, чтобы резервуар R_0 был бесконечно большим. Выбор резервуара R_0 не может сказаться на окончательном результате, в который резервуар R_0 не входит вообще.

6. Применим неравенство (38.5) к уточнению вопроса о верхнем пределе коэффициента полезного действия тепловых машин, который уже разбирался в конце предыдущего параграфа. Как и там, удобно вернуться к прежнему правилу знаков. Элементарное количество теплоты условимся обозначать символом δQ_1, если машина его получает. Элементарное же количество теплоты, отдаваемое машиной, будем обозначать символом δQ_2. Таким образом, по определению δQ_1 и δQ_2 существенно положительны. В этих обозначениях неравенство Клаузиуса запишется в виде

$$\int \frac{\delta Q_1}{T_1} - \int \frac{\delta Q_2}{T_2} \leqslant 0.$$

Здесь T_1 — температура той части окружающей среды, от которой машина получает теплоту δQ_1 — эта часть среды играет роль на-

гревателя. Величина T_2 есть температура «холодильника», т. е. части окружающей среды, которой машина отдает теплоту δQ_2. В отличие от случая, разобранного в предыдущем параграфе, теперь предполагается, что температуры нагревателя и холодильника не остаются постоянными, а меняются в ходе процесса. Пусть $T_{1\,\text{макс}}$ и $T_{2\,\text{мин}}$ означают соответственно максимальную температуру нагревателя и минимальную температуру холодильника. Предыдущее неравенство будет только усилено, если T_1 и T_2 заменить на $T_{1\,\text{макс}}$ и $T_{2\,\text{мин}}$. Следовательно,

$$\int \frac{\delta Q_1}{T_{1\,\text{макс}}} - \int \frac{\delta Q_2}{T_{2\,\text{мин}}} \leqslant 0.$$

Отсюда

$$\frac{Q_1}{T_{1\,\text{макс}}} - \frac{Q_2}{T_{2\,\text{мин}}} \leqslant 0,$$

где Q_1 — полное количество теплоты, полученное машиной от нагревателя за время кругового процесса, а Q_2 — полное количество теплоты, отданное холодильнику. Переписав последнее неравенство в виде

$$-\frac{Q_2}{Q_1} \leqslant \frac{T_{2\,\text{мин}}}{T_{1\,\text{макс}}},$$

а затем прибавив к обеим частям по единице, получим

$$\eta \equiv \frac{Q_1 - Q_2}{Q_1} \leqslant \frac{T_{1\,\text{макс}} - T_{2\,\text{мин}}}{T_{1\,\text{макс}}}. \tag{38.6}$$

Этим неравенством и определяется верхний предел для коэффициента полезного действия тепловой машины η.

ЗАДАЧА

Доказать неравенство Клаузиуса с помощью постулата Клаузиуса.

Доказательство. До соотношений (38.2) включительно рассуждения остаются прежними, а дальше должны быть изменены следующим образом. Наложим дополнительные условия $Q_2 + Q_2' = \ldots = Q_n + Q_n' = 0$ и $A = Q_0 + (Q_1 + Q_1') + \ldots + (Q_n + Q_n') = 0$, или $Q_0 = -(Q_1 + Q_1')$. Тогда получится круговой процесс, в результате которого

резервуар R_0 отдал теплоту Q_0,

резервуар R_1 получил теплоту $-(Q_1 + Q_1') = Q_0$.

Никаких других изменений не произошло. Найдем сначала переданное количество теплоты Q_0. При $i = 1$ из соотношений (38.1) получаем

$$\frac{Q_{01}}{T_0} = \frac{Q_0 + Q_1}{T_1},$$

а при $i = 2, 3, \ldots, n$

$$\frac{Q_{0i}}{T_0} = \frac{Q_i}{T_i}.$$

Складывая эти равенства, находим

$$\frac{Q_0}{T_0} = \frac{Q_0}{T_1} + \sum_{i=1}^{n} \frac{Q_i}{T_i}.$$

Отсюда

$$Q_0 = \frac{T_0 T_1}{T_1 - T_0} \sum_{i=1}^{n} \frac{Q_i}{T_i}.$$

Если $T_0 > T_1$, то должно быть $Q_0 \geqslant 0$. Если же $T_0 < T_1$, то наоборот, $Q_0 \leqslant 0$. Иначе получилось бы противоречие с постулатом Клаузиуса. В обоих случаях мы приходим к неравенству (38.4).

§ 39. Принцип динамического отопления

1. Применим неравенство Клаузиуса к проблеме отопления помещений. При обычных способах отопления теплота, получающаяся в топке от сгорания топлива, непосредственно поступает в отапливаемое помещение. Значительная доля теплоты, уносимая нагретыми газами, бесполезно расходуется на обогревание окружающей атмосферы. Имеются и другие потери теплоты. Но мы отвлечемся от них и будем иметь в виду идеальную отопительную систему, когда все количество теплоты, полученное в топке, поступает на обогревание помещения. Вильям Томсон предложил другую схему отопления, названную *динамическим отоплением*. Теоретически этот способ отопления более выгоден, чем обычный. Хотя динамическое отопление по техническим причинам и не получило распространения, разбор его принципа представляет интерес как любопытный пример применения законов термодинамики. Кроме того, не исключено, что по мере дальнейшей централизации отопления динамическое отопление найдет широкое применение.

2. При динамическом отоплении только часть теплоты, получаемой в топке, поступает в обогреваемое помещение. Остальная часть затрачивается на работу, производимую тепловой машиной (двигателем). Нагревателем в двигателе служит топка, а холодильником — отапливаемое помещение. Эта работа используется для приведения в действие холодильной машины, включаемой между окружающей средой и помещением. Холодильная машина отбирает тепло от окружающей среды и передает его помещению.

Таким образом, помещение получает теплоту и от горячей топки и от холодной окружающей среды. Общее количество теплоты может превзойти теплоту, полученную от топки при обычном способе отопления. В этом и состоит выгода динамического способа отопления.

Пусть T_1, T_2, T_3 — температуры топки, отапливаемого помещения и окружающей среды соответственно. Пусть топка отдала двигателю количество теплоты Q_1. Из этого часть Q_2 поступила на отопление помещения. Двигатель произвел работу $A = Q_1 - Q_2$. Холодильная машина отобрала Q_3 от окружающей среды, передала помещению Q_2'. На это затрачена работа $A' = Q_2' - Q_3$. Если обе машины — идеальные, то вся работа двигателя тратится на приведение в действие холодильной машины. В этом идеальном случае $A = A'$, т. е. $Q_1 - Q_2 = Q_2' - Q_3$. В реальных машинах есть потери на трение и прочие потери. В этом случае $A > A'$, т. е. $Q_1 - Q_2 > Q_2' - Q_3$. Таким образом, всегда $Q_1 - Q_2 \geqslant Q_2' - Q_3$, или $Q_3 \geqslant Q_2 + Q_2' - Q_1$.

Двигатель и холодильную машину вместе можно рассматривать как одну термодинамическую систему, совершившую круговой процесс. В этом круговом процессе система получила:

теплоту Q_1 от топки при температуре T_1,

теплоту $-(Q_2 + Q_2')$ от помещения при температуре T_2,

теплоту Q_3 от окружающей среды при температуре T_3.

На основании неравенства Клаузиуса

$$\frac{Q_1}{T_1} - \frac{Q_2 + Q_2'}{T_2} + \frac{Q_3}{T_3} \leqslant 0.$$

Исключая Q_3 с помощью неравенства

$$Q_3 \geqslant Q_2 + Q_2' - Q_1,$$

получим

$$(Q_2 + Q_2')\Big(\frac{1}{T_3} - \frac{1}{T_2}\Big) - Q_1\Big(\frac{1}{T_3} - \frac{1}{T_1}\Big) \leqslant 0.$$

Принимая во внимание, что $T_1 > T_2 > T_3$, отсюда находим

$$Q_2 + Q_2' \leqslant \frac{1/T_3 - 1/T_1}{1/T_3 - 1/T_2}\, Q_1. \tag{39.1}$$

Величина $q = Q_2 + Q_2'$, есть теплота, поступающая в отапливаемое помещение. В идеальном случае, когда все процессы квазистатические,

$$q = \frac{1/T_3 - 1/T_1}{1/T_3 - 1/T_2}\, Q_1. \tag{39.2}$$

Так как $T_1 > T_2$, то $1/T_3 - 1/T_1 > 1/T_3 - 1/T_2$, и формула (39.2) дает $q > Q_1$. На рис. 29 представлена зависимость q/Q_1 от температуры помещения T_2 при фиксированных температурах «нагревателя» T_1 и «холодильника» T_2. Если процессы — не квазистатические, то количество теплоты, полученное помещением, будет меньше. Однако и в этом случае можно осуществить такие условия, что q будет больше Q_1, так как на коэффициенты полезного действия тепловых машин теоретически не наложено никаких ограничений, помимо ограничения, накладываемого второй теоремой Карно.

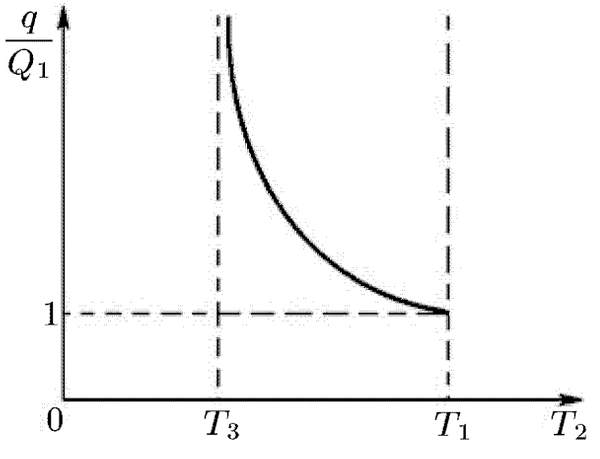


Рис. 29

3. Динамическое отопление может служить примером процесса, в котором теплота от более холодного тела (окружающая среда) переходит к более теплому телу (отапливаемому помещению). Однако такой процесс не противоречит постулату Клаузиуса, так как он сопровождается компенсацией. Компенсация состоит в том, что одновременно теплота переходит от более нагретого тела (топки) к менее нагретому (помещению).

То обстоятельство, что при динамическом способе отопления отапливаемое помещение может получить больше теплоты, чем теплота, развивающаяся в топке, на первый взгляд кажется парадоксальным. Однако в основе такого «парадокса» лежит представление теории теплорода, которое мы интуитивно применяем, быть может, помимо своей воли. В действительности никакого общего закона сохранения количества теплоты не существует, а потому для существования рассматриваемого «парадокса» нет никаких оснований. Поучительно, однако, посмотреть на принцип динамического отопления с точки зрения представления об *обесценивании тепловой энергии* в естественных необратимых процессах, о котором говорилось в § 37. При обычном способе отопления теплота от топки Q_1 при температуре T_1

переходит к помещению в виде такого же количества теплоты, но при более низкой температуре T_2. Это есть процесс *качественного обесценивая теплоты*. При динамическом отоплении, и в этом его преимущество, в идеальном случае, когда все процессы квазистатические, такого обесценивая теплоты нет. На всех стадиях процесса *энергия сохраняется не только количественно, но и качественно*. Так, количество теплоты Q_1 при температуре T_1 эквивалентно количественно и качественно меньшему количеству теплоты Q_2 при температуре T_2 и запасенной работе A. Точно так же количество теплоты Q_3, заимствованное от холодильника при температуре T_3, полностью эквивалентно сумме меньшего количества теплоты Q_2' при температуре помещения T_2 и запаса работы A. Первый принцип термодинамики требует выполнения равенства $Q_3 - Q_2' = A$, но он не накладывает никаких ограничений на величину Q_2'. Ограничения накладываются вторым началом, которое в случае квазистатических процессов требует

$$\frac{Q_3}{T_3} = \frac{Q_2'}{T_2},$$

и, следовательно,

$$Q_2' = \frac{A}{T_3/T_2 - 1}.$$

Когда $T_2 \to T_3$, $Q_2' \to \infty$.

§ 40. Равенство Клаузиуса. Энтропия

1. Допустим, что круговой процесс, совершаемый системой, — *квазистатический*. Неравенство Клаузиуса

$$\oint \frac{\delta Q}{T} \leqslant 0 \tag{40.1}$$

справедливо и для такого процесса. Только под T теперь можно понимать температуру *самой системы*, а не окружающей среды, поскольку обе температуры одинаковы.

Квазистатический процесс обратим и притом в узком смысле слова. Он может идти в противоположном направлении. Для обратного процесса также справедливо неравенство Клаузиуса $\oint \frac{\delta' Q}{T} \leqslant 0$, где $\delta' Q$ обозначают элементарные количества теплоты, получаемые системой на отдельных участках такого обратного процесса. Так как при этом система проходит через те же равновесные состояния, что и в прямом процессе, то $\delta' Q = -\delta Q$, а потому $\oint \frac{\delta Q}{T} \geqslant 0$. Это соотношение совместимо с соотношением (40.1) только в том случае, когда взят знак равенства. Таким образом, *для квазистатического процесса неравенство Клаузиуса переходит в равенство*

$$\oint\limits_{\text{квст}} \frac{\delta Q}{T} = 0. \tag{40.2}$$

На этом равенстве основано введение фундаментального в термодинамике понятия *энтропии*.

2. Пусть система может переходить из начального состояния *1* (рис. 30) в конечное состояние *2* несколькими способами, каждый из которых является квазистатическим процессом. Возьмем два из них —I и II. Эти процессы можно объединить в один квазистатический круговой процесс *1*I*2*II*1*. Применим к нему равенство Клаузиуса:

$$\int\limits_{1\text{I}2} \frac{\delta Q}{T} + \int\limits_{2\text{II}1} \frac{\delta Q}{T} = 0,$$

или

$$\int\limits_{1\text{I}2} \frac{\delta Q}{T} - \int\limits_{1\text{II}2} \frac{\delta Q}{T} = 0,$$

или наконец,

$$\int\limits_{1\text{I}2} \frac{\delta Q}{T} = \int\limits_{1\text{II}2} \frac{\delta Q}{T}. \tag{40.3}$$

Количество теплоты, полученное системой, деленное на абсолютную температуру T, при которой оно было получено, иногда называют *приведенным количеством теплоты*. Величина $\delta Q/T$ есть элементарное приведенное количество теплоты, полученное в бесконечно малом процессе, а интеграл $\int \frac{\delta Q}{T}$ можно назвать приведенным количеством теплоты, полученным в конечном процессе. Пользуясь этой терминологией, равенству Клаузиуса (40.3) можно дать следующую формулировку. *Приведенное количество теплоты, полученное системой при любом квазистатическом круговом процессе, равно нулю.* Эквивалентной является следующая формулировка. *Приведенное количество теплоты, квазистатически полученное системой, не зависит от пути перехода, а определяется лишь начальным и конечным состояниями системы.* Этот важный результат позволяет ввести новую функцию состояния, называемую энтропией.

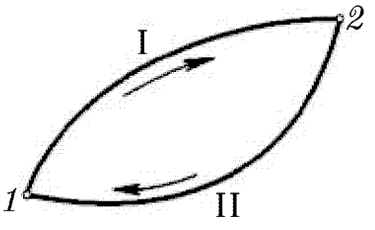


Рис. 30

Энтропия системы есть функция ее состояния, определенная с точностью до произвольной постоянной. Разность энтропий в двух равновесных состояниях 2 и 1, по определению, равна приведенному количеству теплоты, которое надо сообщить системе, чтобы перевести ее из состояния 1 в состояние 2 по любому квазистатическому пути. Таким образом, если энтропии в состояниях *1* и *2* обозначить буквами S_1 и S_2, то по определению

$$S_2 - S_1 = \int\limits_{1\to 2} \frac{\delta Q}{T}. \tag{40.4}$$

Значение произвольной постоянной, с которой определена энтропия, не играет роли. В этом отношении с определением энтропии дело обстоит

так же, как с определением энергии. Физический смысл имеет не сама энтропия, а лишь *разность энтропий*. Условно энтропию системы в каком-либо определенном состоянии можно принять за нуль. Тогда определится и значение произвольной постоянной в выражении для энтропии.

3. Итак, по определению

$$S = \int\limits_{\text{квст}} \frac{\delta Q}{T}, \tag{40.5}$$

где интеграл берется для произвольного квазистатического процесса, переводящего систему в рассматриваемое состояние из другого состояния, условно принятого за начальное. Для дифференциала функции S имеем

$$dS = \left(\frac{\delta Q}{T}\right)_{\text{квст}}. \tag{40.6}$$

Как уже неоднократно подчеркивалось, δQ не является дифференциалом какой бы то ни было функции. Однако формула (40.6) показывает, что *если δQ есть элементарное количество теплоты, квазистатически полученное системой, то после деления на T оно переходит в полный дифференциал функции состояния — энтропии.*

4. В качестве примера вычислим энтропию S одного моля идеального газа. Для всякого бесконечно малого квазистатического процесса с идеальным газом

$$\delta Q = C_V\,dT + P\,dV = C_V(T)\,dT + RT\,\frac{dV}{V}.$$

Отсюда

$$dS = \frac{\delta Q}{T} = C_V(T)\,\frac{dT}{T} + R\,\frac{dV}{V},$$

$$S = \int C_V(T)\,\frac{dT}{T} + R\ln V.$$

Если теплоемкость C_V не зависит от температуры, то интеграл легко берется, и мы получаем

$$S = C_V \ln T + R \ln V + \text{const}. \tag{40.7}$$

Если газ содержит ν молей, то

$$S = \nu C_V \ln T + \nu R \ln V + \text{const}.$$

Надо, однако, иметь в виду, что это выражение было получено в предположении, что число молекул в газе остается постоянным. Поэтому *аддитивная постоянная в выражении для энтропии может зависеть от числа частиц в газе*. Эту постоянную следует определить так, чтобы энтропия S была пропорциональна числу частиц газа или, что то же самое, числу молей ν. Этому условию удовлетворяет выражение

$$S = \nu\left(C_V \ln T + R \ln \frac{V}{\nu} + \text{const}\right), \tag{40.8}$$

или

$$S = \frac{N}{N_A}\left(C_V \ln T + R \ln \frac{V}{N} + \text{const}\right). \tag{40.9}$$

В обоих выражениях аддитивная постоянная в скобках уже не зависит от числа частиц газа. Формулы (40.8) и (40.9) применимы к идеальным газам не только с постоянным, но и с переменным числом частиц.

5. Если квазистатический процесс — адиабатический, то $\delta Q = 0$, а следовательно, $dS = 0$, $S = \text{const}$. Таким образом, всякий квазистатический адиабатический процесс есть процесс, происходящий при постоянной энтропии. Поэтому его можно также назвать *изэнтропическим процессом*.

ЗАДАЧИ

1. Показать, что для любого вещества политропа может пересекать изотерму не более чем в одной точке.

Доказательство. Допустив противоположное, предположим, что A и B — две соседние точки, в которых политропы пересекаются с изотермой (рис. 31). Применим к циклу $ACBDA$ равенство Клаузиуса. На политропе ADB теплоемкость C постоянна, а потому

$$\int\limits_{ADB} \frac{\delta Q}{T} = C \int\limits_{T_A}^{T_B} \frac{dT}{T} = 0.$$

(Интеграл обращается в нуль, так как $T_A = T_B$, поскольку точки A и B лежат на изотерме.) На изотерме ACB

$$\int \frac{\delta Q}{T} = \frac{1}{T} \int \delta Q = \frac{Q}{T}.$$

Таким образом, равенство Клаузиуса сводится к $Q = 0$, где Q — теплота, полученная системой. Но для кругового процесса $Q = A$. Значит, площадь цикла $ACBDA$ равна нулю, что может быть тогда и только тогда, когда между точками A и B политропа и изотерма пересекаются между собой. Это противоречит предположению, что A и B — соседние точки пересечения политропы с изотермой.

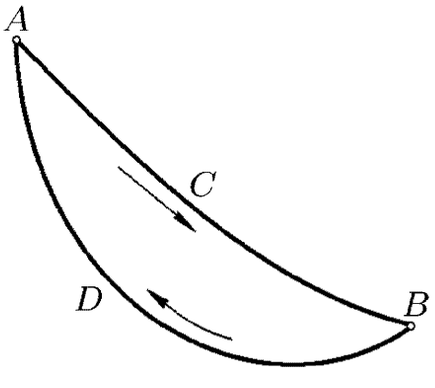


Рис. 31

2. Показать, что для любого вещества адиабата может пересекать изотерму не более чем в одной точке.

3. Цикл состоит из двух изохор и двух изобар (рис. 32). Показать, что для любого вещества с постоянными теплоемкостями C_V и C_P температуры T_1, T_2, T_3, T_4 связаны соотношением $T_1 T_3 = T_2 T_4$.

Решение. Применив к рассматриваемому циклу равенство Клаузиуса, получим

$$(C_P - C_V) \ln \frac{T_2 T_4}{T_1 T_3} = 0,$$

откуда и следует требуемое соотношение, так как $C_P - C_V \neq 0$.

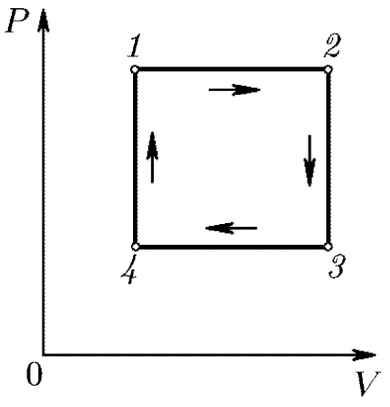


Рис. 32

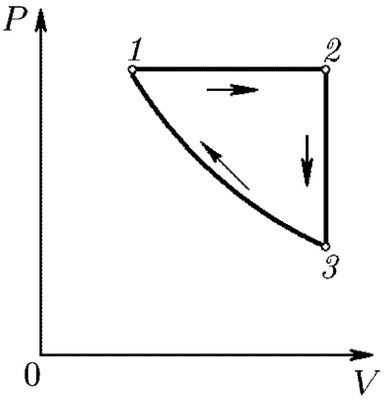


Рис. 33

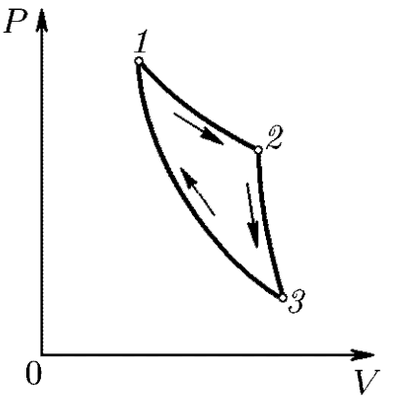


Рис. 34

4. Цикл состоит из изобары *12*, изохоры *23* и адиабаты *31* (рис. 33). Показать, что для любого вещества с постоянными теплоемкостями C_V и C_P температуры T_1, T_2, T_3 связаны соотношением

$$\frac{T_2}{T_3} = \left(\frac{T_2}{T_1}\right)^{\gamma}, \quad \text{где} \quad \gamma = \frac{C_P}{C_V}.$$

5. Определить работу цикла, совершаемого любым веществом и состоящего из изотермы *12*, политропы *23* и адиабаты *31* (рис. 34). Известно, что теплоемкость тела на политропе *23* равна C, а температуры на изотерме *12* и в состоянии *3* равны соответственно T_1 и T_3.

Ответ. $A = C[T_1 \ln(T_1/T_3) - (T_1 - T_3)]$.

6. Цикл состоит из двух изотерм *12*, *34* с температурами T_1 и T_2 и двух изохор *23*, *41* (рис. 35). На изотерме с температурой T_1 получено количество теплоты Q_1. Определить работу цикла, если теплоемкость рабочего вещества C_V зависит только от его температуры, но не зависит от объема.

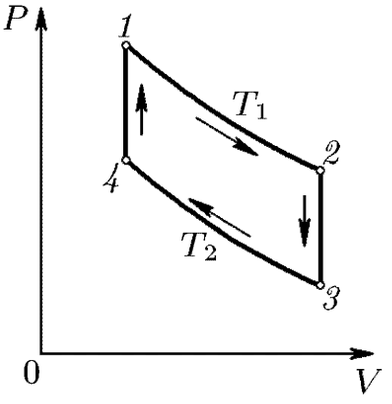


Рис. 35

Ответ. $A = \dfrac{T_1 - T_2}{T_1} Q_1$.

7. Произвольная термодинамическая система квазистатически переходит из равновесного состояния *1* в равновесное состояние *2* двумя способами. В первом способе система адиабатически охлаждается до температуры T_0, затем изотермически получает теплоту и, наконец, адиабатически переходит в состояние *2*. Во втором способе переход осуществляется по произвольному пути, однако так, что на каждом участке этого пути система получает теплоту, а ее температура остается выше T_0. Показать, что в первом способе для перевода системы из состояния *1* в состояние *2* требуется меньшая затрата теплоты, чем во втором.

Решение. Пусть *1342* схематически изображает первый переход, а *152* — второй (рис. 36). Применяя к ним равенство Клаузиуса и учитывая, что на адиабатах *13* и *42* система теплоту не получает, напишем

$$\int\limits_{152} \frac{\delta Q}{T} = \int\limits_{1342} \frac{\delta Q}{T} = \frac{Q_0}{T_0},$$

где Q_0 — теплота, полученная на изотерме *34*. По условию $T > T_0$ и $\delta Q > 0$, а потому

$$\int\limits_{152} \frac{\delta Q}{T} < \int\limits_{152} \frac{\delta Q}{T_0} = \frac{Q}{T_0},$$

где Q — теплота, полученная на пути *152*. Комбинируя последнее неравенство с предыдущим равенством, получаем $Q > Q_0$.

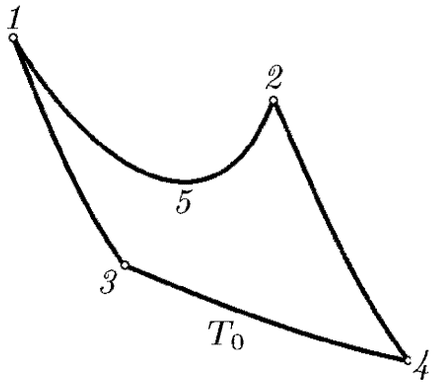


Рис. 36

8. Показать, что разность энтропий системы в состояниях *2* и *1* (при условии, что $S_2 > S_1$) может быть определена как наименьшее количество теплоты, которое требуется сообщить системе, чтобы квазистатически перевести ее из состояния *1* в состояние *2* и притом так, чтобы при переходе температура системы не опускалась ниже одного кельвина.

9. Если во всех точках изотермы коэффициент теплового расширения равен нулю, то такая изотерма совпадает с адиабатой. Доказать.

Решение. Пусть во всех точках изотермы $(\partial V/\partial T)_P = 0$. Тогда из (8.3) следует, что $(\partial P/\partial T)_V = 0$. На изотерме ввиду соотношения (34.2)

$$dU = \left[T\left(\frac{\partial P}{\partial T}\right)_V - P\right] dV = -P\,dV,$$

а потому $\delta Q = dU + P\,dV = 0$. Значит, изотерма во всех точках должна совпадать с адиабатой.

10. В цикле Карно в качестве холодильника выбрана вода при 4 °C. Так как коэффициент теплового расширения при этой температуре равен нулю, то для осуществления цикла Карно не надо сообщать теплоту холодильнику (см. предыдущую задачу), т. е. КПД цикла равен единице. В чем ошибочность этого рассуждения?

Ответ. 1) Если бы коэффициент теплового расширения обращался в нуль на всем протяжении изотермы, то она совпадала бы с адиабатой, и цикл Карно между температурой 4 °C и какой-либо другой температурой осуществить было бы нельзя. 2) На самом деле для воды коэффициент теплового расширения обращается в нуль только в одной точке изотермы.

§ 41. Закон возрастания энтропии

1. Допустим, что система переходит из равновесного состояния *1* в равновесное состояние *2* (рис. 37), но процесс перехода является необратимым — на рисунке он изображен штриховой линией I. Вернем систему из состояния *2* в исходное состояние *1* квазистатически по какому-либо пути II. На основании неравенства Клаузиуса можно написать

$$\oint \frac{\delta Q}{T} \equiv \int\limits_{\text{I}} \frac{\delta Q}{T} + \int\limits_{\text{II}} \frac{\delta Q}{T} \leqslant 0.$$

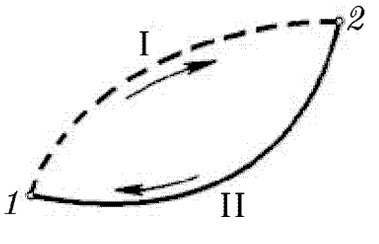


Рис. 37

Так как процесс II квазистатический, то

$$\int\limits_{\text{II}} \frac{\delta Q}{T} = S_1 - S_2.$$

Поэтому неравенство Клаузиуса принимает вид

$$S_2 - S_1 \geqslant \int\limits_{1\to 2} \frac{\delta Q}{T}. \tag{41.1}$$

Здесь под T следует понимать температуру окружающей среды, при которой она отдает системе количество теплоты δQ.

Если система адиабатически изолирована, то $\delta Q = 0$, и интеграл в (41.1) обращается в нуль. Тогда

$$S_2 \geqslant S_1. \tag{41.2}$$

Таким образом, *энтропия адиабатически изолированной системы не может убывать*; *она либо возрастает*, *либо остается постоянной*. Это — *закон возрастания энтропии*. Если $S_2 > S_1$, то переход адиабатически изолированной системы из состояния *2* в состояние *1* невозможен, так как он сопровождался бы убыванием энтропии. Наоборот, адиабатический переход системы из состояния с меньшей энтропией *1* в состояние с большей энтропией *2* не противоречит постулату второго начала термодинамики и в этом смысле является возможным. Второе начало, таким образом, позволяет судить о *направлении процессов*, которые могут происходить в природе.

2. В качестве иллюстрации рассмотрим следующий пример. Пусть один моль идеального газа заключен в цилиндр с твердыми адиабатическими стенками (рис. 38). Цилиндр разделен на две части твердой перегородкой AB. Сначала газ занимает одну из этих частей с объемом V_1. Во второй части — вакуум. Затем перегородка убирается или в ней делается отверстие. Газ начинает перетекать во вторую часть цилиндра, пока не выравняются давления и температуры в обеих частях. Является ли описанный процесс обратимым? Для ответа на этот вопрос вычислим изменение энтропии газа. В результате описанного процесса внутренняя энергия газа не меняется, так как он заключен в жесткую адиабатическую оболочку. Не меняется и температура газа, поскольку она однозначно определяется его внутренней энергией (закон Джоуля). Обозначим через T общую температуру в начале и конце процесса, через V_1 — начальный, а через V_2 — конечный объемы газа. Чтобы вычислить изменение энтропии газа, надо перевести его из начального состояния в конечное *квазистатически*. Это можно сделать, приведя газ в тепловой контакт с нагревателем, имеющим температуру T. Бесконечно медленно уменьшая давление на газ, можно

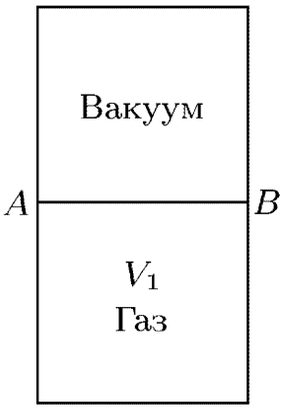


Рис. 38

его изотермически перевести в конечное состояние с объемом V_2. При этом газ будет заимствовать теплоту от нагревателя и превышать ее в эквивалентную работу.

Для изотермического процесса

$$\delta Q = P\,dV = RT\,\frac{dV}{V},$$

$$dS = \frac{\delta Q}{T} = R\,\frac{dV}{V}.$$

Отсюда

$$S_2 - S_1 = R\int\limits_{V_1}^{V_2}\frac{dV}{V} = R\ln\frac{V_2}{V_1} > 0.$$

Энтропия возросла. Поэтому адиабатическое расширение газа в пустоту — необратимый процесс.

Обратим внимание на следующую тривиальную, но довольно распространенную ошибку. Говорят, что в рассматриваемом примере теплота не подводится, т. е. $\delta Q = 0$, а потому также $dS = \delta Q/T = 0$. Отсюда заключают, что $S = \text{const}$, т. е. энтропия в начальном и конечном состояниях одна и та же. Ошибка состоит в том, что равенством $dS = \delta Q/T$ в приведенном примере пользоваться нельзя. Оно относится *только к квазиравновесным процессам*. Для неравновесных процессов это равенство не имеет места. По самому определению энтропии, даваемому интегралом (40.5), интеграл должен быть взят для произвольного, но обязательно *квазистатического процесса*, переводящего систему из начального состояния в конечное.

§ 42. Обобщение понятия энтропии на неравновесные состояния

1. Понятие энтропии, как оно введено выше, относится только к равновесным состояниям. Однако оно может быть обобщено и на неравновесные состояния. Рассмотрим это обобщение для того случая, когда система, находящаяся в неравновесном состоянии, может быть мысленно разбита на отдельные макроскопические подсистемы, каждая из которых практически находится в равновесии, так что ее внутреннее состояние можно характеризовать, например, двумя параметрами T и P. Вообще говоря, подсистемы могут совершать макроскопические движения. Параметры, характеризующие внутреннее состояние подсистем и их макроскопическое движение, могут плавно меняться от подсистемы к подсистеме. В этом случае говорят, что имеет место *локальное термодинамическое равновесие*. Примером может служить жидкость или газ в состоянии макроскопического движения, как оно понимается в гидродинамике и аэродинамике. Можно говорить об энтропии S_i каждой из подсистем в том смысле, в каком это понятие было введено в § 40. Тогда энтропию всей системы можно определить

как сумму энтропий таких подсистем:

$$S = \sum S_i. \tag{42.1}$$

Макроскопические подсистемы, на которые мысленно разбивается вся система, должны быть выбраны настолько малыми, чтобы сумма (42.1) уже не менялась при дальнейшем дроблении системы. Мы не утверждаем, что это можно сделать всегда. Мы подчеркиваем только, что наше расширенное понятие энтропии пока что относится лишь к тем случаям, когда это сделать можно.

2. При таком обобщенном понимании энтропии теорема о ее возрастании остается в силе. Для доказательства рассмотрим сначала адиабатически изолированную систему, заключенную в жесткую оболочку и состоящую из n подсистем с температурами $T_1, T_2, \ldots, T_n$. Предположим сначала, что все подсистемы находятся в покое, химически не реагируют и не перемешиваются друг с другом (например, в результате диффузии), между ними может происходить только теплообмен. Если δQ_i — элементарное количество теплоты, полученное i-й подсистемой в результате теплообмена, то, как было показано в § 16,

$$\sum_{i=1}^{n} \delta Q_i = 0. \tag{42.2}$$

В процессе теплообмена система, поскольку она изолирована, стремится к вполне определенному равновесному состоянию. В этом процессе одни подсистемы нагреваются, получая положительные количества теплоты δQ_i, другие подсистемы охлаждаются — для них δQ_i отрицательны. На основании неравенства Клаузиуса можно написать для i-й подсистемы

$$\Delta S_i \geqslant \int \frac{\delta Q_i}{T_i},$$

а, следовательно, для всей системы

$$\Delta S \geqslant \int \sum_{i=1}^{n} \frac{\delta Q_i}{T_i}. \tag{42.3}$$

Разобьем все подсистемы на две группы: более горячую группу I (с номерами подсистемы от 1 до m), для которой δQ_i отрицательны, и более холодную группу II (с номерами от $m+1$ до n), для которой δQ_i положительны. Запишем неравенство (42.3) в виде

$$\Delta S \geqslant \int \left(\sum_{i=1}^{m} \frac{\delta Q_i}{T_i} + \sum_{i=m+1}^{n} \frac{\delta Q_i}{T_i} \right).$$

Первая сумма отрицательна, вторая — положительна. Очевидно, эти суммы можно представить в виде

$$\sum_{i=1}^{m} \frac{\delta Q_i}{T_i} = \frac{1}{T_{\mathrm{I}}} \sum_{i=1}^{m} \delta Q_i = \frac{\delta Q_{\mathrm{I}}}{T_{\mathrm{I}}},$$

$$\sum_{i=m+1}^{n} \frac{\delta Q_i}{T_i} = \frac{1}{T_{\mathrm{II}}} \sum_{i=m+1}^{n} \delta Q_i = -\frac{\delta Q_{\mathrm{II}}}{T_{\mathrm{II}}}.$$

Здесь δQ_{I} — полное количество теплоты, полученное в элементарном процессе первой, а δQ_{II} — второй группой подсистем, T_{I} и T_{II} — надлежащим образом усредненные температуры подсистем этих групп. Очевидно, что $T_{\mathrm{I}} > T_{\mathrm{II}}$. Кроме того, $\delta Q_{\mathrm{I}} + \delta Q_{\mathrm{II}} = 0$, $\delta Q_{\mathrm{I}} < 0$. Последние два неравенства выражают тот факт, что в процессе теплообмена более теплое тело всегда охлаждается, а менее теплое — нагревается. Этот факт является следствием постулата Клаузиуса, выражающего второе начало термодинамики. Таким образом,

$$\Delta S \geqslant \int \left(\frac{\delta Q_{\mathrm{I}}}{T_{\mathrm{I}}} + \frac{Q_{\mathrm{II}}}{T_{\mathrm{II}}} \right) = \int \left(\frac{1}{T_{\mathrm{II}}} - \frac{1}{T_{\mathrm{I}}} \right) \delta Q_{\mathrm{II}} > 0,$$

т. е. в результате теплообмена энтропия всей системы возрастает.

3. Допустим теперь, что подсистемы не только обмениваются теплом, но и совершают макроскопическое движение. Если система изолирована, то благодаря внутренним процессам трения макроскопическое движение в конце концов прекратится, а его кинетическая энергия перейдет в тепло. Неравенство (42.3) сохраняет силу и в этом случае. Однако теперь все величины δQ_i возрастут, а с ними возрастет и величина, стоящая в правой части неравенства (42.3). Поэтому неравенство $\Delta S \geqslant 0$ только усилится.

4. Освободимся теперь от ограничения, что адиабатическая оболочка, в которую заключена система, жесткая. Пусть адиабатическая оболочка эластична. Тогда при ее растяжениях и сжатиях над системой будет производиться механическая работа, и в системе возникнет макроскопическое движение. Само по себе возникновение макроскопического движения непосредственно не влияет на энтропию системы, поскольку оно не сопровождается сообщением или отнятием теплоты. Его влияние сказывается косвенно — через превращение в теплоту кинетической энергии макроскопического движения. А такое превращение, как мы видели, ведет лишь к возрастанию энтропии. Поэтому нет необходимости предполагать, что оболочка жесткая. Достаточно, чтобы она лишь адиабатически изолировала систему.

5. Случаи, когда в системе происходят химические реакции или процессы перемешивания, в общем виде мы разбирать не будем. Частный случай смешения двух газов будет разобран в следующем параграфе. Здесь же ограничимся следующим замечанием. Доказательство неравенства (41.2) было проведено в общем виде для любых процессов,

происходящих в системе. Единственное ограничение состоит в том, что начальное и конечное состояния системы должны быть равновесными. Естественно ожидать, что возрастание энтропии в неравновесном процессе, переводящем систему из равновесного состояния *1* в равновесное состояние *2*, всегда происходит монотонно. Если это так, то неравенство (41.2) справедливо также для любых неравновесных состояний *1* и *2*, из которых первое предшествует второму. Для доказательства достаточно состояния *1* и *2* рассматривать как промежуточные неравновесные состояния в адиабатическом процессе, переводящем систему из равновесного состояния *3* в равновесное же состояние *4*.

6. Если состояние системы таково, то ее нельзя разбить на макроскопические части, находящиеся в локальном термодинамическом равновесии, то приведенное обобщение понятия энтропии теряет смысл. Однако статистическая физика позволяет распространить понятие энтропии и закон ее возрастания и на такие состояния (см. § 80).

7. Клаузиус, рассматривая всю Вселенную как замкнутую систему, свел содержание второго закона термодинамики к утверждению: «Энтропия Вселенной стремится к максимуму». Когда этот максимум будет достигнут, во Вселенной прекратятся какие бы то ни было процессы. Действительно, каждый процесс приводил бы к возрастанию энтропии, а это невозможно, так как энтропия уже достигла своего предельного — максимального — значения. Таким образом, согласно Клаузиусу, во Вселенной в конце концов должно наступить абсолютно равновесное состояние, в котором никакие процессы уже невозможны. Такое состояние было названо *тепловой смертью Вселенной*. Собственно говоря, для такого вывода совсем не требуется привлекать понятие энтропии и закон ее возрастания. Действительно, этот вывод есть не что иное, как общее начало термодинамики, распространенное на всю Вселенную. Однако и общее начало термодинамики и закон возрастания энтропии получены путем обобщения опытных фактов, относящихся к *ограниченным системам*. Распространение их на всю Вселенную есть экстраполяция, для которой нет оснований. Вселенная в целом может эволюционировать непрерывно и монотонно, никогда не приходя в состояние термодинамического равновесия. Такая возможность допускается, например, общей теорией относительности: благодаря наличию гравитационных полей гигантские космологические системы могут непрерывно эволюционировать в сторону возрастания энтропии, никогда, однако, не приходя в состояние с максимумом энтропии, так как такого состояния вообще не существует. Другая критика концепции тепловой смерти Вселенной была дана Больцманом (см. § 80, п. 6).

ЗАДАЧА

Два твердых тела с постоянными теплоемкостями C_1 и C_2 и температурами T_1 и T_2 образуют замкнутую систему. В результате теплообмена температуры их выравниваются, и система приходит в равновесное состояние. Доказать непосредственным расчетом, что энтропия системы возрастает.

Р е ш е н и е. Пусть T — установившаяся температура системы. Ее можно определить, исходя из требования постоянства внутренней энергии: $C_1T_1 + C_2T_2 = (C_1 + C_2)T$. Приращение энтропии первого тела $\Delta S_1 = C_1 \ln(T/T_1)$, второго — $\Delta S_2 = C_2 \ln(T/T_2)$, всей системы — $\Delta S = \Delta S_1 + \Delta S_2$. Для исследования знака ΔS введем обозначения $\alpha = C_2/C_1$, $x = T_2/T_1$. Величины α и x существенно положительны. После несложных преобразований получаем

$$\Delta S = C_1\left\{\ln\frac{1+\alpha x}{1+\alpha} + \alpha\ln\frac{1+\alpha x}{x(1+\alpha)}\right\}.$$

Отсюда при $x = 1$ (т. е. при $T_1 = T_2$) получаем $\Delta S = 0$, как и должно быть. Для производной имеем

$$\frac{d\Delta S}{dx} = C_1\,\frac{\alpha x - \alpha}{x(1+\alpha x)}.$$

Производная обращается в нуль при $x = 1$, положительна при $x > 1$ и отрицательна при $x < 1$. Значит, при $x = 1$ приращение ΔS минимально и монотонно возрастает по обе стороны этого минимума. Таким образом, при $x \neq 1$ (т. е. при $T_1 \neq T_2$) энтропия системы возрастает.

§ 43. Возрастание энтропии при диффузии газов. Парадокс Гиббса

1. Пусть два идеальных газа *1* и *2* заключены в закрытом сосуде с твердыми адиабатическими стенками, так что объем сосуда V остается неизменным. В начальный момент газы отделены один от другого непроницаемой перегородкой и имеют общую температуру T. Затем перегородка убирается, и начинается необратимый процесс смешение газов. В конце концов он прекращается, и система переходит в равновесное состояние, в котором оба газа равномерно перемешаны. Температура в конечном состоянии будет такая же, что и в начальном состоянии, так как система изолирована, а газы — идеальные. Как изменится энтропия системы после смешения?

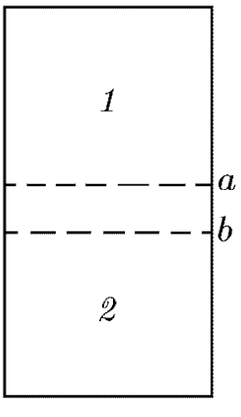


Рис. 39

При термодинамическом определении энтропии задача сводится к вычислению интеграла $\int \frac{\delta Q}{T}$ для процесса, переводящего систему из начального состояния в конечное. Этот процесс может быть любым, но обязательно квазистатическим. Действительно происходящий процесс смешения, возникающий после удаления перегородки, не годится, так как он не квазистатический. Однако принципиально возможно смешать оба газа квазистатически, если только газы не тождественны. Это можно сделать, например, следующим образом.

Допустим, что перегородка, разделяющая газы в начальном состоянии, состоит из двух идеальных полупроницаемых перегородок a и b, сложенных вместе (рис. 39). Перегородка a беспрепятственно пропускает газ *1*, но абсолютно непроницаема для газа *2*. Перегородка

b, напротив, пропускает газ *2*, но непроницаема для газа *1*. Идеальные полупроницаемые перегородки в действительности не существуют, но они допустимы в рассуждениях, применяющихся в мысленных экспериментах [1].

Сложная перегородка, состоящая из перегородок a и b, очевидно, непроницаема и для газа *1* и для газа *2*. Устранив адиабатическую изоляцию системы, приведем ее в тепловой контакт с термостатом, температура которого поддерживается постоянной и равной T. Затем перемещением перегородки b заставим газ *1* квазистатически расширяться от первоначального объема V_1 до конечного объема V. При таком расширении газ *1* совершает работу, и для поддержания температуры постоянной к нему надо подводить тепло. Приращение энтропии газа *1* найдется по формуле (40.7). Поскольку температура остается постоянной, оно равно

$$\Delta_1 S = \nu_1 R \ln \frac{V}{V_1},$$

где ν_1 — число молей первого газа. Состояние газа *2* при этом меняться не будет. Теперь будем таким же образом перемещать полупроницаемую перегородку a, чтобы газ *2* квазистатически заполнил весь объем сосуда. Состояние газа *1* при этом не изменится, а энтропия газа *2* получит приращение

$$\Delta_2 S = \nu_2 R \ln \frac{V}{V_2},$$

где ν_2 — число молей газа *2*, а V_2 — его начальный объем. В результате система придет в то же конечное состояние, что и в описанном выше реальном процессе смешения. Приращение энтропии всей системы

$$\Delta S = R\Big(\nu_1 \ln \frac{V}{V_1} + \nu_2 \ln \frac{V}{V_2}\Big). \tag{43.1}$$

Оно положительно, так как $V_1 < V$ и $V_2 < V$. Энтропия возросла. Это доказывает, что самопроизвольный процесс смешения газов в адиабатической оболочке, описанный в начале этого параграфа, действительно необратим.

2. Формула (43.1) приводит к парадоксальному выводу, называемому *парадоксом Гиббса* (1839–1903). Допустим, что газы *1* и *2* тождественны. Тогда по формуле (43.1) возрастание энтропии ΔS по-прежнему остается. Например, если перегородка делит один и тот же газ в количестве ν молей на две равные части, то $\nu_1 = \nu_2 = \nu/2$, $V_1 = V_2 = V/2$, и формула (43.1) дает

$$\Delta S = \nu R \ln 2.$$

[1] Для водорода наилучшим известным приближением к идеальной полупроницаемой перегородке является горячая палладиевая фольга, пропускающая этот газ.

Однако конечное состояние системы макроскопически ничем не отличается от начального. Энтропия возросла, а состояние системы не изменилось. В этом и состоит парадокс Гиббса.

Для понимания парадокса Гиббса существенно заметить, что формула (43.1) доказана нами только для случая, когда смешивающиеся газы *1* и *2 существенно различны*. Для тождественных газов наши рассуждения неприменимы. Перегородки, проницаемые или непроницаемые для газа *1*, останутся таковыми же и для газа *2*, тождественного с ним. Принципиально невозможно перемешать тождественные газы квазистатическим способом, описанным выше.

Для тождественных газов $\Delta S = 0$, и формула (43.1) неприменима. Но этой формулой можно пользоваться всегда, когда атомы или молекулы смешивающихся газов различны, хотя бы это различие и было сколь угодно малым. Теоретически допустим предельный переход, в котором свойства атомов одного газа неограниченно приближаются к свойствам атомов другого газа. Возрастание энтропии должно было бы сохраниться и при таком предельном переходе. Эйнштейн (1879–1955) видел в этом определенную трудность, свойственную классическому описанию явлений природы. Такой трудности не существует в квантовой физике, где состояния физических систем дискретны. В частности, число различных типов атомов конечно, а потому нельзя выполнить такой предельный переход, в котором свойства одного атома непрерывно переходили в свойства другого.

§ 44. Различные понимания второго начала термодинамики

Термин «второе начало термодинамики» употребляется в физике уже более ста лет. Однако до сих пор разные авторы вкладывают в него различное содержание. Хотя этот вопрос и чисто терминологический, имеет смысл коротко остановиться на нем.

Наиболее логично поступают те авторы, которые понимают под вторым началом основной постулат: постулат Томсона–Планка, постулат Клаузиуса или эквивалентные им утверждения.

Другие авторы сводят содержание второго начала термодинамики к двум положениям, являющимся следствиями основного постулата: 1) существованию энтропии S как функции состояния системы; 2) принципу возрастания энтропии. Эти два положения, как впервые отметила Т. А. Афанасьева-Эренфест (1876–1964), логически независимы друг от друга. В самом деле, существование функции S совершенно не зависит от необратимости естественных процессов, отраженной в формулировке основного постулата. Это видно уже из того, что в основу доказательства существования энтропии S можно было бы положить прямо противоположный постулат, например, такой: «Невозможен круговой процесс, единственным результатом которого было бы нагревание теплового резервуара за счет механической работы». Доказательство же возрастания энтропии существенно опирается именно на основной постулат, а не на обратное ему утверждение. Если бы было справедливо обратное утверждение, то энтропия адиабатически изолированной системы не возрастала бы, а убывала.

Наконец, многие авторы по примеру Афанасьевой-Эренфест понимают под вторым началом термодинамики только одно следствие основного постулата, а именно существование энтропии как функции состояния системы. Основанием для такого понимания может служить замечание, что все соотношения, имеющие характер равенств, выводимые из второго начала термодинамики, используют лишь одно свойство энтропии — ее бесконечно малое приращение является полным дифференциалом.

§ 45. Термодинамические функции

1. Наряду с энтропией можно пользоваться множеством других, связанных с ней функций состояния. Рассмотрим главнейшие из них.

Если процесс квазистатический, то $\delta Q = T\,dS$. Для такого процесса уравнение первого начала

$$\delta Q = dU + P\,dV \tag{45.1}$$

можно переписать в виде

$$dU = T\,dS - P\,dV. \tag{45.2}$$

Если ввести энтальпию $I = U + PV$, то можно исключить U и получить

$$dI = T\,dS + V\,dP. \tag{45.3}$$

Так как $T\,dS = \delta Q$, то при постоянном давлении $dI = \delta Q$. Отсюда видно, что *энтальпия есть такая функция состояния, приращение которой в квазистатическом процессе при постоянном давлении дает количество теплоты, полученное системой.* По этой причине энтальпию называют также *тепловой функцией* или *теплосодержанием.*

Особенно важное значение в термодинамики имеют две функции состояния: *свободная энергия* Ψ, введенная Гельмгольцем, и *термодинамический потенциал* Φ, введенный Гиббсом. Эти функции состояния определяются выражениями

$$\Psi = U - TS, \tag{45.4}$$

$$\Phi = \Psi + PV = U - TS + PV. \tag{45.5}$$

Для их дифференциалов легко получить

$$d\Psi = -S\,dT - P\,dV, \tag{45.6}$$

$$d\Phi = -S\,dT + V\,dP. \tag{45.7}$$

При изотермическом процессе $dT = 0$, а потому $d\Psi = -P\,dV = -\delta A$. Отсюда $A = \Psi_1 - \Psi_2$. Следовательно, *свободная энергия есть функция состояния системы, убыль которой в квазистатическом изотермическом процессе дает работу, произведенную системой.*

Соотношения (45.2), (45.3), (45.6) и (45.7) наводят на мысль рассматривать внутреннюю энергию U как функцию аргументов S и V, энтальпию I — как функцию S и P, свободную энергию Ψ — как функцию T и V, термодинамический потенциал Φ — как функцию T и P:

$$U = U(S, V), \quad I = I(S, P), \quad \Psi = \Psi(T, V), \quad \Phi = \Phi(T, P). \tag{45.8}$$

Такого рода соотношения называются *каноническими уравнениями состояния вещества*. Они были систематически введены в термодинамику Гиббсом. Гиббс отметил, что каждое из канонических уравнений состояния дает более богатую информацию о свойствах вещества, чем термическое или калорическое уравнения состояния в отдельности. *Каноническое уравнение состояния, в какой бы из четырех форм* (45.8) *оно ни было взято, содержит полные сведения о термических и калорических свойствах вещества.* Действительно, из уравнений (45.8) получаем

$$dU = \left(\frac{\partial U}{\partial S}\right)_V dS + \left(\frac{\partial U}{\partial V}\right)_S dV, \quad dI = \left(\frac{\partial I}{\partial S}\right)_P dS + \left(\frac{\partial I}{\partial P}\right)_S dP,$$

$$d\Psi = \left(\frac{\partial \Psi}{\partial T}\right)_V dT + \left(\frac{\partial \Psi}{\partial V}\right)_T dV, \quad d\Phi = \left(\frac{\partial \Phi}{\partial T}\right)_P dT + \left(\frac{\partial \Phi}{\partial P}\right)_T dP.$$

Сравнение этих соотношений с соотношениями (45.2), (45.3), (45.6) и (45.7) дает

$$T = \left(\frac{\partial U}{\partial S}\right)_V, \quad P = -\left(\frac{\partial U}{\partial V}\right)_S, \tag{45.9}$$

$$T = \left(\frac{\partial I}{\partial S}\right)_P, \quad V = \left(\frac{\partial I}{\partial P}\right)_S, \tag{45.10}$$

$$S = -\left(\frac{\partial \Psi}{\partial T}\right)_V, \quad P = -\left(\frac{\partial \Psi}{\partial V}\right)_T, \tag{45.11}$$

$$S = -\left(\frac{\partial \Phi}{\partial T}\right)_P, \quad V = \left(\frac{\partial \Phi}{\partial P}\right)_T. \tag{45.12}$$

Отметим два следствия выведенных уравнений. Из определения функций Ψ и Φ следует $U = \Psi + TS$, $I = \Phi + TS$. Подставив сюда выражения для энтропии из формул (45.11) и (45.12), получим

$$U = \Psi - T\left(\frac{\partial \Psi}{\partial T}\right)_V, \tag{45.13}$$

$$I = \Phi - T\left(\frac{\partial \Phi}{\partial T}\right)_P. \tag{45.14}$$

Эти уравнения называются *уравнениями Гиббса–Гельмгольца*. Сразу можно отметить пользу, которую можно извлечь из этих уравнений. Часто бывает легко найти свободную энергию Ψ с точностью до слагаемого, зависящего только от температуры. Это можно сделать, вычислив изотермическую работу, совершаемую системой. Тогда формула (45.13) позволяет с той же неопределенностью найти и внутреннюю энергию системы.

Если известна функция $U = U(S, V)$, то дифференцированием ее по S и V можно найти температуру и давление системы, т. е. получить полные сведения о ее термических свойствах. Затем по формуле (45.1) можно найти δQ и соответствующие теплоемкости, т. е. получить полные сведения также и о калорических свойствах системы. То же самое

можно сделать с помощью любого из оставшихся трех канонических уравнений состояния.

2. Далее, вторичным дифференцированием из соотношений (45.9) находим

$$\left(\frac{\partial T}{\partial V}\right)_S = \frac{\partial^2 U}{\partial S\,\partial V}, \quad \left(\frac{\partial P}{\partial S}\right)_V = -\frac{\partial^2 V}{\partial V\,\partial S}.$$

Отсюда на основании известной теоремы анализа о перемене порядка дифференцирования следует:

$$\left(\frac{\partial T}{\partial V}\right)_S = -\left(\frac{\partial P}{\partial S}\right)_V. \tag{45.15}$$

Аналогично,

$$\left(\frac{\partial T}{\partial P}\right)_S = \left(\frac{\partial V}{\partial S}\right)_P, \tag{45.16}$$

$$\left(\frac{\partial S}{\partial V}\right)_T = \left(\frac{\partial P}{\partial T}\right)_V, \tag{45.17}$$

$$\left(\frac{\partial S}{\partial P}\right)_T = -\left(\frac{\partial V}{\partial T}\right)_P. \tag{45.18}$$

Эти и подобные им соотношения называются *соотношениями взаимности*, или *соотношениями Максвелла*. Они постоянно используются для вывода различных соотношений между величинами, характеризующими термодинамически равновесные состояния системы. Такой метод вывода называется *методом термодинамических функций*, или *термодинамических потенциалов*, в отличие от *метода циклов*, о котором говорилось в § 34. Поясним его на двух примерах.

Пример 1. Рассмотрим бесконечно малый квазистатический изотермический процесс. Поделив соотношение (45.2) на dV, найдем

$$\left(\frac{\partial U}{\partial V}\right)_T = T\left(\frac{\partial S}{\partial V}\right)_T - P,$$

или на основании соотношения (45.17)

$$\left(\frac{\partial U}{\partial V}\right)_T = T\left(\frac{\partial P}{\partial T}\right)_V - P. \tag{45.19}$$

Пример 2. Для такого же процесса делением на dP из формулы (45.3) получаем

$$\left(\frac{\partial I}{\partial P}\right)_T = T\left(\frac{\partial S}{\partial P}\right)_T + V,$$

или на основании соотношения (45.18)

$$\left(\frac{\partial I}{\partial P}\right)_T = V - T\left(\frac{\partial V}{\partial T}\right)_P. \tag{45.20}$$

Соотношения (45.19) и (45.20) были получены в § 34 методом циклов. Формально метод термодинамических функций проще метода циклов.

3. Термодинамические функции I, Ψ и Φ были введены нами для систем с *двумя степенями свободы*, т. е. таких систем, внутреннее состояние которых определяется двумя параметрами. Все сказанное нетрудно обобщить и на *системы со многими степенями свободы*. Надо только во всех соотношениях выражение для элементарной работы $\delta A = P\,dV$ заменить выражением (12.5). В результате получаются следующие определения:

$$I = U + \sum A_i a_i \quad \text{(энтальпия)}, \tag{45.21}$$

$$\Psi = U - TS \quad \text{(свободная энергия)}, \tag{45.22}$$

$$\Phi = \Psi + \sum A_i a_i \quad \text{(термодинамический потенциал)}. \tag{45.23}$$

Для дифференциалов соответствующих функций получаем

$$dU = T\,dS - \sum A_i\,da_i, \tag{45.24}$$

$$dI = T\,dS + \sum a_i\,dA_i, \tag{45.25}$$

$$d\Psi = -S\,dT - \sum A_i\,da_i, \tag{45.26}$$

$$d\Phi = -S\,dT + \sum a_i\,dA_i. \tag{45.27}$$

4. Если число частиц N в системе может изменяться, то в формуле (45.2) надо добавить член $\mu^*\,dN$, учитывающий изменение внутренней энергии газа за счет изменения числа частиц. Вместо (45.2) следует писать

$$dU = T\,dS - P\,dV + \mu^* dN. \tag{45.28}$$

Такой же член добавится в правых частях формул (45.3), (45.6), (45.7). Величина μ^* в термодинамике называется *химическим потенциалом*. Из этого определения следует

$$\mu^* = \left(\frac{\partial U}{\partial N}\right)_{V,S} = \left(\frac{\partial \Psi}{\partial N}\right)_{T,V} = \left(\frac{\partial \Phi}{\partial N}\right)_{T,P} = \left(\frac{\partial I}{\partial N}\right)_{P,S}. \tag{45.29}$$

Все термодинамические величины можно разделить на *интенсивные и экстенсивные. Интенсивные величины — это такие, которые зависят только от внутреннего состояния тел, но не от их размеров*. К ним относятся, например, температура и давление. *Экстенсивные величины — это такие, которые изменяются пропорционально массе системы, если при этом внутреннее состояние ее не меняется*. Примерами таких величин являются внутренняя и свободная энергия, энтропия, термодинамический потенциал и пр. Найдем общий вид зависимости некоторых из этих величин от числа N частиц в системе. Начнем с термодинамического потенциала Φ. Для систем с переменным числом частиц он является функцией T, P и N, т. е. $\Phi = \Phi(T, P, N)$. Сохраняя T и P неизменными, увеличим число частиц

в α раз. Тогда Φ возрастает в такое же число раз, а потому $\alpha\Phi = \Phi(T, P, \alpha N)$. Выберем теперь α так, чтобы $\alpha N = 1$, т. е. $\alpha = 1/N$. Тогда

$$\Phi = N\Phi(T, P, 1). \tag{45.30}$$

Отсюда

$$\mu^* = \left(\frac{\partial \Phi}{\partial N}\right)_{T,P} = \Phi(T, P, 1), \tag{45.31}$$

$$\Phi = \mu^*(T, P)N. \tag{45.32}$$

Химический потенциал μ^* может быть, таким образом, *истолкован как термодинамический потенциал, отнесенный к одной частице.*

Через остальные термодинамические функции химический потенциал не может быть истолкован столь просто. Например, для свободной энергии можно написать $\Psi = \Psi(T, V, N)$. При увеличении N в α раз в такое же число раз увеличивается не только Ψ, но и V, т. е. $\alpha\Psi = \Psi(T, \alpha V, \alpha N)$. Полагая снова $\alpha = 1/N$, получаем

$$\Psi = N\Psi(T, V/N, 1). \tag{45.33}$$

Число частиц N стоит не только в качестве множителя, но и под знаком функции $\Psi(T, V/N, 1)$. Поэтому $\mu^* = (\partial\Psi/\partial N)_{T,V} \neq \neq \Psi(T, V/N, 1)$.

Термодинамическое определение μ^* не однозначно. Величины U и S определены с точностью до произвольных аддитивных постоянных U_0 и S_0, а, следовательно, Ψ и Φ — с точностью до линейной функции температуры $U_0 - S_0 T$. Если бы U_0 и S_0 не зависели от числа частиц N, то их наличие никак не сказалось бы на величине μ^*. Но U_0 и S_0 могут зависеть от N. В этом случае в выражении для μ^* появится в качестве слагаемого произвольная линейная функция температуры. Для однозначности определения надо фиксировать начала отсчета энергии и энтропии.

§ 46. Термодинамическая теория эффекта Джоуля–Томсона

1. Об эффекте Джоуля–Томсона мы говорили в § 19, но пользовались только первым началом термодинамики. Учтем теперь и второе начало. Пусть по разные стороны пробки в опыте Джоуля–Томсона (см. рис. 18) поддерживается малая разность давлений ΔP. Соответствующая ей разность температур ΔT измеряется в опыте. Течение газа предполагается установившимся. Задача теории состоит в том, чтобы, зная ΔP и уравнение состояния газа, вычислить ΔT. Как было показано в § 19, при прохождении через пробку энтальпия газа I не меняется: $\Delta I = 0$. Пренебрегая величинами высшего порядка малости, можно написать

$$\Delta I = \left(\frac{\partial I}{\partial T}\right)_P \Delta T + \left(\frac{\partial I}{\partial P}\right)_T \Delta P = 0.$$

Так как $(\partial I/\partial T)_P = C_P$, а $(\partial I/\partial P)_T$ определяется выражением (45.20), то отсюда находим

$$\left(\frac{\Delta T}{\Delta P}\right)_I = \frac{T(\partial V/\partial T)_P - V}{C_P}. \tag{46.1}$$

Индекс I в левой части, как всегда, указывает, что в ходе процесса энтальпия I не меняется.

Если газ идеальный, то $V = RT/P$, $T(\partial V/\partial T)_P = V$, а потому $\Delta T = 0$. Таким образом, *для идеальных газов эффект Джоуля–Томсона не имеет места*. Но для реальных газов, вообще говоря, происходит либо нагревание, либо охлаждение. Этот вопрос мы подробно разберем в § 104 после того, как в нашем распоряжении будет приближенное уравнение состояния реальных газов (уравнение Ван-дер-Ваальса).

2. Повышение или понижение температуры реального газа при стационарном течении его через пробку называется *дифференциальным эффектом Джоуля–Томсона*. Этим названием хотят подчеркнуть, что величины ΔT и ΔP, входящие в формулу (46.1), являются малыми, так что их отношение можно заменить частной производной $(\partial T/\partial P)_I$. Наряду с дифференциальным эффектом различают также *интегральный эффект Джоуля–Томсона*. В этом случае разность давлений, проходимая текущим газом, велика (обычно десятки и даже сотни атмосфер). Изменения температуры также значительны. Поэтому интегральный эффект применяется в технике для получения низких температур (см. § 105).

Интегральный эффект осуществляется путем *дросселирования газа*, т. е. протеканием его через вентиль (малое отверстие), по разные стороны которого поддерживается большая разность давлений (см. § 25). В этом случае также сохраняется энтальпия, как это было показано в § 25. Дросселирование не является равновесным процессом. Однако начальное и конечное состояния газа являются равновесными. Они полностью определяются заданием энтальпии и давления. При вычислении изменения температуры реальный процесс можно заменить квазистатическим процессом, происходящим при постоянной энтальпии, т. е. рассуждать так, как если бы интегральный эффект Джоуля–Томсона состоял из непрерывной последовательности дифференциальных эффектов. Таким путем получаем выражение

$$T_2 - T_1 = \int_{P_1}^{P_2} \left(\frac{\partial T}{\partial P}\right)_I dP = \int_{P_1}^{P_2} \frac{T(\partial V/\partial T)_P - V}{C_P}\, dP. \tag{46.2}$$

Разумеется, такая замена реального процесса квазистатическим годится только для вычисления параметров газа в конечном состоянии. Результат получается правильным не потому, что оба процесса тождественны, а потому, что они приводят *к одному и тому же* конечному состоянию.

3. Вильям Томсон использовал дифференциальный эффект Джоуля–Томсона для градуировки термометра в абсолютной термодинамической шкале. Допустим, что проводится опыт Джоуля–Томсона с каким-либо реальным газом. Температура газа измеряется с помощью произвольного термометра — эту эмпирическую температуру будем обозначать буквой τ. Задача сводится к нахождению функции $T = T(\tau)$. Очевидно

$$\frac{\Delta T}{\Delta P} = \frac{\Delta T}{\Delta \tau}\,\frac{\Delta \tau}{\Delta P},$$

или, заменяя конечные приращения дифференциалами,

$$\frac{\Delta T}{\Delta P} = \frac{dT}{d\tau}\,\frac{\Delta \tau}{\Delta P}.$$

Далее,

$$\left(\frac{\partial V}{\partial T}\right)_P = \left(\frac{\partial V}{\partial \tau}\right)_P \frac{d\tau}{dT},$$

$$C_P = \left(\frac{\delta Q}{dT}\right)_P = \left(\frac{\delta Q}{d\tau}\right)_P \frac{d\tau}{dT} = C'_P\,\frac{d\tau}{dT},$$

где C'_P — теплоемкость при постоянном давлении, измеренная в эмпирической температурной шкале, т. е.

$$C'_P = (\delta Q/d\tau)_P.$$

Формула (45.1) переходит в

$$\frac{dT}{d\tau}\,\frac{\Delta \tau}{\Delta P} = \frac{T\left(\dfrac{\partial V}{\partial \tau}\right)_P \dfrac{d\tau}{dT} - V}{C'_P\,\dfrac{d\tau}{dT}}, \quad \text{откуда} \quad \frac{dT}{T} = \frac{\dfrac{1}{V}\left(\dfrac{\partial V}{\partial \tau}\right)_P}{1 + \dfrac{C'_P}{V}\,\dfrac{\Delta \tau}{\Delta P}},$$

или после интегрирования

$$T = T_0 \exp \int_{\tau_0}^{\tau} \frac{\dfrac{1}{V}\left(\dfrac{\partial V}{\partial \tau}\right)_P}{1 + \dfrac{C'_P}{V}\,\dfrac{\Delta \tau}{\Delta P}}\, d\tau. \qquad (46.3)$$

Здесь T_0 и τ_0 — температуры какой-либо одной реперной точки по абсолютной и эмпирической температурным шкалам соответственно. Все величины, стоящие в правой части под знаком интеграла, могут быть измерены экспериментально, так как для этого требуется измерять температуру лишь по эмпирической шкале. Поэтому интеграл может быть вычислен как функция эмпирической температуры τ. Тем самым будет установлена функциональная связь $T = T(\tau)$.

Удобно в качестве эмпирического термометра взять газовый термометр, наполненный тем же газом, с которым производится опыт Джоуля–Томсона. В этом случае по определению газовой абсолютной температуры $PV = A\tau$, где A — постоянная. Следовательно, $\dfrac{1}{V}\left(\dfrac{\partial V}{\partial \tau}\right)_P = \dfrac{A}{PV} = \dfrac{1}{\tau}$. Кроме того, величина $\dfrac{C'_P}{V}\,\dfrac{\Delta \tau}{\Delta P}$ мала по сравнению с единицей, так что квадратом ее можно пренебречь. В этом приближении интеграл, входящий в формулу

(46.3), переходит в

$$\int\limits_{\tau_0}^{\tau}\Big(1-\frac{C'_P}{V}\,\frac{\Delta\tau}{\Delta P}\Big)\frac{d\tau}{\tau}=\ln\frac{\tau}{\tau_0}-\int\limits_{\tau_0}^{\tau}\frac{PC'_P}{A\tau^2}\,\frac{\Delta\tau}{\Delta P}\,d\tau.$$

Интеграл в правой части является малой поправкой. В принятом нами приближении постоянную A, стоящую под знаком этого интеграла, можно заменить на универсальную газовую постоянную R. Это дает

$$T=\frac{T_0}{\tau_0}\,\tau\,\exp\int\limits_{\tau_0}^{\tau}\frac{PC'_P}{R\tau^2}\,\frac{\Delta\tau}{\Delta P}\,d\tau.$$

Входящий сюда интеграл также мал по сравнению с единицей, а потому приближенно

$$T=\frac{T_0}{\tau_0}\,\tau\Big[1+\int\limits_{\tau_0}^{\tau}\frac{PC'_P}{R\tau^2}\,\frac{\Delta\tau}{\Delta P}\,d\tau\Big]. \tag{46.4}$$

При вычислениях по этой формуле за исходную следует взять температуру тройной точки воды $T_0=\tau_0=273{,}16$ К. После этого по формуле (46.4) можно вычислить поправку к показаниям газового термометра для любой температуры τ. При этом для гелия в широком температурном интервале $C'_P/R\approx 5/3$ (1,67).

ЗАДАЧИ

1. Показать, что в процессе Джоуля–Томсона энтропия газа увеличивается.

Р е ш е н и е. Для вычисления изменения энтропии газа заменим реальный процесс Джоуля–Томсона квазистатическим изэнтальпическим процессом, переводящим систему в то же конечное состояние. Для такого процесса $dI=T\,dS+V\,dP=0$, а потому

$$\Big(\frac{\partial S}{\partial P}\Big)_I=-\frac{V}{T}<0.$$

Учитывая, что давление в процессе Джоуля–Томсона понижается, заключаем отсюда, что энтропия S возрастает.

2. Сосуд с твердыми адиабатическими стенками разделен на две части твердой адиабатической перегородкой. По одну сторону перегородки находится газ, по другую — вакуум. Вывести общую терминологическую формулу для температуры газа, которая установится в нем после удаления перегородки. Применить полученную формулу к идеальному газу и показать, что в этом случае изменения температуры не произойдет.

Р е ш е н и е. Так как над газом не производится работа и теплота не подводится, то после удаления перегородки и установления равновесия внутренняя энергия газа не изменится. Реальный процесс, совершаемый газом, является неравновесным и очень сложным. Однако начальное и конечное состояния равновесны, а температура газа в равновесном состоянии определяется двумя параметрами, за которые удобно взять внутреннюю энергию и объем газа. При вычислении изменения температуры реальный процесс можно заменить квазистатическим процессом при постоянной внутренней

энергии. Для такого процесса

$$T_2 - T_1 = \int_{V_1}^{V_2} \left(\frac{\partial T}{\partial V}\right)_U dV. \tag{46.5}$$

Для вычисления частной производной, входящей в этот интеграл, надо дифференциал

$$dU = \left(\frac{\partial U}{\partial T}\right)_V dT + \left(\frac{\partial U}{\partial V}\right)_T dV$$

положить равным нулю. Если еще воспользоваться формулами (18.3) и (34.2), то получится

$$\left(\frac{\partial T}{\partial V}\right)_U = \frac{P - T(\partial P/\partial T)_V}{C_V}. \tag{46.6}$$

Окончательно

$$T_2 - T_1 = \int_{V_1}^{V_2} \frac{P - T(\partial P/\partial T)_V}{C_V}\, dV. \tag{46.7}$$

Для идеального газа эта формула дает $T_2 - T_1 = 0$.

§ 47. Общие замечания о методе термодинамических функций. Примеры

Метод термодинамических функций, в сущности, основан на том, что если некоторая величина f, характеризующая состояние системы при термодинамическом равновесии, есть функция других величин x и y, а ее дифференциал представлен в виде

$$df = X(x, y)\, dx + Y(x, y)\, dy,$$

то

$$X = \left(\frac{\partial f}{\partial x}\right)_y, \quad Y = \left(\frac{\partial f}{\partial y}\right)_x$$

и, следовательно,

$$\left(\frac{\partial Y}{\partial x}\right)_y = \left(\frac{\partial X}{\partial y}\right)_x. \tag{47.1}$$

Из соотношений типа (47.1) и выводятся различные термодинамические равенства. Именно так были получены соотношения (45.19) и (45.20).

Применяя этот метод, необходимо, однако, убедиться, что выражение $X\, dx + Y\, dy$ является именно дифференциалом (полным дифференциалом) функции состояния $f(x, y)$, а не просто какой-то бесконечно малой величиной. Иначе можно прийти к ошибочным выводам. Приведем один пример неправильного рассуждения такого рода. Допустим, что элементарное количество теплоты δQ мы стали рассматривать как дифференциал некоторой функции состояния $Q = Q(T, P)$. По первому началу этот дифференциал

$$dQ = dU + P\, dV = dI - V\, dP,$$

или

$$dQ = \left(\frac{\partial I}{\partial T}\right)_P dT + \left[\left(\frac{\partial I}{\partial P}\right)_T - V\right] dP.$$

Применяя к этому выражению соотношение (47.1), получим

$$\frac{\partial^2 I}{\partial T \partial P} = \frac{\partial}{\partial T}\left[\left(\frac{\partial I}{\partial P}\right)_T - V\right], \quad \text{откуда} \quad \left(\frac{\partial V}{\partial T}\right)_P = 0.$$

Отсюда заключаем, что тепловое расширение тел невозможно, а это находится в резком противоречии с опытом. Противоречие получилось потому, что теплоту Q мы *неправильно* рассматривали как функцию состояния системы, а δQ — как дифференциал этой функции. На самом деле такой функции не существует. Для того чтобы предостеречь читателя от ошибок такого рода, бесконечно малые величины мы обозначали символом δ, если они не были полными дифференциалами функций состояния. Символ же d сохранен только для обозначения таких бесконечно малых величин, которые являются полными дифференциалами.

При выводе некоторых термодинамических теорем встречаются производные, не входящие в соотношения (45.15)–(45.18). В таком случае эти производные надо выразить через величины (45.15)–(45.18). Для примера найдем связь между адиабатической $(\partial V/\partial P)_S$ и изотермической $(\partial V/\partial P)_T$ сжимаемостями изотропного вещества. Используя формулы (45.15)–(45.18), находим

$$\left(\frac{\partial V}{\partial P}\right)_S = \left(\frac{\partial V}{\partial T}\right)_S \left(\frac{\partial T}{\partial P}\right)_S = -\left(\frac{\partial S}{\partial P}\right)_V \left(\frac{\partial V}{\partial S}\right)_P,$$

$$\left(\frac{\partial V}{\partial P}\right)_T = \left(\frac{\partial V}{\partial S}\right)_T \left(\frac{\partial S}{\partial P}\right)_T = -\left(\frac{\partial T}{\partial P}\right)_P \left(\frac{\partial V}{\partial T}\right)_P.$$

Отсюда почленным делением получается

$$\frac{(\partial V/\partial P)_S}{(\partial V/\partial P)_T} = \left[\left(\frac{\partial S}{\partial P}\right)_V \left(\frac{\partial P}{\partial T}\right)_V\right]\left[\left(\frac{\partial T}{\partial V}\right)_P \left(\frac{\partial V}{\partial S}\right)_P\right].$$

Дифференцирование в первой квадратной скобке производится при постоянном объеме V, а во второй — при постоянном давлении P. Поэтому на основании правила дифференцирования сложной функции

$$\frac{(\partial V/\partial P)_S}{(\partial V/\partial P)_T} = \left(\frac{\partial S}{\partial T}\right)_V \left(\frac{\partial T}{\partial S}\right)_P = \frac{C_V}{C_P} = \frac{1}{\gamma},$$

так как $T(\partial S/\partial T)_V = C_V$, $T(\partial S/\partial T)_P = C_P$. Этот результат был уже получен иным путем в задаче к § 23.

В заключение приведем еще несколько примеров на применение методов термодинамических функций.

1. Найдем связь между адиабатическими и изотермическими модулями объемного сжатия K_S и K_T физически однородного и изотропного вещества. По определению этих модулей

$$K_S = -V\left(\frac{\partial P}{\partial V}\right)_S, \quad K_T = -V\left(\frac{\partial P}{\partial V}\right)_T.$$

Преобразуем производную $(\partial P/\partial V)_S$. Величины P, V, T связаны уравнением состояния. Кроме того, в рассматриваемом случае между ними есть еще одно соотношение, выражающее адиабатичность процесса. В адиабатическом процессе переменные P, V, T, S могут рассматриваться как функции одной из них. Возьмем в качестве независимой переменной температуру T. Тогда

$$\Big(\frac{\partial P}{\partial V}\Big)_S = \Big(\frac{\partial P}{\partial T}\Big)_S \Big(\frac{\partial T}{\partial V}\Big)_S.$$

Ввиду уравнения состояния между величинами P, T, S в любом состоянии существует функциональная зависимость. То же относится к величинам T, V, S. Поэтому

$$\Big(\frac{\partial P}{\partial T}\Big)_S = -\Big(\frac{\partial P}{\partial S}\Big)_T \Big(\frac{\partial S}{\partial T}\Big)_P,$$

$$\Big(\frac{\partial T}{\partial V}\Big)_S = -\Big(\frac{\partial T}{\partial S}\Big)_V \Big(\frac{\partial S}{\partial V}\Big)_T$$

и далее

$$\Big(\frac{\partial P}{\partial V}\Big)_S = \Big[\Big(\frac{\partial S}{\partial T}\Big)_P \Big(\frac{\partial T}{\partial S}\Big)_V\Big] \cdot \Big[\Big(\frac{\partial P}{\partial S}\Big)_T \Big(\frac{\partial S}{\partial V}\Big)_T\Big].$$

Правую часть преобразуем с помощью соотношений

$$\Big(\frac{\partial S}{\partial T}\Big)_P \Big(\frac{\partial T}{\partial S}\Big)_V = \frac{T(\partial S/\partial T)_P}{T(\partial S/\partial T)_V} = \frac{(\delta Q/dT)_P}{(\delta Q/dT)_V} = \frac{C_P}{C_V},$$

$$\Big(\frac{\partial P}{\partial S}\Big)_T \Big(\frac{\partial S}{\partial V}\Big)_T = \Big(\frac{\partial P}{\partial V}\Big)_T.$$

В результате получим

$$\Big(\frac{\partial P}{\partial V}\Big)_S = \gamma \Big(\frac{\partial P}{\partial V}\Big)_T, \quad \text{или} \quad K_S = \gamma K_T, \tag{47.2}$$

где $\gamma = C_P/C_V$. Таким образом, адиабатический модуль всестороннего сжатия в γ раз больше изотермического модуля. Этот результат был уже получен иным способом в § 23 (см. задачу к § 23).

2. Выведем еще раз формулу для разности теплоемкостей $C_P - C_V$. По определению энтропии и теплоемкости

$$C_P = T\Big(\frac{\partial S}{\partial T}\Big)_P. \tag{47.3}$$

Рассматривая энтропию S как функцию температуры и объема, можем написать

$$dS = \Big(\frac{\partial S}{\partial T}\Big)_V dT + \Big(\frac{\partial S}{\partial V}\Big)_T dV.$$

Отсюда

$$\Big(\frac{\partial S}{\partial T}\Big)_P = \Big(\frac{\partial S}{\partial T}\Big)_V + \Big(\frac{\partial S}{\partial V}\Big)_T \Big(\frac{\partial V}{\partial T}\Big)_P = \frac{C_V}{T} + \Big(\frac{\partial P}{\partial T}\Big)_V \Big(\frac{\partial V}{\partial T}\Big)_P.$$

Таким образом,

$$C_P - C_V = T\Big(\frac{\partial P}{\partial T}\Big)_V \Big(\frac{\partial V}{\partial T}\Big)_P. \tag{47.4}$$

Приведем еще без подробных доказательств несколько полезных термодинамических соотношений:

$$\left(\frac{\partial C_V}{\partial V}\right)_T = T\left(\frac{\partial^2 P}{\partial T^2}\right)_V, \tag{47.5}$$

$$\left(\frac{\partial C_P}{\partial P}\right)_T = -T\left(\frac{\partial^2 V}{\partial T^2}\right)_P, \tag{47.6}$$

$$\left(\frac{\partial U}{\partial P}\right)_T = T\left(\frac{\partial S}{\partial P}\right)_T - P\left(\frac{\partial V}{\partial P}\right)_T = -T\left(\frac{\partial V}{\partial T}\right)_P - P\left(\frac{\partial V}{\partial P}\right)_T, \tag{47.7}$$

$$\left(\frac{\partial I}{\partial V}\right)_T = T\left(\frac{\partial S}{\partial V}\right)_T + V\left(\frac{\partial P}{\partial V}\right)_T = T\left(\frac{\partial P}{\partial T}\right)_V + V\left(\frac{\partial P}{\partial V}\right)_T, \tag{47.8}$$

$$\left(\frac{\partial U}{\partial T}\right) = C_P - P\left(\frac{\partial V}{\partial T}\right)_P, \tag{47.9}$$

$$\left(\frac{\partial I}{\partial T}\right)_V = C_V + V\left(\frac{\partial P}{\partial T}\right)_V, \tag{47.10}$$

$$\left(\frac{\partial T}{\partial V}\right)_S = -\left(\frac{\partial T}{\partial S}\right)_V\left(\frac{\partial S}{\partial V}\right)_T = -\frac{T}{C_V}\left(\frac{\partial S}{\partial V}\right)_T, \tag{47.11}$$

$$\left(\frac{\partial T}{\partial P}\right)_S = -\left(\frac{\partial T}{\partial S}\right)_P\left(\frac{\partial S}{\partial P}\right)_T = -\frac{T}{C_P}\left(\frac{\partial S}{\partial P}\right)_T, \tag{47.12}$$

$$\left(\frac{\partial V}{\partial T}\right)_P = -\left(\frac{\partial S}{\partial P}\right)_T = \left(\frac{\partial S}{\partial T}\right)_P\left(\frac{\partial T}{\partial P}\right)_S = \frac{C_P}{T}\left(\frac{\partial T}{\partial P}\right)_S = -\frac{C_P}{T}\left(\frac{\partial T}{\partial V}\right)_S\left(\frac{\partial V}{\partial P}\right)_S, \tag{47.13}$$

$$\left(\frac{\partial V}{\partial P}\right)_S = \left(\frac{\partial V}{\partial P}\right)_T + \left(\frac{\partial V}{\partial T}\right)_P\left(\frac{\partial T}{\partial P}\right)_S = \left(\frac{\partial V}{\partial P}\right)_T + \left(\frac{\partial V}{\partial T}\right)_P\left(\frac{\partial V}{\partial S}\right)_P = \left(\frac{\partial V}{\partial P}\right)_T + \frac{T}{C_P}\left(\frac{\partial V}{\partial T}\right)_P^2, \tag{47.14}$$

$$\left(\frac{\partial P}{\partial V}\right)_S = \left(\frac{\partial P}{\partial V}\right)_T - \frac{T}{C_V}\left(\frac{\partial P}{\partial T}\right)_V^2. \tag{47.15}$$

§ 48. Максимальная работа и свободная энергия

1. Рассмотрим какую-либо термодинамическую систему в произвольном, вообще говоря, неравновесном состоянии. Пусть она граничит со средой, температура T_0 которой поддерживается постоянной. Система может обмениваться теплотой только с этой средой. С остальными телами теплообмена нет. Что касается работы, производимой системой, то в общем случае она слагается из двух частей: из работы над рассматриваемой средой и над всеми остальными телами. Эту общую работу, как обычно, будем обозначать буквой A. Пусть система переходит из состояния *1* в состояние *2*. В соответствии с первым началом термодинамики $A = U_1 - U_2 + Q$. Работа A и количество сообщенной теплоты Q зависят от вида процесса, переводящего систему из состояния *1* в

состояние *2*. Второе начало термодинамики позволяет указать верхний предел для работы A. Согласно неравенству Клаузиуса

$$S_2 - S_1 \geqslant \int\limits_{1\to 2} \frac{\delta Q}{T_0},$$

или ввиду предполагаемого постоянства температуры T_0

$$S_2 - S_1 \geqslant \frac{Q}{T_0}.$$

Подставив сюда $Q = A - U_1 + U_2$ и введя обозначение

$$Y = U - T_0 S, \tag{48.1}$$

получим

$$A \leqslant Y_1 - Y_2. \tag{48.2}$$

Знак равенства относится к обратимым процессам. Таким образом, работа, которую может совершить система, не может превзойти убыли функции $Y \equiv U - T_0 S$.

В частности, когда температура системы T все время равна температуре окружающей среды T_0, функция Y переходит в свободную энергию системы $Y = \Psi = U - TS$. В этом случае $A \leqslant \Psi_1 - \Psi_2$. Знак равенства относится к обратимым процессам. Таким образом, для обратимого процесса $T = T_0$,

$$A_{\text{макс}} = \Psi_1 - \Psi_2. \tag{48.3}$$

Если воспользоваться уравнением Гиббса–Гельмгольца (45.13), то выражение (48.3) легко преобразовать к виду

$$A_{\text{макс}} = U_1 - U_2 + T\Big(\frac{\partial(\Psi_1 - \Psi_2)}{\partial T}\Big)_V,$$

или

$$A_{\text{макс}} = U_1 - U_2 + T\Big(\frac{\partial A_{\text{макс}}}{\partial T}\Big)_V. \tag{48.4}$$

Эта формула называется *уравнением Гиббса–Гельмгольца для максимальной работы*. Она имеет многочисленные применения.

2. Из изложенного становятся ясными мотивы, которыми руководствовался Гельмгольц, назвавший величину $\Psi = U - TS$ свободной энергией системы. Величина U есть внутренняя, или полная, энергия системы. Но если система находится в тепловом контакте со средой, температура которой T поддерживается постоянной, то только часть этой энергии, а именно $U - TS$, может быть использована для получения работы. В этом смысле она и является свободной. Оставшаяся часть при сохранении неизменной температуры среды, с которой установлен тепловой контакт, в работу превращена быть не может. Она называется *связанной энергией*. В предельном случае $T \to 0$ различие между внутренней и свободной энергиями пропадает.

3. Наряду с полной работой A в термодинамике часто рассматривают так называемую *полезную работу*. Хотя это понятие для физических приложений и не необходимо, приведем нужные разъяснения. Понятие полезной работы вводится в тех случаях, когда рассматриваемая (вообще говоря, неравновесная) термодинамическая система помещена в среду, находящуюся в равновесии и поддерживаемую при постоянной температуре T_0 и постоянном давлении P_0. Предполагается, что система может производить работу не только против давления среды P_0, но и над другими телами. Эта последняя составная часть работы и называется полезной работой. Работа, производимая против давления среды P_0, представляется выражением $P_0(V_2-V_1)$. Эту часть нужно вычесть из полной работы A, чтобы получить полезную работу $A^{\text{полезн}}$. Для нее из формулы (48.2) получаем

$$A^{\text{полезн}} \leqslant Z_1 - Z_2, \tag{48.5}$$

где

$$Z = Y + P_0V = U - T_0S + P_0V. \tag{48.6}$$

В частности, когда температура и давление среды равны температуре и давлению самой системы, функция Z переходит в термодинамический потенциал $Z = \Phi = U - TS + PV$. В этом случае

$$A^{\text{полезн}} \leqslant \Phi_1 - \Phi_2. \tag{48.7}$$

Максимальная полезная работа получается при обратном процессе и равна $A^{\text{полезн}}_{\text{макс}} = \Phi_1 - \Phi_2$. Для этой работы по аналогии с формулой (48.4) получаем *вторую формулу Гиббса–Гельмгольца*

$$A^{\text{полезн}}_{\text{макс}} = I_1 - I_1 + T\Big(\frac{\partial A^{\text{полезн}}_{\text{макс}}}{\partial T}\Big)_P. \tag{48.8}$$

Когда термодинамическая система состоит только из твердых и жидких тел, изменением ее объема при всех процессах, как правило, можно пренебречь. В этих случаях различие между полной работой A и работой $A^{\text{полезн}}$ пропадает. Для газообразных систем, напротив, это различие может быть существенным.

ЗАДАЧА

Используя результаты этого параграфа, дать другое решение задачи 2 § 37.

§ 49. Электродвижущая сила гальванического элемента

1. В качестве примера применения термодинамики к электрическим явлениям рассмотрим вопрос об электродвижущей силе гальванического элемента.

Разумеется, термодинамика не может ответить на вопрос, как и почему в гальваническом элементе возникает электрический ток. Опираясь на опыт, она констатирует лишь, что гальванический элемент есть неравновесная термодинамическая система, приближающаяся к состоянию равновесия с

возникновением электрического тока. Этого оказывается достаточно для того, чтобы термодинамическими средствами установить количественное соотношение между электродвижущей силой гальванического элемента и некоторыми другими его характеристиками.

Будем считать, что процесс приближения к состоянию равновесия происходит *квазистатически*. Как и всякий квазистатический процесс, этот процесс является обратимым в узком смысле слова. Изменение направления протекающего тока вызывает химические реакции, противоположные тем, которые происходят в элементе при нормальном направлении тока. Если это условие выполнено, то гальванический элемент называют *обратимым*. Такой элемент всегда можно привести к начальному состоянию, изменив в нем направление тока на обратное.

Для обратимости гальванического элемента необходимо, чтобы джоулево тепло, выделяющееся в нем, было пренебрежимо мало. Поэтому в дальнейшем предполагается, что ток I, протекающий через элемент, сколь угодно мал, т. е. полное сопротивление цепи бесконечно велико. Это не накладывает никаких ограничений на значение проходящего заряда q, так как время прохождения электрического тока может быть сколь угодно большим.

При таких условиях можно полностью пренебречь джоулевым количеством теплоты по сравнению с работой, совершаемой электродвижущей силой $\mathscr{E}$ элемента. Действительно, работа электродвижущей силы в течение времени t равна $\mathscr{E}It$, тогда как джоулева теплота, выделяющаяся внутри элемента за то же время, равна R_iI^2t, где R_i — внутреннее сопротивление элемента, т. е. величина конечная. Если силу тока рассматривать как бесконечно малую первого порядка, то работа элемента будет также первого порядка малости, тогда как джоулева теплота — второго порядка малости. Ясно поэтому, что при $I \to 0$ джоулевой теплотой по сравнению с работой можно пренебречь.

2. В основу термодинамической теории гальванического элемента можно положить различные термодинамические соотношения. Мы будем предполагать, что элемент — *электролитический*, т. е. состоит из одних только твердых и жидких тел. Газовые элементы рассматривать не будем. Таким образом, наша система механической работы в обычном смысле этого слова, т. е. работы по перемещению макроскопических тел, не совершает. Вся работа системы сводится к работе электродвижущей силы элемента, поддерживающей в цепи электрический ток. При прохождении через цепь заряда dq гальванический элемент совершает электрическую работу $\delta A = \mathscr{E}\,dq$. Элемент представляет собой термодинамическую систему с *одной степенью свободы*, в которой роль обобщенной координаты a_i играет заряд q, а обобщенной силы — электродвижущая сила $\mathscr{E}$. Поэтому, чтобы прийти к искомому соотношению, достаточно в уравнении (34.9) положить $a_i = q$, $A_i = \mathscr{E}$. При этом можно считать, что электродвижущая сила $\mathscr{E}$ может зависеть лишь от температуры электролита, так как в обычных условиях внешнее давление и объем электролита в элементе практически остаются постоянными. Имея это в виду, можно написать

$$\frac{\partial U}{\partial q} = T\frac{d\mathscr{E}}{dT} - \mathscr{E}, \tag{49.1}$$

или

$$\mathscr{E} - T\frac{d\mathscr{E}}{dt} = u, \tag{49.2}$$

где $u = -\partial U/\partial q$ есть уменьшение внутренней энергии элемента при прохождении через него единицы количества электричества. Уравнение (49.2), установленное Гельмгольцем в 1882 г., и решает поставленную задачу. Его можно переписать в виде

$$T^2 \frac{d}{dt}\left(\frac{\mathscr{E}}{T}\right) = -u. \tag{49.3}$$

Отсюда интегрированием находим

$$\mathscr{E}(T) = -T \int_{T_0}^{T} \frac{u(T)}{T^2}\, dT + \mathscr{E}(T_0). \tag{49.4}$$

Эта формула позволяет вычислить электродвижущую силу гальванического элемента при любой температуре T, если известно ее значение при какой-либо одной температуре T_0.

3. Вильям Томсон в 1851 г. дал иную формулу для электродвижущей силы обратимого гальванического элемента. Рассуждения его основывались на законе сохранения энергии. При прохождении единичного заряда гальванический элемент совершает работу $\mathscr{E}$. Работа совершается за счет убыли внутренней энергии элемента. Это приводит к формуле Томсона $\mathscr{E} = u$. Сравнение с формулой Гельмгольца (49.2) показывает, что формула Томсона дает верный результат только в том случае, когда электродвижущаяся сила $\mathscr{E}$ не зависит от температуры. В общем случае равенство $\mathscr{E} = u$ не имеет места. Выясним, в чем недостаточность рассуждений Томсона, и попутно дадим новый вывод формулы Гельмгольца (49.2). Элементарная работа гальванического элемента при прохождении бесконечно малого заряда dq всегда дается выражением $\mathscr{E}\, dq$. Но если через элемент проходит конечный заряд q (например, $q = 1$), то работа может быть представлена произведением $\mathscr{E}q$ только в том случае, когда значение $\mathscr{E}$ *поддерживается постоянным*. Поскольку $\mathscr{E}$ зависит от T, для этого необходимо поддерживать температуру T постоянной, т. е. подводить или отводить теплоту. Это количество теплоты и не было учтено Томсоном.

Будем поддерживать температуру T постоянной. Тогда работа при прохождении единичного заряда будет равна $\mathscr{E}$. С другой стороны, в обратимом изотермическом процессе та же работа равна убыли свободной энергии системы. В соответствии с формулой (48.3) это означает, что $\mathscr{E} = A_{\text{макс}}$, и формула (49.2) получится из формулы (48.4), если в последней положить $A_{\text{макс}} = \mathscr{E}$, $U_1 - U_2 = u$.

4. Уравнение (49.2) можно записать в другом виде, отнеся уменьшение внутренней энергии не к единице прошедшего электричества, а к одному молю вещества, выделившегося на одном из электродов. Обозначим это уменьшение внутренней энергии через $u_{\text{моль}}$. Связь между u и $u_{\text{моль}}$ легко установить с помощью *закона электролиза Фарадея* (1791–1867). Согласно этому закону число молей вещества, выделившегося на одном из электродов при прохождении заряда q, определяется выражением

$$\nu = \frac{q}{nF}, \tag{49.5}$$

где $F \approx 96500$ Кл/моль — универсальная постоянная (постоянная Фарадея), а n — валентность. Отсюда следует, что единичный заряд $q = 1$ выделяет на

электроде $1/nF$ молей. Внутренняя энергия при этом уменьшается на u. Поэтому при выделении одного моля вещества уменьшение внутренней энергии будет в nF раз больше, т. е. $u_{\text{моль}} = nFu$. С помощью этого соотношения уравнение (49.2) преобразуется в

$$\mathscr{E} - T\frac{d\mathscr{E}}{dT} = \frac{1}{nF}\,u_{\text{моль}}. \tag{49.6}$$

§ 50. Общие критерии термодинамической устойчивости

Допустим, что адиабатически изолированная система находится в термодинамическом равновесии, причем ее энтропия S в рассматриваемом состоянии максимальна, т. е. больше энтропий всех возможных бесконечно близких состояний, в которые система может перейти без подвода или отвода тепла. Тогда можно утверждать, что самопроизвольный адиабатический переход системы во все эти состояния невозможен, т. е. система находится в *устойчивом термодинамическом равновесии*. Действительно, если бы такой переход был возможен, то энтропии начального 1 и конечного 2 состояний были бы связаны соотношением $S_1 > S_2$. Но это соотношение находится в противоречии с принципом возрастания энтропии, согласно которому при адиабатических переходах должно быть $S_1 \leqslant S_2$. Таким образом, мы приходим к следующему критерию термодинамической устойчивости.

Если система адиабатически изолирована и ее энтропия в некотором равновесном состоянии максимальна, то это состояние термодинамически устойчиво. Это значит, что система, оставаясь адиабатически изолированной, не может самопроизвольно перейти ни в какое другое состояние.

В приложениях термодинамики к конкретным вопросам часто бывает удобно вместо адиабатической изоляции системы накладывать на ее поведение другие ограничения. Тогда критерии термодинамической устойчивости изменятся. Особенно удобны два критерия.

1. Пусть система окружена средой, температура которой поддерживается постоянной. Кроме того, объем системы V также поддерживается постоянным, например, система заключена в жесткую оболочку. В этих условиях работа системы A всегда равна нулю и соотношение (48.2) переходит в $Y_1 - Y_2 \geqslant 0$. Следовательно, функция $Y \equiv \equiv U - T_0 S$ может только уменьшаться или оставаться неизменной. Отсюда, рассуждая, как и раньше, получаем следующий критерий термодинамической устойчивости.

Если температура окружающей среды T_0 и объем системы V поддерживаются постоянными и в рассматриваемом состоянии функция $Y = U - T_0 S$ минимальна, то состояние системы термодинамически устойчиво. В частности, если температура среды равна температуре системы, роль функции Y выполняет свободная энергия $\Psi = U - TS$.

2. Допустим теперь, что система со всех сторон окружена средой, температура T_0 и давление P_0 которой поддерживаются постоянными. Никакой работы, помимо работы против внешнего давления P_0, система совершать не может. Иными словами, полезная работа системы всегда равна нулю, так что соотношение (48.5) дает $Z_2 \leqslant Z_1$. Все самопроизвольные процессы в системе могут идти только с уменьшением функции $Z \equiv Y + P_0V$. Поэтому *если функция Z в некотором равновесном состоянии достигла минимума, то равновесие будет устойчивым. В частности, когда $P = P_0$, это утверждение относится к термодинамическому потенциалу системы $\Phi = F + PV$.*

Приведем еще два, менее употребительные, условия термодинамической устойчивости. В них роль потенциальных функций выполняют внутренняя энергия U и энтальпия I.

3. Перепишем неравенство Клаузиуса (41.1) в виде

$$S_2 - S_1 \geqslant \int \frac{dU + \delta A}{T}.$$

Пусть энтропия и объем системы поддерживаются постоянными. Тогда $S_2 - S_1 = 0$ и $\delta A = P\,V = 0$, поэтому предыдущее неравенство дает

$$\int \frac{dU}{T} \leqslant 0.$$

Так как $T > 0$, то отсюда следует, что $dU \leqslant 0$. Если объем и энтропию системы поддерживать постоянными, то самопроизвольные процессы в ней могут идти лишь с уменьшением внутренней энергии. Если внутренняя энергия системы достигла минимума, то дальнейшие процессы в системе становятся невозможными. Это приводит к следующему критерию термодинамической устойчивости.

Если объем и энтропия системы поддерживаются постоянными и система в некотором равновесном состоянии достигла минимума внутренней энергии, то равновесие термодинамически устойчиво.

4. *Если давление и энтропия системы поддерживаются постоянными и система в некотором равновесном состоянии достигла минимума энтальпии, то равновесие термодинамически устойчиво.*

Для доказательства этого положения следует переписать неравенство Клаузиуса в виде

$$S_2 - S_1 \geqslant \int \frac{dI - V\,P}{T}$$

и повторить предыдущие рассуждения.

§ 51. Принцип Ле-Шателье–Брауна и устойчивость термодинамического равновесия

1. В заключение этой главы рассмотрим принцип, сформулированный французским ученым Ле-Шателье (1850–1936) в 1884 г. и, в расширенном виде, немецким физиком Брауном (1850–1918) в 1887 г.

Этот принцип позволяет предвидеть *направление течения процесса* в системе, когда она выведена внешним воздействием из состояния устойчивого равновесия. *Принцип Ле-Шателье–Брауна* не является столь всеобъемлющим, как второе начало термодинамики. В частности, он не позволяет высказывать никаких количественных заключений о поведении системы. Необходимым условием применимости принципа Ле-Шателье–Брауна является наличие *устойчивости равновесия*, из которого система выводится внешним воздействием. Он неприменим к процессам, переводящим систему в более устойчивое состояние, например, к взрывам. Принцип Ле-Шателье–Брауна был сформулирован как обобщение знаменитого и всем хорошо известного электродинамического *правила Ленца* (1804–1865), определяющего направление индукционного тока. Он гласит:

Если система находится в устойчивом равновесии, то всякий процесс, вызванный в ней внешним воздействием или другим первичным процессом, всегда бывает направлен таким образом, что он стремится уничтожить изменения, произведенные внешним воздействием или первичным процессом.

Ле-Шателье и Браун применяли главным образом индуктивный метод, рассмотрев большое число примеров, которые, по их мнению, являются частными случаями сформулированного ими общего правила. Данная ими формулировка была, однако, столь туманной, что не допускала в каждом конкретном случае однозначного применения правила. Неопределенность можно устранить и получить точные математические формулы, выражающие принцип Ле-Шателье–Брауна, если к рассматриваемой проблеме привлечь критерии устойчивости термодинамического равновесия, сформулированные в предыдущем параграфе.

2. Последующие результаты основаны на том, что устойчивость равновесия системы формулируется как условие *минимума или максимума* некоторой функции состояния, которую мы будем в дальнейшем обозначать через f. Эти результаты применимы поэтому не только к проблемам термодинамики, но и к проблемам механики или электродинамики, в которых устойчивость равновесия также связывается с минимумом или максимумом некоторых функций. При этом всегда можно пользоваться либо только условием минимума, либо только условием максимума. Действительно, если в положении равновесия функция f максимальна, то вместо нее можно взять функцию $-f$, которая будет уже минимальна. Условимся всегда так выбирать функцию f, чтобы в положении равновесия она была *минимальна*. Функция f должна зависеть от внутренних параметров, определяющих состояние системы. Часть из этих параметров может быть фиксирована, т. е. не должна меняться. Остальные параметры могут меняться в результате внешних процессов. Эти параметры мы будем называть *свободными* и обозначать буквами $x, y, z, \ldots$ В качестве функции f можно взять, например, величину Z, определяемую выражением (48.6). Если рассматриваемая система физически однородна и изотропна, то свободных параметров будет два. В качестве этих параметров можно взять, например, S и V. Но если система неоднородна, то ее внутрення энергия U может зависеть не только от S и V, но и от других параметров. Например, если система состоит из

двух фаз: жидкости и ее пара, то параметров будет три. В качестве третьего параметра можно взять, например, массу пара или массу жидкости.

3. Закрепим все свободные параметры, за исключением двух x и y, которым предоставим возможность изменяться. Тогда f может рассматриваться как функция только двух аргументов x и y. Разумеется, в положении равновесия она будет также минимальна, как и функция $f(x, y, z, \ldots)$ всех свободных параметров. Поэтому в этом положении ее частные производные первого порядка должны обращаться в нуль. Обозначив их через $X(x, y)$ и $Y(x, y)$, можем написать в положении равновесия

$$\Big(\frac{\partial f}{\partial x}\Big)_y \equiv X(x, y) = 0, \quad \Big(\frac{\partial f}{\partial y}\Big)_x \equiv Y(x, y) = 0. \tag{51.1}$$

Величины X и Y играют роль *обобщенных сил*, действующих в системе. При этом по свойством частных производных имеет место соотношение

$$\Big(\frac{\partial X}{\partial y}\Big)_x = \Big(\frac{\partial Y}{\partial x}\Big)_y, \tag{51.2}$$

выполняющееся при любых значениях x и y.

4. Соотношения (51.1) являются *необходимыми условиями равновесия*. Однако при их выполнении равновесие может быть и неустойчивым. Они могут соблюдаться и в точке максимума. *Условием устойчивости является минимум функции* f. Значит, в точке равновесия второй дифференциал

$$d^2 f = \frac{\partial^2 f}{\partial x^2}\, dx^2 + 2\,\frac{\partial^2 f}{\partial x\, \partial y}\, dx\, dy + \frac{\partial^2 f}{\partial y^2}\, dy^2 = \\ = \Big(\frac{\partial X}{\partial x}\Big)_y dx^2 + 2\Big(\frac{\partial X}{\partial y}\Big)_x dx\, dy + \Big(\frac{\partial Y}{\partial y}\Big)_x dy^2$$

должен быть положительным, каковы бы ни были бесконечно малые приращения аргументов dx и dy. Для этого в положении равновесия должны выполняться условия

$$\Big(\frac{\partial X}{\partial x}\Big)_y > 0, \quad \Big(\frac{\partial Y}{\partial y}\Big)_x > 0,$$

$$\begin{vmatrix} (\partial X/\partial x)_y & (\partial X/\partial y)_x \\ (\partial X/\partial y)_x & (\partial Y/\partial y)_x \end{vmatrix} \geqslant 0. \tag{51.3}$$

Эти три условия не независимы. Каждое из первых двух условий является следствием другого и последнего условия. Ввиду соотношения (51.2) последнему условию можно придать следующую, более симметричную форму:

$$\begin{vmatrix} (\partial X/\partial x)_y & (\partial X/\partial y)_x \\ (\partial Y/\partial x)_y & (\partial Y/\partial y)_x \end{vmatrix} \geqslant 0. \tag{51.4}$$

Разделив обе части этого неравенства на существенно положительную величину $(\partial X/\partial x)_y$ и раскрыв детерминант, придадим ему вид

$$\Big(\frac{\partial Y}{\partial y}\Big)_x - \Big(\frac{\partial Y}{\partial x}\Big)_y \Big(\frac{\partial x}{\partial X}\Big)_y \Big(\frac{\partial X}{\partial y}\Big)_x \geqslant 0.$$

Обобщенная сила X является функцией параметров x и y, т. е. величины X, x, y функционально связаны. Поэтому к ним применимо тождество (8.9),

которое дает

$$\left(\frac{\partial x}{\partial X}\right)_y\left(\frac{\partial X}{\partial y}\right)_x = -\left(\frac{\partial x}{\partial y}\right)_X, \quad \left(\frac{\partial Y}{\partial y}\right)_x + \left(\frac{\partial Y}{\partial x}\right)_y\left(\frac{\partial x}{\partial y}\right)_{X=0} \geqslant 0. \tag{51.5}$$

В левой части последнего неравенства стоит частная производная величины Y по y при постоянном X (точнее, при $X = 0$, поскольку соотношение относится к точке равновесия). Действительно, рассматривая Y сначала как функцию x, y, а затем как функцию X, y можем написать

$$dY = \left(\frac{\partial Y}{\partial x}\right)_y dx + \left(\frac{\partial Y}{\partial y}\right)_x dy = \left(\frac{\partial Y}{\partial X}\right)_y dX + \left(\frac{\partial Y}{\partial y}\right)_X dy.$$

Полагая здесь $X = \text{const}$, $dX = 0$ и поделив обе части равенства на dy, получим тождество

$$\left(\frac{\partial Y}{\partial y}\right)_X = \left(\frac{\partial Y}{\partial y}\right)_x + \left(\frac{\partial Y}{\partial x}\right)_y\left(\frac{\partial x}{\partial y}\right)_X. \tag{51.6}$$

Следовательно, третье условие принимает вид

$$\left(\frac{\partial Y}{\partial y}\right)_{X=0} \geqslant 0. \tag{51.7}$$

Аналогично

$$\left(\frac{\partial X}{\partial x}\right)_{Y=0} \geqslant 0. \tag{51.8}$$

5. Сравним теперь значения производных (51.7) и (51.8) с значениями производных (51.4) и (51.3). Подставив в (51.6) значение производной $(\partial x/\partial y)_X$ из (51.5), получим

$$\left(\frac{\partial Y}{\partial y}\right)_X = \left(\frac{\partial Y}{\partial y}\right)_x - \left(\frac{\partial Y}{\partial x}\right)_y\left(\frac{\partial x}{\partial X}\right)_y\left(\frac{\partial X}{\partial y}\right)_x,$$

или на основании соотношения (51.2)

$$\left(\frac{\partial Y}{\partial y}\right)_X = \left(\frac{\partial Y}{\partial y}\right)_x - \frac{(\partial X/\partial y)_x^2}{(\partial X/\partial x)_y}.$$

Числитель последней дроби, как всякий квадрат, не может быть отрицательным. Знаменатель, ввиду соотношения (51.3), существенно положителен. Значит, сама дробь не отрицательна, а потому должно иметь место неравенство

$$\left(\frac{\partial Y}{\partial y}\right)_{X=0} \leqslant \left(\frac{\partial Y}{\partial y}\right)_x. \tag{51.9}$$

Аналогично

$$\left(\frac{\partial X}{\partial x}\right)_{Y=0} \leqslant \left(\frac{\partial X}{\partial x}\right)_y. \tag{51.10}$$

6. Воспользуемся неравенствами (51.3), (51.4), (51.7) и (51.8) для вывода некоторых соотношений, в которых речь идет о сравнении знаков различных физических величин в состоянии устойчивого равновесия. Мы исходим из соотношения взаимности (51.2). Величина X есть функция x и y. Однако можно не конкретизировать независимые переменные, а сказать только, что X, x, y находятся в функциональной связи между собой. Отсюда следует, что имеет место тождество

$$\left(\frac{\partial X}{\partial y}\right)_x = -\left(\frac{\partial X}{\partial x}\right)_y\left(\frac{\partial x}{\partial y}\right)_X.$$

Из четырех величин X, Y, x, y только две могут меняться независимо. Но если в процессе величина X поддерживается постоянной, то из оставшихся трех величин Y, x, y независимо может меняться только одна. Возьмем в качестве таковой величину Y. Тогда, применяя правило дифференцирования функции от функции, можем написать

$$\left(\frac{\partial x}{\partial y}\right)_X = \left(\frac{\partial x}{\partial Y}\right)_X \left(\frac{\partial Y}{\partial y}\right)_X.$$

В результате получим

$$\left(\frac{\partial X}{\partial y}\right)_x = -\left(\frac{\partial X}{\partial x}\right)_y \left(\frac{\partial x}{\partial y}\right)_X = -\left(\frac{\partial X}{\partial x}\right)_y \left(\frac{\partial x}{\partial Y}\right)_X \left(\frac{\partial Y}{\partial y}\right)_X.$$

Аналогично

$$\left(\frac{\partial Y}{\partial x}\right)_y = -\left(\frac{\partial Y}{\partial y}\right)_x \left(\frac{\partial y}{\partial x}\right)_Y = -\left(\frac{\partial Y}{\partial y}\right)_x \left(\frac{\partial y}{\partial X}\right)_Y \left(\frac{\partial X}{\partial x}\right)_Y.$$

Производные $(\partial X/\partial y)_x$ и $(\partial Y/\partial x)_y$ имеют одинаковые знаки, так как в силу соотношения (51.2) они равны между собой. В состоянии устойчивого равновесия, как доказано выше, производные $(\partial X/\partial x)_y$, $(\partial Y/\partial y)_x$, $(\partial X/\partial x)_Y$, $(\partial Y/\partial y)_X$ существенно положительны. В результате получается следующий результат, который мы назовем *теоремой о знаках* [1]).

В состоянии устойчивого равновесия совпадают знаки следующих шести производных:

$$\left(\frac{\partial X}{\partial y}\right)_x, \quad -\left(\frac{\partial x}{\partial y}\right)_{X=0}, \quad -\left(\frac{\partial x}{\partial Y}\right)_X, \quad \left(\frac{\partial Y}{\partial x}\right)_y, \quad -\left(\frac{\partial y}{\partial x}\right)_{Y=0}, \quad -\left(\frac{\partial y}{\partial X}\right)_Y. \tag{51.11}$$

7. Доказанная теорема имеет прямое отношение к принципу Ле-Шателье–Брауна. Допустим, что нарушилось состояние равновесия системы, в результате которого параметр x получил бесконечно малое приращение $\Delta_1 x$, тогда как параметр y остался неизменным. Это вызовет изменение обобщенной силы на

$$\Delta Y = \left(\frac{\partial Y}{\partial x}\right)_y \Delta_1 x.$$

Но изменение силы Y на ΔY вблизи состояния равновесия влечет за собой изменение того же параметра x на

$$\Delta_2 x = \left(\frac{\partial x}{\partial Y}\right)_{X=0} \Delta Y = \left(\frac{\partial x}{\partial Y}\right)_{X=0} \left(\frac{\partial Y}{\partial x}\right)_y \Delta_1 x.$$

По теореме о знаках знаки производных $(\partial Y/\partial x)_y$ и $(\partial x/\partial Y)_{X=0}$ противоположны, а потому противоположны и знаки бесконечно малых приращений $\Delta_1 x$ и $\Delta_2 x$. Таким образом, изменение параметра x влечет за собой такие процессы, которые препятствуют этому изменению. Этого и требует принцип Ле-Шателье–Брауна.

Неравенства (51.9) и (51.10) вместе с условиями положительности входящих в них производных также могут быть истолкованы в смысле принципа Ле-Шателье–Брауна. Действительно, рассмотрим такое нарушение равновесия, при котором параметр x получил бесконечно малое приращение Δx,

[1]) По своей форме теорема о знаках имеет характер неравенства, поскольку совпадение знаков величин a и b можно записать в виде $a/b > 0$.

тогда как параметр y остался неизменным. При таком нарушении равновесия обобщенная сила X получит приращение:

$$\Delta_1 X = \left(\frac{\partial X}{\partial x}\right)_y \Delta x.$$

Это, вообще говоря, нарушает условие равновесия $Y = 0$. Для того чтобы оно не нарушалось, обобщенная сила X при том же самом Δx должна была бы получить приращение

$$\Delta_2 X = \left(\frac{\partial X}{\partial x}\right)_{Y=0} \Delta x.$$

Это приращение, согласно соотношению (51.10), по модулю меньше приращения $\Delta_1 X$. Поэтому для восстановления равновесия в системе должны существовать процессы, препятствующие нарастанию модуля Δx.

8. Применим теперь полученные результаты к вопросам термодинамики. Для этого надо конкретизировать потенциальную функцию f. Ограничимся рассмотрением физически однородных и изотропных тел. Подходящей потенциальной функцией может служить $Z = U - T_0 S + P_0 V$, в которой T_0 и P_0 — температура и давление среды, окружающей рассматриваемое тело. Эта функция минимальна в положении равновесия и содержит два свободных параметра, за которые можно принять энтропию S и объем V. Прочие функции U, I, F, Φ, Y для наших целей не годятся, так как они не имеют нужного числа свободных параметров. Например, условие равновесия, формулируемое с помощью потенциальной функции $U(S, V)$, требует минимума этой функции при постоянных S и V. Но если S и V фиксированы, то у функции U не остается ни одного свободного параметра. То же относится к функциям I, F, Φ. Функция $Y = U - T_0 S$ имеет один свободный параметр. Всеми этими функциями можно пользоваться, когда число параметров, определяющих внутреннее состояние системы, превышает два.

Роль «рассматриваемого тела» может играть какая-либо произвольно выбранная малая часть самого тела. Остальную часть можно рассматривать как «окружающую среду».

Если положить $F = Z(S, V)$, то обобщенными координатами будут $x = S$, $y = V$, а обобщенными силами —

$$X = \left(\frac{\partial Z}{\partial S}\right)_V = \left(\frac{\partial U}{\partial S}\right)_V - T_0, \quad Y = \left(\frac{\partial Z}{\partial V}\right)_S = \left(\frac{\partial U}{\partial V}\right)_S + P_0,$$

или ввиду соотношений (45.9)

$$X = T - T_0, \quad Y = P_0 - P. \tag{51.12}$$

Необходимые условия равновесия (51.1) требуют $T - T_0 = 0$, $P - P_0 = 0$. *При равновесии температура и давление тела равны температуре и давлению окружающей среды.*

Неравенства (51.3), (51.4), (51.7) и (51.8) переходят в

$$\left(\frac{\partial T}{\partial S}\right)_V > 0, \quad \left(\frac{\partial T}{\partial S}\right)_P > 0, \tag{51.13}$$

$$\left(\frac{\partial P}{\partial V}\right)_S < 0, \quad \left(\frac{\partial P}{\partial V}\right)_T < 0. \tag{51.14}$$

Физический смысл последних двух неравенств очевиден. Они показывают, *что объем тела уменьшается как при адиабатическом, так и при изотермическом повышении давления*. Первые два неравенства также имеют простой физический смысл. Для квазистатических процессов $\delta Q = T\, dS$.

Поэтому упомянутые неравенства можно переписать в виде

$$\left(\frac{\delta Q}{dT}\right)_V > 0, \quad \left(\frac{\delta Q}{dT}\right)_P > 0, \tag{51.15}$$

или

$$C_V > 0, \quad C_P > 0. \tag{51.16}$$

Неравенства (51.9) и (51.10) переходят в

$$-\left(\frac{\partial P}{\partial V}\right)_T \leqslant -\left(\frac{\partial P}{\partial V}\right)_S, \quad \left(\frac{\partial T}{\partial S}\right)_P \leqslant \left(\frac{\partial T}{\partial S}\right)_V, \tag{51.17}$$

или

$$C_P \geqslant C_V. \tag{51.18}$$

В случае идеальных газов последнее неравенство объясняется тем, что при нагревании газа при постоянном давлении он расширяется, и часть получаемой теплоты затрачивается на работу против внешнего давления. Однако такое объяснение не годится для тел, объем которых при нагревании уменьшается, например для воды между 0 и 4 °C. В то же время неравенство (51.18) справедливо в обоих случаях. Отсюда следует, что при нагревании при постоянном P затрачивается бóльшая работа против молекулярных сил, чем при нагревании при постоянном V.

9. Неравенства (51.13)–(51.18) являются необходимыми условиями устойчивости термодинамического равновесия. Допустим, например, что для некоторого физически однородного и изотропного вещества нарушено одно или оба условия (51.14). Заключим вещество в цилиндр, закрытый поршнем, который может в нем свободно перемещаться. Положим на поршень груз, создающий постоянное внешнее давление P_0. В состоянии равновесия это давление должно уравновешиваться внутренним давлением P: $P = P_0$. Всю систему поместим либо в термостат, температура которого поддерживается постоянной, либо адиабатически изолируем. Допустим, что поршень немного сместили вверх, так что объем тела V несколько увеличился. Это поведет к увеличению внутреннего давления P, так как по нашему предположению $\partial P/\partial V > 0$. Возникает разность давлений $P - P_0$, которая заставит поршень еще больше сместиться вверх. Это вызовет дальнейшее возрастание разности давлений $P - P_0$, и поршень все с большим и большим ускорением будет двигаться вверх. Рассуждая аналогично, придем к заключению, что смещение поршня вниз также поведет к появлению разности давлений, которая заставит вещество сжиматься, пока не будет нерушено условие $\partial P/\partial V > 0$. Таким образом, равновесных состояний, для которых условия (51.13) и (51.14) не соблюдаются, существовать не может. Такие состояния были бы абсолютно неустойчивы. Не то будет, если $\partial P/\partial V < 0$, как это имеет место в реальных условиях. Тогда при смещении поршня из положения равновесия всегда возникает разность давлений, препятствующая такому смещению. Это можно рассматривать как пример, подтверждающий принцип Ле-Шателье–Брауна.

10. Посмотрим теперь, что получилось бы, если бы теплоемкости вещества C_V и C_P были отрицательны. Поместим такое вещество в теплопроводящую оболочку, окруженную средой, температура T_0 которой поддерживается постоянной. Оболочка должна быть либо абсолютно жесткой, обеспечивающей постоянство объема тела, либо эластичной, не оказывающей никакого сопротивления расширению и сжатию тела. В последнем случае давление окружающей среды должно поддерживаться постоянным.

В состоянии равновесия температура тела T должна равняться температуре среды T_0. Допустим, что по какой-либо причине температура тела немного понизилась. Теплота самопроизвольно переходит всегда от тела с более высокой температурой к телу с менее высокой температурой. Поэтому при понижении температуры тела часть теплоты δQ перейдет от среды к телу. Это вызовет дальнейшее изменение температуры тела на dT. Величина dT должна быть отрицательной, так как по нашему предположению теплоемкость $\delta Q/dT$ отрицательна. Таким образом, температура тела еще более понизится. Это вызовет дальнейший переход теплоты от среды к телу и новое понижение его температуры. В результате температура тела будет неограниченно понижаться. Рассуждая аналогично, найдем, что всякое случайное повышение температуры тела приведет к неограниченному нагреванию его. Следовательно, при отрицательных теплоемкостях C_V и C_P устойчивое тепловое равновесие тела с окружающей средой невозможно. Напротив, когда указанные теплоемкости положительны, как это имеет место в действительности, всякое изменение температуры тела вызывает такие потоки теплоты, при которых возникшие разности температур сглаживаются, т. е. равновесие восстанавливается. Этого и требует принцип Ле-Шателье–Брауна.

11. Остается применить к рассматриваемому нами вопросу теорему о знаках. Ряд производных (51.11) теперь переходит в

$$\Big(\frac{\partial T}{\partial V}\Big)_S,\quad -\Big(\frac{\partial S}{\partial V}\Big)_T,\quad \Big(\frac{\partial S}{\partial P}\Big)_T,\quad -\Big(\frac{\partial P}{\partial S}\Big)_V,\quad -\Big(\frac{\partial V}{\partial S}\Big)_P,\quad \Big(\frac{\partial V}{\partial T}\Big)_P.$$

Эти величины могут быть как положительными, так и отрицательными. Но все они непременно должны иметь одинаковые знаки. Если принять во внимание соотношения Максвелла (45.15)–(45.18), то к указанным шести производным можно добавить еще две:

$$-\Big(\frac{\partial T}{\partial P}\Big)_S \quad \text{и} \quad -\Big(\frac{\partial P}{\partial T}\Big)_V,$$

которые должны иметь те же самые знаки.

Учтем теперь, что для обратимых процессов $\delta Q = T\,dS$. Тогда теорему о знаках для физически однородного тела можно сформулировать следующим образом.

Восемь величин

$$\begin{gathered}-\Big(\frac{\partial T}{\partial V}\Big)_S,\quad \Big(\frac{\partial P}{\partial T}\Big)_V,\quad \Big(\frac{\partial V}{\partial T}\Big)_P,\quad \Big(\frac{\partial T}{\partial P}\Big)_S,\\ \Big(\frac{\delta Q}{\partial V}\Big)_T,\quad -\Big(\frac{\delta Q}{\partial P}\Big)_T,\quad \Big(\frac{\delta Q}{\partial P}\Big)_V,\quad \Big(\frac{\delta Q}{\partial V}\Big)_P\end{gathered} \tag{51.19}$$

всегда имеют одинаковые знаки.

Совпадение знаков $(\partial V/\partial T)_P$ и $(\partial T/\partial P)_S$ физически означает следующее. *Если температурный коэффициент объемного расширения тела положителен, то при адиабатическом сжатии его температура будет повышаться. Если же он отрицателен, то при адиабатическом сжатии температура будет понижаться.*

Аналогично, *если температурный коэффициент давления положителен*: $(\partial P/\partial T)_V > 0$, *то при адиабатическом расширении температура тела будет понижаться, а при адиабатическом сжатии — повышаться.*

Величины $(\delta Q/\partial V)_T$ и $(\delta Q/\partial P)_T$ называются *скрытыми теплотами изменения объема и изменения давления. Эти величины всегда имеют противоположные знаки.* Физический смысл $(\delta Q/\partial P)_V$ и $(\delta Q/\partial V)_P$ ясен, но они не получили специальных названий. Их знаки всегда одинаковы.

12. Остановимся в заключение еще на одном любопытном явлении, теоретически предсказанном В. Томсоном и экспериментально подтвержденном Джоулем. Последний экспериментально обнаружил, что резиновый жгут нагревается, если его быстро (адиабатически) растянуть. Отсюда Томсон сделал вывод, что при нагревании натянутого резинового жгута (при постоянном натяжении) его длина должна сокращаться. Этот вывод и был проверен на опыте Джоулем.

Теория этого явления содержится в общих положениях, изложенных в настоящей параграфе. Элементарная работа при расширении жгута представляется выражением $\delta A = -\tau dl$. Роль объема V играет длина жгута l, роль давления — натяжение τ, взятое с противоположным знаком. Поэтому ясно, что вместо функции $Z = U - T_0 S + P_0 V$ надо пользоваться функцией $U - T_0 S - \tau_0 l$. Тогда по теореме о знаках восемь величин

$$-\Big(\frac{\partial T}{\partial l}\Big)_S, \quad -\Big(\frac{\partial \tau}{\partial T}\Big)_l, \quad \Big(\frac{\partial l}{\partial T}\Big)_\tau, \quad -\Big(\frac{\partial T}{\partial \tau}\Big)_S,$$

$$\Big(\frac{\delta Q}{\partial l}\Big)_T, \quad \Big(\frac{\delta Q}{\partial \tau}\Big)_T, \quad -\Big(\frac{\delta Q}{\partial \tau}\Big)_l, \quad \Big(\frac{\delta Q}{\partial l}\Big)_\tau$$

будут всегда иметь одинаковые знаки. Согласно опытам Джоуля производная $(\partial T/\partial l)_S$ положительна. Поэтому для резины должно быть $(\partial l/\partial T)_\tau < 0$. А это и значит, что если натяжение τ поддерживать постоянным, то при нагревании резинового жгута его длина должна уменьшаться. Большинство тел ведет себя иначе — при нагревании тела обычно расширяются. Такие тела при адиабатическом растяжении должны охлаждаться.

Заметим, что опытами П. Н. Лебедева (1866–1912) было показано, что коэффициент объемного расширения натянутой резины положителен. Отсюда следует, что при нагревании натянутого резинового жгута поперечные размеры его увеличиваются. Натянутая резина, таким образом, есть тело анизотропное. Температурный коэффициент линейного расширения ее в направлении натяжения отрицателен, а в перпендикулярном направлении — положителен.

Глава IV
ТЕПЛОПРОВОДНОСТЬ

§ 52. Уравнение теплопроводности

1. В этой главе будут рассмотрены элементы *математической теории теплопроводности*. Основы этой теории были заложены французским математиком Фурье (1768–1830) в первой четверти XIX века. Естественно, что Фурье исходил из представлений теории теплорода, которой тогда пытались объяснить все тепловые явления. Эти представления неверны. Но мы видели (см. § 16), что если объем системы или давление поддерживаются постоянными, то явления протекают так, как если бы теплота была каким-то веществом, которое может только перемещаться в пространстве, но не может создаваться или уничтожаться. Если постоянен объем системы, то количество теплоты следует отождествить с внутренней энергией, а если постоянно давление, то — с энтальпией системы. В обоих случаях математические основы теории теплопроводности Фурье остаются верными, хотя их физическое обоснование не имеет ничего общего с представлениями, из которых исходил сам Фурье.

В дальнейшем предполагается, что передача теплоты осуществляется исключительно путем теплообмена. Предполагается, что конвекции нет. В твердых телах это осуществляется само собой. В жидкостях же и газах надо позаботиться, чтобы конвекция была устранена; например, нагревать эти тела сверху. Точно так же предполагается, что потерями теплоты на лучеиспускание можно пренебречь. Кроме того, будем предполагать, что объем системы остается постоянным, так что никаких перемещений вещества в процессе передачи теплоты не возникает. Ограничимся, наконец, рассмотрением только одномерных задач, когда температура тела, помимо времени, зависит только от одной пространственной координаты.

2. В математической теории теплопроводности распространение теплоты рассматривается подобно течению жидкости. *Плотностью потока теплоты называется вектор* $\mathbf{j}$, *совпадающий по направлению с направлением распространения теплоты и численно равный количеству теплоты, проходящему в одну секунду через площадку в один квадратный сантиметр, перпендикулярную к направлению потока теплоты.* Найдем дифференциальное уравнение, которому удовлетворяет вектор $\mathbf{j}$ в одномерных задачах.

Пусть имеется неограниченная среда, в которой возникает поток теплоты в направлении, параллельном оси X. В одномерном общем случае свойства среды и величи́ны, характеризующие тепловой поток, могут меняться в том же направлении. Кроме того, они могут

меняться во времени. Поэтому плотность потока теплоты $\mathbf{j}$ следует рассматривать как функцию координаты x и времени t: $\mathbf{j} = \mathbf{j}(x, t)$. Выделим мысленно в среде бесконечно длинную призму или цилиндр с образующими, параллельными оси X, и рассмотрим бесконечно малый участок такого цилиндра AB длины dx (рис. 40). Пусть S — площадь поперечного сечения цилиндра. Количество теплоты, поступающее в цилиндр AB за время dt через основание A с координатой x, равно $j(x)S\,dt$. Количество теплоты, уходящее за то же время через основание B, будет $j(x+dx)S\,dt$. Так как через боковую поверхность цилиндра теплота не поступает, то полное количество теплоты, поступающее за время dt через рассматриваемый участок цилиндра, равно

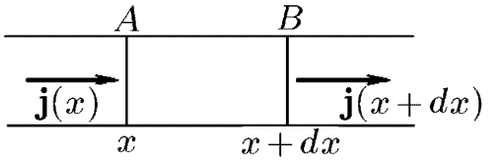


Рис. 40

$$[j(x) - j(x+dx)]\,S\,dt = -(\partial j/\partial x)S\,dx\,dt.$$

Но эту теплоту можно представить в виде $dM \cdot c_v\,dT$, где $dM = \rho S\,dx$ — масса цилиндра AB, c_v — удельная теплоемкость, dT — повышение температуры. Приравнивая оба выражения и производя сокращение, получим

$$\rho c_v \frac{\partial T}{\partial t} = -\frac{\partial j}{\partial x}. \tag{52.1}$$

3. Теперь надо установить связь между плотностью потока теплоты и температурой среды T. Опыт показывает, что поток теплоты имеет место только тогда, когда температура среды меняется от точки к точке. Теплота течет всегда в направлении от высшей температуры к низшей. Простейшим является случай бесконечной однородной пластинки толщины l. Если на одной плоской границе пластинки поддерживается температура T_1, а на другой — температура T_2, причем $T_1 > T_2$, то опыт показывает, что *поток теплоты пропорционален разности температур* $T_1 - T_2$ *и обратно пропорционален толщине пластинки* l. Математически это можно представить в виде

$$j = \varkappa \frac{T_1 - T_2}{l}, \tag{52.2}$$

где $\varkappa$ — положительная постоянная, зависящая только от материала пластинки и его физического состояния. Эта постоянная называется теплопроводностью материала пластинки [1].

[1] Термин «теплопроводность» употребляется в двух смыслах: то как *явление передачи теплоты*, то как *коэффициент для количественного описания этого явления*. Из контекста будет ясно, в каком смысле употребляется этот термин. Аналогичное замечание по терминологии относится и к другим понятиям, как-то: *температуропроводность, диффузия, термодиффузия, вязкость, поверхностное натяжение* и пр. Впрочем, перед некоторыми аналогичными понятиями слово «коэффициент» необходимо сохранить. Примером может служить *коэффициент концентрации диффузии* (см. § 92).

Допустим, что пластинка бесконечно тонкая. Если ось X направлена в сторону понижения температуры, то $l = dx$, $T_1 = T(x)$, $T_2 = T(x + dx)$,

$$\frac{T_2 - T_1}{l} = \frac{T(x + dx) - T(x)}{dx} = \frac{\partial T}{\partial x},$$

и формула (52.1) переходит в

$$j = -\varkappa \frac{\partial T}{\partial x}. \tag{52.3}$$

Выражение (52.3) остается верным и в том случае, когда ось X направлена в сторону повышения температуры, так как в этом случае $l = -dx$, $T_1 = T(x + dx)$, $T_2 = T(x)$. Оно также справедливо и в общем случае неоднородной среды с совершенно произвольным распределением температуры, и притом не только слоистой среды, но и такой, свойства и температура которой являются функциями всех трех пространственных координат x, y, z. Достаточно в рассматриваемой точке пространства направить ось X в сторону максимального понижения или повышения температуры и рассмотреть бесконечно тонкий слой, перпендикулярный к этому направлению. Такой слой может считаться однородным, и к нему применима формула (52.3). Теплопроводность $\varkappa$ будет функцией всех трех пространственных координат x, y, z. В нашей одномерной задаче она будет зависеть только от одной пространственной координаты x: $\varkappa = \varkappa(x)$.

Если выражение (52.3) подставить в формулу (52.1), то получится

$$\rho c_v \frac{\partial T}{\partial x} = \frac{\partial}{\partial x}\left(\varkappa \frac{\partial T}{\partial x}\right). \tag{52.4}$$

Это уравнение называется *уравнением теплопроводности*. В частном случае, когда среда однородна, теплопроводность $\varkappa$ не зависит от температуры, уравнение принимает вид

$$\rho c_v \frac{\partial T}{\partial t} = \varkappa \frac{\partial^2 T}{\partial x^2}, \tag{52.5}$$

или

$$\frac{\partial T}{\partial t} = \chi \frac{\partial^2 T}{\partial x^2}, \tag{52.6}$$

где введено обозначение

$$\chi = \frac{\varkappa}{\rho c_v}. \tag{52.7}$$

Постоянная χ называется *температуропроводностью среды*.

В среде могут оказаться *источники теплоты*. Например, теплота может выделяться в результате прохождения электрического тока или радиоактивного распада. Такие источники мы не принимали во внимание. Чтобы их учесть, введем величину q, равную количеству теплоты, выделяемому источниками в единице объема среды в одну

секунду. Тогда вместо уравнения (52.1) следует писать

$$\rho c_v \frac{\partial T}{\partial t} = -\frac{\partial j}{\partial x} + q. \tag{52.8}$$

В соответствии с этим изменятся и остальные уравнения.

4. В общем случае, когда свойства и температура среды зависят от всех трех пространственных координат x, y, z, уравнение теплопроводности, выражающее баланс теплоты в теле, имеет вид

$$\rho c_v \frac{\partial T}{\partial t} = -\left(\frac{\partial j_x}{\partial x} + \frac{\partial j_y}{\partial y} + \frac{\partial j_z}{\partial z}\right) + q. \tag{52.9}$$

Однако решения такого уравнения аналитически можно получить только в простейших случаях. Наиболее важными являются случаи, когда среда и распределение температуры в ней обладают *сферической* или *цилиндрической симметрией*. Поэтому мы не будем выводить уравнение (52.9) а ограничимся случаями сферической и цилиндрической симметрии. В этих случаях, вместо прямоугольной системы координат, более удобными являются *сферическая и цилиндрическая координатные системы*.

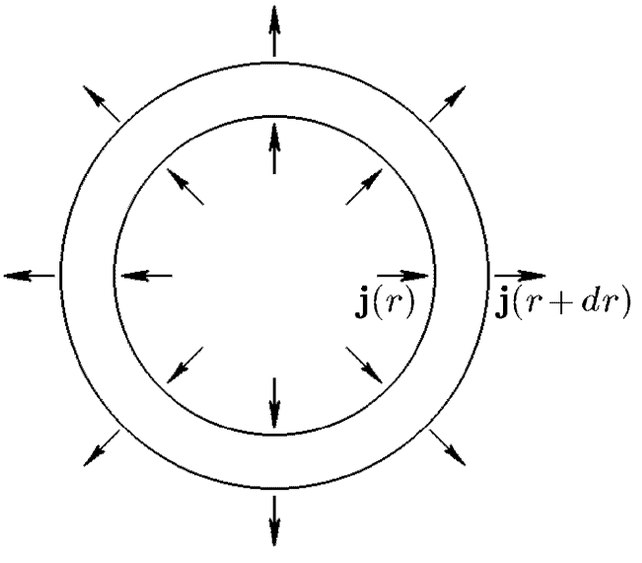


Рис. 41

Рассмотрим сначала случай сферической симметрии. Вектор плотности потока теплоты $\mathbf{j}$ направлен вдоль радиуса, причем $\mathbf{j}$, помимо времени, зависит только от r. Опишем вокруг центра симметрии две концентрические сферы с радиусами r и $r + dr$ (рис. 41). Количество теплоты, поступающее за время dt в пространство между этими сферами через первую из них, равно $j(r) \cdot 4\pi r^2 dt$. Количество теплоты, вытекающее за то же время через вторую сферу, будет $j(r+dr)\cdot 4\pi(r+dr)^2\,dt$. Эти два количества удобно писать в виде $4\pi(jr^2)_r\,dt$ и $4\pi(j^2)_{r+dr}\,dt$, чтобы подчеркнуть, что речь идет об одной и той же функции jr^2, но при разных значениях аргумента: r и $r+dr$. Разность между ними

$$4\pi[(jr^2)_r - (jr^2)_{r+dr}]\,dt = -4\pi\frac{\partial}{\partial r}(r^2 j)dr\,dt$$

дает количество теплоты, втекающее за время dt в рассматриваемый сферический слой из окружающего пространства. При наличии источников сюда надо добавить количество теплоты

$$4\pi q r^2\,dr\,dt,$$

поставляемое источниками. Но изменение количества теплоты в слое можно представить в виде $\rho \cdot 4\pi r^2\, dr \cdot c_v\, dT$. Поэтому уравнение баланса теплоты будет

$$\rho c_v \frac{\partial T}{\partial t} = -\frac{1}{r^2}\frac{\partial}{\partial r}(r^2 j) + q. \tag{52.10}$$

Вместо соотношения (52.3) следует писать $j = -\varkappa \partial T/\partial r$, так что

$$\rho c_v \frac{\partial T}{\partial t} = \frac{1}{r^2}\frac{\partial}{\partial r}\left(\varkappa r^2 \frac{\partial T}{\partial r}\right) + q. \tag{52.11}$$

Аналогичные рассуждения проводятся и в случае цилиндрической симметрии. Понимая теперь под r расстояние до оси симметрии, получим

$$\rho c_v \frac{\partial T}{\partial t} = -\frac{1}{r}\frac{\partial}{\partial r}(rj) + q, \tag{52.12}$$

$$\rho c_v \frac{\partial T}{\partial t} = \frac{1}{r}\frac{\partial}{\partial r}\left(\varkappa r \frac{\partial T}{\partial r}\right) + q. \tag{52.13}$$

5. К уравнению теплопроводности надо добавить общее соотношение, которое должно выполняться на границе раздела двух произвольных сред. Это *граничное условие* состоит в том, что по обе стороны указанной границы должны быть одинаковы нормальные составляющие вектора $\mathbf{j}$. Действительно, пусть AB — граница раздела, а $\mathbf{n}$ — единичный вектор нормали к ней, проведенный например, от первой среды ко второй (рис. 42). Вырежем мысленно бесконечно малый цилиндр с образующими, перпендикулярными к границе раздела, и основаниями по разные стороны от нее. Высота цилиндра h должна быть бесконечно малой высшего порядка по сравнению с линейными размерами оснований. Тогда потоком теплоты через боковую поверхность цилиндра можно пренебречь. Если S — площадь основания цилиндра, то количество теплоты, вступающее в него в 1 с, будет равно

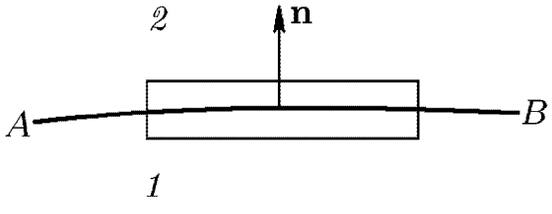


Рис. 42

$$[j_n^{(1)} - j_n^{(2)}]S.$$

Но эта величина, как и количество теплоты, содержащееся в цилиндре, должна быть пропорциональна его объему Sh, т. е. в пределе при $h \to$ $\to 0$ должна обращаться в нуль. Таким образом, в пределе, когда оба основания цилиндра сливаются друг с другом на границе AB, должно быть

$$j_n^{(1)} = j_n^{(2)}. \tag{52.14}$$

Это значит, что *на любой границе нормальная составляющая вектора потока теплоты непрерывна*. Доказательство предполагает, что на

границе раздела сред нет источников теплоты с конечной поверхностной плотностью. При наличии таковых нормальная составляющая вектора $\mathbf{j}$ может претерпевать разрыв.

§ 53. Простейшие стационарные задачи на теплопроводность

Все задачи на теплопроводность могут быть разделены на *стационарные* и *нестационарные*. *Стационарными называются такие задачи, в которых температура T не меняется во времени.* Она является функцией только пространственных координат. В этом случае $\partial T/\partial t = 0$. В одномерных задачах T зависит только от одной пространственной координаты, так что отпадает надобность в символе для частных производных. Рассмотрим простейшие стационарные одномерные задачи.

1. Стационарное распределение температуры в бесконечной плоскопараллельной пластинке. Допустим, что имеется бесконечная пластинка толщины l, поверхности которой поддерживаются при постоянных температурах T_1 и T_2. Требуется найти распределение температуры T внутри такой пластинки. Примем за ось X прямую, перпендикулярную к пластинке. Начало координат поместим на плоскости *1*, ограничивающей пластинку. Теплопроводность $\varkappa$ может зависеть от координаты x. Уравнение (52.4) переходит в

$$\frac{d}{dx}\left(\varkappa \frac{dT}{dx}\right) = 0.$$

Из него следует, что $\varkappa\, dT/dx = \text{const}$, или, ввиду (52.3), $j = \text{const}$. Постоянство плотности потока теплоты справедливо независимо от того, однородна пластинка или нет. Рассмотрим теперь простейший случай *однородной пластинки*. В этом случае теплопроводность $\varkappa$ постоянна, а потому $dT/dx = \text{const}$. Обозначая постоянную буквой A и интегрируя, получим

$$T = Ax + B,$$

где B — вторая постоянная интегрирования. Температура поперек пластинки меняется с координатой x по линейному закону. Постоянные A и B совершенно не зависят от теплопроводности. Они определяются из граничных условий. При $x = 0$ должно быть $T = T_1$, а при $x = l$ $T = T_2$. Это приводит к системе уравнений

$$T_1 = B, \quad T_2 = Al + B.$$

Определив из нее постоянные A и B, найдем распределение температуры:

$$T = \frac{T_2 - T_1}{l}\, x + T_1. \tag{53.1}$$

2. Стационарное распределение температуры между двумя концентрическими сферами. Обозначим радиус внутренней сферы через r_1, а внешней — r_2. Пространство между сферами запол-

нено средой, теплопроводность которой может зависеть от r. Из (52.11) следует, что при отсутствии в среде источников тепла распределение описывается уравнением

$$\frac{d}{dr}\left(\varkappa r^2 \frac{dT}{dr}\right) = 0.$$

Оно дает $\varkappa r^2\, dT/dr = \text{const}$. Таким образом, *плотность потока теплоты* $j = -\varkappa\, dT/dr$ *меняется обратно пропорционально квадрату расстояния* r. Так и должно быть, так как поток теплоты через сферу радиуса r равен $4\pi r^2 j$, а этот поток должен быть одним и тем же для всех сфер. Допустим теперь, что среда между сферами *однородна*. Тогда теплопроводность $\varkappa$ будет постоянной, а потому $r^2\, dT/dr = \text{const}$. Обозначая постоянную $-A$, получим $dT/dr = -A/r^2$, или после интегрирования

$$T = \frac{A}{\mathrm{r}} + B.$$

Постоянные интегрирования A и B определятся из значений, которые принимает температура T на границах сферического слоя. Это приводит к системе уравнений:

$$T_1 = \frac{A}{r_1} + B, \quad T_2 = \frac{A}{r_2} + B.$$

Определив из нее постоянные A и B, находим распределение температуры между сферами:

$$T = \frac{r_2 T_2 - r_1 T_1}{r_2 - r_1} + \frac{r_1 r_2 (T_1 - T_2)}{r(r_2 - r_1)}. \tag{53.2}$$

3. Стационарное распределение температуры между двумя концентрическими бесконечно длинными цилиндрами. Радиус внутреннего цилиндра обозначим через r_1, внешнего — r_2. Температуры их поддерживаются при постоянных значениях T_1 и T_2. Стационарное распределение температуры между цилиндрами находится так же, как и в предыдущем случае. Если среда между цилиндрами *однородна*, то получается

$$T = \frac{T_1 \ln r_2 - T_2 \ln r_1}{\ln(r_2/r_1)} + \frac{T_2 - T_1}{\ln(r_2/r_1)} \ln r. \tag{53.3}$$

ЗАДАЧИ

1. Урановый шар радиуса $R = 10$ см, помещенный в сосуд с водой, облучается равномерным потоком нейтронов. В результате реакций деления ядер урана в шаре выделяется энергия $q = 100\ \text{Вт}/\text{см}^3$. Температура воды $T = 373$ К, теплопроводность урана $\varkappa = 400\ \text{Вт}/(\text{м}\cdot\text{К})$. Найти стационарное распределение температуры в шаре, а также температуру в его центре.

Решение. В стационарном случае $\partial T/\partial t = 0$. В этом случае после однократного интегрирования уравнения (52.11) ($q = \text{const}$) получим

$$\frac{dT}{dr} = -\frac{q}{3\varkappa} r + \frac{C}{r^2}.$$

Постоянная интегрирования C должна равняться нулю, так как в противном случае в центре шара мы получили бы бесконечное значение для производной dT/dr. Интегрируя вторично с учетом граничного условия $T = T_0$ при $r = R$, найдем

$$T = T_0 + \frac{q}{6\varkappa}(R^2 - r^2).$$

Температура в центре шара

$$T_C = T_0 + \frac{qR^2}{6\varkappa} = 790 \text{ K}.$$

2. По однородному цилиндрическому проводу без изоляции течет постоянный электрический ток. Определить стационарное распределение температуры в проводе, если его поверхность поддерживается при постоянной температуре T_0.

Ответ. $T = T_0 + \dfrac{I^2\rho}{4\pi^2 R^4 \varkappa}(R^2 - r^2)$, где I — сила тока, ρ — удельное сопротивление провода, R — радиус провода, r — расстояние до его оси. Все величины выражаются в единицах СГС.

§ 54. Нестационарные задачи. Теорема единственности

1. Будем предполагать, что среда, в которой распространяется теплота, *однородна*, т. е. все параметры среды $\varkappa, \rho, c_v$ не зависят от координаты. Будем считать также, что они не зависят от времени и температуры, т. е. являются *постоянными*. Когда температура T зависит только от одной пространственной координаты x и времени, уравнение теплопроводности при наличии источников теплоты имеет вид (52.8) или с учетом (52.3)

$$\rho c_v \frac{\partial T}{\partial t} = \varkappa \frac{\partial^2 T}{\partial x^2} + q(x, t). \tag{54.1}$$

«Плотность мощности» источников теплоты q должна считаться заданной функцией координаты и времени. Но заданием источников решение уравнения (54.1) еще не определяется однозначно. К нему необходимо добавить так называемые *начальные и граничные условия*. Типичные начальные и граничные условия состоят в следующем.

Начальное условие определяет температуру во всем теле в какой-то один момент времени, который удобно принять за начало отсчета времени. Это условие можно записать в виде

$$T_{i=0} = f(x), \tag{54.2}$$

где $f(x)$ — заданная функция координаты x. Граничные условия определяют температуру тела на границе тела во все моменты времени. В одномерном случае тело имеет вид плоскопараллельной пластинки, ограниченной плоскостями $x = 0$ и $x = l$. Поэтому граничные условия запишутся в виде

$$T_{x=0} = \varphi_1(t), \quad T_{x=l} = \varphi_2(t), \tag{54.3}$$

где $\varphi_1(t)$ и $\varphi_2(t)$ — заданные функции времени.

2. Единственность решения сформулированной краевой задачи обусловлена тем, что температуропроводность χ есть величина *существенно положительная*. Для доказательства единственности решения допустим, что уравнение (54.1) имеет два решения: $T_1(x,t)$ и $T_2(x,t)$, удовлетворяющие начальному условию (54.2) и краевым условиям (54.3). Тогда

$$\frac{\partial T_1}{\partial t} = \chi \frac{\partial^2 T_1}{\partial x^2} + \frac{q}{\rho c}, \quad \frac{\partial T_2}{\partial t} = \chi \frac{\partial^2 T_2}{\partial x^2} + \frac{q}{\rho c}.$$

Вычитая почленно и вводя обозначение $\Theta = T_1 - T_2$, получим

$$\frac{\partial \Theta}{\partial t} = \chi \frac{\partial^2 \Theta}{\partial x^2}, \tag{54.4}$$

т. е. функция $\Theta(x,t)$ удовлетворяет уравнению теплопроводности без источников. Кроме того, ясно, что эта функция удовлетворяет «нулевым» начальным и граничным условиям:

$$\Theta_{t=0} = 0 \quad \text{при любых } x; \tag{54.5}$$

$$\Theta_{x=0} = 0, \quad \Theta_{x=l} = 0 \quad \text{при любых } t. \tag{54.6}$$

Рассмотрим интеграл $I(t) = \int\limits_0^l \Theta^2\, dx$. Ясно, что он не может быть отрицательным. Кроме того, ввиду (54.5), $I(0) = 0$. Найдем производную интеграла $I(t)$ по времени:

$$\frac{dI}{dt} = 2\int\limits_0^l \Theta \frac{\partial \Theta}{\partial t}\, dx = 2\chi \int\limits_0^l \Theta \frac{\partial^2 \Theta}{\partial x^2}\, dx.$$

Интегрируя по частям, получим

$$\frac{dI}{dt} = 2\chi\Theta \frac{\partial \Theta}{\partial x}\Big|_0^l - 2\chi \int\limits_0^l \Big(\frac{\partial \Theta}{\partial x}\Big)^2 dx.$$

Первое слагаемое в правой части обращается в нуль ввиду граничных условий (54.6). Второе слагаемое отрицательно или нуль, так как $\chi > 0$. Таким образом, $dI/dt \leqslant 0$. С течением времени интеграл I может только убывать или оставаться постоянным. Первое невозможно, так как должно быть $I(0) = 0$, $I(t) \geqslant 0$. Остается единственная возможность $dI/dt = 0$, т. е. $I(t) = \text{const} = I(0) = 0$. Это возможно тогда и только тогда, когда $\Theta(x,t) \equiv 0$, т. е. $T_1(x,t) \equiv T_2(x,t)$. Единственность решения доказана.

Рассуждая так же, легко доказать, что теорема единственности справедлива и для задач со сферической или цилиндрической симметрией. Она остается справедливой и для тел произвольной формы,

когда T зависит от всех трех пространственных координат. Доказательство проводится так же, только вместо простых интегралов надо пользоваться объемными и поверхностными интегралами. Это доказательство выходит за пределы нашего курса.

Если каким-либо способом удается найти или угадать решение уравнения теплопроводности, удовлетворяющее требуемым начальным и граничным условиям, то теорема единственности позволяет утверждать, что это и будет искомым решением задачи. Примеры на использование теоремы единственности будут приведены в § 56.

3. Могут быть и такие задачи, в которых единственность решения обусловлена другими причинами. В качестве примера рассмотрим следующую задачу.

Два теплоизолированных тела 1 и 2 с разными температурами соединены между собой однородным теплопроводящим стержнем, боковая поверхность которого также теплоизолирована. Начальные температуры тел равны соответственно T_{10} и T_{20}. Требуется найти закон изменения температуры этих тел во времени.

В такой формулировке задача содержит еще слишком много неопределенного. Для устранения неопределенности предположим прежде всего, что теплопроводность обоих тел очень велика (математически — бесконечно велика). Тогда выравнивание температур между различными частями тел будет происходить практически мгновенно. Поэтому в каждый момент времени t можно ввести определенные температуры $T_1(t)$ и $T_2(t)$, характеризующие тела 1 и 2 в целом. Но этого еще недостаточно, чтобы задача стала полностью определенной. Необходимо еще ввести дополнительно некоторые предположения относительно стержня. Поток теплоты через поперечное сечение стержня будет зависеть от начального распределения температуры в нем. Если начальная температура стержня равна T_{10}, то на границе с телом 1 в стержне в начальный момент времени не будет никакого теплового потока, тогда как на границе с телом 2 поток теплоты будет максимальным. Если стержень имел промежуточную температуру между T_{10} и T_{20}, то начальный поток теплоты будет как-то меняться вдоль стержня от сечения к сечению. Допустим, однако, что теплоемкость стержня пренебрежимо мала по сравнению с теплоемкостями тел C_1 и C_2. По истечении некоторого времени в стержне возникает равномерное падение температуры, при котором поток теплоты не будет изменяться вдоль стержня. За это время температуры тел 1 и 2, ввиду больших значений их теплоемкостей, практически не изменятся. Поэтому от процесса установления потока теплоты в стержне можно отвлечься и считать, что с самого начала поток теплоты вдоль стержня один и тот же во всех его сечениях. Тогда задача становится математически определенной, т. е. однозначной. Допустим для определенности, что $T_1 > T_2$. Поток теплоты вдоль стержня от тела 1 к телу 2 равен

$$\varkappa S \frac{T_1 - T_2}{l},$$

где S площадь поперечного сечения стержня, l — его длина. Этот поток численно равен скорости убывания $-dQ_1/dt$ теплоты в теле 1 или скорости приращения $+dQ_2/dt$ теплоты в теле 2. Считая теплоемкости C_1 и C_2 постоянными, можно написать $Q_1 = C_1T_1$, $Q_2 = C_2T_2$. Это приводит к уравнениям

$$C_1 \frac{dT_1}{dt} = -\varkappa S \frac{T_1 - T_2}{l}, \quad C_2 \frac{dT_2}{dt} = \varkappa S \frac{T_1 - T_2}{l}. \tag{54.7}$$

Почленное сложение уравнений (54.7) дает

$$C_1 \frac{dT_1}{dt} + C_2 \frac{dT_2}{dt} = 0,$$

или после интегрирования $C_1T_1 + C_2T_2 = \text{const}$. Это уравнение выражает сохранение общего количества теплоты, содержащегося в телах 1 и 2. В начальный момент $T_1 = T_{10}$, $T_2 = T_{20}$, а потому

$$C_1T_1 + C_2T_2 = C_1T_{10} + C_2T_{20}. \tag{54.8}$$

Этого уравнения недостаточно для определения неизвестных T_1 и T_2. Для нахождения недостающего уравнения разрешим уравнения (54.7) относительно производных dT_1/dt и dT_2/dt и вычтем почленно из одного уравнения другое. Тогда получим

$$\frac{d(T_1 - T_2)}{dt} = -\frac{T_1 - T_2}{\tau}, \tag{54.9}$$

где введено обозначение

$$\frac{1}{\tau} = \frac{\varkappa S}{l}\left(\frac{1}{C_1} + \frac{1}{C_2}\right). \tag{54.10}$$

Постоянная τ имеет размерность времени. Интегрируя уравнение (54.9), получим

$$T_1 - T_2 = Ae^{-t/\tau}.$$

Разность температур $T_1 - T_2$ убывает во времени по экспоненциальному закону. За время τ эта разность убывает в e раз. Поэтому τ характеризует время установления теплового равновесия между телами 1 и 2. Оно называется *временем релаксации или временем выравнивания температур* рассматриваемых тел. Постоянная интегрирования A найдется из начальных условий: $T_1 = T_{10}$, $T_2 = T_{20}$ при $t = 0$. Это дает

$$T_1 - T_2 = (T_{10} - T_{20})e^{-t/\tau}. \tag{54.11}$$

Решая теперь систему уравнений (54.8) и (54.11), найдем

$$T_1 = \frac{C_1T_{10} + C_2T_{20}}{C_1 + C_2} + \frac{C_2}{C_1 + C_2}(T_{10} - T_{20})\,e^{-t/\tau}, \tag{54.12}$$

$$T_2 = \frac{C_1T_{10} + C_2T_{20}}{C_1 + C_2} - \frac{C_1}{C_1 + C_2}(T_{10} - T_{20})\,e^{-t/\tau}.$$

При $t \gg \tau$ экспоненциальные члены в этих выражениях пренебрежимо малы, и формулы (54.12) переходят в общеизвестное выражение, определяющее «температуру смеси».

ЗАДАЧИ

1. Определить толщину льда, образующегося в течение заданного времени на спокойной поверхности озера. Считать, что температура T окружающего воздуха все время постоянна и равна температуре наружной поверхности льда ($T < T_{\text{пл}}$, где $T_{\text{пл}}$ — температура плавления льда).

Р е ш е н и е. Обозначим буквой x толщину образовавшегося слоя льда к моменту времени t. Если замерзание идет не очень быстро, как это в действительности имеет место в естественных условиях, то в слое льда установится линейное падение температуры от $T_{\text{пл}}$ до T. В этом случае количество теплоты, уходящее наружу от единицы поверхности льда за

время dt, представится выражением

$$\varkappa \frac{T_{\text{пл}} - T}{x}\, dt.$$

Но эту величину можно представить в виде $q\rho\, dx$, где dx — толщина слоя льда, образовавшегося за время dt, ρ — плотность льда, q — удельная теплота плавления льда. Это приводит к уравнению

$$\varkappa \frac{T_{\text{пл}} - T}{x}\, dt = q\rho\, dx.$$

Умножая на x и интегрируя, получим

$$\varkappa(T_{\text{пл}} - T)\, t = \frac{q\rho x^2}{2} + A.$$

Примем за начало отсчета времени момент, когда образование льда на поверхности воды только что началось. Тогда $x = 0$ при $t = 0$, а потому $A = 0$. В результате получим

$$x = \sqrt{\frac{2\varkappa(T_{\text{пл}} - T)t}{q\rho}}. \tag{54.13}$$

Для льда $\varkappa = 2{,}22 \cdot 10^5$ эрг/(с · см · К), $q = 3{,}35 \cdot 10^9$ эрг/г, $\rho = 0{,}9$ г/см^3. Допустим, что температура окружающего воздуха равна $-10\,^\circ$С. Пользуясь этими данными, нетрудно вычислить, что за сутки ($t = 86\,400$ с) образуется слой льда толщиной $x \approx 11{,}3$ см.

2. Сферический кусок льда (с начальным радиусом $R_0 = 1$ см) погружен в большую массу воды с температурой $10\,^\circ$С. Предполагая, что теплопередача в жидкости связана только с ее теплопроводностью, определить время τ, в течение которого лед полностью растает. Теплопроводность воды $\varkappa = 6 \cdot 10^{-3}$ Вт/(см · К), удельная теплота плавления льда $q = 330$ Дж/г.

Р е ш е н и е. Если таяние льда идет очень быстро, то мгновенное распределение температуры в окружающей воде будет таким же, что и в стационарном случае при тех же граничных значениях температуры. Согласно (53.2) оно в рассматриваемом случае имеет вид

$$T = T_\infty + \frac{R}{r}(T_0 - T_\infty),$$

где R — мгновенное значение радиуса куска льда, T_0 и T_∞ — постоянные температуры воды на поверхности шара и в бесконечности (по условию задачи $T_\infty - T_0 = 10$ К). Количество теплоты, поступающее к шару от окружающей воды за время dt,

$$4\pi r^2 \varkappa \frac{dT}{dr}\, dt = 4\pi\varkappa R(T_\infty - T_0)\, dt.$$

Эта теплота идет на расплавление льда и потому может быть также представлена выражением

$$-q\, dm = -4\pi R^2\, \rho_{\text{л}} q\, dR.$$

Приравнивая оба выражения, получим

$$\varkappa(T_\infty - T_0)dt = -\rho_{\text{л}} q R\, dR.$$

Отсюда интегрированием находим искомое время таяния льда:

$$\tau = \frac{\rho_{\text{л}} q R_0^2}{2\varkappa(T_\infty - T_0)} \approx 2480\,\text{с} \approx 40\,\text{мин}.$$

§ 55. Принцип суперпозиции температур. Температурные волны

1. Уравнение теплопроводности (52.6) *линейно* и *однородно*. Следствием этого является важное свойство его решений, называемое *принципом суперпозиции* температурных возмущений. Пусть $T_1(x,t)$ и $T_2(x,t)$ — какие-либо два решения уравнения (52.6), т. е.

$$\frac{\partial T_1}{\partial t}=\chi\frac{\partial^2 T_1}{\partial x^2},\quad \frac{\partial T_2}{\partial t}=\chi\frac{\partial^2 T_2}{\partial x^2}.$$

Если почленно сложить эти соотношения, то получится

$$\frac{\partial (T_1+T_2)}{\partial t}=\chi\frac{\partial^2 (T_1+T_2)}{\partial x^2}.$$

Отсюда видно, что сумма $T=T_1+T_2$ также является решением уравнения (52.6). Вообще, *сумма произвольного числа решений уравнения теплопроводности сама является решением того же уравнения*. Эта математическая теорема выражает следующий физический факт. *Пусть* $T_1(x,t), T_2(x,t), \ldots$ — *какие-либо возможные произвольные распределения температуры в среде. Тогда их сумма* $T=T_1(x,t)+T_2(x,t)+\ldots$ *дает также некоторое возможное распределение температуры в той же среде*. Это положение и называется принципом суперпозиции (наложения) температурных возмущений.

Для правильного понимания и применения суперпозиции температур необходимо иметь в виду, что свойства реальных сред, в том числе и температуропроводность χ, меняются с температурой. Этого при доказательстве мы не учитывали. Температура $T=T_1+T_2+\ldots$ может оказаться, например, настолько высокой, что твердое тело расплавится или испарится. Тогда решение $T=T_1+T_2+\ldots$ потеряет всякий смысл. Таким образом, свойства линейности и однородности уравнение теплопроводности сохраняет лишь приближенно в каком-то температурном интервале, в котором температуропроводность постоянна. Ширина интервала зависит от самой среды, а также от степени точности, предъявляемой расчету. Принцип суперпозиции сохраняет силу только тогда, когда все температуры $T_1, T_2, \ldots$, а также их сумма не выходят за пределы этого интервала. Вне этих пределов принцип суперпозиции несправедлив. Основное значение принципа суперпозиции состоит в том, что он позволяет по известным решениям уравнения теплопроводности «конструировать» новые решения.

2. Теорема, обратная только что доказанной, конечно, несправедлива. Сумма $T=T_1+T_2$ может быть решением уравнения теплопроводности (52.6), но слагаемые T_1 и T_2 могут и не быть таковыми. Однако формально математически можно ввести *комплексные решения*. Пусть T — комплексная функция, удовлетворяющая уравнению (52.6). Разобьем ее на вещественную и мнимую части: $T=T_1+iT_2$, где T_1 и T_2 — величины вещественные. Подставляя это выражение в уравнение (52.6) и отделяя вещественную часть от мнимой, получим

$$\Big(\frac{\partial T_1}{\partial t}-\chi\frac{\partial^2 T_1}{\partial x^2}\Big)+i\Big(\frac{\partial T_2}{\partial t}-\chi\frac{\partial^2 T_2}{\partial x^2}\Big)=0.$$

Но комплексное число тогда и только тогда равно нулю, когда в отдельности равны нулю его вещественная и мнимая части, т. е.

$$\frac{\partial T_1}{\partial t}-\chi\frac{\partial^2 T_1}{\partial x^2}=0,\quad \frac{\partial T_2}{\partial t}-\chi\frac{\partial^2 T_2}{\partial x^2}=0.$$

Значит, если комплексная функция $T = T_1 + iT_2$ является решением уравнения теплопроводности, то вещественные функции T_1 и T_2 также являются решениями того же уравнения. Справедливость этого утверждения связана с тем, что переменные x и t, а также температуропроводность χ — *величины вещественные*. Оно остается в силе для любых линейных однородных дифференциальных уравнений с вещественными коэффициентами и часто дает удобный метод отыскания вещественных решений таких уравнений. Проиллюстрируем это на примере так называемых *температурных* волн. Этот вопрос можно было бы рассмотреть и не выходя за пределы класса вещественных функций, но такой метод был бы громоздким и противоестественным.

3. Если в каком-либо месте среды температура периодически меняется во времени, то это приведет к периодическим изменениям температуры и во всех остальных точках среды. Рассмотрим простейший случай, когда среда однородна и заполняет пространство, ограниченное плоскостью $x = 0$. Ось X направим внутрь среды перпендикулярно к ее границе. Пусть температура на поверхности среды меняется во времени по синусоидальному или косинусоидальному закону, колеблясь вокруг некоторого среднего значения. Это среднее значение можно принять равным нулю, если условиться отсчитывать от него температуру. Так мы и поступим. При отыскании периодических решений уравнения теплопроводности вместо синуса или косинуса удобнее пользоваться комплексной показательной функцией, а затем с помощью известной формулы Эйлера (1707–1783)

$$e^{i\alpha} = \cos\alpha + i\sin\alpha \tag{55.1}$$

перейти к вещественной форме решения. Рассмотрим комплексную функцию

$$T = T_0 e^{i(\omega t - kx)}, \tag{55.2}$$

где T_0, ω и k — постоянные. Посмотрим, при каких значениях этих постоянных функция (55.2) будет решением уравнения теплопроводности. Дифференцирование дает

$$\frac{\partial T}{\partial t} = i\omega T_0 e^{i(\omega t - kx)} = i\omega T, \quad \frac{\partial^2 T}{\partial x^2} = -k^2 T_0 e^{i(\omega t - kx)} = -k^2 T.$$

Подставляя эти выражения в уравнение (52.6) и сокращая, получим

$$i\omega = -\chi k^2. \tag{55.3}$$

Если выполнено это условие, то функция (55.2) будет решением уравнения (52.6), какова бы ни была постоянная T_0. Постоянную ω мы выберем вещественной и положительной. Тогда постоянная k будет комплексной и может иметь два значения:

$$k = \sqrt{-i\,\frac{\omega}{\chi}} = \pm\sqrt{\frac{\omega}{2\chi}}(1 - i). \tag{55.4}$$

В результате выражение (55.2) преобразуется в

$$T = T_0 \exp\Big(\mp\sqrt{\frac{\omega}{2\chi}}\,x\Big) \exp\Big[i\Big(\omega t \mp \sqrt{\frac{\omega}{2\chi}}\,x\Big)\Big]. \tag{55.5}$$

Здесь содержатся два, а не четыре решения, так как верхний знак «−» должен комбинироваться с верхним же знаком «−», а нижний — с нижним. Из этих двух решений одно надо отбросить по физическим соображениям.

Колебания температуры начинают возбуждаться на поверхности среды и передаются внутрь нее. Естественно, что эти колебания должны *затухать*, а не нарастать по мере удаления от поверхности среды. Между тем знаку «+» в формуле (55.5) соответствует экспоненциально растущий множитель $\exp\left(+\sqrt{\omega/2\chi}\,x\right)$, стремящийся к бесконечности при $x \to \infty$. Этот знак не удовлетворяет условиям задачи, и надо сохранить только знак «−». Далее, необходимо перейти к вещественной форме решения, так как только такие решения имеют физический смысл. Как показано выше, всякое комплексное решение эквивалентно двум вещественным решениям. Из комплексного решения (55.5) описанным выше способом получаются два решения в вещественной форме:

$$T \equiv T_1 = T_0 \exp\left(-\sqrt{\frac{\omega}{2\chi}}\,x\right) \cos\left(\omega t - \sqrt{\frac{\omega}{2\chi}}\,x\right), \tag{55.6}$$

$$T \equiv T_2 = T_0 \exp\left(-\sqrt{\frac{\omega}{2\chi}}\,x\right) \sin\left(\omega t - \sqrt{\frac{\omega}{2\chi}}\,x\right). \tag{55.7}$$

Можно было бы убедиться непосредственной подстановкой, что найденные выражения являются решениями уравнения (52.6), удовлетворяющими граничному условию на поверхности среды. Тогда отпала бы необходимость в получении вспомогательного решения в комплексной форме (55.5). Но такой способ, как уже отмечалось, сложен и противоестествен.

4. Выясним теперь физический смысл полученных решений. Оба решения (55.6) и (55.7) однотипны — синус всегда можно преобразовать в косинус путем изменения начала отсчета времени. Поэтому достаточно ограничиться исследованием одного из них. Остановимся, например, на решении (55.6).

Если фиксировать x, то видно, что в каждой точке пространства температура T совершает во времени гармонические колебания с одним и тем же периодом $\tau = 2\pi/\omega$. Фаза этих колебаний меняется от точки к точке. Поверхность равной фазы

$$\omega t - \sqrt{\frac{\omega}{2\chi}}\,x = \text{const} \tag{55.8}$$

есть плоскость, параллельная поверхности среды. Она не остается на месте, а перемещается в направлении оси X с определенной скоростью v. Поэтому возмущения, описываемые решением (55.6), называют *температурной волной*, а постоянную v — *фазовой скоростью* или просто *скоростью этой волны*. Скорость v легко найти дифференцированием уравнения (55.8). Это дает

$$v = \frac{dx}{dt} = \sqrt{2\chi\omega} = 2\sqrt{\frac{\pi\chi}{\tau}}. \tag{55.9}$$

Длина температурной волны λ есть расстояние, проходимое ею за период τ,

$$\lambda = v\tau = 2\pi\sqrt{\pi\chi\tau}. \tag{55.10}$$

Амплитуда A температурной волны, как видно из формулы (55.6), затухает в направлении распространения по экспоненциальному закону:

$$A = T_0 e^{-\alpha x}, \tag{55.11}$$

где

$$\alpha = \sqrt{\frac{\omega}{2\chi}} = \sqrt{\frac{\pi}{\chi\tau}} = \frac{2\pi}{\lambda}. \tag{55.12}$$

Постоянная α называется *коэффициентом затухания температурной волны*. На протяжении длины $l = 1/\alpha = \lambda/2\pi$ амплитуда волны убывает в e раз.

5. Легко найти, каким начальным и граничным условиям удовлетворяет решение (55.6). Эти условия получатся, если в формуле (55.6) положить сначала $x = 0$, а затем $t = 0$. Таким путем находим

$$T_{x=0} = T_0 \cos \omega t, \tag{55.13}$$

$$T_{t=0} = T_0 \exp\Big(-\sqrt{\frac{\omega}{2\chi}}\, x\Big) \cos \sqrt{\frac{\omega}{2\chi}}\, x. \tag{55.14}$$

На основании теоремы единственности (см. § 54) делаем вывод, что единственным решением, удовлетворяющим этим условиям, является решение (55.6). В противоположность граничному условию (55.13) начальное условие (55.14) имеет весьма искусственный характер, и реальная физическая задача должна ставиться не как задача с граничными условиями, а иначе. Возможна, например, следующая постановка. На поверхности среды в момент времени $t = 0$ возбуждаются, а затем поддерживаются неограниченно долго гармонические колебания, представляемые выражением (55.13). Никаких источников теплоты внутри среды нет, начальное распределение температуры может быть каким угодно. Требуется определить, какие колебания температуры установятся в среде по прошествии достаточно длинного промежутка времени. Ответ дает формула (55.6). Действительно, по прошествии очень длинного промежутка времени все колебания температуры в среде затухнут, за исключением вынужденных колебаний, поддерживаемых внешними источниками, причем эти вынужденные колебания должны обладать той же периодичностью во времени, что и колебания температуры на поверхности среды.

6. Применим выведенные результаты к тепловым волнам, возбуждаемым в поверхностном слое Земли суточными и годовыми колебаниями температуры ее поверхности. Для простоты будем считать, что колебания являются гармоническими. Реальные колебания, конечно, не гармонические. Но это мало существенно. Дело в том, что любое периодическое колебание можно представить в виде наложения гармонических колебаний кратных периодов, причем основное значение имеют низкочастотные колебания, поскольку коэффициент затухания растет пропорционально квадратному корню из частоты. Периодами таких низкочастотных колебаний в нашей задаче являются соответственно год и сутки. Глубины проникновения суточных и годовых температурных волн, согласно формуле (55.12), должны быть связаны соотношением

$$l_{\text{год}}/l_{\text{сут}} = \sqrt{\tau_{\text{год}}/\tau_{\text{сут}}} = \sqrt{365} \approx 19.$$

И действительно, экспериментально было найдено, что колебания температуры, вызываемые нагреванием земной поверхности днем и охлаждением ночью, не влияют на температуру Земли уже на глубине около 1 м. Годовые же колебания земной поверхности, связанные с нагреванием ее летом и охлаждением зимой, перестают наблюдаться на глубине около 20 м. Глубже температура Земли совершено не зависит от температурных колебаний ее поверхности. Все это находится в полном соответствии с теоретической оценкой, приведенной выше. Вместе с тем мы видим, что глубина проникновения температурных волн пренебрежимо мала по сравнению с радиусом Земли. Вот почему при вычислениях можно было совсем пренебречь сферичностью Земли и считать ее плоской.

Другое подтверждение теории дают наблюдения по скорости распространения тепловых волн вблизи земной поверхности. Наблюдения показали, что скорость распространения тепловых волн с периодом в одни сутки $v_{\text{сут}} \approx 1$ м/сут, а скорость волн с годичным периодом $v_{\text{год}} \approx 0{,}046$ м/сут. Отношение этих скоростей

$$v_{\text{сут}}/v_{\text{год}} \approx 1/0{,}046 \approx 22,$$

тогда как по теории оно должно быть

$$v_{\text{сут}}/v_{\text{год}} = \sqrt{\tau_{\text{год}}/\tau_{\text{сут}}} = \sqrt{365} \approx 19.$$

Полного согласия ожидать трудно хотя бы потому, что Земля не является однородной средой, как это предполагает теория.

§ 56. Задача об остывании полупространства

1. Пусть однородная среда заполняет полупространство, ограниченное плоскостью $x = 0$. В начальный момент времени $t = 0$ температура среды всюду одинакова и равна T_0. Температура на поверхности среды все время поддерживается постоянной и равна $T_1 \neq T_0$. Таким образом, в начальный момент на границе среды температура испытывает скачок. Требуется найти распределение температуры $T(x, t)$ в среде во все последующие моменты времени. Эта задача была поставлена и решена В. Томсоном. Она является типичной краевой задачей, к которой применима теорема единственности, доказанная в § 54.

Направим ось X внутрь среды перпендикулярно к ее границе. Распределение температуры описывается уравнением теплопроводности (52.6). Чтобы найти его решение, удовлетворяющее требуемым начальным и краевым условиям, воспользуемся сначала методом размерности. Задача состоит в нахождении связи между переменными T, x, t и параметрами T_0, T_1, χ. Как видно из уравнения (52.6), температуропроводность χ имеет размерность квадрата длины, деленного на время. Учитывая это, нетрудно стандартным способом показать, что из шести величин T, x, t, T_0, T_1, χ можно составить только три независимые безразмерные комбинации, например T/T_0, T_1/T_0, $x/\sqrt{\chi t}$. Согласно правилу размерности распределение температуры в среде может быть записано в виде функциональной связи между этими безразмерными комбинациями. Но вторая из них T_1/T_0 есть просто постоянное число и, следовательно, может не учитываться при написании искомой функциональной связи. Таким образом, должно быть

$$T/T_0 = F(x/\sqrt{\chi t}), \quad \text{или} \quad T = f(\xi), \tag{56.1}$$

где введено обозначение

$$\xi = x/\sqrt{\chi t}. \tag{56.2}$$

Явный вид функции f можно определить из уравнения теплопроводности (52.6). Дифференцированием находим:

$$\frac{\partial T}{\partial x} = \frac{df}{d\xi}\frac{\partial \xi}{\partial x} = \frac{df}{d\xi}\frac{1}{2\sqrt{\chi t}},$$

$$\frac{\partial^2 T}{\partial x^2} = \frac{d^2 f}{d\xi^2}\frac{1}{2\sqrt{\chi t}}\frac{\partial \xi}{\partial x} = \frac{d^2 f}{d\xi^2}\frac{1}{4\chi t},$$

$$\frac{\partial T}{\partial t} = \frac{df}{d\xi}\frac{\partial \xi}{\partial t} = -\frac{1}{4}\frac{df}{d\xi}\frac{x}{t\sqrt{\chi t}}.$$

Подставляя эти значения в уравнение (52.6), получим после сокращения

$$d^2 f/d\xi^2 = -2\xi\, df/d\xi. \tag{56.3}$$

Обозначая дифференцирование по ξ штрихом и разделяя переменные, запишем это уравнение в виде

$$df'/f' = -2\xi\, d\xi, \quad \text{или} \quad df'/f' = -d\xi^2.$$

Интегрирование дает

$$f' = Ae^{-\xi^2}.$$

Интегрируя вторично и имея в виду, что $f = T$, получим

$$T = A \int_0^{x/2\sqrt{\chi t}} e^{-\xi^2}\, d\xi + B.$$

Остается найти постоянные интегрирования A и B. Полагая $x = 0$, $t \neq \neq 0$, получаем $T = B$. Таким образом, постоянная B дает температуру поверхности среды во все моменты времени $t \neq 0$. По условию задачи она постоянна и равна T_1. Для определения постоянной A воспользуемся начальным условием: $T = T_0$ при $t = 0$. Это дает

$$T_0 = A \int_0^{\infty} e^{-\xi^2}\, d\xi + T_1.$$

В интегральном исчислении доказывается, что входящий сюда интеграл равен $\sqrt{\pi/2}$. Таким образом, $T_0 = (A+2)\sqrt{\pi} + T_1$. Окончательное решение задачи имеет вид

$$T = 2\,\frac{T_0 - T_1}{\sqrt{\pi}} \int_0^{x/2\sqrt{\chi t}} e^{-\xi^2}\, d\xi + T_1. \tag{56.4}$$

Из этой формулы дифференцированием по x получаем значение температурного градиента

$$\frac{\partial T}{\partial x} = \frac{T_0 - T_1}{\sqrt{\pi\chi t}} \exp\Big(-\frac{x^2}{4\chi t}\Big). \tag{56.5}$$

В частности, на поверхности среды, т. е. при $x = 0$,

$$\frac{\partial T}{\partial x} = \frac{T_0 - T_1}{\sqrt{\pi\chi t}}. \tag{56.6}$$

Если по формуле (55.9) ввести сюда скорость распространения тепловых волн v с периодом τ, то получится

$$\frac{\partial T}{\partial x} = 2\,\frac{T_0 - T_1}{v\sqrt{\tau t}}. \tag{56.7}$$

2. Формула (56.7) интересна в том отношении, что с ее помощью Томсон вычислил возраст Земли. В то время считалось, что первоначально Земля находилась в огненно-жидком состоянии. В недрах Земли происходили интенсивные процессы перемешивания, приводившие к выравниванию температур. Приближенно можно было считать температуру Земли одной и той же во всех ее точках. Поэтому задача об остывании Земли аналогична рассмотренной нами задаче об остывании полупространства. Сферичность

Земли не может играть существенной роли, если нас интересует температурный градиент внутри тонкого поверхностного слоя Земли. В этом случае можно воспользоваться формулой (56.7) без всяких изменений. По мере остывания Земли образовывалась твердая земная кора. Начало этого процесса и принимается за момент времени, от которого отсчитывается возраст Земли. Томсон предположил, что температуропроводность Земли все время оставалась постоянной, и для вычисления возраста Земли t воспользовался формулой (56.7). Как мы уже говорили, на глубине 20 и более метров на температуру Земли уже не оказывают влияния температурные колебания окружающей атмосферы. Измерения показали, что на таких глубинах температура повышается приблизительно на 1° при углублении на каждые 25 м. Далее, Томсон условно принял, что температура поверхности Земли T_1 равна $0\,^\circ\mathrm{C}$, а в качестве T_0 взял температуру затвердевания горных пород: $T_0 \approx 4000\,^\circ\mathrm{C}$. Если за период τ взять одни сутки, то, как мы видели, наблюдения дают $v \approx 1$ м/сутки. Подставляя эти значения в формулу (56.7), получим

$$t = \frac{4(T_0 - T_1)^2}{v^2 \tau (\partial T/\partial x)^2} = \frac{4 \cdot 4000^2}{(1/25)^2} \text{ сут} \approx 10^8 \text{ лет}.$$

Приведенная оценка дает сильно заниженное значение для возраста Земли. Это и понятно. Томсон не учитывал и не мог учитывать интенсивное выделение теплоты в недрах Земли в результате происходящих в ней радиоактивных процессов. Кроме того, модель огненно-жидкой Земли не согласуется со многими фактами и в настоящее время не считается правильной. В настоящее время не существует общепризнанной теории происхождения Земли. А без такой теории трудно говорить об определенном возрасте Земли. В настоящее время возраст Земли по геологическим оценкам считается равным $(4\text{–}5) \cdot 10^9$ лет.

§ 57. Внешняя теплопередача

1. Формула (52.3), определяющая плотность потока теплоты j, относится к случаю, когда распределение температуры в среде непрерывно, а теплопроводность $\varkappa$ также является непрерывной функцией координат. В этом случае говорят о *внутренней теплопроводности*. В сущности, распределение температуры в пространстве всегда непрерывно. Однако для упрощения математических расчетов иногда бывает целесообразно ввести идеализированное представление о *скачке температур* на границе раздела двух различных тел, не находящихся в тепловом равновесии друг с другом. Допустим, например, что нагретое металлическое тело охлаждается потоком воды или воздуха. Ввиду большой теплопроводности металлов происходит быстрое выравнивание температур между различными частями металлического тела. Идеализируя задачу, можно принять, что в каждый момент времени тело имеет одну и ту же температуру. Точно так же окружающей среде, ввиду происходящих в ней процессов перемешивания, можно приписать в каждый момент времени одну и ту же температуру, отличную, однако, от температуры самого тела.

Благодаря процессам теплообмена возникает тепловой поток через границу тел, обусловленный скачком температуры на этой границе. Нормальная составляющая этого потока зависит от материала обеих

сред, а также от их температур. Простейшее предположение, введенное Ньютоном, состоит в том, что *величина* j_n *пропорциональна разности температур тел на границе*. Обычно предполагается, что одно из тел целиком окружает другое тело. Его мы будем называть *окружающей средой*. Таким образом,

$$j_n = \alpha(T - T_0), \tag{57.1}$$

где T — температура тела, а T_0 — температура окружающей среды. Нормаль n проведена от тела к среде. Постоянная α называется *коэффициентом теплопередачи*. При $\alpha = \infty$ температура на границе всегда непрерывна, т. е. $T = T_0$; при $\alpha = 0$ тело адиабатически изолировано.

Опыты показали, что закон Ньютона (57.1) выполняется приближенно и притом лишь при небольших разностях температур. Поэтому коэффициент теплопередачи не имеет такого же важного значения, какое имеет теплопроводность среды. Коэффициент теплопередачи является сугубо эмпирическим коэффициентом, которым можно пользоваться только в грубых расчетах.

2. Допустим теперь, что тело имеет форму тонкого бесконечно длинного стержня, ориентированного в направлении оси X. Поперечное сечение стержня может быть каким угодно, однако одним и тем же при любых x. Теплопроводность материала стержня $\varkappa$ должна быть достаточно большой, а сам стержень тонким, чтобы его температура T не менялась с координатами y и z. Она может зависеть только от времени t и координаты x. От тех же аргументов может зависеть и температура T_0 окружающей среды на поверхности стержня.

В этих предположениях выведем уравнение теплопроводности с учетом внешней теплопередачи.

Рассуждения будут такими же, что и при выводе уравнения (52.4). Только в балансе теплоты необходимо учесть дополнительный тепловой поток через боковую поверхность стержня. Для бесконечно малого элемента AB стержня (см. рис. 40) этот дополнительный тепловой поток, направленный к окружающей среде, равен $\alpha p(T - T_0)dx$, где p — периметр поперечного сечения стержня. Поэтому вместо уравнения (52.4) получится

$$\rho c_v S \frac{\partial T}{\partial t} = S \frac{\partial}{\partial x}\left(\varkappa \frac{\partial T}{\partial x}\right) - \alpha p(T - T_0), \tag{57.2}$$

где S — площадь поперечного сечения стержня. Предполагая $\varkappa$ постоянным и вводя обозначение

$$b^2 = \frac{\alpha p}{\rho c_v S}, \tag{57.3}$$

получим

$$\frac{\partial T}{\partial t} = \chi \frac{\partial^2 T}{\partial x^2} - b^2(T - T_0). \tag{57.4}$$

ЗАДАЧИ

1. Найти стационарное распределение температуры в тонком однородном стержне, концы которого поддерживаются при постоянных температурах T_1 и T_2, а температура окружающей среды T_0 также постоянна.

Р е ш е н и е. Удобно за нуль температуры принять температуру T_0 окружающей среды. В таком случае уравнение (57.4) переходит в

$$\frac{\partial^2 T}{\partial x^2} - \beta^2 T = 0, \tag{57.5}$$

где β — положительная постоянная, определяемая выражением

$$\beta = b/\sqrt{\chi}. \tag{57.6}$$

Общее решение уравнения (57.5) есть

$$T = Ae^{\beta x} + Be^{-\beta x}. \tag{57.7}$$

Постоянные интегрирования A и B определятся из граничных условий: $T = T_1$ при $x = 0$, $T = T_2$ при $x = l$. (Длина стержня обозначена через l, начало координат помещено на конце стержня.) После элементарных вычислений получим

$$T = \frac{T_1 \operatorname{sh}[\beta(l-x)] + T_2 \operatorname{sh}(\beta x)}{\operatorname{sh}(\beta l)}. \tag{57.8}$$

2. Пусть $T_1, T_2, T_3, \ldots$ — температуры последовательных равноотстоящих точек стержня в стационарном состоянии. Температура окружающей среды принята за нуль (см. предыдущую задачу). Показать, что эти температуры удовлетворяют соотношению

$$\frac{T_1+T_3}{T_2} = \frac{T_2+T_4}{T_3} = \frac{T_3+T_5}{T_4} = \ldots = \text{const} = e^{\beta\Delta x} + e^{-\beta\Delta x}, \tag{57.9}$$

где Δx — расстояние между двумя соседними точками рассматриваемого ряда равноотстоящих точек.

Глава V

ПРОСТЕЙШИЕ ВОПРОСЫ МОЛЕКУЛЯРНО-КИНЕТИЧЕСКОЙ ТЕОРИИ ВЕЩЕСТВА

§ 58. Введение

1. Идея об атомном строении вещества возникла в глубокой древности. Она в различной форме высказывалась и развивалась многими исследователями на протяжении веков. Однако до начала XX века атомная гипотеза оставалась лишь гениальной догадкой, не имевшей прямых экспериментальных доказательств. До середины XIX века конкретные представления об атомах и молекулах, а также физические теории, создавшиеся на их основе, были довольно наивными и относились скорее к области фантазии, чем к науке. Атомы представляли себе, например, как шарики с крючками, посредством которых они сцепляются друг с другом, или как шестеренки с зубцами, с помощью которых передается вращение от одного атома к другому. Законы механики Ньютона, по существу, не использовались. Все рассуждения носили качественный характер и основывались на сомнительных предположениях. Исключением была работа Даниила Бернулли (1738 г.), давшего не только качественное, но и количественное объяснение давления газа, по существу не отличающееся от современного. Горячим сторонником и пропагандистом молекулярно-кинетической теории вещества был М. В. Ломоносов (1711–1765). На основе этой теории Ломоносов предсказал существование абсолютного нуля температуры; указал, что закон Бойля–Мариотта не может быть абсолютно точным законом — от него при достаточно больших сжатиях газов должны наблюдаться отступления.

2. В начале XIX века атомно-молекулярная гипотеза получила убедительное подтверждение в химии в результате открытия *закона постоянства состава и закона кратных отношений*. Согласно первому из этих законов *массы химических элементов, из которых состоит то или иное химическое соединение, находятся во вполне определенном отношении*. Так, 1 г водорода, чтобы образовать воду, должен соединиться с 8 г кислорода. Если отношение масс водорода и кислорода не равно $1 : 8$, то излишек одного из этих химических элементов в реакцию не вступает.

Закон кратных отношений относится к таким реакциям, в которых два химических элемента A и B могут образовывать друг с другом не одно, а несколько химических соединений. Он утверждает, что *массы элемента B, вступающие в химические соединения с одной и той же массой элемента A, находятся в отношении небольших целых чисел.*

Так, водород при соединении с кислородом может образовывать не только воду H_2O, но и перекись водорода H_2O_2. Массы кислорода, приходящиеся на одно и то же количество водорода в этих двух соединениях, относятся как 1 : 2. Еще лучшую иллюстрацию закона кратных отношений дают соединения азота с кислородом:

				7 г азота, соединяясь с
4 г	кислорода,	образуют		11 г закиси азота N_2O,
8 г	»	»	»	15 г окиси азота NO,
12 г	»	»	»	19 г азотистого ангидрида N_2O_3,
16 г	»	»	»	23 г двуокиси азота NO_2,
20 г	»	»	»	27 г азотного ангидрида N_2O_5.

Массы кислорода, приходящиеся на одну и ту же массу азота (7 г), в этих соединениях относятся как

$$4 : 8 : 12 : 16 : 20 = 1 : 2 : 3 : 4 : 5.$$

Вряд ли возможно объяснить эти закономерности, не пользуясь представлением об атомах вещества. Объяснение, данное английским химиком Дальтоном (1766–1844), состоит в том, что в химических реакциях атомы различных веществ соединяются друг с другом в более сложные образования — молекулы, причем все молекулы химически чистого вещества построены совершенно одинаково. На вопрос о числе атомов в молекулах различных химических соединений удалось удовлетворительно ответить на основе эмпирически найденного закона Гей-Люссака и предложенной для его объяснения гипотезы, получившей позднее название закона Авогадро. Согласно закону Гей-Люссака *объемы газов (при одинаковых давлениях и температурах), вступающие в химические реакции друг с другом, а также объемы получающихся химических соединений в газообразном состоянии (при тех же давлениях и температурах) относятся между собой как целые и притом небольшие числа.* Для объяснения этого закона Авогадро выдвинул гипотезу, по которой *равные объемы различных газов при одинаковых давлениях и температурах содержат одно и то же число молекул.* На этой гипотезе основаны в химии методы определения *относительных атомных и молекулярных масс.* Ее строгое доказательство позднее было дано в кинетической теории газов. Значение атомно-молекулярной гипотезы в химии трудно переоценить. Без нее был бы невозможен быстрый прогресс этой науки в XIX веке.

3. В физике строго научное развитие молекулярной теории началось примерно со второй половины XIX века, главным образом благодаря трудам Клаузиуса, Максвелла (1831–1879) и Больцмана (1844–1906), в которых были заложены основы *кинетической теории газов.* О внутреннем строении атомов и молекул, а также о силах, с которыми они взаимодействуют между собой, в то время ничего не было известно. Основоположники кинетической теории газов пользовались упрощенными, идеализированными моделями этих частиц. Молекулы и атомы они рассматривали как идеально твердые шарики или как

материальные точки, взаимодействующие друг с другом центральными силами. Успехи теории были связаны не с этими идеализированными моделями, имеющими ограниченную область применимости, а с тем, что теория строилась на основе общих принципов механики Ньютона: законах сохранения импульса и энергии, оправдавших себя не только для макроскопических тел, но и для макроскопических объектов, подчиняющихся квантовой механике. Широко использовались математические методы, в частности методы *математической теории вероятности*. Существенно также, что теория развивалась под постоянным контролем опыта. Такая теория выгодно отличалась от наивных умозрительных построений предшествовавших атомистов.

4. До XX столетия на атомы смотрели как на мельчайшие неделимые частицы вещества. Это представление оказалось неверным. Атом является сложной системой, состоящей из ядра и окружающей его электронной оболочки. Атомизм проявляется не в том, что атомы неделимы, а в том, что все атомы, равно как и все простейшие (так называемые элементарные) частицы рассматриваемого вида, *абсолютно тождественны* и характеризуются *вполне определенными признаками — массой*, *зарядом ядра*, *излучаемым спектром* и пр. Подобной тождественности в области макромира не существует — в макромире нет двух абсолютно одинаковых тел. Атомизм проявляется также в том, что внутренние состояния атомов не непрерывны, а *дискретны*. Энергия атома, например, может принимать не непрерывный, а лишь дискретный ряд значений. Дискретные значения энергии атома называются его *энергетическими уровнями*. Обычно атом находится в так называемом *нормальном состоянии*, в котором его энергия минимальна. Для того чтобы перевести атом в ближайшее *возбужденное состояние*, требуется внешнее воздействие и затрата энергии. Если этой энергии недостаточно, то после прекращения воздействия внутреннее состояние атома окажется в точности таким же, каким оно было до воздействия. Дискретность возможных состояний атомных систем и является той физической, хотя ранее и не осознававшейся причиной, которая позволила химикам прийти к представлению о неделимости атомов и дала возможность физикам в кинетической теории газов рассматривать атомы и молекулы как неизменяемые материальные точки или идеально твердые шарики. Однако при увеличении энергии внешних воздействий, например при повышении температуры газа, такие представления становятся недействительными. Так, при температурах порядка 1000–3000 К молекулы начинают *диссоциировать*, т. е. распадаться на атомы. При температурах порядка 10 000 К и выше начинается *ионизация*, т. е. распад атомов на ионы и электроны. При температурах порядка десятков и сотен миллионов кельвинов начинаются *термоядерные реакции*, т. е. процессы слияния и распада атомных ядер. Мы широко будем пользоваться классическими моделями кинетической теории газов. Однако при этом необходимо иметь в виду границы применимости таких моделей.

§ 59. Давление газа с точки зрения молекулярно-кинетической теории

1. Молекулы взаимодействуют друг с другом посредством молекулярных сил. На далеких расстояниях это силы притяжения, убывающие с увеличением расстояния, на близких — силы отталкивания, быстро возрастающие при сближении молекул. Расстояние между центрами сблизившихся молекул, на котором силы притяжения переходят в силы отталкивания, принимается за диаметр молекулы. В газах при нормальных условиях средние расстояния между молекулами еще велики по сравнению с их диаметрами. На таких расстояниях молекулярные силы очень слабы и не играют существенной роли. Молекулярные силы проявляются лишь на близких расстояниях порядка диаметров молекул. Под действием этих сил скорости сблизившихся молекул претерпевают значительные изменения как по модулю, так и по направлению. Взаимодействия молекул на близких расстояниях называют *столкновениями*. Между двумя последовательными столкновениями молекула газа движется практически свободно, т. е. прямолинейно и равномерно. При каждом столкновении молекула газа почти мгновенно меняет направление своего движения, а затем движется с новой скоростью опять прямолинейно и равномерно, пока не произойдет следующее столкновение. Если газ в целом находится в покое (например, заключен в закрытом сосуде), то в результате столкновений устанавливается хаотическое движение, в котором все направления движения молекул равновероятны. Оно называется *тепловым движением*. Чем более разрежен газ, тем длинее средний путь, проходимый молекулой между двумя последовательными столкновениями. Для достаточно разряженного газа, заключенного в сосуд, можно в первом приближении пренебречь размерами молекул и столкновениями их друг с другом. Надо учесть только столкновения молекул со стенками сосуда, в который газ заключен. В этом приближении молекулы газа могут рассматриваться как материальные точки, не взаимодействующие между собой и движущиеся прямолинейно и равномерно между каждыми двумя последовательными столкновениями со стенками сосуда. Такая простейшая модель приводит к законам идеальных газов. Чтобы показать это, надо выяснить молекулярный смысл давления, температуры и внутренней энергии газа.

2. Давление газа на стенку сосуда есть результат ударов молекул газа об эту стенку. При каждом ударе молекула газа действует на стенку с определенной (с макроскопической точки зрения бесконечно малой) силой. Обратно направленная сила, с которой действует на молекулу стенка сосуда, заставляет молекулу отражаться от стенки. Если бы в сосуде содержалось всего несколько молекул, то их удары следовали бы друг за другом редко и беспорядочно и нельзя было бы говорить ни о какой регулярной силе давления, действующей на стенку. Мы имели бы дело с отдельными практически мгновенными бесконечно малыми толчками, которым время от времени подвергалась бы стенка. Если же

число молекул в сосуде очень велико, то будет велико и число ударов их о стенку сосуда. Удары станут следовать непрерывно друг за другом. Одновременно о стенку сосуда будет ударяться громадное количество молекул. Бесконечно малые силы отдельных ударов складываются в конечную и почти постоянную силу, действующую на стенку. Эта сила, усредненная по времени, и есть давление газа, с которым имеет дело макроскопическая физика.

3. Вычислим давление газа на стенку сосуда. Пусть газ заключен в закрытый сосуд и все молекулы одинаковы. Вообще говоря, они движутся с различными скоростями, отличающимися друг от друга как по величине, так и по направлению. Разделим все молекулы на группы так, чтобы молекулы одной и той же группы в рассматриваемый момент времени имели приблизительно одинаковые по величине и направлению скорости. Скорость молекул i-й группы обозначим через v_i, а число таких молекул в единице объема — через n_i. Возьмем на стенке сосуда малую площадку σ (рис. 43). Если молекулы движутся по направлению к площадке σ, то они могут столкнуться с ней. Если же они движутся от площадки, то столкновений не будет. Предположим, что молекулы i-й группы движутся по направлению к площадке σ, и подсчитаем число z_i молекул такой группы, ударяющихся об эту площадку за малое время dt. Построим на площадке σ, как на основании, косой цилиндр с образующими $v_i\,dt$, расположенный внутри сосуда. Всякая молекула i-й группы, находящаяся в этом цилиндре, за время dt успеет достигнуть площадки σ и удариться о нее. Поэтому число ударов z_i будет равно числу молекул i-й группы внутри построенного цилиндра, т. е. $z_i = n_i\,dV$, где dV — объем цилиндра. Направим координатную ось X вдоль внешней нормали к площадке σ. Тогда высота цилиндра будет равна $v_{ix}\,dt$, а его объем $dV = \sigma v_{ix}\,dt$. Следовательно,

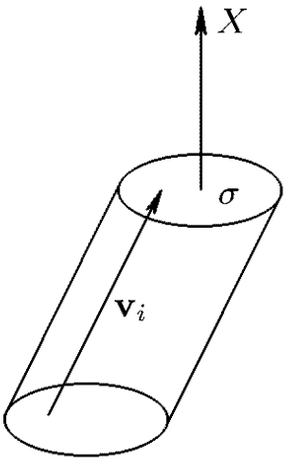


Рис. 43

$$z_i = \sigma n_i v_{ix}\,dt.$$

Дальнейший ход вычислений зависит от характера взаимодействия ударяющихся молекул со стенкой. Обычно при вычислениях считают, что стенка гладкая, а молекулы при ударе отражаются от нее зеркально, т. е. по законам удара идеально упругих шаров: абсолютная величина скорости при отражении не изменяется, угол падения равен углу отражения. Затем доказывается, что эти предположения не являются существенными. Однако в действительности стенка сосуда для ударяющейся молекулы не может быть идеальным зеркалом — ведь она сама состоит из молекул. Благодаря этому молекулы i-й группы после отражения будут иметь, вообще говоря, самые разнообразные по величине и направлению скорости, направленные от стенки, и распределятся по различным скоростным группам. Поэтому мы проведем дальнейшие

вычисления, не вводя никаких специальных предположений относительно законов отражения молекул от стенки сосуда. Единственное предположение, которое будет использовано в вычислениях, состоит в том, что *при отражении от стенки молекула в среднем не теряет и не приобретает кинетическую энергию*. В дальнейшем будет показано, что это предположение означает, что *температура газа должна быть равна температуре стенки*. Для целей вычисления процесс взаимодействия молекулы со стенкой удобно мысленно разбить на два этапа. На первом этапе молекула замедляется и останавливается, как бы прилипая к стенке. На втором этапе молекула отталкивается стенкой, ускоряется и отскакивает от нее. Вычислим сначала силу $\mathbf{F}_1$, которая действовала бы на площадку σ со стороны газа, если бы весь процесс взаимодействия молекул газа со стенкой ограничивался только первым этапом, т. е. в предположении, что после ударов молекулы газа как бы прилипают к стенке. Молекулы i-й группы, ударившиеся о площадку σ за время dt, от удара обладали количеством движения $z_i \mathbf{p}_i = \sigma n_i v_{ix} \mathbf{p}_i \, dt$, где $\mathbf{p}_i$ — количество движения одной молекулы. Чтобы остановить эти молекулы, стенка должна действовать на них с силой $\mathbf{f}_i'$, импульс которой равен $\mathbf{f}_i' \, dt = -\sigma n_i v_{ix} \mathbf{p}_i \, dt$. Изменив направление вектора $\mathbf{f}_i'$ на противоположное, мы найдем силу $\mathbf{f}_i = -\mathbf{f}_i' = \sigma n_i v_{ix} \mathbf{p}_i$, с которой действуют на площадку σ молекулы i-й группы на первом этапе. Сила $\mathbf{F}_1$, действующая на эту площадку со стороны всего газа, найдется суммированием этих выражений по всем группам молекул, летящих по направлению к стенке (для них $v_{ix} > 0$), т. е.

$$\mathbf{F}_1 = \sum_{v_{ix}>0} \sigma n_i v_{ix} \mathbf{p}_i.$$

К силе $\mathbf{F}_1$ следует прибавить силу $\mathbf{F}_2$, которая действует на площадку σ на втором этапе. Сила $\mathbf{F}_2$ вполне аналогична силе отдачи, испытываемой орудием при выстреле. Роль снаряда играют молекулы, летящие от площадки σ, т. е. молекулы, для которых $v_{ix} < 0$. Сила

$$\mathbf{F}_2 = \sum_{v_{ix}<0} \sigma n_i v_{ix} \mathbf{p}_i.$$

Разделение взаимодействия на два этапа, конечно, является только искусственным вычислительным приемом. На самом деле силы $\mathbf{F}_1$ и $\mathbf{F}_2$ действуют одновременно и складываются в одну результирующую силу

$$\mathbf{F} = \mathbf{F}_1 + \mathbf{F}_2 = \sigma \sum n_i v_{ix} \mathbf{p}_i.$$

Здесь суммирование производится уже по всем группам молекул, летящим как к стенке, так и от нее.

Сила $\mathbf{F}$ направлена нормально к площадке σ. Это является следствием хаотичности теплового движения молекул. Действительно, составляющая силы $\mathbf{F}$ в направлении оси Y равна

$$F_y = \sigma \sum n_i v_{ix} p_{iy}.$$

Ввиду хаотичности теплового движения среди слагаемых входящей сюда суммы встретится примерно столько же положительных членов, сколько и отрицательных. В среднем положительные слагаемые будут скомпенсированы отрицательными, так что сумма обратится в нуль. То же справедливо и для составляющей F_z. Этого не будет только для нормальной составляющей F_x, представляемой суммой

$$F_x = \sigma \sum n_i v_{ix} p_{ix},$$

все члены которой существенно положительны, так как знаки проекций v_{ix} и p_{ix} всегда одинаковы. Разделив слагающую F_x на площадь σ, получим давление газа на стенку сосуда:

$$P = \sum n_i v_{ix} p_{ix}.$$

Это выражение можно упростить, если ввести среднее значение произведения $v_x p_x$. Сумма таких произведений для молекул газа, находящихся в единице объема, равна $\sum n_i v_{ix} p_{ix}$. Чтобы найти среднее, надо эту сумму разделить на общее число молекул n в единице объема. Это дает

$$\langle v_x p_x \rangle = \frac{1}{n} \sum n_i v_{ix} p_{ix} \tag{59.1}$$

(угловые скобки означают усреднения по совокупности всех молекул). Давление P теперь можно представить в виде

$$P = n \langle v_x p_x \rangle. \tag{59.2}$$

По определению скалярного произведения

$$\mathbf{vp} = v_x p_x + v_y p_y + v_z p_z.$$

Усредняя это соотношение, получим

$$\langle \mathbf{vp} \rangle = \langle v_x p_x \rangle + \langle v_y p_y \rangle + \langle v_z p_z \rangle.$$

При хаотическом движении, каковым является тепловое движение молекул газа, все направления скоростей молекул равновероятны, а потому

$$\langle v_x p_x \rangle = \langle v_y p_y \rangle = \langle v_z p_z \rangle = \frac{1}{3} \langle \mathbf{vp} \rangle. \tag{59.3}$$

Это дает

$$P = \frac{1}{3} n \langle \mathbf{v} p \rangle. \tag{59.4}$$

Если объем сосуда, в котором заключен газ, равен V, а полное число молекул в этом объеме равно N, то $n = N/V$. Подставляя это значение в предыдущую формулу, получим

$$PV = \frac{1}{3} N \langle \mathbf{v} p \rangle. \tag{59.5}$$

4. При выводе формул (59.4) и (59.5) не учитывались столкновения молекул друг с другом. Для не слишком плотных газов межмолекулярные столкновения практически не влияют на окончательный

результат. При столкновениях молекулы переходят лишь из одной скоростной группы в другую. Состав каждой скоростной группы поэтому быстро и непрерывно меняется. Но для вычисления давления P несущественно, какие именно индивидуальные молекулы входят в каждую скоростную группу. Существенны лишь *средние числа* молекул в группах. Если состояние газа — установившееся, что должно предполагаться при выводе формул (59.4) и (59.5), то среднее число молекул в каждой из скоростных групп остается неизменным. Остается в среднем неизменной и сумма $\sum n_i(\mathbf{v}_i\mathbf{p}_i)$, а с ней и давление газа P.

Однако столкновения вносят качественные изменения в физическую интерпретацию давления P. Пока не было столкновений, молекулы газа совершенно не взаимодействовали друг с другом. Величина P имела только один смысл: она давала *давление газа на стенку сосуда*. При наличии столкновений появляется силовое взаимодействие между макроскопическими частями газа. Роль стенки для любой макроскопической части газа может играть граничащая с ней другая макроскопическая часть того же газа. В этих условиях величина P имеет также смысл *внутреннего давления*, посредством которого осуществляется силовое взаимодействие между примыкающими друг к другу макроскопическими частями газа. Именно такой смысл имеет давление P в гидродинамике и аэродинамике.

5. Формулы (59.4) и (59.5) применимы как к нерелятивистским, так и к релятивистским движениям молекул. В случае нерелятивистских движений масса молекулы m может считаться постоянной. Полагая в формулах (59.4) и (59.5) $p = mv$, получим для этого случая

$$P = \frac{1}{3}\, nm\langle v^2\rangle, \tag{59.6}$$

$$PV = \frac{1}{3}\, Nm\langle v^2\rangle. \tag{59.7}$$

При выводе этих формул молекулы рассматривались как *бесструктурные материальные точки*. Не принималось во внимание вращение молекул, а также внутримолекулярное движение. При столкновении могут меняться скорости вращения молекул. Молекула может перейти в возбужденное состояние или из возбужденного состояния вернуться в нормальное. Но все эти процессы не играют роли, когда речь идет о вычислении давления газа. Существенно только изменение *поступательного количества движения молекулы* при столкновениях ее со стенкой. Оно равно массе молекулы, умноженной на изменение скорости ее центра масс. Поэтому формулы (59.6) и (59.7) остаются в силе. Надо только понимать под v скорость поступательного движения молекулы (точнее, ее центра масс). Таким образом, формуле (59.7) можно придать вид

$$PV = \frac{2}{3}\, \langle \overline{E}_{\text{пост}}\rangle, \tag{59.8}$$

где $\langle \overline{E}_{\text{пост}}\rangle$ — среднее значение суммы кинетических энергий поступательного движения всех молекул газа. При столкновениях энергии

вращательного и внутримолекулярного движений могут переходить в энергию поступательного движения и наоборот. Однако в установившемся состоянии среднее значение величины $\overline{E}_{\text{пост}}$ остается неизменным.

Формула (59.8), как ясно из ее вывода, справедлива не только для однородного газа, но и для смеси различных газов. В этом случае под $\overline{E}_{\text{пост}}$ по-прежнему следует понимать сумму кинетических энергий поступательного движения молекул всех газов, содержащихся в сосуде. Из вывода ясно также, что для нашей модели газа, состоящей из невзаимодействующих молекул, справедлив закон Дальтона: *давление смеси газов равно сумме принципиальных давлений этих газов*.

§ 60. Скорости теплового движения газовых молекул

1. Выведенные формулы позволяют составить представление о скоростях теплового движения молекул газа. Не все молекулы газа движутся с одинаковыми скоростями. Встречаются медленные молекулы, скорости которых близки к нулю. Встречаются очень быстрые молекулы, скорости которых во много раз превосходят средние скорости молекулярного движения. Между этими пределами скорости молекул с различной степенью вероятности принимают всевозможные значения. Закон распределения скоростей газовых молекул будет рассмотрен в § 72. Для грубого представления о скоростях молекул газа могут служить некоторые средние величины, вычисляемые по определенным правилам. Рассмотрим прежде всего *среднюю квадратичную скорость*. Так называется величина

$$\overline{v}_{\text{кв}} = \sqrt{\langle v^2 \rangle}, \tag{60.1}$$

т. е. квадратный корень из среднего значения квадрата скорости поступательного движения молекулы. Напомним, что для вычисления $\langle v^2 \rangle$ надо скорость каждой молекулы возвести в квадрат, сложить полученные значения и сумму разделить на общее число молекул. От средней квадратичной скорости надо отличать *среднюю арифметическую* или, короче, просто *среднюю скорость* молекулы $\overline{v}$. Она определяется как сумма абсолютных скоростей всех молекул газа, деленная на их общее число. Как будет показано в § 73, величины $\overline{v}_{\text{кв}}$ и $\overline{v}$ отличаются друг от друга только численным множителем порядка единицы. Для $\overline{v}_{\text{кв}}$ формула (59.6) дает

$$\overline{v}_{\text{кв}} = \sqrt{3P/\rho}. \tag{60.2}$$

Скорость $\overline{v}_{\text{кв}}$ того же порядка что и скорость звука в газе $c = \sqrt{\gamma P/\rho}$ (см. т. I, § 86). Обе скорости связаны соотношением

$$\overline{v}_{\text{кв}} = c\sqrt{3/\gamma}. \tag{60.3}$$

Соотношения именно такого типа и следовало ожидать. Передача возмущений в звуковой волне осуществляется молекулами, движущимися

с тепловыми скоростями. Поэтому скорость звука по порядку величины должна совпадать со средней скоростью теплового движения молекулы. То же относится и к скорости истечения газа в вакуум, выражение для которой было получено в § 26.

2. Зная P и ρ при какой-либо температуре, легко вычислить среднюю квадратичную скорость $\overline{v}_{\text{кв}}$ при той же температуре. Однако для удобства вычислений формулу (60.2) лучше преобразовать с помощью уравнения состояния идеальных газов $P/\rho = RT/\mu$. Тогда получится

$$\overline{v}_{\text{кв}} = \sqrt{3RT/\mu}. \tag{60.4}$$

Так, для молекулярного водорода ($\mu = 2 \cdot 1{,}008$) при температуре 0 °C эта формула дает

$$\overline{v}_{\text{кв}} = \sqrt{\frac{3 \cdot 8{,}3143 \cdot 10^7 \cdot 273{,}15}{2 \cdot 1{,}008}} = 183\,800 \text{ см/с} = 1838 \text{ м/с}.$$

Аналогично, для азота $\overline{v}_{\text{кв}} = 493$ м/с, для кислорода $\overline{v}_{\text{кв}} = 461$ м/с и т. д.

3. Скорости того же порядка получены в опытах с молекулярными и атомными пучками. Средняя длина свободного пробега молекулы в газах, т. е. среднее расстояние, проходимое ею от одного столкновения до следующего, при нормальном давлении порядка 10^{-5} см. При давлении в 1 мм рт. ст. это величина порядка 10^{-2} см; при давлении в 10^{-6} мм рт. ст. — порядка 10^4 см $= 100$ м(см. § 86). В высоком вакууме молекулы газа движутся практически без столкновений между собой. Они сталкиваются лишь со стенками сосуда. Этим и пользуются для получения молекулярных и атомных пучков. Пучки получаются испарением металлов и других веществ в высоком вакууме.

Прямое измерение скоростей атомов в атомном пучке впервые было выполнено О. Штерном (1888–1969) в 1920 г. Упрощенная схема его опыта, ставшего классическим, изображена на рис. 44. Платиновая нить A, покрытая снаружи тонким слоем серебра, располагалась вдоль оси цилиндра C. Пространство внутри цилиндра откачивалось, и вакуум поддерживался непрерывно работающим насосом до давления порядка 10^{-5}–10^{-6} мм рт. ст. При пропускании электрического тока через платиновую проволоку она разогревалась до температуры выше точки плавления серебра (961,9 °C). Серебро интенсивно испарялось, и его атомы летели прямолинейно и равномерно от нити A к внутренней поверхности цилиндра C. Стенки последнего охлаждались, чтобы атомы серебра лучше конденсировались на них. На пути летящих атомов помещался экран с узкой щелью B, вырезавшей узкий атомный пучок. Пучок конденсировался на принимающей пластинке, прикрепленной к внутренней поверхности цилиндра

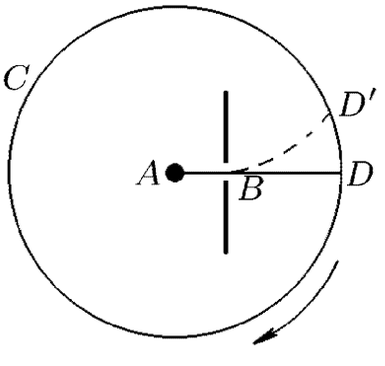


Рис. 44

(последняя на рис. 44 не изображена). Цилиндр вместе с экраном и нитью можно было приводить в быстрое вращение с круговой частотой порядка 2500–2700 мин$^{-1}$. Когда вся система была неподвижна, атомы серебра, пройдя через щель B, попадали на принимающую пластинку и, конденсируясь с ней, давали резкое изображение щели B в виде плоскости D, расположенной в одной плоскости с нитью A и щелью B. Затем система приводилась во вращение. В результате изображение щели смещалось в D'. Обозначим буквой s расстояние между изображениями D и D', измеренное вдоль вогнутой поверхности принимающей пластинки. Оно, очевидно, равно $s = V\tau$, где $V = \omega R$ — линейная скорость точек поверхности вращающегося цилиндра, R — его радиус, ω — угловая скорость вращения. Величина τ есть время прохождения атомами серебра расстояния BD. Обозначим это расстояние буквой l. Тогда $\tau = l/v$, где v — скорость атомов серебра. Таким образом, $s = \omega Rl/v$, откуда

$$v = \omega Rl/s. \tag{60.5}$$

В опытах Штерна изображение D получалось резким, тогда как изображение D' было всегда размытым. Это указывает на то, что атомы серебра движутся с различными скоростями. Формула (60.5) дает некоторую среднюю скорость, если под s понимать расстояние между центрами полосок D и D', измеренное вдоль дуги соответствующего круга. Практически для измерения такой скорости удобнее привести прибор во вращение сначала в одном направлении, а затем в противоположном и измерить расстояние между центрами получившихся изображений щели B. Максимальная температура нити в опытах Штерна составляла около 1200 °C. Для v получались значения от 560 до 640 м/с, близкие к средней квадратичной скорости 584 м/с, вычисленной по формуле (60.4), что находится в качественном согласии с выводами кинетической теории газов.

§ 61. Давление фотонного газа

Формулы (59.6) и (59.7) являются существенно нерелятивистскими, т. е. применимы только в тех случаях, когда средние скорости теплового движения молекул пренебрежимо малы по сравнению со скоростью света. Напротив, применимость формул (59.4) и (59.5) не связана с этим ограничением. Когда скорость частиц газа сравнима со скоростью света, газ называется *релятивистским*. В земных условиях такой случай осуществляется прежде всего для *фотонного газа*, т. е. газа, состоящего из фотонов, хаотически движущихся во всевозможных направлениях. *Фотонный газ всегда релятивистский*, поскольку фотоны всегда движутся со скоростью света.

Допустим, что имеется полость, стенки которой изготовлены из произвольного материала и поддерживаются при постоянной температуре. Стенки излучают и поглощают фотоны, в результате чего в полости и образуется фотонный газ. Каждый фотон, поглощаясь стенкой или отражаясь от нее, передает ей некоторый импульс. При излучении фотона стенка испытывает отдачу. В результате этих про-

цессов возникает давление фотонного газа на стенки полости. Так как фотонный газ предполагается изотропным, т. е. все направления движения фотонов в нем представлены с равной вероятностью, то для вычисления давления фотонного газа на стенку сосуда можно воспользоваться общей формулой (59.5). Энергия фотона ε связана с его импульсом соотношением $p = \varepsilon/c$, скорость фотона $v = c$, где c — скорость света. Поэтому формула (59.5) дает

$$PV = \frac{1}{3}\langle N\varepsilon\rangle = \frac{1}{3}\overline{E}, \tag{61.1}$$

где N — общее число фотонов полости, а $\overline{E}$ — средняя энергия фотонного газа. *Давление фотонного газа равно одной трети плотности энергии излучения в полости.* Формула (61.1) аналогична формуле (59.8), но отличается от нее коэффициентом. Разница обусловлена различием соотношений между энергией и импульсом для нерелятивистской частицы и для фотона.

В отличие от обычного газа, в котором молекулы не могут возникать и уничтожаться, число фотонов N в полости — величина непостоянная. Фотоны могут излучаться и поглощаться стенками полости. Поэтому для средней энергии излучения в полости нельзя писать $\overline{E} = N\overline{\varepsilon}$, а следует писать $\overline{E} = \langle N\varepsilon\rangle$, как это и сделано в формуле (61.1).

Коэффициент $1/3$ получился потому, что излучение в полости предполагалось *изотропным*. Как распределена энергия излучения по спектру частот — это не имеет значения. Существенна только изотропия излучения. Если излучение не изотропно, то формула (61.1) сохранит свой вид, но коэффициент $1/3$ заменится другим. Например, если излучение падает на стенку нормально и отражается от нее также нормально, то $PV = \overline{E}$.

Формула (61.1) играет важную роль в теории теплового излучения.

§ 62. Молекулярно-кинетический смысл температуры. Равномерное распределение кинетической энергии теплового движения по поступательным степеням свободы

1. Выясним физический смысл температуры в молекулярно-кинетической теории. Для этого возьмем цилиндр с поршнем AB (рис. 45), который может свободно без трения перемещаться вдоль цилиндра. По разные стороны поршня находятся одинаковые или различные идеальные газы. Величины, характеризующие первый газ, будем отмечать индексом 1, характеризующие второй газ — индексом 2. Для механического

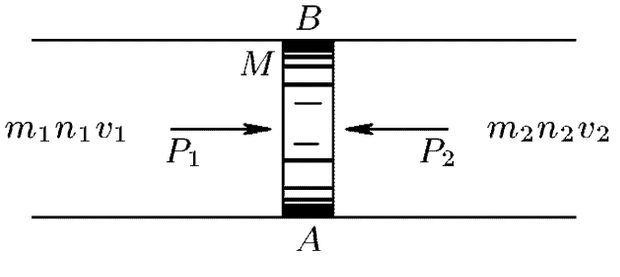


Рис. 45

равновесия поршня необходимо, чтобы давления газов были одинаковы:

$$P_1 = P_2 \quad \text{или} \quad \frac{1}{3}\, n_1 m_1 \overline{v_1^2} = \frac{1}{3}\, n_2 m_2 \overline{v_2^2}.$$

Но для того чтобы равновесие сохранялось длительно, необходимо еще равенство температур обоих газов: $T_1 = T_2$. В самом деле, допустим, что $T_1 > T_2$. Тогда начнется процесс выравнивания температур, в результате которого первый газ будет охлаждаться, а второй — нагреваться. Давление на поршень слева станет понижаться, а справа — повышаться, и поршень придет в движение справа налево. В процессе теплообмена молекулы газов обмениваются друг с другом кинетическими энергиями. Физический смысл макроскопического параметра — температуры — можно установить, рассмотрев процесс теплообмена с молекулярной точки зрения.

2. Скорость и другие характеристики теплообмена меняются с изменением материала и размеров поршня. Но конечный результат теплообмена, который сейчас нас только и интересует, от этого не зависит. Поэтому в целях упрощения вычислений можно идеализировать задачу, совершенно отвлекаясь от молекулярного строения поршня. Поршень мы будем рассматривать как сплошное идеально гладкое тело, с которым молекулы газов могут претерпевать упругие столкновения. Удары со стороны молекул, которым подвергается поршень слева и справа, в среднем уравновешивают друг друга. Но в каждый момент времени мгновенные силы ударов, вообще говоря, не уравновешиваются. В результате поршень непрерывно совершает беспорядочное тепловое движение туда и обратно. С этим явлением в рассматриваемой идеализированной модели и связана возможность обмена кинетическими энергиями теплового движения газов.

Предположим, что газы по обе стороны поршня настолько разряжены, что в каждый момент времени с поршнем сталкивается всего лишь одна молекула. Процессы, в которых с поршнем одновременно сталкиваются две или несколько молекул, настолько редки, что ими можно полностью пренебречь. Окончательные результаты, к которым мы придем, не связаны с этим ограничением. В следующем параграфе мы от него освободимся.

Рассмотрим столкновение какой-либо молекулы первого газа с движущимся поршнем. Поршень может двигаться только вдоль оси цилиндра, которую мы примем за ось X. Пусть u — скорость поршня до удара, u' — после удара. Соответствующие компоненты скорости молекулы обозначим через v_{1x} и v'_{1x}; массу поршня — через M. При ударе соблюдается закон сохранения импульса, а так как удар упругий, то имеет место также и сохранение кинетической энергии:

$$m_1 v_{1x} + Mu = m_1 v'_{1x} + Mu',$$

$$\frac{m_1}{2}\, v_{1x}^2 + \frac{M}{2}\, u^2 = \frac{m_1}{2}\, v_{1x}'^2 + \frac{M}{2}\, u'^2.$$

Это в точности такие же уравнения, какие используются в механике при решении задачи о столкновении идеально упругих шаров. Из них находим

$$u'_{1x} = \frac{2Mu - (M - m_1)v_{1x}}{M + m_1},$$

а для кинетической энергии движения молекулы вдоль оси X после удара

$$\frac{m_1 v'^2_{1x}}{2} = \frac{m_1}{2}\,\frac{4M^2u^2 - 4M(M - m_1)uv_{1x} + (M - m_1)^2 v^2_{1x}}{(M + m_1)^2}.$$

Напишем такое соотношение для каждой из молекул первого газа, сталкивающейся с поршнем, просуммируем по всем столкновениям и разделим на число столкновений. Короче говоря, произведем усреднение по всем столкновениям. Если состояние всей системы установилось, т. е. макроскопический процесс теплообмена закончился, то средняя скорость поршня равна нулю. Поршень совершает беспорядочные дрожания около положения равновесия, его скорость u с одинаковой вероятностью принимает положительные и отрицательные значения. Поэтому в результате усреднения произведения uv_{1x} получится нуль, и для средней кинетической энергии молекулы после столкновения можно написать

$$\frac{m_1}{2}\,\langle v'^2_{1x}\rangle = \frac{m_1}{2}\,\frac{4M^2\langle u^2\rangle + (M - m_1)^2\langle v^2_{1x}\rangle}{(M + m_1)^2}.$$

Теплообмена между газами не будет, когда средняя кинетическая энергия молекулы в результате отражения от поршня не меняется. Поэтому в установившемся состоянии написанное выражение должно быть равно средней кинетической энергии молекулы до удара $(m_1/2)\langle v^2_{1x}\rangle$. Это дает

$$\frac{4M^2\langle u^2\rangle + (M - m_1)^2\langle v^2_{1x}\rangle}{(M + m_1)^2} = \langle v^2_{1x}\rangle.$$

Отсюда после элементарных преобразований находим

$$\frac{m_1\langle v^2_{1x}\rangle}{2} = \frac{M\langle u^2\rangle}{2}. \tag{62.1}$$

Приведенное рассуждение, разумеется, применимо и ко второму газу. Следовательно,

$$\frac{m_2}{2}\langle v^2_{2x}\rangle = \frac{M}{2}\,\langle u^2\rangle, \tag{62.2}$$

а потому

$$\frac{m_1}{2}\,\langle v^2_{1x}\rangle = \frac{m_2}{2}\,\langle v^2_{2x}\rangle. \tag{62.3}$$

Ввиду хаотичности теплового движения молекул газа в нем нет никаких избранных направлений движения — все направления одинаково вероятны. Поэтому

$$\langle v^2_x\rangle = \langle v^2_y\rangle = \langle v^2_z\rangle,$$

а следовательно,

$$\frac{m_1}{2}\langle v_1^2\rangle = \frac{m_2}{2}\langle v_2^2\rangle. \tag{62.4}$$

Мы доказали, что *в состоянии теплового равновесия средние кинетические энергии всех молекул газа одинаковы.*

3. Средняя кинетическая энергия $\overline{\varepsilon}_{\text{пост}}$ поступательного движения молекулы газа, таким образом, обладает *основным свойством температуры* — в состоянии теплового равновесия она одинакова для всех молекул газов, находящихся в тепловом контакте, а также для различных молекул газовой смеси. Она не зависит от массы и внутренней структуры молекулы. Поэтому величину $\overline{\varepsilon}_{\text{пост}}$ или любую монотонную функцию ее можно принять за меру температуры газа, а также тела, находящегося с ним в тепловом равновесии. Удобно за меру температуры взять величину

$$\Theta = \frac{2}{3}\overline{\varepsilon}_{\text{пост}}. \tag{62.5}$$

Преимущество такого выбора заключается в том, что тогда формула (59.8) принимает вид

$$PV = \frac{2}{3}N\overline{\varepsilon}_{\text{пост}} = N\Theta, \tag{62.6}$$

напоминающий уравнение Клапейрона–Менделеева $PV = RT$.

Из молекулярно-кинетического толкования температуры можно вывести закон Авогадро. Возьмем два идеальных газа 1 и 2. Для них можно написать

$$P_1V_1 = N_1\Theta_1, \quad P_2V_2 = N_2\Theta_2.$$

Если $P_1 = P_2$, $V_1 = V_2$, $\Theta_1 = \Theta_2$, то из этих уравнений следует $N_1 = N_2$. В равных объемах идеальных газов при одинаковых давлениях и температурах содержится одинаковое число молекул. Это и есть закон Авогадро.

Величина Θ, определяемая формулой (62.5), называется *энергетической* или *кинетической температурой*. Она измеряется в тех же единицах, что и энергия, например, в джоулях и эргах. Для установления связи между кинетической температурой Θ и абсолютной термодинамической температурой T можно воспользоваться циклом Карно с идеальным одноатомным газом. Внутренняя энергия U такого газа состоит только из кинетической энергии поступательного движения его молекулы. Она равна $U = N\overline{\varepsilon}_{\text{пост}} = (3/2)N\Theta$, т. е. зависит только от температуры Θ. Поэтому можно повторить без всяких изменений рассуждения, приведенные в § 32 при установлении связи между термодинамической и идеально-газовой шкалами температур. В результате мы придем к соотношению

$$\frac{\Theta_1}{T_1} = \frac{\Theta_2}{T_2}.$$

Следовательно, отношение Θ/T есть универсальная постоянная, зависящая только от выбора единиц для Θ и T. Она называется *постоянной Больцмана* и является одной из важнейших фундаментальных постоянных физики. Эту постоянную принято обозначать буквой k. Таким образом, по определению

$$\Theta = kT. \tag{62.7}$$

Некоторые из методов экспериментального определения постоянной Больцмана будут изложены в дальнейшем. По современным данным

$$k = (1{,}380622 \pm 0{,}000044) \cdot 10^{-23} \text{ Дж/К} =$$
$$= (1{,}380622 \pm 0{,}000044) \cdot 10^{-16} \text{ эрг/К}.$$

4. Обозначим буквой N_A число молекул в одном моле. Эта универсальная величина называется *постоянной Авогадро*. Возьмем один моль идеального газа. Тогда, с одной стороны, имеет место соотношение (62.6), которое с учетом формулы (62.7) можно переписать в виде

$$PV = N_\text{A} kT. \tag{62.8}$$

С другой стороны, по уравнению Клапейрона–Менделеева $PV = RT$. Сравнивая эти уравнения, получим

$$R = N_\text{A} k. \tag{62.9}$$

Это соотношение позволяет определить постоянную Больцмана k как *универсальную газовую постоянную, отнесенную к одной молекуле газа*. Если известны значения R и k, то по формуле (62.9) можно вычислить постоянную Авогадро. По современным данным $N_\text{A} = R/k = (6{,}022045 \pm 0{,}000031) \cdot 10^{23}$ моль$^{-1}$.

5. Энергетическая шкала температур, в которой за температуру принимается величина Θ, теоретически является наиболее совершенной температурной шкалой. Она отличается от термодинамической шкалы только размерностью и единицей температуры. Температура в энергетической шкале измеряется теми же единицами, что и энергия. То обстоятельство, что для температуры введена особая единица — градус, объясняется историческими причинами. Кроме того, энергетические единицы температуры — эрг и джоуль — для измерения обычно встречающихся температур слишком велики. Впрочем, для измерения сверхвысоких температур очень удобна единица энергии — *электронвольт*. Как уже говорилось, при температурах порядка 1000–3000 К молекулы газа диссоциируют. При температурах порядка 10 000 К и выше происходит ионизация атомов. *Под сверхвысокими температурами подразумеваются температуры, когда процессы ионизации становятся существенными*. Энергию ионизации принято измерять в электронвольтах. *Электронвольт есть энергия, приобретенная электроном при прохождении разности потенциалов в один вольт*. Для атома водорода энергия ионизации равна 13,56 эВ. Для других атомов она того же порядка. Наибольшей энергией ионизации обладают атомы благородных газов, а наименьшей — атомы щелочных металлов.

Таким образом, энергия ионизации порядка десятка электронвольт. Поэтому электронвольт является удобной единицей для измерения сверхвысоких температур. Так как заряд электрона $e = 1{,}60 \cdot 10^{-19}$ Кл, то 1 эВ $= 1{,}6 \cdot 10^{-19}$ Дж $= 1{,}6 \cdot 10^{-12}$ эрг. Используя значение постоянной Больцмана, отсюда получаем

$$1 \text{ эВ} = \frac{1{,}6 \cdot 10^{-12}}{1{,}38 \cdot 10^{-16}} = 1{,}16 \cdot 10^4 \text{ К}.$$

Тысяча электронвольт называется *килоэлектронвольтом*. Температуры, развивающиеся в момент взрыва атомных и водородных бомб, порядка 10 кэВ $\approx 10^8$ К. Примерно до таких же температур надо нагреть плазму, т. е. проводящий ионизированный газ, чтобы в ней начались *термоядерные реакции*. Так называются процессы слияния или распада атомных ядер, обусловленные их взаимными столкновениями при сверхвысоких температурах.

ЗАДАЧИ

1. Сколько молекул находится в одном грамме воды?

Ответ. $3{,}34 \cdot 10^{22}$.

2. Сколько молекул находится в одном кубическом сантиметре воздуха при нормальном давлении и температуре 0 °С?

Ответ. $2{,}7 \cdot 10^{19}$.

3. Допустим, что все молекулы воды в стакане как-то отмечены. После этого вода была вылита в водопроводный сток. По прошествии длительного времени вылитая вода равномерно перемешалась со всей водой, имеющейся на Земле. Какое количество отмеченных молекул окажется в стакане, если его вновь наполнить водопроводной водой?

Ответ. Примерно 10^{-4}.

§ 63. Равномерное распределение кинетической энергии по степеням свободы

1. Формулы (62.1), (62.2) и (62.3) показывают, что в состоянии теплового равновесия средняя кинетическая энергия движения поршня вдоль оси цилиндра равна средней кинетической энергии движения молекулы газа в том же направлении. Поршень, если отвлечься от его молекулярного строения, является механической системой с одной степенью свободы — его положение определяется одной координатой x. Молекула, если также отвлечься от ее внутреннего строения, имеет *три поступательных степени свободы* — ее положение в пространстве можно задать тремя координатами x, y, z. Ввиду хаотичности теплового движения все направления скорости молекулы равновероятны. Кинетические энергии движения молекулы вдоль координатных осей X, Y, Z в среднем одинаковы. Таким образом, в состоянии теплового равновесия на каждую поступательную степень свободы молекулы и поршня приходится одна и та же средняя кинетическая энергия. Ее легко найти, заметив, что полная кинетическая энергия молекулы,

согласно формуле (62.5), есть $\overline{\varepsilon}_{\text{пост}} = (3/2)\Theta = (3/2)kT$. Эта энергия равномерно распределяется по трем степеням свободы молекулы. Поэтому *на одну поступательную степень свободы молекулы в среднем приходится кинетическая энергия* $\overline{\varepsilon}_{\text{кин}} = (1/2)\Theta = (1/2)kT$.

2. При выводе предполагалось, что поршень может двигаться вдоль оси цилиндра совершенно свободно. Однако для окончательного результата это несущественно. Можно представить себе, например, что поршень удерживается в положении равновесия пружиной. В состоянии теплового равновесия средняя кинетическая энергия поршня по-прежнему будет равна $(1/2)kT$. Действительно, собственный период колебаний поршня на пружине очень велик по сравнению с длительностью столкновения молекулы с поршнем. Поэтому наличие пружины никак не влияет на акт столкновения молекулы с поршнем — последний ведет себя так, как если бы он был свободным. Однако при наличии пружины поршень будет обладать также *потенциальной энергией*, испытывающей быстрые и нерегулярные изменения под действием ударов окружающих молекул. Предположим, что сила — квазиупругая, т. е. пропорциональна смещению поршня из положения равновесия. Найдем для этого случая среднее значение потенциальной энергии поршня при тепловом равновесии. Пусть $\varkappa$ — жесткость пружины. Свободные колебания поршня будут гармоническими: $x = a\cos(\omega t + \delta)$ с круговой частотой $\omega = \sqrt{\varkappa/M}$. Потенциальная энергия поршня

$$\varepsilon_{\text{пот}} = \frac{1}{2}\varkappa x^2 = \frac{1}{2}\varkappa a^2\cos^2(\omega t + \delta),$$

кинетическая

$$\varepsilon_{\text{кин}} = \frac{1}{2}M\dot{x}^2 = \frac{1}{2}Ma^2\omega^2\sin^2(\omega t + \delta) = \frac{1}{2}\varkappa a^2\sin^2(\omega t + \delta).$$

Запишем эти выражения в виде

$$\varepsilon_{\text{пот}} = \frac{1}{4}\varkappa a^2\{1 + \cos[2(\omega t + \delta)]\},$$

$$\varepsilon_{\text{кин}} = \frac{1}{4}\varkappa a^2\{1 - \cos[2(\omega t + \delta)]\}.$$

Косинус и синус с равной вероятностью принимают как положительные, так и отрицательные значения и при усреднении обращаются в нуль. Поэтому

$$\overline{\varepsilon}_{\text{пот}} = \overline{\varepsilon}_{\text{кин}} = \frac{1}{4}\varkappa\overline{a}^2. \tag{63.1}$$

Отсюда $\langle\varepsilon_{\text{пот}}\rangle = (1/2)kT$. Таким образом, *средние значения кинетической и потенциальной энергий одинаковы*. Если сила, удерживающая поршень, не квазиупругая, то этот результат, вообще говоря, неверен.

3. Наличие поршня не может сказаться на окончательном распределении энергии между газами, находящимися по разные стороны от него. Если убрать поршень, то обмен энергиями между ними будет

осуществляться путем непосредственных столкновений молекул газа 1 с молекулами газа 2. Явление осложняется перемешиванием (диффузией) газов. Но средние кинетические энергии поступательного движения молекул обоих газов в состоянии теплового равновесия останутся одинаковыми. То же справедливо и для смеси произвольного числа различных газов.

4. Учтем, наконец, молекулярную структуру поршня и определим среднюю кинетическую энергию поступательного движения его молекул. Если u — скорость центра масс поршня, то

$$u = \frac{1}{M}\sum m_i u_i,$$

где m_i — масса, а u_i — скорость молекулы поршня в направлении оси X. (Вместо u_i было бы логичнее писать u_{ix}, но мы опускаем индекс x, чтобы не загромождать формулы.) Возведя в квадрат, получим

$$\frac{1}{2}Mu^2 = \frac{1}{2M}\sum m_i m_j u_i u_j.$$

Усредним это соотношение по времени. Ввиду хаотичности теплового движения молекул поршня $\langle u_i u_j\rangle = 0$ при $i \neq j$. В предыдущей сумме надо учитывать только слагаемые с $i = j$. В результате получится

$$\frac{1}{2}M\langle u^2\rangle = \frac{1}{2M}\sum m_i^2\langle u_i^2\rangle. \tag{63.2}$$

По доказанному выше $(1/2)M\langle u^2\rangle = (1/2)kT$, следовательно,

$$\frac{1}{2M}\sum m_i^2\langle u_i^2\rangle = \frac{1}{2}kT. \tag{63.3}$$

Допустим теперь, что все молекулы поршня, а потому и все массы m_i, одинаковы. Тогда $\sum m_i^2\langle u_i^2\rangle = Nm_i^2\langle u_i^2\rangle$, где $N = M/m_i$ — общее число молекул поршня. В результате находим

$$\frac{1}{2}m_i\langle u_i^2\rangle = \frac{1}{2}kT. \tag{63.4}$$

Таким образом, и для молекул поршня имеет место равномерное распределение кинетической энергии по степеням свободы: на каждую поступательную степень свободы приходится в среднем кинетическая энергия $(1/2)kT$. Разумеется, это справедливо не только для энергии движения вдоль оси цилиндра, но, ввиду хаотичности теплового движения, также и для энергии движения молекулы в любом направлении. Сделанное при выводе предположение об одинаковости молекул поршня не играет роли.

5. Приведенное рассуждение позволяет снять ограничение, наложенное в предыдущем параграфе на плотности газов. Действительно, возьмем в качестве поршня сколь угодно плотный газ, заключенный между двумя твердыми стенками. К молекулам газа применим результат (63.4). Это показывает, что для справедливости теоремы о равномерном распределении кинетической энергии по степеням свободы предположение о малости плотности газов совершенно несущественно.

6. Несущественно также и то обстоятельство, что телом, к которому относилось приведенное рассуждение, является поршень. Для любого тела, если оно находится в состоянии теплового равновесия, на каждую поступательную степень свободы приходится в среднем одна и та же кинетическая энергия $(1/2)kT$. Используя эту теорему и проводя рассуждения, приведшие нас к формуле (63.2), в обратном порядке, можно получить новый существенный результат. Пусть произвольное макроскопическое тело находится в жидкой или газообразной среде, в которой оно может свободно двигаться в любом направлении. Можно предположить, что сила тяжести и другие поля отсутствуют. Можно также предположить, что тело удерживается в положении равновесия какими-либо силами, например архимедовой подъемной силой, упругой силой пружины и т. п. Во всех этих случаях центр масс тела должен совершать беспорядочные тепловые движения, для скорости V которых можно написать

$$\frac{1}{2}M\langle V^2\rangle = \frac{1}{2M}\sum m_i^2\langle v_i^2\rangle.$$

В приведенном ранее рассуждении считалась известной левая часть этого равенства. Теперь, наоборот, известна правая часть и нужно найти левую. Так как молекула имеет три поступательных степени свободы, то $(1/2)m_i\langle v_i^2\rangle = (3/2)kT$, а потому

$$\frac{1}{2M}\sum m_i^2\,\langle v_i^2\rangle = \frac{3}{2}\,kT\,\frac{\sum m_i}{M} = \frac{3}{2}\,kT.$$

Это дает

$$\frac{1}{2}M\langle V^2\rangle = \frac{3}{2}\,kT. \tag{63.5}$$

Таким образом, *на поступательное движение центра масс макроскопического тела в среднем приходится та же энергия* $(3/2)kT$, *что и на поступательное движение одной молекулы*. В этом отношении всякое макроскопическое тело ведет себя как гигантская молекула.

Видно, что и на вращение тела как целого вокруг неподвижной оси при тепловом равновесии приходится в среднем кинетическая энергия $(1/2)kT$. Чтобы это доказать, достаточно заметить, что угловая скорость вращения тела Ω вокруг неподвижной оси равна моменту количества движения тела, деленному на его момент инерции I относительно той же оси, т. е.

$$\Omega = \frac{1}{I}\sum m_i r_i u_i,$$

где u_i — составляющая скорости i-й молекулы, перпендикулярная к оси вращения и к радиусу-вектору r_i. По аналогии с формулой (63.2) получаем

$$\frac{1}{2}I\langle\Omega^2\rangle = \frac{1}{2I}\sum m_i^2 r_i^2\,\langle u_i^2\rangle,$$

откуда

$$\frac{1}{2}\,I\langle\Omega^2\rangle = \frac{1}{2}\,kT\,\frac{1}{I}\sum m_i r_i^2 = \frac{1}{2}\,kT.$$

7. Приведенные рассуждения могут рассматриваться как убедительные аргументы, доказывающие *классическую теорему о равномерном распределении кинетической энергии по степеням свободы* и разъясняющие ее смысл в частных случаях. Приведем теперь без доказательства общую формулировку этой теоремы. Предварительно напомним некоторые сведения из классической механики.

В классической теории атомы рассматриваются как материальные точки, а всякое макроскопическое тело — как система материальных точек. Если число материальных точек в системе равно N и на систему не наложены никакие дополнительные связи, ограничивающие свободу ее движения, то требуется $3N$ координат, чтобы однозначно задать положение всех точек системы. Классическая теория, однако, пользуется и такими механическими моделями, в которых на движение материальных точек наложены определенные ограничения — связи. При наличии связей число независимых координат, заданием которых однозначно определяется конфигурация, т. е. положение всех точек системы, уменьшается. В качестве таких независимых координат можно взять те прямоугольные координаты материальных точек, через которые выражаются все остальные координаты. Число этих независимых координат f называется *числом степеней свободы системы*. Не обязательно пользоваться прямолинейными координатами. Можно взять f любых других величин $q_1, q_2, \ldots, q_f$, однозначно определяющих конфигурацию системы. Они называются *обобщенными координатами*, а их производные по времени $\dot{q}_1, \dot{q}_2, \ldots, \dot{q}_f$ — *обобщенными скоростями*.

Радиусы-векторы $\mathbf{r}_i$ материальных точек системы являются функциями обобщенных координат:

$$\mathbf{r}_i = \mathbf{r}_i(q_1, q_2, \ldots, q_f).$$

Следовательно,

$$\mathbf{v}_i = \dot{\mathbf{r}}_i = \frac{\partial \mathbf{r}_i}{\partial q_1}\,\dot{q}_1 + \frac{\partial \mathbf{r}_i}{\partial q_2}\,\dot{q}_2 + \ldots + \frac{\partial \mathbf{r}_i}{\partial q_f}\,\dot{q}_f,$$

т. е. обычные скорости $\mathbf{v}_i$ материальных точек системы являются линейными однородными функциями обобщенных скоростей $\dot{q}_1, \dot{q}_2, \ldots, \dot{q}_f$. Коэффициенты, входящие в эти функции, зависят, вообще говоря, от всех обобщенных координат механической системы. Используя полученное выражение, для кинетической энергии системы находим

$$E_{\text{кин}} = \frac{1}{2}\sum_{i=1}^{N} m_i v_i^2 = \frac{1}{2}\sum_{i=1}^{f}\sum_{k=1}^{f} a_{ik}\dot{q}_i\dot{q}_k. \tag{63.6}$$

Кинетическая энергия представляется квадратичной формой обобщенных скоростей $\dot{q}_i$. Коэффициенты этой формы a_{ik}, вообще говоря, зависят от обобщенных координат $q_1, q_2, q_3, \ldots$

В общем случае в сумму (63.6) входят члены с попарными произведениями различных обобщенных скоростей. По этой причине слагаемые указанной суммы, вообще говоря, не могут быть интерпретированы как кинетические энергии, приходящиеся на соответствующие степени свободы системы. Однако обобщенные координаты всегда можно выбрать так, чтобы такая интерпретация сделалась возможной. Действительно, в математике доказывается, что надлежащим выбором обобщенных координат квадратичную форму (63.6) всегда можно привести к так называемому *диагональному виду*, т. е. к такому виду, в котором она содержит только квадратичные члены и не содержит членов с попарными произведениями обобщенных скоростей. При таком выборе обобщенных координат

$$\varepsilon_{\text{кин}} = \frac{1}{2} \sum_{i=1}^{f} a_i \dot{q}_i^2, \tag{63.7}$$

где коэффициенты a_i являются функциями обобщенных координат. Если возбуждена только одна i-я степень свободы, то сумма (63.7) сводится к одному слагаемому $a_i \dot{q}_i^2/2$. Это слагаемое поэтому можно интерпретировать как кинетическую энергию, приходящуюся на i-ю степень свободы. Таким образом, при указанном выборе обобщенных координат полная кинетическая энергия системы представляется в виде суммы кинетических энергий, приходящихся на отдельные степени свободы. Так, если за координатные оси выбрать главные оси вращения твердого тела, то его кинетическая энергия в любой момент времени может быть представлена в виде

$$\varepsilon_{\text{кин}} = \frac{1}{2} M(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) + \frac{1}{2} (I_x \dot{\varphi}_x^2 + I_y \dot{\varphi}_y^2 + I_z \dot{\varphi}_z^2),$$

где M — масса тела, x, y, z — прямоугольные координаты его центра масс, I_x, I_y, I_z — моменты инерции тела относительно координатных осей, $\dot{\varphi}_x, \dot{\varphi}_y, \dot{\varphi}_z$ — угловые скорости вращения тела относительно тех же осей.

В дальнейшем при изложении теории теплоемкости предполагается, что обобщенные координаты выбраны так, что кинетическая энергия представляется выражением (63.7), т. е. в виде суммы квадратичных членов.

8. Так как между частицами системы есть силовое взаимодействие, то при тепловом движении энергия каждой частицы быстро и беспорядочно меняется во времени. Беспорядочно меняются во времени и слагаемые суммы (63.7). В молекулярно-кинетической теории представляет большой интерес знание средних значений таких слагаемых. Основная теорема, применимая к классическим системам, состоит в том, что *в состоянии теплового равновесия на каждую степень свободы приходится в среднем одна и та же кинетическая энергия.* Это положение называется *теоремой о равномерном распределении*

кинетической энергии по степеням свободы. Его первоначальные доказательства для частных случаев были даны Максвеллом и Больцманом. Общее доказательство дается в статической механике, однако оно выходит за рамки нашего курса, и мы ограничимся лишь замечанием, что *в основе доказательства лежит предположение о применимости законов классической механики к атомно-молекулярным системам*, а также одно общее предположение вероятностного характера (так называемая *эргодическая гипотеза*), принять которое необходимо для согласования статистической физики с аксиоматической термодинамикой.

Средняя кинетическая энергия, приходящаяся при тепловом равновесии на одну степень свободы любой атомно-молекулярной системы равна $(1/2)kT$. В этом легко убедиться, если представить, что рассматриваемая система находится в тепловом контакте с одноатомным газом той же температуры. Так как для газа эта энергия равна $(1/2)kT$, то по теореме о равномерном распределении то же будет и для любой степени свободы рассматриваемой системы.

Когда обобщенные координаты выбраны так, что в выражение (63.6) входят также попарные произведения обобщенных скоростей, то говорить о распределении кинетической энергии по степеням свободы не имеет смысла. В этом случае теорема о равномерном распределении кинетической энергии по степеням свободы обобщается. Так как $\varepsilon_{\text{кин}}$ является однородной функцией обобщенных скоростей второй степени, то по теореме Эйлера

$$\sum_i \frac{\partial \varepsilon_{\text{кин}}}{\partial \dot{q}_i}\, \dot{q}_i = 2\varepsilon_{\text{кин}}. \tag{63.8}$$

В статистической механике доказывается, что при термодинамическом равновесии средние значения всех слагаемых в левой части одинаковы. Это приводит к результату

$$\frac{1}{2}\left\langle \frac{\partial \varepsilon_{\text{кин}}}{\partial \dot{q}_i}\, \dot{q}_i \right\rangle = \frac{1}{2}\, kT, \tag{63.9}$$

являющемуся обобщением теоремы о равномерном распределении кинетической энергии по степеням свободы.

9. Если смешать два химически не реагирующих идеальных газа с одинаковыми температурами, то при этом средние кинетические энергии поступательного движения молекул каждого газа не изменятся. Иными словами, в результате смешения не изменятся температуры газов. Это утверждение для многоатомных газов совсем не тривиально. Оно является следствием теоремы о равномерном распределении кинетической энергии по степеням свободы. Действительно, температура газа определяется средней кинетической энергией поступательного движения его молекул. Если газ многоатомный, то внутренняя энергия вполне определенным образом распределяется между кинетической энергией поступательного движения, энергией вращения и внутреннего движения молекул. Неизменность температуры означает, что в результате смешения такое распределение остается неизменным

для каждого газа. А это непосредственно вытекает из теоремы о равномерном распределении кинетической энергии по степеням свободы.

Доказательство закона Дальтона для многоатомных газов также основано на той же теореме. Рассмотрим два химически не реагирующих газа. Пусть $\overline{E}_{1\,\text{пост}}$ и $\overline{E}_{2\,\text{пост}}$ — средние кинетические энергии поступательного движения всех молекул этих газов. Пусть до и после смешения газы занимали один и тот же объем V. Тогда до смешения $P_1V = (2/3)\overline{E}_{1\,\text{пост}}$, $P_2V = (2/3)\overline{E}_{2\,\text{пост}}$. Если до смешения температуры газов были одинаковы, то после смешения энергии $\overline{E}_{1\,\text{пост}}$ и $\overline{E}_{2\,\text{пост}}$ не изменятся. Поэтому давление смеси газов P будет определяться соотношением

$$PV = \frac{2}{3}\,\overline{E}_{\text{пост}} = \frac{2}{3}\,(\overline{E}_{1\,\text{пост}} + \overline{E}_{2\,\text{пост}}) = (P_1 + P_2)V.$$

Отсюда $P = P_1 + P_2$, т. е. давление смеси идеальных газов равно сумме парциальных давлений этих газов.

§ 64. Броуновское движение

1. Результаты, изложенные в предыдущем параграфе, нашли блестящее экспериментальное подтверждение в явлении *броуновского движения*. Это явление было открыто в 1827 г. английским ботаником Броуном (1773–1858) во время испытания только что вошедших тогда в употребление ахроматических объективов. Оно заключается в том, что все мельчайшие частицы, взвешенные в жидкости, находятся в непрерывном дрожании. Это движение никогда не прекращается. В кювете, закрытой со всех сторон (во избежание испарения), его можно наблюдать днями, месяцами, годами. Оно обнаруживается в жидких включениях кварца, которым насчитываются тысячи лет. Движение вечно и самопроизвольно.

Броуновское движение в жидкости тем оживленнее, чем меньше вязкость жидкости. Его едва удается подметить в глицерине, а в газах оно, напротив, чрезвычайно интенсивно. Перрену удалось наблюдать броуновское движение капелек, лежащих на «черных пятнах» мыльных пузырей (т. е. на самых тонких участках мыльной пленки). Диаметр этих капелек в 100–1000 раз больше толщины мыльной пленки. Броуновским движением перпендикулярно к пленке можно пренебречь, но в плоскости самой пленки оно происходит чрезвычайно интенсивно (почти так же, как в газе). В одной и той же жидкости броуновское движение происходит тем интенсивнее, чем меньше размеры броуновских частиц. Интенсивность движения увеличивается с повышением температуры жидкости. Две частицы движутся в одной и той же жидкости совершенно одинаково, если одинаковы их размеры и форма: ни вещество частиц, ни его плотность не играют здесь никакой роли.

Движения броуновских частиц, расположенных даже весьма близко друг к другу, совершенно независимы, так что о каких-либо течениях, т. е. конвективном происхождении движения, не может быть

речи. Броуновское движение вызывается толчками, испытываемыми взвешенными частицами со стороны окружающих молекул, совершающих тепловое движение. Толчки иногда в точности не уравновешивают друг друга. В каждый момент времени частица движется в определенном направлении. Спустя короткое время направление равнодействующей силы ударов со стороны окружающих молекул меняется, и частица начинает двигаться в другом направлении. Таким образом, под влиянием ударов молекул окружающей среды скорость броуновской частицы непрерывно и беспорядочно меняется по величине и направлению. Это и есть броуновское движение. Любопытно отметить, что Лукреций (в поэме «О природе вещей») предвидел и описал это явление, но, конечно, не имел возможности его наблюдать.

2. Формула (63.5) лежит в основе количественной теории броуновского движения. Если бы можно было измерить мгновенную скорость броуновской частицы, то по этой формуле можно было бы вычислить постоянную Больцмана k, а по ней и постоянную Авогадро $N_A \approx R/k$. Попытки таких измерений предпринимались, но неизменно приводили к противоречивым результатам. Дело в том, что практически невозможно точно измерить мгновенную скорость частицы V. Если измерить расстояние между двумя положениями броуновской частицы и разделить его на время τ, которое она затрачивает на прохождение из одного положения в другое, то таким путем получится скорость порядка нескольких микрометров в секунду. Это дает для кинетической энергии движения броуновской частицы величину, примерно в 10^5 раз меньшую, чем следует. Как бы мал ни был промежуток времени τ, путь броуновской частицы между рассматриваемыми положениями не прямолинеен, а очень запутан. Он состоит из громадного множества зигзагов, непрерывно и беспорядочно следующих один за другим.

Проверка молекулярно-кинетического объяснения броуновского движения и вычисление из этого явления постоянных k и N_A стали возможными лишь после того, как в 1905 г. Эйнштейн разработал математическую теорию броуновского движения, в которую мгновенная скорость броуновской частицы не входит. Вместо нее входит длина прямолинейного отрезка, соединяющего положение частицы в два различные момента времени, — величина, доступная измерению на опыте. Любопытно отметить, что при разработке своей теории Эйнштейн ничего не знал о существовании броуновского движения. Он предсказал это явление и построил его полную количественную теорию. Польский физик Мариан Смолуховский (1872–1917) в 1906 г. независимо от Эйнштейна также построил количественную теорию броуновского движения, хотя его окончательная формула и является приближенной — она отличается от формулы Эйнштейна числовым коэффициентом порядка единицы. Приведем здесь упрощенный вывод формулы Эйнштейна. В § 93 будет приведен другой вывод, близкий к выводу самого Эйнштейна.

3. Будем считать, что броуновская частица имеет форму шарика радиуса a. Рассмотрим движение ее в жидкости. Если небольшой шар

радиуса a равномерно движется в жидкости со скоростью V, то как показывают опыт и теория, на него действует сила сопротивления F, пропорциональная скорости V. Коэффициент пропорциональности в формуле

$$V = BF \tag{64.1}$$

называется *подвижностью частицы*. Для шарообразной частицы подвижность была теоретически вычислена Стоксом (1819–1903):

$$B = \frac{1}{6\pi\eta a}, \tag{64.2}$$

где η — вязкость жидкости (см. т. I, § 101). Таким образом, подвижность сферической частицы обратно пропорциональна ее радиусу. Она может быть измерена по скорости установившегося движения частицы под действием силы тяжести (точнее, под действием разности силы тяжести и архимедовой подъемной силы). Достаточно измерить подвижность для какой-либо одной частицы при малых числах Рейнольдса ($R \ll 1$). Если радиус ее равен a_0, а подвижность B_0, то подвижность другой частицы радиуса a найдется по формуле $B = (a_0/a)B_0$.

Уравнение движения броуновской частицы в направлении оси X имеет вид

$$M\ddot{x} = -\frac{1}{B}\,\dot{x} + X.$$

Первое слагаемое в правой части есть *регулярная сила трения*, обусловленная движением броуновской частицы со скоростью $\dot{x}$. Второе слагаемое X учитывает *беспорядочно действующие толчки*, которым подвергается броуновская частица со стороны окружающих молекул. В сущности, и первое слагаемое — сила трения — также обусловлено толчками молекул. Однако если частица уже движется, то в среднем толчки, действующие против движения, сильнее толчков, действующих в направлении движения. Это обстоятельство и учитывается слагаемым $-\dot{x}/B$. Слагаемое же X есть сила толчков, которая действовала бы на частицу, если бы она была неподвижна. Среднее значение такой силы равно нулю.

Умножим предыдущее уравнение на x и преобразуем его, пользуясь следующими тождествами:

$$\frac{d}{dt}\,x^2 = 2x\dot{x}, \qquad \frac{d^2}{dt^2}\,x^2 = 2\dot{x}^2 + 2x\dot{x}.$$

Получим

$$M\,\frac{d^2}{dt^2}\,x^2 + \frac{1}{B}\,\frac{d}{dt}\,x^2 - 2M\dot{x}^2 = 2Xx.$$

Будем отсчитывать координату x от положения, которое частица занимала в момент времени $t = 0$. Напишем предыдущее уравнение для каждой из множества тождественных броуновских частиц, сложим и разделим на число всех частиц. Короче говоря, усредним предыдущее уравнение по всем частицам. Ввиду хаотичности молекулярного

движения $\langle Xx\rangle = 0$. Далее, согласно формуле (63.5), $\langle M\dot{x}^2\rangle = kT$. Поэтому

$$M\frac{d^2}{dt^2}\langle x^2\rangle + \frac{1}{B}\frac{d}{dt}\langle x^2\rangle - 2kT = 0. \tag{64.3}$$

4. Нет необходимости решать это уравнение в общем виде. Логичнее пойти по более короткому пути. Докажем, что средний квадрат смещения броуновской частицы $\langle x^2\rangle$ пропорционален времени t. Для этого заметим, что все положения броуновской частицы и все моменты времени совершенно равноправны. Отсюда следует, что смещение броуновской частицы за время $t_2 - t_1$ между двумя моментами времени t_1 и t_2 есть *случайная функция* только разности $t_2 - t_1$, не зависящая ни от t_1, ни от t_2. Слово «случайная» означает, что эта функция еще не определяется значением аргумента $t_2 - t_1$. При одном и том же значении $t_2 - t_1$ смещение частицы может принимать любые значения, но с различной вероятностью. Аргументом $t_2 - t_1$ определяются не сами смещения, а их *вероятности*. Смещения мы будем обозначать через $x_{t_2-t_1}$, т. е. будем писать аргумент $t_2 - t_1$ в виде индекса. Ясно, что сумма смещений частицы за два последовательные промежутка времени — от 0 до t и от t до $t+\tau$ — равна смещению ее за время от 0 до $t+\tau$, т. е.

$$x_{t+\tau} = x_t + x_\tau.$$

Возведем это соотношение в квадрат, усредним и примем во внимание, что $\langle x_t x_\tau\rangle = 0$. Тогда получим

$$\langle x^2_{t+\tau}\rangle = \langle x^2_t\rangle + \langle x^2_\tau\rangle.$$

Усредненная величина $\langle x^2_t\rangle$ есть, очевидно, обычная регулярная функция аргумента t, однозначно определяющаяся значением этого аргумента. Обозначая ее через $f(t)$, запишем предыдущее соотношение в виде

$$f(t+\tau) = f(t) + f(\tau).$$

Из этого функционального уравнения следует, что $f(t)$, т. е. $\langle x^2_t\rangle$, есть линейная однородная функция времени t. Доказанное, очевидно, справедливо для броуновских частиц любой формы, а не только сферических. Итак, должно быть $\langle x^2\rangle = At$. Постоянная A определяется подстановкой этого выражения в уравнение (64.3). В результате получится

$$\langle x^2\rangle = 2kTBt. \tag{64.4}$$

Это и есть *формула Эйнштейна* [1]. В ней x означает смещение частицы только в одном избранном направлении (принятом нами за направление оси X), т. е. x есть проекция полного смещения $\mathbf{r}$ на это

[1] Заметим, что формула, выведенная Смолуховским, отличается от формулы Эйнштейна (64.4) только тем, что вместо множителя 2 стоит множитель 64/27.

направление. Очевидно, $r^2 = x^2 + y^2 + z^2$. Усредняя и принимая во внимание, что $\langle x^2 \rangle = \langle y^2 \rangle = \langle z^2 \rangle$, получим $\langle r^2 \rangle = 3\langle x^2 \rangle$. Поэтому формулу Эйнштейна можно также записать в виде

$$\langle r^2 \rangle = 6kTBt. \tag{64.5}$$

5. Формула (64.4) была со всей возможной тщательностью подтверждена экспериментально французским физиком Жаном Перреном (1870–1942) в ряде работ, начатых в 1908 г. Перрен отмечал через равные промежутки времени ($t = 30$ с) последовательные положения одной какой-либо определенной броуновской частицы в поле зрения микроскопа и соединял эти положения прямолинейными отрезками. Мы воспроизводим один из оригинальных рисунков Перрена (рис. 46). На нем описанным способом зафиксированы пути трех броуновских

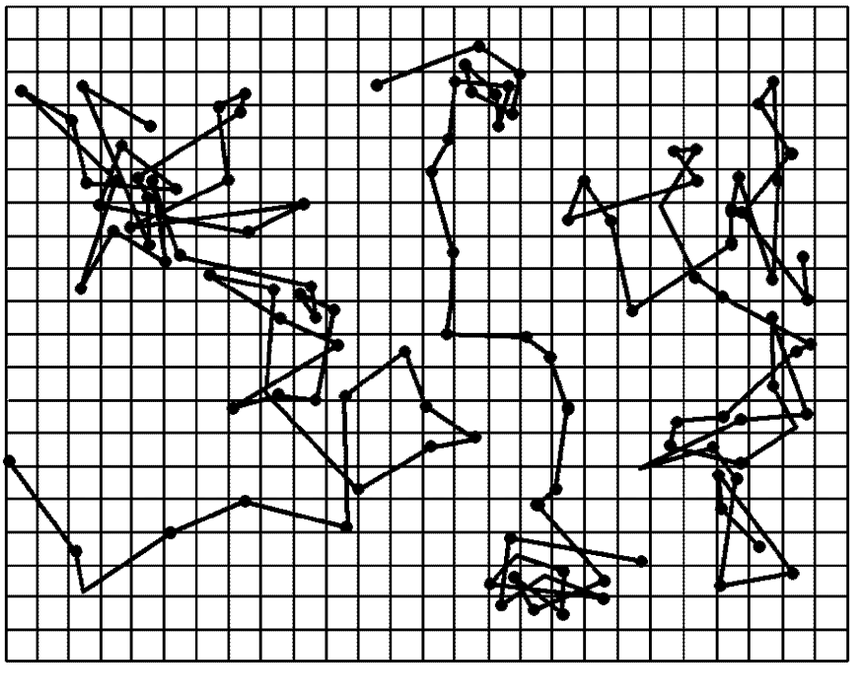


Рис. 46

частиц. Длина 16 клеток рисунка составляет 50 мкм, диаметр броуновской частицы равен 0,53 мкм. Конечно, приведенный рисунок дает только отдельный намек на причудливые изломы действительной траектории частицы. Если бы, например, нанести положения частицы через промежутки времени, в 100 раз более мелкие, то каждый прямолинейный отрезок на рисунке заменился бы соответствующей зигзагообразной ломаной, которая была бы стол же сложна, как и весь рисунок. Отсюда ясно, насколько безнадежно найти истинную скорость броуновской частицы по длине прямолинейного отрезка, проходимого ею за определенный, даже очень короткий, промежуток времени. На рисунке легко измерить проекции рассматриваемых перемещений броуновской частицы на какое-либо направление, например на горизонтальную ось координатной сетки. После этого можно вычислить значение среднего квадрата смещения $\langle x^2 \rangle$ и по формуле (64.4)

найти постоянную Больцмана k и постоянную Авогадро N_A. Перрен получил для этих постоянных значения, согласующиеся в пределах ошибок измерений с другими методами.

§ 65. Вращательное броуновское движение

Вращательное броуновское движение в теоретическом отношении проще поступательного и легче поддается опытному исследованию. Опыт ставится следующим образом.

На очень тонкой кварцевой нити подвешивается маленькое зеркальце. Под действием ударов молекул окружающего газа зеркальце совершает беспорядочные крутильные колебания около положения равновесия. Это и *есть вращательное броуновское движение*. Для его наблюдения на зеркальце направляется световой луч. После отражения от зеркальца луч падает на шкалу. По положению светового зайчика на шкале можно определить угловое положение зеркальца. При повороте зеркальца на некоторый угол на такой же угол закручивается нить. Закрученная нить обладает потенциальной энергией $(1/2)f\varphi^2$, где f — модуль кручения нити, а φ — угол поворота зеркальца из положения равновесия. Если бы не было никаких других сил, то под действием закрученной нити зеркальце совершало бы гармонические крутильные колебания.

При гармонических колебаниях средние значения потенциальной и кинетической энергий равны $(1/2)kT$. Это приводит к формуле $f\langle\varphi^2\rangle = kT$, из которой следует

$$k = \frac{f}{T}\langle\varphi^2\rangle. \tag{65.1}$$

С помощью этой формулы можно вычислить постоянную Больцмана k. Справа стоят величины, которые можно измерить экспериментально. Величину $\langle\varphi^2\rangle$ можно найти, если отмечать положения светового зайчика на шкале через равные промежутки времени. По этим положениям определятся угловые координаты зеркальца, т. е. углы $\alpha_1, \alpha_2, \ldots, \alpha_n$, образуемые плоскостью зеркальца с некоторой фиксированной вертикальной плоскостью. При достаточно большом числе n угловая координата α_0 плоскости зеркальца в положении равновесия найдется как среднее арифметическое углов $\alpha_1, \alpha_2, \ldots, \alpha_n$. После этого найдутся угловые смещения зеркальца из положения равновесия: $\varphi_1 = \alpha_1 - \alpha_0, \ldots, \varphi_n = \alpha_n - \alpha_0$, а затем и

$$\langle\varphi^2\rangle = \frac{\varphi_1^2 + \varphi_2^2 + \ldots + \varphi_n^2}{n}.$$

Для определения модуля кручения нити f поворачивают зеркальце из положения равновесия на угол, большой по сравнению с $\sqrt{\langle\varphi^2\rangle}$. В результате зеркальце начнет совершать правильные крутильные колебания, на которые накладывается броуновское дрожание. Измерив

период этих крутильных колебаний τ, найдем f по формуле

$$\tau = 2\pi\sqrt{I/f},$$

где I — момент инерции зеркальца (см. т. I, § 79). Последний не входит в формулу (65.1). Теоретически зеркальце можно взять сколь угодно большим, его масса, размеры и форма совсем не влияют на величину $\langle\varphi^2\rangle$. Масса зеркальца ограничена лишь прочностью нити, на которой оно подвешено. Кроме того, для справедливости теории необходимо, чтобы масса нити была пренебрежимо мала по сравнению с массой зеркальца. Опыт был поставлен Капплером в 1932 г. Приведем результаты одного из его опытов:

$$T = 287\ \text{К}, \quad f = 9{,}43\cdot 10^{-16}\ \text{Н}\cdot\text{м} = 9{,}43\cdot 10^{-9}\ \text{дин}\cdot\text{см},$$

$$\langle\varphi^2\rangle = 4{,}18\cdot 10^{-6}.$$

Пользуясь ими, находим

$$k = \frac{f}{T}\langle\varphi^2\rangle = \frac{9{,}43\cdot 10^{-16}\cdot 4{,}18\cdot 10^{-6}}{287} = 1{,}37\cdot 10^{-23}\ \text{Дж/К} = 1{,}37\cdot 10^{-16}\ \text{эрг/К}.$$

Это дает для постоянной Авогадро $N_\text{А} = R/k = 6{,}05\cdot 10^{23}\ \text{моль}^{-1}$.

§ 66. Классическая теория теплоемкости идеальных газов

1. *Классическая теория теплоемкости основана на предположении, что к атомно-молекулярным системам применимы законы классической ньютоновой механики.* В действительности применимость ньютоновой механики к атомно-молекулярным системам ограничена. По этой причине классическая теория не смогла дать полного удовлетворительного решения проблемы теплоемкости и была заменена более общей *квантовой теорией*. Однако во многих случаях классическая теория приводила к удивительно хорошему согласию с опытом. Причина этого в том, что классическая теория является приближенным предельным случаем квантовой и, следовательно, имеет определенную область применимости. В пределах этой области выводы классической теории практически не отличаются от выводов квантовой. Мы начинаем изложение с классической теории. Она проще квантовой. При таком порядке изложения отчетливее выявятся принципиальные затруднения классической физики, преодоление которых привело к замене классических представлений квантовыми.

Для классических систем справедлива теорема о равномерном распределении кинетической энергии по степеням свободы. На основе этой теоремы можно построить классическую теорию теплоемкостей газов и твердых тел. Начнем с теплоемкости газов. В § 24 было показано,

что для идеальных газов

$$C_V = \frac{R}{\gamma - 1}, \quad C_P = \gamma \frac{R}{\gamma - 1}. \tag{66.1}$$

Отсюда видно, что постоянная адиабаты γ однозначно определяет обе молярные теплоемкости C_P и C_V идеального газа. Поэтому для сопоставления теории с опытом достаточно сравнивать между собой опытные и теоретические значения только постоянной адиабаты γ.

Внутренняя энергия газа состоит из кинетической энергии поступательного, вращательного и внутреннего движения молекул и атомов, а также из потенциальной энергии их взаимодействия. Для идеальных газов, когда молекулярные силы пренебрежимо малы, потенциальной энергией взаимодействия молекул можно пренебречь.

2. Теплоемкость одноатомных газов. Будем рассматривать молекулы одноатомного газа как материальные точки. Они могут совершать только поступательные движения. Вся внутренняя энергия газа сводится к кинетической энергии поступательного движения атомов. Средняя кинетическая энергия, приходящаяся на один атом, равна $(3/2)\Theta = (3/2)kT$. Для внутренней энергии одного моля газа получаем

$$U = N_\text{A} \cdot \frac{3}{2} kT = \frac{3}{2} RT, \tag{66.2}$$

где N_A — постоянная Авогадро. Отсюда находим молярную теплоемкость при постоянном объеме:

$$C_V = \frac{dU}{dT} = \frac{3}{2} R \approx 12{,}5 \text{ Дж/(моль} \cdot \text{K)} \approx 3 \text{ кал/(моль} \cdot \text{K)} \tag{66.3}$$

и при постоянном давлении:

$$C_P = C_V + R = \frac{5}{2} R \approx 20{,}8 \text{ Дж/(моль} \cdot \text{K)} \approx 5 \text{ кал/(моль} \cdot \text{K)}. \tag{66.4}$$

Показатель адиабаты

$$\gamma = C_P/C_V = 5/3 = 1{,}67. \tag{66.5}$$

Для одноатомных газов экспериментальные значения γ приведены в табл. 3. Согласие с экспериментом очень хорошее.

Таблица 3

Газ	T, K	γ	Газ	T, K	γ
Hg	527	1,666	Ne	292	1,64
He	290	1,660	At	288	1,65
	93	1,673		98	1,69

3. Теплоемкость двухатомных газов. В качестве модели молекулы двухатомного газа принимают две материальные точки *1* и *2*, жестко связанные друг с другом (рис. 47). Такая модель напо-

минает гантель. Для определения ее положения в пространстве достаточно задать пять независимых координат. Действительно, положение первой материальной точки можно задать ее прямоугольными координатами x_1, y_1, z_1; положение второй — прямоугольными координатами x_2, y_2, z_2. Эти шесть величин, однако, не независимы, а связаны соотношением

$$(x_2-x_1)^2+(y_2-y_1)^2+(z_2-z_1)^2 = l_{12}^2 = \text{const},$$

которое означает, что расстояние l_{12} между точками *1* и *2* остается неизменным. Получается, таким образом, пять независимых координат. Значит, наша модель двухатомной молекулы имеет пять степеней свободы.

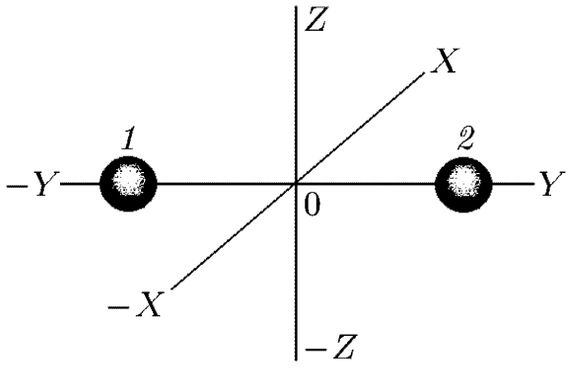


Рис. 47

В классической теории нет необходимости конкретизировать координаты, определяющие конфигурацию молекулы. Надо знать только среднюю кинетическую энергию всей молекулы. А для ее вычисления можно воспользоваться общими формулами (63.8) и (63.9). Они показывают, что средняя кинетическая энергия молекулы равна $(1/2)fkT$, где f — число степеней свободы молекулы (для двухатомной молекулы $f = 5$). Для целей квантовой теории теплоемкостей необходимо, однако, распределить полную кинетическую энергию молекулы по вполне определенным степеням свободы. Удобно в качестве обобщенных координат взять три прямоугольные координаты центра масс двухатомной молекулы и два угла, определяющие направление оси *12*. Кинетическая энергия молекулы слагается из кинетической энергии поступательного движения ее центра масс и кинетической энергии вращения вокруг него:

$$\varepsilon_{\text{кин}} = \frac{1}{2}\,mv^2 + \frac{1}{2}\,I\omega^2.$$

Здесь I — момент инерции молекулы относительно оси, проходящей через 0 перпендикулярно к прямой *12*. Разлагая v и ω на их компоненты, представим $\varepsilon_{\text{кин}}$ в виде суммы пяти членов:

$$E_{\text{кин}} = \frac{1}{2}\,mv_x^2 + \frac{1}{2}\,mv_y^2 + \frac{1}{2}\,mv_z^2 + \frac{1}{2}\,I_x\omega_x^2 + \frac{1}{2}\,I_z\omega_z^2.$$

Эта формула дает разложение величины $E_{\text{кин}}$ на кинетические энергии, соответствующие трем поступательным и двум вращательным степеням свободы. На каждую из этих степеней свободы приходится в среднем кинетическая энергия $(1/2)kT$, и мы приходим к прежнему результату $\overline{\varepsilon}_{\text{кин}} = (5/2)kT$. Внутренняя энергия моля двухатомного газа по классической теории определяется выражением

$$U = N\cdot\frac{5}{2}\,kT = \frac{5}{2}\,Rt. \qquad (66.6)$$

Отсюда находим

$$C_V = \frac{dU}{dt} = \frac{5}{7}\,R \approx 20{,}8 \text{ Дж/(моль} \cdot \text{K)} \approx 5 \text{ кал/(моль} \cdot \text{K)}, \qquad (66.7)$$

$$C_P = C_V + R = \frac{7}{2}\,R \approx 29{,}1 \text{ Дж/(моль} \cdot \text{K)} \approx 7 \text{ кал/(моль} \cdot \text{K)}, \qquad (66.8)$$

$$\gamma = C_P/C_V = 7/5 = 1{,}4. \qquad (66.9)$$

В табл. 4 приведены экспериментальные значения γ для некоторых двухатомных газов.

Таблица 4

Газ	T, K	γ	Газ	T, K	γ
H_2	280	1,407		293	1,398
N_2, 293	1,398		O_2	197	1,411
N_2, 92	1,419			92	1,404

4. Теплоемкость многоатомных газов. Если молекулу рассматривать как твердое тело, то такая модель будет обладать шестью степенями свободы: тремя поступательными и тремя вращательными. Ее средняя кинетическая энергия равна $6 \cdot (1/2)kT = 3kT$. Поэтому для многоатомных газов $U = N \cdot 3kT = 3RT$,

$$C_V = 3R \approx 24{,}9 \text{ Дж/(моль} \cdot \text{K)} \approx 6 \text{ кал/(моль} \cdot \text{K)},$$

$$C_P = 4R \approx 33{,}3 \text{ Дж/(моль} \cdot \text{K)} \approx 8 \text{ кал/(моль} \cdot \text{K)}, \qquad (66.10)$$

$$\gamma = C_P/C_V = 4/3 = 1{,}33.$$

Опыт дает при температуре 292 K для CH_4 $\gamma = 1{,}320$, для SO_2 $\gamma = 1{,}260$.

Допустим теперь, что молекула имеет f степеней свободы и вся энергия ее — кинетическая. Тогда

$$U = \frac{f}{2}\,kTN = \frac{f}{2}\,RT,$$

$$C_V = \frac{f}{2}\,R, \quad C_P = \frac{f+2}{2}\,R, \qquad (66.11)$$

$$\gamma = \frac{f+2}{f}.$$

Кинетическая энергия поступательного движения всех молекул

$$\langle E_{\text{пост}} \rangle = N \cdot \frac{3}{2}\,kT = \frac{3}{f}\,U.$$

Поэтому

$$PV = \frac{2}{3}\,\langle E_{\text{пост}} \rangle = \frac{2}{f}\,U = RT. \qquad (66.12)$$

ЗАДАЧА

Вычислить по классической теории удельные теплоемкости c_v и c_P смеси идеальных газов, состоящей из ν_1 молей одноатомного, ν_2 молей двухатомного и ν_3 молей многоатомных газов. Молекулярные массы газов равны соответственно M_1, M_2, M_3.

Ответ.

$$c_v = \frac{3\nu_1 + 5\nu_2 + 6\nu_3}{2(\nu_1 M_1 + \nu_2 M_2 + \nu_3 M_3)} R, \quad c_P = \frac{5\nu_1 + 7\nu_2 + 8\nu_3}{2(\nu_1 M_1 + \nu_2 M_2 + \nu_3 M_3)} R.$$

§ 67. Адиабатическое нагревание и охлаждение газа с точки зрения молекулярно-кинетической теории

1. Применим полученные результаты к адиабатическому сжатию и расширению идеального газа. Это явление было уже рассмотрено с точки зрения формальной термодинамики (см. § 21), и было показано, что при адиабатическом сжатии газ нагревается, а при адиабатическом расширении — охлаждается. Раскрытие физического механизма этого, как и всякого другого, явления находится вне компетенции формальной термодинамики. Это дело молекулярно-кинетической теории. Чтобы разобраться в механизме явления, допустим, что адиабатическое сжатие или расширение осуществляется перемещением поршня в цилиндре, в котором заключен газ. Если бы поршень оставался неподвижным, то молекулы газа отражались бы от него в среднем с такими же по величине скоростями, какими они обладали до отражения. Для движущегося поршня этого уже не будет. Молекулы, отраженные от движущегося поршня, будут сохранять величину средней скорости только в системе отсчета, в которой поршень покоится. Средние скорости молекул относительно неподвижных стенок цилиндра изменятся. Если поршень вдвигается в цилиндр, то при отражении от него средние скорости молекул увеличиваются — газ нагревается. Если же поршень выдвигается из цилиндра, то они уменьшаются — газ охлаждается. Явление аналогично изменению скорости идеально упругого мяча при отражении от движущейся стенки. Если стенка и мяч движутся навстречу друг другу, то при отражении скорость мяча увеличивается; если же они движутся в одну сторону, то скорость мяча при отражении уменьшается. Так просто и наглядно молекулярно-кинетическая теория объясняет нагревание и охлаждение газа при адиабатическом сжатии и расширении.

2. Нетрудно облечь эти качественные рассуждения в количественную форму и таким путем получить уравнение адиабаты Пуассона. Если поршень вдвигается или выдвигается очень быстро, то термодинамическое равновесие газа уменьшается. При ударе о поршень заметно меняется кинетическая энергия только поступательного движения молекулы; вращательная энергия и энергия внутримолекулярных или внутриатомных движений в среднем остается без изменения. Поэтому при быстром движении поршня в газе возникает макроскопическое

движение — при вдвигании поршня на долю поступательной степени свободы будет приходиться в среднем бóльшая кинетическая энергия, чем на долю вращательной или колебательной степени свободы, а при выдвигании — меньшая. Равномерное распределение кинетической энергии по степеням свободы нарушается. Если остановить поршень, то в результате столкновений между молекулами начнется процесс приближения газа к состоянию термодинамического равновесия. При этом происходит перераспределение кинетической энергии между различными степенями свободы, пока не будет достигнуто равномерное распределение. Однако когда поршень движется медленно (в пределе бесконечно медленно), этот процесс перераспределения можно считать закончившимся в каждый момент времени. Иными словами, в любой момент времени состояние газа может считаться равновесным, а происходящий с ним процесс — квазистатическим.

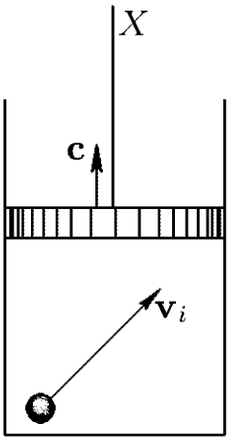


Рис. 48

3. Итак, допустим, что поршень движется бесконечно медленно со скоростью $\mathbf{c}$ (рис. 48). Для простоты будем считать поршень идеально гладким, а отражение молекул от него — зеркальным. Пусть молекула подлетает к поршню со скоростью $\mathbf{v}_i$. Относительно поршня ее скорость будет $\mathbf{v}_{i\,\text{отн}} = \mathbf{v}_i - \mathbf{c}$. Нормальная составляющая относительной скорости $(v_{i\,\text{отн}})_x = v_{ix} - c$. Обозначим через $\mathbf{v}'_{i\,\text{отн}}$ скорость i-й молекулы относительно поршня после отражения. Касательная составляющая относительной скорости в результате отражения не изменится, а нормальная изменит знак, так что

$$(v'_{i\,\text{отн}})_x = -(v_{i\,\text{отн}})_x = -v_{ix} + c.$$

Обозначим далее через $\mathbf{v}'_i$ скорость молекулы относительно неподвижных стенок цилиндра после отражения. Ее нормальная составляющая $v'_{ix} = (v'_{i\,\text{отн}})_x + c = -v_{ix} + 2c$, а касательная составляющая такая же, что и у скорости $\mathbf{v}_i$. В результате отражения от поршня кинетическая энергия молекулы получит приращение

$$\frac{1}{2}\,m(-v_{ix} + 2c)^2 - \frac{1}{2}\,mv_{ix}^2 = -2mcv_{ix} + 2mc^2.$$

Слагаемым $2mc^2$ можно пренебречь как величиной второго порядка малости по c. Обозначим через n_i число молекул в единице объема, скорости которых равны или, лучше, приблизительно равны $\mathbf{v}_i$. Число ударов таких молекул о поршень за время dt равно $z_i = Sn_i(v_{ix} - c)\,dt$, где S — площадь поршня. Здесь также можно пренебречь скоростью c как величиной, бесконечно малой по сравнению с v_{ix}, т. е. положить $z_i = Sn_iv_{ix}\,dt$. В результате кинетическая энергия молекул рассматриваемой группы за время dt получит приращение

$$-2mcv_{ix}z_i = -2mn_iv_{ix}^2Sc\,dt = -2mn_iv_{ix}^2\,dV,$$

где $dV = Sc\,dt$ — приращение объема газа за то же время. Приращение кинетической энергии всего газа

$$dE_{\text{кин}} = dU = -dV \sum_{v_{ix}>0} 2mn_i v_{ix}^2.$$

Здесь суммирование ведется только по тем группам молекул, которые движутся по направлению к поршню. Если же суммировать по всем группам молекул, движущимся как к поршню, так и от него, то сумму надо разделить пополам. При таком понимании суммирования

$$dU = -dV \sum mn_i v_{ix}^2.$$

В § 59 было показано, что входящая сюда сумма равна давлению газа P. Поэтому

$$dU + P\,dV = 0.$$

Подставив сюда значение U из формулы (66.12), получим

$$\left(\frac{f}{2} + 1\right) P\,dV + \frac{f}{2}\,V\,dP = 0,$$

или на основании последнего из соотношений (66.11)

$$\gamma P\,V + V\,P = 0.$$

Это — *дифференциальное уравнение* адиабаты, полученное в § 21 чисто термодинамически. Так как по классической теории теплоемкости γ — величина постоянная, то интегрированием этого уравнения получается уравнение Пуассона

$$PV^\gamma = \text{const}.$$

4. При бесконечно медленном движении поршня каждое отражение молекулы сопровождается бесконечно малым изменением ее скорости. Возникает вопрос, каким образом в этих условиях может получиться конечное изменение температуры газа. На этот вопрос легко ответить. Дело в том, что чем медленнее движется поршень, тем больше требуется времени, чтобы объем газа изменился на заданную величину. За это время произойдет бóльшее число ударов молекул о поршень, чем при более быстром движении. При каждом отражении изменение энергии молекулы тем меньше, чем медленнее движется поршень. Однако произведение числа ударов на среднее изменение энергии молекулы при одном отражении от скорости движения поршня не зависит (если только процесс может считаться квазистатическим). Оно определяется только величинами начального и конечного объемов газа. Поэтому приращение кинетической энергии теплового движения газа определяется только приращением его объема и совершенно не зависит от того, быстро или медленно двигался поршень (при условии, что процесс — квазистатический).

5. Термин «адиабатический» применяется в физике в двух смыслах. В термодинамике адиабатическим называют процесс, происходящий без подвода и отвода тепла. Он может быть как равновесным, так и неравновесным. *В механике под адиабатическим воздействием на систему понимают такое воздействие, при котором ее внешние*

параметры меняются бесконечно медленно. Рассмотрим, например, математический маятник — шарик, подвешенный на нерастяжимой нити, перекинутой через гвоздь. Внешними параметрами такой системы являются длина нити l и ускорение свободного падения g. Потянув рукой за свободный конец нити, можно менять l. Можно также менять значение g, перемещая маятник вверх или вниз. Если эти изменения производятся достаточно медленно, то воздействия на маятник будут адиабатическими. *Функции динамических переменных систем, остающиеся постоянными при адиабатических воздействиях на нее, называются адиабатическими инвариантами* (см. т. I, § 43). В этом смысле величина PV^γ является адиабатическим инвариантом теплоизолированной системы, состоящей из идеального газа. Если газ в цилиндре теплоизолирован, но поршень движется быстро, то величина PV^γ в ходе процесса, вообще говоря, не будет оставаться постоянной. Более того, поскольку при быстрых движениях поршня процесс будет неравновесным, газу в целом нельзя будет приписать какое-либо определенное давление P. Если остановить поршень и подождать, чтобы газ пришел в состояние равновесия, то даже в этом случае PV^γ, вообще говоря, изменится. Допустим, например, что поршень выдвигается со скоростью, в несколько раз прерывающей среднюю скорость теплового движения молекул. Тогда подавляющая доля молекул не сможет «догнать» поршень и отразиться от него. Процесс будет происходить так же, как расширение газа в пустоту. При этом сохраниться постоянной внутренняя энергия газа, а с ней и произведение PV. Величина же PV^γ изменится.

6. Из приведенного рассуждения ясно, что *при одном и том же увеличении объема понижение температуры газа будет наибольшим, когда расширение производится квазистатически*. С формально термодинамической точки зрения понижение температуры газа объясняется работой, которую он вынужден совершать при расширении. В технике квазистатическое адиабатическое расширение газа с производством внешней работы используется для получения низких температур (см. § 105). В отношении этого способа необходимо заметить, что по мере понижения температуры газа его давление уменьшается и может оказаться недостаточным для преодоления внешних сил. Казалось бы, что дальнейшее понижение температуры указанным методом становится невозможным. Это неверно. Для преодоления внешних сил вовсе не обязательно использовать давление самого газа. Важно только заставить газ расширяться, приведя в движение одну из стенок сосуда (поршень), в котором он заключен. А для этого можно воспользоваться каким-либо двигателем. При отражении от движущейся стенки, как ясно из молекулярно-кинетического рассмотрения, приведенного выше, газ будет продолжать охлаждаться. И это охлаждение в принципе может продолжаться до тех пор, пока газ не перейдет в жидкое состояние. Конечно, и в этом случае охлаждение происходит за счет работы, производимой газом. Но это есть *вынужденная работа*, возможная только потому, что двигатель приводит в движение

поршень, от которого отражаются молекулы газа. В отсутствие двигателя разреженный газ не смог бы произвести работу, так как его давление недостаточно для преодоления внешнего давления и различного рода вредных сопротивлений.

ЗАДАЧА

Оценить порядок величины максимальной скорости, с которой артиллерийский снаряд может вылететь из ствола орудия. Какие требования надо предъявлять к пороху, чтобы эта скорость была возможно большей?

Решение. Когда снаряд движется в стволе орудия со скоростью, превышающей скорость теплового движения молекул пороховых газов, последние почти перестают оказывать давление на дно снаряда и ускорять его. Отсюда следует, что максимально достижимая скорость снаряда при вылете из ствола орудия будет порядка средней скорости теплового движения молекул пороховых газов. Она тем больше, чем выше температура пороховых газов и чем меньше их молекулярная масса.

§ 68. Классическая теория теплоемкости твердых тел (кристаллов)

1. Простейшей моделью кристалла является правильно построенная кристаллическая решетка, в узлах которой помещаются атомы, принимаемые за материальные точки. Атомы совершают тепловые колебания около положений равновесия. Если колебания малы, то они будут гармоническими. Энергия каждого атома слагается из кинетической и потенциальной. На каждую степень свободы приходится в среднем кинетическая энергия $(1/2)kT$. Как было показано в § 63, при гармонических колебаниях на одну степень свободы приходится в среднем такая же потенциальная энергия, т. е. $(1/2)kT$. Таким образом, среднее значение полной энергии, приходящейся на одну колебательную степень свободы, равно

$$\overline{\varepsilon}_{\text{кол}} = \overline{\varepsilon}_{\text{кин}} + \overline{\varepsilon}_{\text{пот}} = kT. \tag{68.1}$$

Теперь легко рассчитать теплоемкость кристаллической решетки. Для простоты будем считать, что все атомы одинаковы. Каждый атом обладает тремя колебательными степенями свободы, а потому на него приходится средняя энергия $3kT$. Умножив эту величину на постоянную Авогадро $N_{\text{А}}$, найдем внутреннюю энергию моля твердого тела $U = N \cdot 3kN_{\text{А}} = 3RT$. Отсюда для молярной теплоемкости твердого тела получаем

$$C_V = dU/dt = 3R \approx 24{,}9 \text{ Дж/(моль} \cdot \text{К)} \approx 6 \text{ кал/(моль} \cdot \text{К)}. \tag{68.2}$$

Еще в 1819 г. Дюлонг (1785–1838) и Пти (1791–1820) установили эмпирическое правило, согласно которому *произведение удельной теплоемкости химического элемента в твердом состоянии на его атомную массу приблизительно одинаково для всех элементов и составляет около* 6 кал/(моль · К). Мы видим, что правило Дюлонга и Пти находит простое объяснение в классической теории теплоемкостей. Вывод

показывает, что в правиле Дюлонга и Пти речь идет о молярной теплоемкости при постоянном объеме. В табл. 5 приведены молярные

Таблица 5

Элемент	C_V, кал/(моль · К)	Элемент	C_V, кал/(моль · К)
C	1,44	Pt	6,11
B	2,44	Au	5,99
Al	5,51	Pb	5,94
Ca	5,60	U	6,47
Al	6,11		

теплоемкости некоторых элементов в твердом состоянии в температурном интервале от 15 до 100 °C.

2. Пусть теперь твердое тело является химическим соединением, например NaCl. Его кристаллическая решетка построена из атомов различных типов. Очевидно, молекулярная масса химического соединения равна сумме атомных масс всех атомов, из которых состоит молекула этого соединения. Для применимости теоремы о равномерном распределении энергии по степеням свободы не имеет значения, одинаковы или различны атомы. Каждый атом обладает тремя колебательными степенями свободы, и на него в среднем приходится энергия $3kT$. Если в молекуле n атомов, то на молекулу придется в среднем энергия $3nkT$. Молярная теплоемкость будет $3nkN = 3nR$, т. е. она в n раз больше, чем у того же вещества, если бы его молекулы были отдноатомны. Иными словами, *молярная теплоемкость твердого соединения равна сумме молярных теплоемкостей элементов, из которых оно состоит*. Это правило было найдено эмпирически и называется законом Джоуля и Коппа. Джоуль высказал его в 1844 г. Но только в 1864 г. закон был окончательно сформулирован Коппом и подтвержден множеством фактов, полученных самим Коппом. Заметим, что закон Джоуля–Коппа в приведенной выше формулировке: «Молярная теплоемкость твердого соединения приблизительно равна сумме молярных теплоемкостей элементов, из которых оно состоит» — обладает большей общностью, чем правило Дюлонга и Пти. Правило Дюлонга и Пти может нарушаться, т. е. молярные теплоемкости химических элементов, входящих в соединение, могут отличаться друг от друга, но тем не менее закон Джоуля и Коппа может оставаться справедливым. Именно это и было установлено Коппом.

§ 69. Недостаточность классической теории теплоемкостей. Понятие о квантовой теории (качественное рассмотрение)

1. Сравнение классической теории теплоемкости с опытом показывает, что она в основном правильно описывает определенный круг явлений. Однако многие явления она не объясняет. Ряд опытных фактов находится в резком противоречии с этой теорией.

Прежде всего, *классическая теория не дает объяснения зависимости теплоемкости тел от температуры*. В табл. 6 приведены для

Таблица 6

T, К	C_V, кал/(моль · К)	T, К	C_V, кал/(моль · К)
35	2,98	1200	5,49
100	3,10	1600	5,76
290	4,90	1800	5,92
600	5,08	2000	6,06
800	5,22	2500	6,40
1000	5,36		

примера молярные теплоемкости газообразного водорода при различных температурах.

Можно было бы попытаться объяснить зависимость теплоемкости от температуры негармоничностью колебательных степеней свободы при больших амплитудах колебаний. Например, в случае молекулы водорода взаимодействие атомов приводит к их колебаниям вдоль оси YY (см. рис. 47). Для негармонических колебаний средняя кинетическая энергия, приходящаяся на одну степень свободы, уже не равна соответствующей средней потенциальной энергии. Соотношение между ними зависит от амплитуды колебаний, т. е. в конце концов от температуры газа. При высоких температурах учет колебаний улучшает дело, поскольку теплоемкость, связанная с колебательной степенью свободы, меняется с изменением температуры. Однако эти соображения теряют силу при низких температурах, где расхождения классической теории с опытом проявляются особенно резко. При низких температурах, согласно представлениям классической теории, амплитуды колебаний малы, а потому сами колебания могут считаться гармоническими. В этом случае для теплоемкости C_V по классической теории мы получили бы 7 кал/(моль · К) вместо экспериментального значения 3 кал/(моль · К). Опыт показывает, что ниже 100 К водород начинает вести себя как одноатомный газ. К этому следует добавить экспериментально установленный факт, что при приближении к абсолютному нулю теплоемкости C_V и C_P всех тел стремятся к нулю.

2. Классическая теория непоследовательна. По теореме о равномерном распределении кинетической энергии все степени свободы равноправны. Поэтому требуется лишь подсчитать *полное число* степеней свободы, совсем не обращая внимания на их природу. Между тем классическая теория по каким-то непонятным причинам учитывает одни и отбрасывает другие степени свободы. Так, атом одноатомного газа классическая теория рассматривает как материальную точку с тремя степенями свободы и этим достигает известного согласия с опытом. Но атом — не точка. Если принять его за твердое тело, то получится шесть степеней свободы — три поступательные и три вращательные. Теоретическая теплоемкость C_V одноатомного газа возрастает до 6 кал/(моль · К). Атом не является также твердым телом,

а имеет внутреннюю структуру. Число степеней свободы его больше шести. Поэтому при последовательном рассмотрении теплоемкость C_V одноатомного газа по классической теории должна была бы быть много больше 6 кал/(моль · К), а это противоречит фактам.

Возьмем, далее, двухатомную молекулу. Классическая теория учитывает вращения ее вокруг осей XX и ZZ (см. рис. 47), но не принимает во внимание вращение вокруг оси YY. Последнее в классической теории исключается на том основании, что не имеет смысла говорить о вращении материальных точек *1* и *2* вокруг соединяющей прямой *12*. Но такая аргументация формальна и неубедительна, так как атомы не являются материальными точками. Если атомы *1* и *2* рассматривать как твердые тела, то каждый из них будет иметь 6, а вся молекула — 12 степеней свободы. Им соответствует в среднем кинетическая энергия $6kT$. К этой энергии следует прибавить еще среднюю потенциальную энергию колеблющихся атомов вдоль прямой *12*, которая равна $(1/2)kT$. Всего мы получаем, следовательно, $(13/2)kT$, а потому теплоемкость C_V для такой модели двухатомного газа должна быть равна 13 кал/(моль · К). При учете внутренней структуры атомов это значение еще больше возрастает.

3. Определенное согласие с опытом достигается в классической теории благодаря использованию механических моделей с наложенными связями, ограничивающими свободу движения. Идея связей заимствована из теоретической механики. Так она служит искусственным приемом для решения различных задач о равновесии и движении идеализированных макроскопических механических систем. В действительности в макроскопических телах никаких связей не существует. Тем более их не существует в атомных системах. Двухатомная молекула, например, часто рассматривается как неизменяемая система двух материальных точек, скрепленных жестким невесомым стержнем. На самом деле это, конечно, не так — никаких жестких стержней, скрепляющих атомы в молекулах, нет. Речь идет об идеализированной модели атомной системы. Для макроскопических систем физика в состоянии раскрыть физическую природу связей и установить, когда можно пользоваться такой идеализацией. Но классическая физика не может ответить на вопрос, почему в определенной области явлений атомные системы приближенно ведут себя как макроскопические модели с наложенными связями.

4. Рассмотрим, наконец, теплоемкость металлов. Металл состоит из положительно заряженных ионов, совершающих тепловые колебания вокруг узлов кристаллической решетки. Между ними движутся так называемые свободные электроны, т. е. электроны, сравнительно слабо связанные с ионами решетки. Они ведут себя подобно электронному газу. Наличием свободных электронов объясняется высокая электрическая проводимость металлов. По значению электрической проводимости можно оценить концентрацию свободных электронов. Она оказалась того же порядка, что и концентрация ионов, образующих

кристаллическую решетку. Классическая теория теплоемкости отвлекается от наличия электронного газа. Она учитывает тепловые колебания одних только ионов и благодаря этому приходит к правильному значению для теплоемкости $C_V = 6$ кал/(моль · К) (правило Дюлонга и Пти). Между тем следовало бы учесть также вклад в теплоемкость, вносимый электронами. Если электроны принять за материальные точки, то на каждый свободный электрон будет приходиться средняя кинетическая энергия $(3/2)kT$. Поэтому по классической теории теплоемкость электронного газа должна была бы быть сравнимой с теплоемкостью решетки. Опыт показывает, что свободные электроны практически не вносят никакого вклада в теплоемкость металлов.

5. Таким образом, опытные факты приводят к заключению, что всегда явления протекают так, что *эффективный вклад в теплоемкость вносят не все, а только некоторые степени свободы*. При понижении температуры некоторые степени свободы становятся малоэффективными и, наконец, совсем выпадают из игры. Про такие степени свободы говорят, что они «заморожены». Наоборот, при повышении температуры начинают проявляться все новые и новые степени свободы, которые ранее были либо малоэффективны, либо совсем заморожены. Это значит, что классический закон равномерного распределения кинетической энергии по степеням свободы не справедлив и требует уточнения. Парадоксально, что успехи самой классической теории связаны с тем, что она в применении к конкретным системам фактически отказывалась от этого закона. Она исключала некоторые степени свободы путем наложения на систему идеально жестких связей. Фактически это эквивалентно тому, что такие степени свободы считаются замороженными.

6. Трудности такого рода были преодолены после того, как теория теплоемкости была построена на *квантовой основе*. Не вдаваясь в детали, ограничимся здесь немногими качественными указаниями. В § 58 уже говорилось, что внутренняя энергия атомных систем может принимать лишь *дискретные значения*. Приведем в качестве примера гармонический осциллятор, т. е. частицу, которая по классическим представлениям может совершать гармонические колебания под действием квазиупругой силы — силы, пропорциональной отклонению частицы от положения равновесия. Двухатомная молекула, когда речь идет о малых колебаниях ее атомов относительно друг друга, может рассматриваться как гармонический осциллятор. Как доказывается в квантовой механике, возможные значения колебательной энергии такой системы представляются формулой

$$\varepsilon_n = (n + 1/2)\, h\nu, \tag{69.1}$$

где ν — частота осциллятора, n — целое число, которое может принимать значения $0, 1, 2, \ldots$, h — универсальная постоянная, называемая

постоянной Планка. По современным данным

$$h = (6{,}626176 \pm 0{,}000036) \cdot 10^{-34} \text{ Дж} \cdot \text{с} = \\ = (6{,}626176 \pm 0{,}000036) \cdot 10^{-27} \text{ эрг} \cdot \text{с}.$$

Таким образом, энергетический спектр гармонического осциллятора состоит из бесконечного множества равноотстоящих уровней. Расстояние между соседними уровнями равно $h\nu$. Самому низкому уровню соответствует энергия $\varepsilon_0 = (1/2)h\nu$. Она называется *нулевой энергией*. Наличие нулевой энергии означает, что даже в состоянии с наименьшей энергией колебания осциллятора не прекращаются. Такие колебания называются *нулевыми*. Воздействуя на осциллятор, его можно возбудить, т. е. перевести на один из возможных более высоких уровней энергии. Ближайшим является энергетический уровень с $n = 1$ и энергией $\varepsilon_1 = (3/2)h\nu$.

7. Допустим теперь, что газ состоит из гармонических осцилляторов, например двухатомных молекул. Предположим, что температура газа настолько низка, что $kT \leqslant h\nu$. Средняя энергия теплового движения молекулы порядка kT. Такой энергии недостаточно, чтобы возбудить осциллятор, т. е. перевести его с нулевого уровня на ближайший — первый — энергетический уровень. Возбуждение может происходить только при столкновениях с молекулами, энергия которых значительно больше средней. Однако таких молекул относительно мало, так что практически все осцилляторы останутся на нижнем энергетическом уровне. Эта картина сохранится при дальнейшем повышении температуры газа, пока соблюдается условие $kT \leqslant h\nu$. При выполнении этого условия колебательная энергия осцилляторов практически не зависит от температуры и по этой причине не влияет на теплоемкость газа. Вот почему при условии $kT \ll h\nu$ в теории теплоемкости двухатомных газов можно не учитывать колебания атомов *1* и *2* (см. рис. 47) вдоль соединяющей их прямой. Отличие квантовой от формальной классической теории состоит здесь в том, что согласно квантовой теории должны происходить нулевые колебания, тогда как формальная классическая теория совсем исключает их, считая молекулы абсолютно жесткими. Однако это различие не может сказаться на величине теплоемкости, так как *энергия нулевых колебаний не зависит от температуры*. Моделью жесткой двухатомной молекулы в теории теплоемкости можно пользоваться лишь при соблюдении условия $kT \leqslant h\nu$. Если это условие не соблюдается, то классическая модель становится неприменимой. При повышении температуры, когда величина kT становится сравнимой с $h\nu$, начинает возбуждаться первый, а затем и более высокие колебательные уровни двухатомной молекулы. Температура

$$T_v = \frac{h\nu}{k} \tag{69.2}$$

называется *характеристической температурой*. При $T \gtrsim T_v$ колебания существенно влияют на теплоемкость двухатомного газа. При

$T \ll T_v$ происходят только нулевые колебания, не сказывающиеся на теплоемкости. Для водорода $T_v \sim 6000$ К.

8. Совершенно аналогично влияет на теплоемкость газов и вращение молекул. Энергия вращения также квантуется. Ее возможные значения по квантовой механике определяются формулой

$$\varepsilon_l = \frac{h^2}{8\pi^2 I}\, l(l+1), \tag{69.3}$$

где I — момент инерции молекулы, а l — целое число, которое может принимать значения $0, 1, 2, \ldots$ Расстояния между уровнями энергии не постоянны, а возрастают в арифметической прогрессии. В состоянии с наименьшей энергией $l = 0$ вращения не возбуждены. В состоянии с $l = 1$ энергия вращения

$$\varepsilon_1 = \frac{h^2}{4\pi^2 I}. \tag{69.4}$$

Если $kT \ll \varepsilon_1$, то средней тепловой энергии молекул недостаточно для возбуждения вращений. В этом случае при вычислении теплоемкости вращения можно не учитывать. В противоположном случае, когда $kT \gg \varepsilon_1$, возбуждается много вращательных уровней. Тогда дискретность энергетических уровней слабо сказывается на теплоемкости. Для вращений становится приближенно применимой классическая теорема о равномерном распределении кинетической энергии по степеням свободы. Характеристическая температура для вращения молекул определяется формулой

$$T_r = \frac{\varepsilon_1}{k} = \frac{h^2}{4\pi^2 I k}. \tag{69.5}$$

При $T \gg T_r$ справедлива классическая теория, при $T \ll T_r$ вращения не возбуждены и не оказывают влияния на теплоемкость. Для водорода при вращении молекул вокруг осей XX и ZZ (см. рис. 47) характеристическая температура $T_r \sim 175$ К. Вращения вокруг оси XX не возбуждены из-за малости соответствующего момента инерции. По той же причине не возбуждены вращения атомов одноатомных газов.

9. Изложенные соображения применимы не только к колебаниям гармонических осцилляторов и вращениям жестких молекул, но и к любым квантовым системам. Они показывают, что *дискретность энергетических уровней несовместима с классической теоремой о равномерном распределении энергии по степеням свободы*. Только тогда, когда средняя энергия теплового движения kT велика по сравнению с разностями между высшими энергетическими уровнями и наинизшими из них, возбуждается много энергетических уровней. При таком условии дискретность уровней становится малосущественной и атомная система ведет себя как классическая, в которой энергия меняется непрерывно. Отсюда следует, что *чем выше температура, тем лучше оправдывается классическая теорема о равномерном распределении энергии по степеням свободы*.

10. Мы не рассматриваем здесь квантовую теорию теплоемкости с количественной стороны. Об этом будет идти речь в гл. VI, § 85 и более подробно в т. V, § 54. Однако уже качественное рассмотрение показывает, что одного представления о дискретности энергетических уровней достаточно, чтобы выяснить, в каких случаях можно и в каких случаях нельзя пользоваться классической теорией теплоемкости и ее грубыми механическими моделями. В качестве первого примера оценим молярную теплоемкость кислорода, нагретого до температуры 1000 эВ и выше. При комнатной температуре все атомы находятся в основном — низшем — состоянии, электронные уровни не возбуждены. Не возбуждены также колебания атомов в молекуле — молекула кислорода ведет себя как жесткая двухатомная молекула. При нагревании молекулы кислорода сначала диссоциируют, т. е. распадаются на два атома. Затем начинается процесс ионизации, т. е. отрыв электронов от атомов. Сначала оторвутся внешние электроны, наименее прочно связанные с атомными ядрами. При дальнейшем нагревании начнется отрыв и внутренних электронов. Для отрыва последнего электрона требуется энергия порядка 870 эВ. При температурах 1000 эВ и выше практически все электроны окажутся оторванными от атомных ядер. Вещество перейдет в состояние полностью ионизованной плазмы, состоящей из электронов и «голых» атомных ядер. Из каждого атома образуется 9 частиц: ядро и 8 электронов; из каждой молекулы — 18 частиц: два ядра и 16 электронов. Если пренебречь потенциальной энергией взаимодействия частиц, то вся внутренняя энергия сведется к кинетической энергии теплового движения электронов и атомных ядер. Средняя энергия одной частицы $(3/2)kT$, средняя энергия частиц, образовавшихся из молекулы, $18 \cdot (3/2)kT = 27kT$, внутренняя энергия одного моля $U = 27NkT = 27RT$, а молярная теплоемкость $C_V = 27R \approx 54$ кал/(моль · К).

11. В качестве второго примера возьмем вращательное броуновское движение, рассмотренное в § 65. Выясним, применима ли к такому движению классическая теорема о равномерном распределении кинетической энергии по степеням свободы. Зеркальце можно рассматривать как гармонический осциллятор с собственной частотой

$$\nu = \frac{1}{2\pi}\sqrt{\frac{f}{I}}.$$

Если взять $I \sim 0{,}01$ г · см^2, то получится $\nu \sim 1{,}5 \cdot 10^{-4}$ с$^{-1}$, $h\nu \sim$ $\sim 10^{-31}$ эрг, тогда как средняя энергия теплового движения $kT \sim$ $\sim 4 \cdot 10^{-14}$ эрг. Таким образом, $h\nu/kT \sim 2{,}5 \cdot 10^{-18}$, т. е. с громадным запасом $kT \gg h\nu$. Применимость теоремы о равномерном распределении энергии здесь не вызывает сомнений.

Глава VI

СТАТИСТИЧЕСКИЕ РАСПРЕДЕЛЕНИЯ

§ 70. Элементарные сведения из теории вероятностей

1. С молекулярной точки зрения физические величины, встречающиеся в термодинамике, как и в любом другом отделе макроскопической физики, имеют смысл *средних значений*, которые принимают при определенных условиях какие-то функции микросостояния системы (см. § 9). Про величины такого рода говорят, что они имеют *статистический характер* или являются *статистическими*. Примеры таких величин (давление, плотность, температура, средний квадрат смещения частицы при броуновском движении и пр.) и способы их вычисления приведены в предыдущей главе. То обстоятельство, что эти величины подчиняются определенным закономерностям, не свойственным отдельным атомам и молекулам, связано с колоссальным количеством таких частиц в макроскопических телах. Такие закономерности, равно как и любые закономерности, обусловленные массовостью участвующих в их возникновении ингредиентов называются *статистическими* или *вероятностными закономерностями*.

Допустим, например, что бросается монета. Что выпадет в результате бросания — герб или решка — это предсказать невозможно. Хотя движение монеты и строго подчиняется законам механики, но на это движение, а также на начальные условия влияет множество случайных и неконтролируемых факторов, которые делают результат бросания непредсказуемым. Однако если бросаний произведено очень много, то числа выпавших гербов и решек окажутся почти равными. И это равенство будет выполняться тем точнее, чем больше произведено бросаний. В приведенном примере и проявляется статистическая закономерность. На том же примере видно, что предсказания, которые делаются на основе статистических законов, не являются абсолютно достоверными, а носят характер прогнозов, которые могут и не оправдываться. Почти все законы макроскопической физики — статистические. Однако колоссальность количества молекул и атомов в макроскопических телах превращает статистические законы физики и основанные на них предсказания в практически абсолютно достоверные.

Классическая физика считала, что за статистическими или вероятностными законами, управляющими поведением макроскопических систем, стоят точные *динамические законы*, которым подчиняются отдельные атомы, молекулы и составляющие их частицы. Квантовая физика идет дальше. Она утверждает, что и элементарные законы

микромира являются также законами статистическими. С ее точки зрения не существует строго динамических законов — все законы статистические. Однако здесь нет необходимости вдаваться в обсуждение этих вопросов. Для наших целей пока достаточна классическая точка зрения.

С чисто математической точки зрения, отвлекающейся от конкретного смысла рассматриваемых величин, статистические закономерности изучаются *теорией вероятностей*. Ниже приводятся самые элементарные сведения из теории вероятностей, необходимые для дальнейшего изложения.

2. Современная математическая теория вероятностей строится как абстрактная аксиоматическая наука. Под вероятностями понимают некоторые числа, подчиняющиеся определенной системе аксиом, из которых формально логически выводится все остальное. Вопрос о конкретном смысле вероятностей в абстрактной теории не ставится. Он решается отдельно в конкретных случаях, когда от теории переходят к ее приложениям. В физике, как и во всех прикладных вопросах, более предпочтителен другой подход к теории вероятностей, в котором вероятность органически связана с ее конкретной интерпретацией. Такой подход характерен для всей теории вероятностей, какой она существовала примерно до 20-х годов прошлого столетия. Этот подход встретил серьезные и обоснованные возражения со стороны математиков. Однако при первоначальном знакомстве с элементами теории вероятностей всякий иной подход нецелесообразен.

3. *Событиями или случаями в теории вероятностей называют всякие явления, относительно которых имеет смысл ставить вопрос, могут они происходить или нет.* Опыт или совокупность условий, в результате которых появляется то или иное событие, в теории вероятностей называется *испытанием*.

Если при данных условиях событие обязательно произойдет, то оно называется *достоверным*. Если же оно произойти не может, то его называют *невозможным*. Допустим, например, что мы чертим треугольник на бумаге. Событие, состоящее в том, что при этом получится треугольник, у которого каждая сторона короче суммы двух других сторон, есть достоверное событие. Появление треугольника, у которого одна из сторон длиннее суммы двух других, есть также событие, хотя и невозможное.

Событие называется случайным, если в результате испытания оно может как произойти, так и не произойти. Например, при игре в орлянку может выпасть либо герб, либо решка. Это — случайные события. Положим в урну несколько занумерованных, в остальном совершенно одинаковых шаров и тщательно перемешаем их. Если наугад вынуть один шар, то появление шара с определенным номером будет также случайным событием.

4. *Суммой двух событий A и B называется событие, состоящее в появлении либо события A, либо события B (без указания, какого именно).* Так, если в урне лежат красный, зеленый и белый шары, то

появление при вынимании цветного шара есть сумма двух событий: 1) появление красного шара и 2) появление зеленого шара. Сумма событий A и B обозначается $A + B$. *Произведением событий A и B называется событие, состоящее в появлении как события A, так и события B.* Так, если монета бросается два раза, то появление при первом бросании герба, а при втором решки есть произведение двух событий: 1) появление при первом бросании герба и 2) появление при втором бросании решки. Произведение событий A и B обозначается через AB.

События $A_1, A_2, \ldots, A_n$ называются единственно возможными, если при данном испытании одно из них (неизвестно — какое) обязательно должно произойти. Очевидно, сумма всех единственно возможных событий есть событие достоверное. *События $A_1, A_2, \ldots, A_n$ называются несовместимыми, если появление одного из них исключает появление любого из остальных.* Очевидно, произведение всех несовместимых событий есть событие невозможное. *Два случайных события называются равновозможными или равновероятностными, если нет никаких оснований ожидать, что при испытаниях одно из них будет появляться чаще другого.* Например, выпадение герба или решки при бросании монеты — равновозможные события. *Несколько событий называются равновозможными, если каждые два из них равновозможны.*

5. *Вероятность случайного события есть количественная мера ожидаемой возможности его появления.* Для введения этой меры рассмотрим сначала n единственно возможных, несовместимых и равновозможных событий $A_1, A_2, \ldots, A_n$. Вероятностью каждого из них называют дробь $1/n$. Например, если в урне лежат 100 тщательно перемешанных одинаковых занумерованных шаров, то вероятность вынуть наугад шар с номером 1 равна $1/100$.

Распространим теперь понятие вероятности на случай, когда единственно возможные и несовместимые события $A_1, A_2, \ldots, A_n$ не равновозможны, но могут быть представлены в виде суммы равновозможных событий, представляющих их частные случаи. Пусть, например, событие A_i разложено на m_i единственно возможных несовместимых и равновозможных событий $A_{i\,1}, A_{i\,2}, \ldots, A_{n m_n}$. Очевидно, все события $A_{1\,1}, A_{1\,2}, \ldots, A_{1\,n}, \ldots, A_{n1}, A_{n2}, A_{n m_n}$ будут единственно возможны, несовместимы и равновозможны. Вероятностью события A_i называют дробь

$$P(A_i) = \frac{m_i}{m + m_2 + \ldots + m_n}. \tag{70.1}$$

Условимся называть события $A_{i\,1}, A_{i\,2}, \ldots, A_{i\,m_1}$, при которых наступает событие A_i, *благоприятными случаями для A_i.* Тогда определение вероятности, можно формулировать следующим образом. *Вероятностью события называется отношение числа равновозможных случаев, благоприятных этому событию, к числу всех равновозможных случаев, которые могут встретиться при испытании.*

Достоверное и невозможное событие можно рассматривать как предельные варианты случайных событий. Вероятность достоверного события равна единице, вероятность невозможного события — нулю.

Пример 1. В урне лежат 100 тщательно перемешанных шаров, отличающихся друг от друга только цветом: 30 белых, 25 красных и 45 зеленых. Какова вероятность вынуть белый шар? Число равновозможных случаев, которые могут встретиться при испытании, равно $30 + 25 + 45 = 100$. Число равновозможных случаев, благоприятных выниманию белого шара, есть 30. Поэтому вероятность вынуть белый шар будет $P_{\text{бел}} = 30/100 = 3/10$. Аналогично, для красных и зеленых шаров $P_{\text{кр}} = 25/100 = 1/4$, $P_{\text{зел}} = 45/100 = 9/20$.

6. Определение вероятности (70.1) предполагает, что еще до испытания имеются какие-то основания (например, соображения симметрии, однородности и пр.) оценивать равновозможность событий, а также представлять события в виде сумм равновозможных событий. Поэтому так определенную вероятность называют *априорной вероятностью*, т. е. такой вероятностью, о которой мы судим до опыта. Судить о равновозможности событий, даже в простейших случаях, не так легко, как это может показаться на первый взгляд. Приведем пример.

Пример 2. Монета бросается два раза. Какова вероятность того, что при двух бросаниях по крайней мере один раз выпадает герб? Равновозможных случаев, которые могут представиться при таком испытании (т. е. при двух бросаниях), четыре, а именно:

1) герб — герб,
2) герб — решка,
3) решка — герб,
4) решка — решка. (A)

Из них первые три — благоприятные рассматриваемому событию, т. е. появлению герба по крайней мере один раз. Поэтому искомая вероятность $3/4$. Даламбер (1717–1783) оспаривал этот результат. Он писал, что если при первом бросании выпал герб, то второе бросание становится ненужным, так как и без того ясно, что мы имеем дело с благоприятным случаем. Поэтому вместо четырех различных возможностей, перечисленных выше, Даламбер берет только три, а именно:

1) герб,
2) решка — герб,
3) решка — решка. (A')

Из них благоприятных два, и искомая вероятность, по Даламберу, равна $2/3$. Аналогично дело обстоит и при трех бросаниях. Какова вероятность того, что при трех бросаниях монеты появится герб, по крайней мере один раз? Равновозможных случаев всего восемь:

1) герб — герб — герб,
2) герб — герб — решка,
3) герб — решка — герб,
4) герб — решка — решка,
5) решка — герб — герб,
6) решка — герб — решка,
7) решка — решка — герб,
8) решка — решка — решка. (B)

Благоприятных случаев семь, и искомая вероятность равна 7/8. По Даламберу же появление герба делает дальнейшие бросания уже ненужными, а потому он перечисляет только четыре различных случая, а именно:

1) герб,
2) решка — герб,
3) решка — решка — герб,
4) решка — решка — решка. (B')

Из них благоприятных три, и искомая вероятность, по Даламберу, равна 3/4. Ошибка Даламбера состоит в том, что случаи (A'), а также (B') он принял за равновозможные, тогда как в действительности они не являются таковыми.

7. Если пользоваться одним только определением вероятности (70.1), то вычисление вероятности в каждом конкретном случае требует разложения событий на равновозможные. Необходимость этого устраняется основными теоремами теории вероятностей, известными под названием *теоремы сложения* и *теоремы умножения вероятностей*.

Теорема сложения вероятностей. *Вероятность суммы несовместимых событий равна сумме вероятностей этих событий.*

Действительно, разложим единственно возможные и несовместимые события $A_1, A_2, \ldots, A_n$ на равновозможные, как это делалось при введении определения (70.1). Пусть событие B является суммой событий A_1 и A_2, т. е. состоит в появлении либо события A_1, либо события A_2 (безразлично какого). Так как события A_1 и A_2 несовместимы, то число равновозможных случаев, благоприятных событию B, будет равно сумме равновозможных случаев, благоприятных событиям A_1 и A_2, т. е. $m_1 + m_2$. Вероятность же события B будет

$$P(B) = \frac{m_1 + m_2}{m_1 + m_2 + \ldots + m_n} = P(A_1) + P(A_2).$$

Таким образом, если события A_1 и A_2 несовместимы, то

$$P(A_1 + A_2) = P(A_1) + P(A_2). \tag{70.2}$$

Пример 3. Вероятность вынуть красный шар в примере 1 равна $P_{\text{кр}} = 25/100$, вероятность вынуть зеленый $P_{\text{зел}} = 45/100$, вероятность вынуть цветной шар

$$P_{\text{цв}} = \frac{25 + 45}{1000} = P_{\text{кр}} + P_{\text{зел}} = \frac{70}{100} = \frac{7}{10}.$$

Пример 4. В примере 2 вероятность события 1) из группы (A') равна 1/2, а вероятности событий 2) и 3) из той же группы равны 1/4. Действительно, событие 1) из группы (A') есть сумма несовместимых равновозможных событий 1) и 2) из группы (A). Поэтому для вероятности (при двух бросаниях) появления герба хотя бы один раз мы получаем на основании теоремы сложения вероятностей

$$P = \frac{1}{2} + \frac{1}{4} = \frac{3}{4},$$

т. е. верный результат. Аналогично разбирается случай трех бросаний. Вероятности событий из группы (B') равны соответственно 1/2, 1/4, 1/8, 1/8,

и для вероятности, о которой идет речь в примере 2, находим

$$P = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} = \frac{7}{8},$$

т. е. снова верный результат.

Сумма вероятностей всех единственно возможных и несовместимых событий равна единице:

$$P_1 + P_2 + \ldots + P_n = 1. \tag{70.3}$$

Это утверждение является непосредственным следствием теоремы сложения вероятностей. Действительно, так как события единственно возможны, то появление одного из них (безразлично, какого) есть событие достоверное. Вероятность такого события равна единице. С другой стороны, по теореме сложения вероятностей вероятность того же события может быть представлена суммой $P_1 + P_2 + \ldots + P_n$. В результате и получается соотношение (70.3).

Соотношение (70.3) часто называют *условием нормировки вероятности*. Вероятность в принципе можно было бы определить не выражением (70.1), а выражением, ему пропорциональным, т. е. тем же выражением, умноженным на произвольный постоянный коэффициент k. Тогда соотношение (70.3) не имело бы места. Только при $k = 1$ условие нормировки приводится к виду (70.3).

Если число единственно возможных несовместимых событий равно двум, то события называются противоположными. Каждому событию соответствует противоположное, состоящее в том, что первое событие не произойдет. Очевидно, сумма вероятностей противоположных событий равна единице.

8. Теорема умножения вероятностей. *Вероятность произведения двух событий A и B равна произведению вероятности одного из них $P(A)$ на вероятность другого, вычисленную в предположении, что первое событие произошло.*

Последнюю вероятность, о которой говорится в теореме, называют *условной вероятностью события B при условии, что событие A произошло*, и обозначают $P(B|A)$. Таким образом,

$$P(AB) = P(A)P(B|A). \tag{70.4}$$

Для доказательства допустим, что из n единственно возможных, несовместимых и равновозможных случаев

$$C_1, C_2, \ldots, C_l, C_{l+1}, C_m, C_{m+1}, \ldots, C_n$$

событию A благоприятствуют первые m случаев, остальные же ему не благоприятствуют. Пусть, далее, из m случаев

$$C_1, C_2, \ldots, C_l, C_{l+1}, \ldots, C_m$$

первые l случаев благоприятствуют событию B, остальные же ему не благоприятствуют. Значит, число случаев, благоприятствующих и A, и B, равно l, а потому $P(AB) = l/n$. Далее, очевидно, $P(A) = m/n$. Наконец, если событие A произошло, то случаи

$C_{m+1}, \ldots, C_n$ становятся невозможными, а все остальные случаи $C_1, C_2, \ldots, C_l, C_{l+1}, \ldots, C_m$ по-прежнему продолжают оставаться равновозможными. Поэтому $P(B|A) = l/m$. Таким образом,

$$P(A)P(B|A) = \frac{m}{n}\,\frac{l}{m} = \frac{l}{n} = P(AB),$$

что и доказывает теорему.

Располагая при доказательстве события A и B в обратном порядке, получим также

$$P(AB) = P(B)P(A|B). \tag{79.5}$$

Отметим важный частный случай теоремы умножения вероятностей. Допустим, что вероятность каждого из двух событий A и B не зависит от того, произошло второе событие или не произошло. В этом случае события A и B называются *независимыми* или *статистически независимыми*. Для независимых событий

$$P(AB) = P(A)P(B), \tag{70.6}$$

т. е. *вероятность произведения двух независимых событий равна произведению их вероятностей.*

Пример 5. В урне лежат четыре одинаковых шара, занумерованных цифрами 1, 2, 3, 4. Какова вероятность того, что при последовательном вынимании двух шаров они окажутся с номерами 1 и 2? Вынем один шар. Вероятность того, что он окажется с номером либо 1, либо 2 (событие A), равна по теореме сложения вероятностей

$$P(A) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

Если событие A произошло, то в урне останется три шара, один из которых будет иметь номер 1, либо 2. Вероятность вынуть шар с таким номером (событие B) равна $P(B/A) = 1/3$. Искомая вероятность по теореме умножения вероятностей равна

$$P(AB) = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}.$$

Проверим решение непосредственным подсчетом равновозможных случаев. Единственно возможно, несовместимые и равновозможные события в рассматриваемом примере изобразим таблицей

<u>12</u>	<u>21</u>	31	41
13	23	32	42
14	24	34	43

Здесь первая цифра означает номер при первом вынимании, а вторая — при втором. Число всех равновозможных случаев двенадцать. Из них благоприятных искомому событию — два, а именно 12 и 21 (эти случаи подчеркнуты). Искомая вероятность есть $2/12 = 1/6$.

Изменим теперь постановку задачи. Вынув шар и определив его номер, положим его обратно в урну и все шары тщательно перемешаем. Второе вынимание производим, следовательно, при таких же условиях, т. е. при том же количестве шаров в урне, что и первое. Теперь события A и B становятся независимыми. Вероятность $P(A)$ останется прежней, т. е. равной $1/2$. Найдем $P(B)$. Если при первом вынимании появился шар с номером 1 (2), то

событие B состоит в том, что при втором вынимании должен появиться шар с номером 2 (1). Вероятность этого события $P(B) = 1/4$. Таким образом, по теореме умножения вероятностей $P(AB) = P(A)P(B) = (1/2) \cdot (1/4) = 1/8$. В правильности результата нетрудно убедиться также, написав все равновозможные случаи, которые могут встречаться, а именно:

11	<u>21</u>	31	41
<u>12</u>	22	32	42
13	23	33	43
14	24	34	44

Благоприятные случаи подчеркнуты.

9. Определение априорной вероятности может встретить принципиальные трудности. Допустим, например, что бросается игральная кость, грани которой занумерованы цифрами 1, 2, 3, 4, 5, 6. Если кость совершенно однородна и имеет форму идеально правильного куба, то появления всех этих шести цифр при бросании будут равновозможными событиями. Но если кость неоднородна или не является правильным кубом, то это уже не будет. Тогда хотя понятие вероятности и сохраняет смысл, однако трудно представить себе, как в этом случае различные события можно разложить на равновозможные. Надо указать какой-то другой способ, с помощью которого можно было бы, хотя бы принципиально, найти вероятность и в указанном случае. Один из способов состоит в следующем.

Допустим, что игральная кость бросается n раз и при этом грань с номером 1 выпала n_1 раз. Отношение $\nu_1 = n_1/n$ называется *относительной частотой* появления рассматриваемого события. Опыт показывает, и в этом проявляется статистическая закономерность, что при неограниченном возрастании n относительная частота ν_1 стремится к вполне определенному пределу. Априори ясно, что в случае идеальной игральной кости этот предел должен быть равен $1/6$, т. е. вероятности рассматриваемого события, как она была определена выше. Поэтому представляется естественным и в общем случае определить вероятность события с помощью соотношения

$$P_1 = \lim_{n\to\infty} \nu_1 \equiv \lim_{n\to\infty} \frac{n_1}{n}. \tag{70.7}$$

Конечно, применимость этого определения не ограничивается случаем бросания игральной кости. Оно распространяется без всяких изменений на все случаи, когда в результате испытаний получается конечное число различных возможностей.

10. Существует еще одна интерпретация вероятности, применяющаяся в физике, вполне аналогичная (70.7). Поясним ее на простейшем примере. Пусть в закрытом сосуде имеется одна молекула. Сталкиваясь со стенками (имеющими молекулярную структуру), молекула претерпевает беспорядочные отражения от них, следующие друг за другом. При этом она побывает в различных местах сосуда. Выделим мысленно в сосуде какой-либо неподвижный объем v. Как определить

вероятность нахождения молекулы в этом объеме? С этой целью будем наблюдать за молекулой в течение длительного времени T. Пусть часть времени t молекула проводит в объеме v. Отношение t/T называется относительным временем пребывания молекулы в объеме v. Предел этого отношения

$$P = \lim_{T\to\infty} \frac{t}{T} \tag{70.8}$$

и есть вероятность нахождения молекулы в объеме v. Статистическая закономерность проявляется опять в том, что предел (70.8) существует, как это показывает опыт.

11. Важными понятиями в теории вероятностей и ее приложениях являются понятия *среднего значения* и *математического ожидания*. Разъясним эти понятия на конкретном примере. Пусть произведено N однотипных измерений одной и той же величины a при неизменных условиях. Пусть в n_1 случаях измеренное значение величины a оказалось равным a_1, в n_2 случаях — $a_2, \ldots$, в n_m случаях — a_m ($n_1+n_2+\ldots+n_m = N$). Среднее значение измеряемой величины определяется выражением

$$\langle a\rangle = \frac{n_1a_1 + n_2a_2 + \ldots + n_ma_m}{N} = \nu_1a_1 + \ldots + \nu_ma_m. \tag{70.9}$$

Допустим для простоты, что никаких других результатов, кроме $a_1, a_2, \ldots, a_m$, при измерениях появиться не может, так что эти результаты являются единственно возможными и несовместимыми. Тогда если неограниченно увеличивать число измерений N, то частоты $\nu_1, \nu_2, \ldots, \nu_m$ перейдут в свои предельные значения $P_1, P_2, \ldots, P_m$ — вероятности появления при измерениях значений $a_1, a_2, \ldots, a_m$. Выражение (70.9) при этом переходит в

$$M(a) = P_1a_1 + P_2a_2 + \ldots + P_ma_m. \tag{70.10}$$

Сумма (70.10) называется *математическим ожиданием величины* a.

Истинное значение измеряемой величины a, как правило, определить невозможно, так как измерения, сколь бы точны они ни были, сопровождаются ошибками. (Исключения имеют место только при счете конечного числа предметов.) Например, число жителей в доме или число деревьев в саду можно сосчитать совершенно точно. Систематические ошибки могут быть исключены путем тщательного изучения приборов и методов измерения. Но случайные ошибки всегда остаются. Влияние их уменьшают путем многократного повторения измерений. Идеальной целью, к которой следовало бы стремиться на этом пути, является нахождение математического ожидания измеряемой величины. Оно и представляло бы окончательный результат измерения, выдаваемый экспериментатором за истинное значение измеряемой величины. Но нахождение математического ожидания требовало бы бесконечного повторения измерений, а потому на практике вместо него приходится довольствоваться средним значением, полученным в результате как можно большего числа измерений.

Можно сказать, что математическое ожидание является пределом, к которому стремится среднее значение $\langle a \rangle$ при неограниченном возрастании числа N. Различать эти понятия крайне необходимо, когда требуется точность в рассуждениях. Однако, когда такой необходимости нет, термином «математическое ожидание» обычно не пользуются и называют средним значением как величину (70.9), так и величину (70.10).

12. Затронем попутно со всей возможной краткостью некоторые вопросы теории ошибок. Хотя непосредственно для нашего курса они не нужны, но вопрос об ошибках измерений является основным при статистической обработке результатов любых измерений. Поэтому имеет смысл на нем остановиться. Здесь речь пойдет только о *случайных ошибках*.

Ошибкой называется разность между измеренным и истинным значениями измеряемой величины. Если в результате N однотипных измерений получено N значений измеряемой величины $a_1, a_2, \ldots, a_N$, то ошибки этих отдельных измерений будут

$$x_i = a_i - a \quad (i = 1, 2, \ldots, N). \tag{70.11}$$

Для характеристики средней степени точности прибора и метода измерения применяют обычно так называемую *среднюю квадратичную ошибку отдельного измерения*. Это есть квадратный корень из среднего квадрата ошибки отдельного измерения, т. е. величина

$$\Delta_{\text{кв}} = \sqrt{\langle x^2 \rangle} = \sqrt{\frac{x_1^2 + x_2^2 + \ldots + x_N^2}{N}}. \tag{70.12}$$

Точное вычисление ошибок $x_1, x_2, \ldots$, а с ними и величины $\Delta_{\text{кв}}$ невозможно, так как истинное значение измеряемой величины a неизвестно. Вместо точного вычисления приходится довольствоваться вероятностной оценкой величины $\Delta_{\text{кв}}$. С этой целью введем понятие *отклонения результатов отдельных измерений от среднего значения* $\langle a \rangle$, т. е. величины

$$y_i = a_i - \langle a \rangle \quad (i = 1, 2, \ldots, N). \tag{70.13}$$

Эти отклонения удовлетворяют тождеству

$$\sum y_i = 0, \tag{70.14}$$

которое непосредственно следует из определения среднего арифметического $\langle a \rangle$. Подчеркнем здесь, что для ошибок x_i подобное равенство места не имеет. Истинное значение суммы ошибок $\sum x_i$, конечно, неизвестно. Однако если рассматриваемую серию из N измерений повторять многократно, устремляя число таких повторений к бесконечности, то можно утверждать, что математическое ожидание указанной суммы будет равно нулю:

$$M\Big(\sum x_i\Big) = 0. \tag{70.15}$$

В этом проявляется случайный характер ошибок.

Из (70.11) и (70.13) следует, что $x_i = y_i + \delta$, где δ — постоянная: $\delta = \langle a \rangle - a$. Она имеет смысл ошибки среднего результата. Ее точное вычисление, конечно, невозможно. Но можно дать вероятностную оценку абсолютного значения величины δ или, лучше, ее квадрата. Величина $\delta_{\text{кв}} = \sqrt{\langle \delta^2 \rangle}$ называется *средней квадратичной ошибкой среднего результата*. Ее вычисление и является главной целью теории ошибок. Возведя равенство $x_i = y_i + \delta$ в

квадрат и просуммировав по всем i, получим ввиду соотношения (70.14)

$$\sum x_i^2 = \sum y_i^2 + \sum \delta^2$$

или

$$N\,\langle x^2\rangle = \sum y_i^2 + N\,\delta^2.$$

Далее,

$$\delta = \frac{\sum a_i}{N} - a = \frac{\sum (a_i - a)}{N} = \frac{\sum x_i}{N},$$

откуда

$$N\delta^2 = \frac{1}{N}\sum x_i^2 + \frac{1}{N}\sum\sum_{i\neq j} x_i x_j.$$

Первое слагаемое в правой части существенно положительно и равно $\langle x^2\rangle$. Что касается двойной суммы, то о ее значении сказать ничего нельзя. Можно утверждать только, что если рассматриваемую серию из N измерений повторять неограниченно, то двойная сумма с равной вероятностью будет принимать как положительные, так и отрицательные значения. Ее математическое ожидание будет равно нулю, подобно математическому ожиданию (70.15). Для вероятностной оценки среднего квадрата $\langle\delta^2\rangle$ заменим двойную сумму $\sum\sum x_i x_j$ ее математическим ожиданием. Таким путем получим $N\langle\delta^2\rangle = \langle x^2\rangle$, а потому

$$N\langle x^2\rangle = \sum y_i^2 + \langle x^2\rangle.$$

Отсюда

$$\Delta_{\text{кв}} = \sqrt{\langle x^2\rangle} = \sqrt{\frac{\sum y_i^2}{N-1}}. \tag{70.16}$$

$$\delta_{\text{кв}} = \sqrt{\frac{\langle x^2\rangle}{N}} = \sqrt{\frac{\sum y_i^2}{N(N-1)}} = \frac{\Delta_{\text{кв}}}{\sqrt{N}}. \tag{70.17}$$

В правые части этих формул входят только известные величины — отклонения результатов отдельных измерений от среднего значения $\langle a\rangle$. Поэтому (70.16) и (70.17) могут служить для фактического вычисления средних квадратичных ошибок $\Delta_{\text{кв}}$ и $\delta_{\text{кв}}$ или, лучше, их вероятностных оценок.

Окончательный результат измерения принято записывать в виде

$$a = \langle a\rangle \pm \delta_{\text{кв}}.$$

Величина $\delta_{\text{кв}}$ определяет число достоверных десятичных знаков, с которыми может быть получено значение измеряемой величины. Величина $\Delta_{\text{кв}}$ от числа измерений не зависит. Увеличивая число измерений, мы только уточняем значение этой величины. Поэтому $\delta_{\text{кв}} \sim 1/\sqrt{N}$. Для того чтобы повысить точность результата на один порядок, оставляя точность отдельных измерений неизменной, надо увеличить число измерений в 100 раз. Повышение точности на два порядка потребовало бы увеличения числа измерений в 10 000 раз. Отсюда видно, что метод многократного повторения измерений эффективен лишь при сравнительно небольших значениях N.

13. Понятие вероятности мы разъяснили применительно к случаям, когда множество различных событий, которые могут появиться при испытании, *конечно*. Но могут быть и такие случаи, когда это множество *бесконечно* и даже *непрерывно*. С такими случаями мы встречаемся, например, при измерении величин, могущих принимать непрерывный ряд значений. Можно, например, ввести вероятность

dP того, что числовое значение измеряемой величины, полученное в результате измерения, будет заключено в пределах от a до $a+da$. Эта вероятность пропорциональна ширине бесконечно узкого интервала da, так что она может быть представлена в виде

$$dP = \rho(a)\,da,$$

причем коэффициент пропорциональности ρ, вообще говоря, зависит от a. Функция $\rho(a)$ называется *плотностью вероятности*. Условие нормировки (70.3) принимает вид

$$\int \rho(a)\,da = 1, \tag{70.18}$$

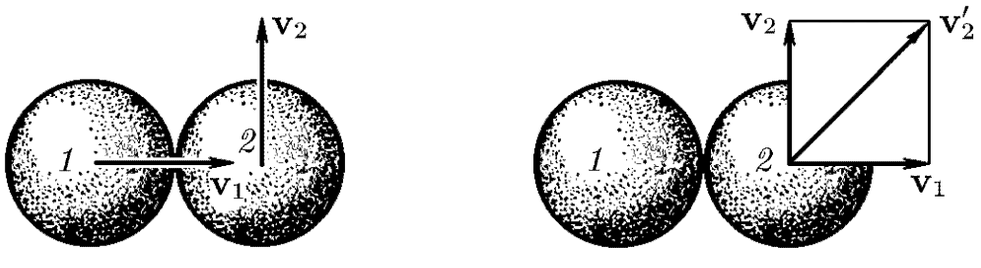
а формула (70.10) для математического ожидания переходит в

$$M(a) = \int a\rho(a)\,da. \tag{70.19}$$

Интегралы берутся по всем значениям, которые может принимать a. Однако во всех случаях в качестве пределов интегрирования можно поставить $-\infty$ и $+\infty$, считая, что вне области изменения a плотность вероятности $\rho(a)$ равна нулю.

§ 71. Распределение скоростей молекулы газа. Постановка задачи

1. В состоянии статистического равновесия все направления скоростей молекул при тепловом движении равновероятны. Если бы это было не так, то тепловое движение газа не было бы вполне беспорядочным. Модули всех скоростей молекул в том же состоянии также не могут быть одинаковыми. Даже если бы случайно они и оказались одинаковыми в какой-то момент времени, то в дальнейшем такое состояние быстро нарушилось бы из-за столкновений молекул между собой. Рассмотрим, например, простейшую модель газа, состоящую из идеально упругих и гладких шариков, взаимодействующих меж-



До удара После удара

Рис. 49

ду собой лишь в моменты столкновений. Допустим, что столкнулись молекулы *1* и *2*, скорости которых до столкновения $\mathbf{v}_1$ и $\mathbf{v}_2$ были взаимно перпендикулярны (рис. 49). Первая молекула двигалась вдоль

линии центров *12*, вторая — перпендикулярно к ней. Так как шары — абсолютно гладкие, то касательные составляющие их скоростей в результате столкновения не изменятся. Однако шары, как известно из элементарной теории удара, должны обмениваться нормальными скоростями. После столкновения первый шар остановится, скорость второго получит приращение $\Delta\mathbf{v}_2 = \mathbf{v}_1$, т. е. обратится в $\mathbf{v}_2' = \mathbf{v}_2 + \mathbf{v}_1$. Она изобразится диагональю параллелограмма, построенного на векторах $\mathbf{v}_1$ и $\mathbf{v}_2$. Если $v_1 = v_2$, то $v_2' = v_1\sqrt{2}$. Этот пример показывает, что при столкновениях меняются не только направления движения молекул, но и абсолютные значения их скоростей. Рассмотренное столкновение является только одним из возможных. На самом деле столкновения бесконечно разнообразны. Они сопровождаются всевозможными изменениями скоростей и приводят в конце концов к вполне определенному статистическому распределению молекул по скоростям.

2. Задача о распределении молекул газа по скоростям была поставлена и решена Максвеллом в 1859 г. Уясним сначала постановку задачи. Допустим, что в закрытом сосуде содержится большое число N молекул газа и что внешних силовых полей, действующих на газ, нет. Примем произвольную точку пространства O за начало координат (рис. 50). Отложим от нее в какой-то момент времени t векторы скоростей всех молекул газа: $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N$. Концы этих векторов называются *скоростными*, или *изображающими*, *точками*. Совокупность всех изображающих точек образует трехмерное пространство, называемое *пространством скоростей*. В нем можно ввести прямоугольные оси. Координатами скоростной точки являются проекции v_x, v_y, v_z вектора $\mathbf{v}$ на эти оси. Задание скоростей всех молекул газа эквивалентно заданию положения их скоростных точек в пространство скоростей. С чисто динамической точки зрения задача о распределении скоростей молекул сводится к определению положения скоростных точек в пространстве скоростей в любой момент времени. Но, как уже указывалось в § 9, для систем с колоссальным числом молекул в такой динамической постановке задача неразрешима и не представляет интереса. Распределение молекул по скоростям должно рассматриваться как статистическая задача. Ее можно формулировать следующим образом.

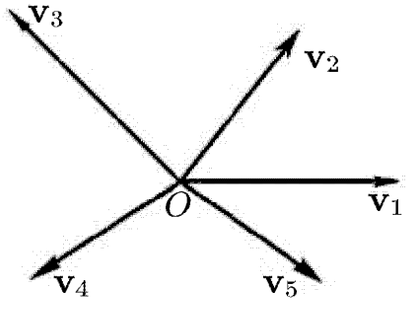


Рис. 50

3. Возьмем в пространстве скоростей физически бесконечно малый элемент объема, имеющий, например, форму прямоугольного параллелепипеда с ребрами dv_x, dv_y, dv_z и с центром в точке v_x, v_y, v_z. Объем этого параллелепипеда равен $d\omega = dv_x\, dv_y\, dv_z$, число изображающих точек в нем обозначим через dN. Из-за взаимодействия молекул меняются их скорости. На геометрическом языке это означает, что изображающие точки одних молекул уходят из элемента объема

$d\omega$, изображающие точки других молекул — вступают в него. Число скоростных точек dN внутри объема $d\omega$, таким образом, не сохраняется постоянным. Если элемент $d\omega$ выбрать очень малым, то в нем окажется мало изображающих точек. Может случиться, например, что в одни моменты времени в объеме $d\omega$ окажется одна или две скоростных точки, а в другие моменты — ни одной. Число скоростных точек dN, таким образом, будет резко и нерегулярно меняться от одного момента времени к другому. Во избежание этого надо объем $d\omega$ выбрать достаточно большим, чтобы в нем находилось еще очень много изображающих точек. Тогда в установившемся состоянии числа dN будут меняться относительно мало, колеблясь вокруг некоторого среднего значения $\langle dN\rangle$, а поведение самих средних значений $\langle dN\rangle$ будет подчиняться определенным статистическим закономерностям, которые мы и должны установить. Но объем $d\omega$ в то же время должен быть настолько малым, чтобы распределение изображающих точек в пространстве скоростей было описано достаточно детально, и настолько малым, чтобы с величинами dv_x, dv_y, dv_z, а также $d\omega$ и dN можно было обращаться как с бесконечно малыми дифференциалами. Обоим требованиям удается удовлетворить практически всегда благодаря колоссальности чисел молекул, содержащихся в газах.

4. Отношение $Nf(\mathbf{v}) = \langle dN\rangle/d\omega$ имеет смысл средней концентрации скоростных точек в пространстве скоростей и вполне аналогично концентрации частиц в обычном (координатном) пространстве. Величина $f(\mathbf{v})$ называется *функцией распределения молекул по скоростям*. Она может быть интерпретирована как плотность вероятности распределения изображающих точек молекул в пространстве скоростей. К определению функции $f(\mathbf{v})$ и сводится статистическая задача о распределении скоростей молекул. Если dN или $\langle dN\rangle$ проинтегрировать по всему скоростному пространству, то в результате, очевидно, получится полное число молекул N. Отсюда следует условие нормировки

$$\int f(\mathbf{v})\, d\omega = 1, \tag{71.1}$$

которому должна удовлетворять функция $f(\mathbf{v})$.

В дальнейшем мы будем иметь дело только со средними числами $\langle dN\rangle$, а не с мгновенными значениями dN. Поэтому для сокращения письма знак усреднения $\langle\ \rangle$ обычно будет опускаться. Иными словами, в дальнейшем всюду под dN будет пониматься среднее или вероятное значение этого числа, т. е. $\langle dN\rangle$.

По смыслу функции распределения

$$dN = Nf(\mathbf{v})\, d\omega = Nf(v_x, v_y, v_z)\, dv_x\, dv_y\, dv_z \tag{71.2}$$

дает среднее число молекул, компоненты скоростей которых лежат между v_x и $v_x + dv_x$, v_y и $v_y + dv_y$, v_z и $v_z + dv_z$ соответственно. Функция распределения $f(\mathbf{v})$ меняется непрерывно и плавно с изменением скорости $\mathbf{v}$. Она описывает не истинное, а только *сглаженное*, *среднее*, или *вероятное*, распределение молекул по скоростям. Это

видно из того, что истинное число скоростных точек в любом элементе $d\omega$ скоростного пространства может быть только целым. Величина же dN, определяемая выражением (71.2), может принимать любые значения.

При статистической постановке задачи не имеет также смысла спрашивать, сколько молекул газа имеют вполне определенную скорость $\mathbf{v}$ с компонентами v_x, v_y, v_z. Речь может идти только о среднем числе dN молекул в элементе объема скоростного пространства $d\omega = dv_x\, dv_y\, dv_z$. При стремлении $d\omega$ к нулю стремится к нулю и число dN. Таким образом, среднее число молекул со строго определенной скоростью $\mathbf{v}$ равно нулю. Это станет очевидным, если заметить, что множество всех векторов $\mathbf{v}$, т. е. всех точек скоростного пространства бесконечно, тогда как полное число молекул газа N конечно. По той же причине равно нулю среднее число молекул со строго определенным направлением вектора скорости.

5. Для лучшего уяснения статистического характера задачи о распределении скоростей молекул может служить демонстрационный прибор, называемый *доской Гальтона*. Эта доска, с передней стороны прикрытая стеклом, в которую в шахматном порядке достаточно часто вбиты гвозди (рис. 51). Внизу под гвоздями установлены равноотстоящие вертикальные

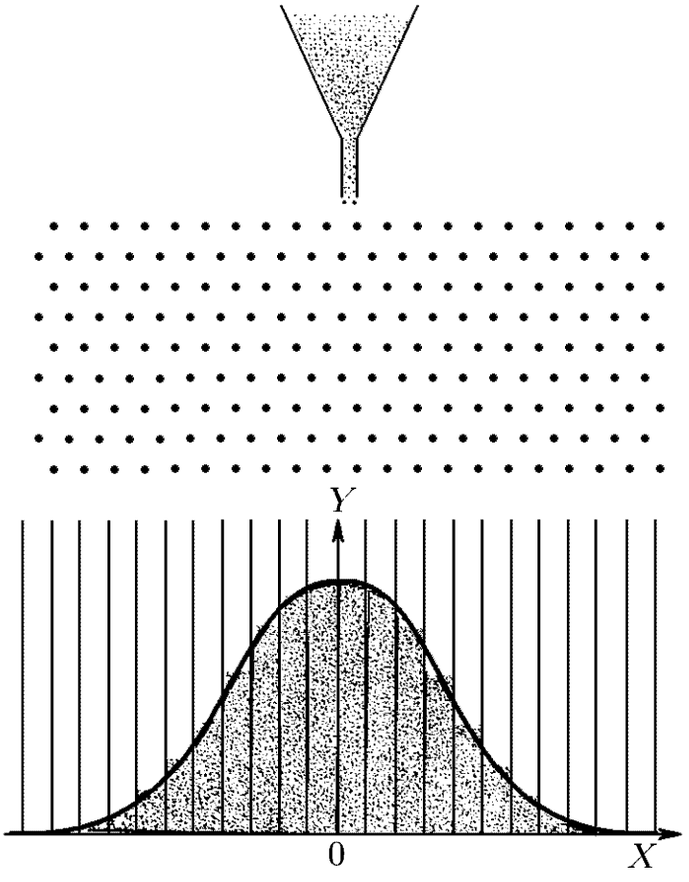


Рис. 51

перегородки из металлических пластинок. Они делят пространство между доской и стеклом на равные объемы, называемые ниже для краткости *ячейками*. Вверху над гвоздями в средней части доски помещена воронка, в которую можно сыпать песок, пшено или другие частицы. Если бросить в воронку одну частицу, то при падении вниз она испытает множество

столкновений с гвоздями и в конце концов попадает в одну из ячеек. В какую ячейку попадает частица — это предсказать невозможно из-за множества случайных факторов, влияющих на ее движение. Можно говорить лишь о вероятности попадания частицы в ту или другую ячейку. Естественно ожидать, что попадание частицы в центральные ячейки более вероятно, чем в крайние. И действительно, если непрерывно сыпать частицы через воронку, то оказывается, что в центральные ячейки, находящиеся под отверстием воронки, попадает частиц больше всего, а в крайние — меньше всего. При очень большом количестве частиц, прошедших через воронку вырисовывается вполне определенная статистическая закономерность распределения их по ячейкам. Эту закономерность можно даже представить в пределе аналитической формулой. С этой целью соединим плавной кривой вершины столбиков, которые образуют частицы в ячейках. Оказывается, что при очень большом числе частиц кривая асимптотически приближается к кривой вида

$$y = \varphi(x) \equiv Ae^{-\alpha x^2}, \tag{71.3}$$

где A и α — положительные постоянные. Из них постоянная α зависит только от параметров прибора, но не зависит от числа частиц. Постоянная A пропорциональная числу частиц и связана с постоянной α условием нормировки.

6. Формула (71.3) выражает так называемый *нормальный закон распределения ошибок Гаусса* (1777–1855), а соответствующая ей кривая *кривой ошибок Гаусса*. Величина $\varphi(x)\,dx$ дает вероятность того, что при измерении будет сделана ошибка, заключенная в пределах от x до $x+dx$. Конечно, при такой интерпретации плотность вероятности $\varphi(x)$ должна быть нормирована условием

$$\int\limits_{-\infty}^{+\infty} \varphi(x)\,dx \equiv A \int\limits_{-\infty}^{+\infty} e^{-\alpha x^2}\,dx = 1, \tag{71.4}$$

с помощью которого постоянную A можно выразить через постоянную α (см. задачу 2 к § 72). Чем больше α, тем более острым будет максимум на кривой ошибок, тем точнее измерения. Поэтому постоянная α должна быть связана со средней квадратичной или средней арифметической ошибками, характеризующими точность измерения (см. задачу 3 к § 72). Что касается доказательства закона ошибок Гаусса (71.3), то оно может быть проведено аналогично тому, как в следующем параграфе доказывается закон распределения скоростей Максвелла.

7. Задача о распределении молекул газа по скоростям, как она сформулирована выше, а также методы решения ее, приводимые дальше, являются чисто *классическими*. Поэтому необходимо прежде всего выяснить границы применимости такого классического рассмотрения. Если не обращать внимания на малосущественные числовые коэффициенты порядка единицы, то ответ можно получить из наглядных соображений, хотя бы с помощью *принципа неопределенности Гейзенберга* (1901–1976). Выделим в газе маленький кубик со сторонами x, y, z, на который в среднем приходится одна частица. Если выполнены условия

$$xp_x \gg h, \quad yp_y \gg h, \quad zp_z \gg h,$$

то движение частицы в этом кубике можно рассматривать классически. Действительно, согласно принципу неопределенности Гейзенберга, произведение квантовых неопределенностей координаты и импульсы частицы порядка h, а потому при соблюдении указанных условий эти неопределенности существенной роли не играют. Перемножив эти три неравенства, можно заменить их одним: $Vp^3 \gg h^3$, где $V = xyz$ — объем кубика, а p — некоторый средний импульс, характеризующий движение частиц газа. Если n — число частиц в единице объема, то $Vn = 1$, в результате $n(h/p)^3 \ll 1$. Величина

$$\lambda = \frac{h}{p} = \frac{h}{mv} \tag{71.5}$$

имеет размерность длины. Она называется *длиной волны де-Бройля* (род. 1882) и играет исключительно важную роль в квантовой механике. С введением λ условие применимости классического рассмотрения газа принимает вид

$$n\lambda^3 \ll 1. \tag{71.6}$$

Оно означает, что среднее число частиц в объеме λ^3 должно быть мало по сравнению с единицей.

Для оценки порядка величины λ воспользуемся какой-либо средней скоростью, характеризующей тепловое движение молекул газа. Пока в нашем распоряжении есть только одна из таких скоростей, а именно средняя квадратичная скорость $\overline{v}_{\text{кв}} = \sqrt{3kT/m}$ (см. § 60). Используя ее, придадим условию (71.6) вид

$$T \gg T_g, \tag{71.6}$$

где введено обозначение

$$T_g = \frac{h^2}{3km}\, n^{2/3}. \tag{71.7}$$

Величина T_g называется *температурой вырождения* газа. Таким образом, *классический способ рассмотрения применим при температурах, значительно превышающих температуру вырождения*. При более строгом рассмотрении за температуру вырождения принимают

$$T_g = \frac{h^2}{2km}\left(\frac{3n}{8\pi}\right)^{2/3}. \tag{71.8}$$

Она примерно втрое меньше предыдущего выражения. Газы ниже температуры вырождения называются *вырожденными*. К ним классический способ рассмотрения неприменим.

Вычислим температуру вырождения для двух крайних случаев. Для электронного газа в серебре $n \approx 6 \cdot 10^{22}$ см$^{-3}$, масса электрона $m = 9{,}11 \cdot 10^{-28}$ г. По формуле (71.8) получаем для электронного газа в серебре $T_g \approx 6{,}5 \cdot 10^4$ К. Подобные же значения получаются для всех других хорошо проводящих металлов. При таких высоких температурах ни один металл в твердом состоянии существовать не

может. Отсюда следует, что электронный газ в хорошо проводящих металлах всегда полностью вырожден. Возьмем теперь другой крайний случай — гелий. Масса атома гелия $m = 6{,}6 \cdot 10^{-24}$ г, а концентрация при нормальных условиях $n = 2{,}7 \cdot 10^{19}$ см$^{-3}$. Вычисление по формуле (71.8) дает $T_g = 0{,}05$ К. У всех остальных газов (за исключением водорода) масса атомов еще больше, а следовательно, T_g ниже, чем у гелия (концентрация n при нормальных условиях одна и та же для всех идеальных газов). Имея столь низкие температуры вырождения, ни одно вещество не может находиться в газообразном состоянии при нормальных условиях. Поэтому все молекулярные газы достаточно далеки от вырождения, и их не только можно, но и следует рассматривать как классические системы.

§ 72. Закон распределения скоростей Максвелла

1. Изложим теперь рассуждения Максвелла, которые привели его в 1859 г. к открытию закона распределения молекул газа по скоростям. Пусть газ состоит из очень большого числа N тождественных молекул, находящихся в состоянии беспорядочного теплового движения при определенной температуре. Предполагается, что силовые поля, действующие на газ, отсутствуют. Какова вероятность того, что x-составляющая скорости молекулы лежит между v_x и $v_x + dv_x$, а остальные две составляющие могут быть какими угодно? Ясно, что эта вероятность должна быть пропорциональна ширине рассматриваемого скоростного интервала dv_x, а коэффициент пропорциональности зависит от v_x. Обозначим эту вероятность через $\varphi(v_x)\,dv_x$. Величина $\varphi(v_x)$ также называется *функцией распределения*. Однако она характеризует распределение молекул не по полной скорости $\mathbf{v}$, а только по ее проекции v_x на ось X. Величина $f(\mathbf{v})$ имеет смысл *объемной* (*трехмерной*), или *совместной*, *функции распределения* в пространстве скоростей, величина $\varphi(v_x)$ — *одномерной* (*линейной*), или *парциальной*, *функции распределения* в том же пространстве.

Аналогично, $\varphi(v_y)\,dv_y$ будет вероятностью того, что y-составляющая скорости молекулы заключена между v_y и $v_y + dv_y$, а составляющие v_x и v_z могут быть какими угодно. Ввиду полного равноправия всех направлений скоростей функция φ будет той же самой, что и в предыдущем случае, но от другого аргумента v_y. Наконец, $\varphi(v_z)\,dv_z$ есть вероятность того, что z-составляющая скорости молекулы лежит между v_z и $v_z + dv_z$, причем остальные составляющие v_x и v_y могут быть какими угодно.

2. Назовем ради краткости попадания молекулы в скоростные интервалы (v_x, v_x+dv_x), (v_y, v_y+dv_y), (v_z, v_z+dv_z), о которых шла речь выше, событиями A, B, C соответственно. Определим вероятность $f(\mathbf{v})\,d\omega$ того, что молекула попадает в элемент объема скоростного пространства $d\omega = dv_x\,dv_y\,dv_z$. Такое попадание есть сложное событие, являющееся произведением событий A, B, C. Его вероятность можно определить по теореме умножения вероятностей. Для этого

надо вероятность события A умножить на условную вероятность события B при условии, что событие A произошло, а затем результат умножить на условную вероятность события C при условии, что произошли события A и B. Максвелл ввел предположение, что события A, B, C *независимы*. Тогда указание условий, при которых должны вычисляться вероятности событий B и C, становится не нужным и теорему умножения вероятностей можно применять в ее простейшей форме, какую она принимает для независимых событий. Это предположение, а с ним и первый вывод закона распределения скоростей, данный Максвеллом, подверглись критике со стороны некоторых математиков и физиков. Указывалось, в частности, что скорости молекул после столкновения не могут быть независимыми от их скоростей до столкновения, поскольку эти скорости связаны между собой законами сохранения энергии и импульса. Однако дальнейшие исследования Максвелла, Больцмана и других ученых показали, что предположение Максвелла правильно, хотя и нуждается в обосновании. Мы примем его пока без обоснований. По сравнению с другими доказательствами, данными самим Максвеллом, а затем Больцманом, первое доказательство Максвелла обладает тем преимуществом, что оно не вводит никаких специальных представлений относительно структуры молекул и сил взаимодействия между ними. Поэтому оно применимо не только к газам, но и к жидкостям и к твердым телам. Требуется только выполнение условия (71.7), чтобы задачу о распределении скоростей молекул можно было трактовать классически.

Итак, мы принимаем, что вероятность того, что изображающая точка молекулы одновременно окажется внутри трех интервалов $(v_x, v_x + dv_x)$, $(v_y, v_y + dv_y)$, $(v_z, v_z + dv_z)$, должна выражаться произведением

$$\varphi(v_x)\varphi(v_y)\varphi(v_z)\,dv_x\,dv_y\,dv_z.$$

Но для той же вероятности мы писали $f(\mathbf{v})\,d\omega$, где $d\omega = dv_x\,dv_y\,dv_z$. Сравнивая оба выражения, находим, что функция распределения $f(\mathbf{v})$ должна иметь вид

$$f(\mathbf{v}) = \varphi(v_x)\varphi(v_y)\varphi(v_z). \tag{72.1}$$

3. Положительные и отрицательные направления координатных осей в газе совершенно эквивалентны. Поэтому должно быть $\varphi(v_x) = \varphi(-v_x)$. Значит, функция φ может зависеть только от модуля или, что то же самое, от квадрата скорости v_x. Точно так же ввиду изотропии газа, функция f может зависеть только от квадрата полной скорости $\mathbf{v}$, но не от ее направления. Вместо квадратов скоростей удобнее взять в качестве аргументов соответствующие кинетические энергии: $\varepsilon_x = mv_x^2/2$, $\varepsilon_y = mv_y^2/2$, $\varepsilon_z = mv_z^2/2$, $\varepsilon = mv^2/2 = \varepsilon_x + \varepsilon_y + \varepsilon_z$. При переходе к новым аргументам сами функции условимся обозначать прежними буквами φ и f, хотя это — аналитически совсем другие функции. Уравнение (72.1) запишется в виде

$$\varphi(\varepsilon_x)\varphi(\varepsilon_y)\varphi(\varepsilon_z) = f(\varepsilon_x + \varepsilon_y + \varepsilon_z), \tag{72.2}$$

причем оно справедливо, каковы бы ни были (положительные) значения аргументов $\varepsilon_x, \varepsilon_y, \varepsilon_z$.

4. Функциональным уравнением (72.2) и определится вид функции φ, а с ней и функции f. Действительно, рассмотрим такие изменения аргументов ε_x, ε_y, ε_z, которые удовлетворяют двум условиям: 1) $\varepsilon_z = \text{const}$, 2) $\varepsilon_x + \varepsilon_y = \text{const}$. При таких условиях уравнение (72.2) все еще остается верным. Из него следует, что

$$\varphi(\varepsilon_x)\varphi(\varepsilon_y) = \text{const}$$

при условии

$$\varepsilon_x + \varepsilon_y = \text{const}.$$

Логарифмируя, а затем дифференцируя первое соотношение, получим

$$\frac{\varphi'(\varepsilon_x)}{\varphi(\varepsilon_x)}\,d\varepsilon_x + \frac{\varphi'(\varepsilon_y)}{\varphi(\varepsilon_y)}\,d\varepsilon_y = 0$$

при условии

$$d\varepsilon_x + d\varepsilon_y = 0.$$

Отсюда

$$\frac{\varphi'(\varepsilon_x)}{\varphi(\varepsilon_x)} = \frac{\varphi'(\varepsilon_y)}{\varphi(\varepsilon_y)}.$$

При выводе предполагалось, что изменения аргументов ε_x и ε_z связаны условием $\varepsilon_x + \varepsilon_y = C = \text{const}$. Однако значения постоянной C, а с ней и аргументов ε_x и ε_y могут быть какими угодно. Поэтому условие $\varepsilon_x + \varepsilon_y = \text{const}$ фактически не накладывает никаких ограничений на значения, которые могут принимать аргументы ε_x и ε_y. Значит, в предыдущем соотношении ε_x и ε_y могут независимо принимать любые значения. Но слева стоит функция только ε_x, а справа — только ε_y. Равенство между ними возможно тогда и только тогда, когда отношения $\varphi'(\varepsilon_x)/\varphi(\varepsilon_x)$ и $\varphi'(\varepsilon_y)/\varphi(\varepsilon_y)$ равны одной и той же постоянной. Обозначив эту постоянную через $-\alpha$, получим

$$\frac{\varphi'(\varepsilon_x)}{\varphi(\varepsilon_x)} = \frac{\varphi'(\varepsilon_y)}{\varphi(\varepsilon_y)} = -\alpha, \quad \text{или} \quad \frac{d\varphi(\varepsilon_x)}{\varphi(\varepsilon_x)} = -\alpha d\varepsilon_x.$$

Интегрирование дает

$$\varphi(\varepsilon_x) = A_1 e^{-\alpha\varepsilon_x}, \quad \varphi(\varepsilon_y) = A_1 e^{-\alpha\varepsilon_y}, \quad \varphi(\varepsilon_z) = A_1 e^{-\alpha\varepsilon_z}, \tag{72.3}$$

где A_1 — новая постоянная, значение которой будет определено ниже. Что касается постоянной α, то она должна быть положительной, так как в противном случае $\varphi(\varepsilon_x)$ неограниченно возрастала бы при неограниченном возрастании кинетической энергии ε_x, что физически невозможно.

5. Для функции распределения $f(\varepsilon) = f(\varepsilon_x + \varepsilon_y + \varepsilon_z)$ из (72.3) получаем

$$f(\varepsilon) = Ae^{-\alpha\varepsilon}, \tag{72.4}$$

причем $A = A_1^3$. Эта формула, отличающаяся исключительной простотой, и выражает максвелловский закон распределения скоростей. Для того чтобы придать ему окончательный вид, необходимо еще определить постоянные A и α. Для этого проще начать не с функции f, а с функции φ. Последняя функция в зависимости от скорости v_x представлена на рис. 52. Она тождественна с гауссовой кривой ошибок. Площадь элементарной полоски, заштрихованной на рисунке, дает вероятность того, что x-составляющая скорости молекулы лежит внутри интервала $(v_x, v_x + dv_x)$, а умноженная на N, она дает вероятное число молекул со скоростями в том же интервале. Функция $\varphi(v_x)$ должна быть нормирована условием

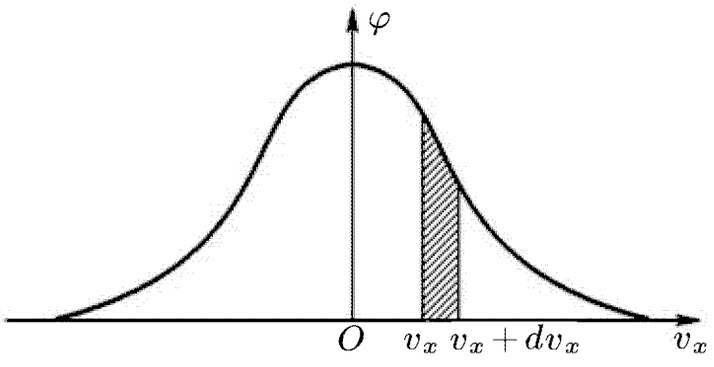


Рис. 52

$$\int\limits_{-\infty}^{+\infty} \varphi(v_x)\, dv_x \equiv A_1 \int\limits_{-\infty}^{+\infty} \exp\Big(-\frac{\alpha m v_x^2}{2}\Big)\, dv_x = 1. \tag{72.5}$$

Интегрирование в пределах от $-\infty$ до $+\infty$ не означает, что в газе есть молекулы с бесконечно большими скоростями. В действительности при достаточно больших скоростях формулы (72.3) и (72.4) становятся неприменимыми. Кинетическая энергия молекулы ε не может превосходить кинетической энергии всего газа K. Поэтому при $\varepsilon > K$ формула (72.4) заведомо не имеет смысла. Но она становится неприменимой уже при много меньших значениях ε, когда перестают выполняться условия, накладываемые на элементы объема скоростного пространства $d\omega$, необходимые для введения самого понятия функции распределения (см. § 71, п. 3). Интегрирование в бесконечных пределах следует рассматривать только как вычислительный прием. Он возможен потому, что молекул со скоростями, удовлетворяющими условию $\alpha m v_x^2 \gg 1$, очень мало, и такие молекулы практически не вносят сколько-нибудь заметного вклада в нормировочный интеграл (72.5). Значение этого интеграла практически не изменится, если истинное распределение молекул по скоростям заменить экспоненциальными (72.3) не только в области его применимости, но и при больших скоростях, где оно не действовало. Это и сделано в формуле (72.5).

Введем в качестве переменной интегрирования величину $\xi = \sqrt{\alpha m/2}\, v_x$. Тогда условие (72.5) примет вид

$$A_1 \sqrt{\frac{2}{\alpha m}} \int\limits_{-\infty}^{+\infty} e^{-\xi^2}\, d\xi = 1. \tag{72.6}$$

Входящий сюда интеграл называется интегралом Пуассона. В курсах математического анализа доказывается, что

$$\int_{-\infty}^{+\infty} e^{-\xi^2}\, d\xi = \sqrt{\pi}. \tag{72.7}$$

С использованием этого результата получаем

$$A_1 = \sqrt{m\alpha/2\pi}. \tag{72.8}$$

6. Задача свелась к вычислению одной только постоянной α. Для этого замечаем, что средняя кинетическая энергия $\langle \varepsilon_x \rangle$ теплового движения вдоль оси X выражается через функцию распределения φ соотношением

$$\langle \varepsilon_x \rangle = \int_{-\infty}^{+\infty} \varepsilon_x \varphi(\varepsilon_x)\, dv_x,$$

или более подробно

$$\langle \varepsilon_x \rangle = A_1 \int_{-\infty}^{+\infty} \frac{mv_x^2}{2} \exp\Big(-\frac{\alpha m v_x^2}{2}\Big)\, dv_x.$$

Введем прежнюю переменную интегрирования ξ. Тогда получим

$$\langle \varepsilon_x \rangle = \frac{A_1}{\alpha} \sqrt{\frac{2}{\alpha m}} \int_{-\infty}^{+\infty} \xi^2 e^{-\xi^2}\, d\xi.$$

Интегрированием по частям находим

$$\int_{-\infty}^{+\infty} \xi^2 e^{-\xi^2}\, d\xi = -\frac{1}{2}\,[\xi e^{-\xi^2}]_{-\infty}^{+\infty} + \frac{1}{2} \int_{-\infty}^{+\infty} e^{-\xi^2}\, d\xi.$$

Первое слагаемое в правой части обращается в нуль, так как при $\xi \to$ $\to \infty$ показательная функция $e^{-\xi^2}$ стремится к нулю быстрее, чем стремится к бесконечности любая степень ξ. В результате получаем

$$\langle \varepsilon_x \rangle = \frac{A_1}{2\alpha} \sqrt{\frac{2}{\alpha m}} \int_{-\infty}^{+\infty} e^{-\xi^2}\, d\xi,$$

или, учитывая (72.6), $\langle \varepsilon_x \rangle = (1/2)\alpha$. (Заметим, что для получения этого результата знания интеграла Пуассона (72.7) не требуется.) Но по определению кинетической температуры $\langle \varepsilon_x \rangle = (1/2)\Theta = (1/2)kT$. Это дает

$$\alpha = \frac{1}{kT}, \tag{72.9}$$

$$A_1 = \sqrt{\frac{m}{2\pi kT}}, \quad A = \Big(\frac{m}{2\pi kT}\Big)^{3/2}. \tag{72.10}$$

В результате для функции распределения $\varphi(v_x)$ и $f(v)$ получаем

$$\varphi(v_x) = \left(\frac{m}{2\pi kT}\right)^{1/2} \exp\left(-\frac{\varepsilon_x}{kT}\right), \tag{72.11}$$

$$f(\mathbf{v}) = \left(\frac{m}{2\pi kT}\right)^{3/2} \exp\left(-\frac{\varepsilon}{kT}\right), \tag{72.12}$$

или более подробно

$$\varphi(v_x) = \left(\frac{m}{2\pi kT}\right)^{1/2} \exp\left(-\frac{mv_x^2}{2kT}\right), \tag{72.13}$$

$$f(\mathbf{v}) = \left(\frac{m}{2\pi kT}\right)^{3/2} \exp\left(-\frac{mv^2}{2kT}\right). \tag{72.14}$$

Это и есть окончательная формула, выражающая максвелловский закон распределения скоростей. Она применима не только к газам, но и к жидкостям и к твердым телам во всех случаях, когда еще можно пользоваться классическим способом описания движения.

ЗАДАЧИ

1. Применяя метод Максвелла, с помощью которого были выведены формулы (72.3), получить формулу (71.3), определяющую распределение частиц по ячейкам в демонстрационном опыте с доской Гальтона. С этой целью рассмотреть двухметровую доску Гальтона с ячейками, имеющими форму прямоугольных параллелепипедов, боковые грани которых параллельны координатным плоскостям XZ и YZ. Аналогичным путем получить закон ошибок Гаусса.

2. В законе ошибок Гаусса (71.4) выразить постоянную A через постоянную α.

Ответ. $A = \sqrt{\alpha/\pi}$.

3. Выразить через α среднюю и среднюю квадратичную ошибки при гауссовом законе распределения ошибок. Найти связь между этими ошибками.

Ответ.

$$\langle|x|\rangle = 1/\sqrt{\pi\alpha}, \quad \Delta_{\text{кв}} = \sqrt{\langle x^2\rangle} = 1/\sqrt{2\alpha}, \quad \Delta_{\text{кв}} = \sqrt{\pi/2}\,\langle|x|\rangle. \tag{72.15}$$

Непосредственное вычисление $\langle x^2\rangle$ довольно кропотливо и требует большой затраты времени. Последняя формула сводит вычисление этой величины к вычислению $\langle|x|\rangle$, что значительно проще. Такой прием целесообразно применять, например, при обработке результатов наблюдения броуновского движения (см. § 64).

§ 73. Распределение молекул по абсолютным значениям скоростей. Средние скорости молекул

1. Функция $f(\varepsilon)$ имеет смысл *объемной плотности вероятности*, с которой изображающие точки молекулы газа распределены по пространству скоростей. Умноженная на полное число молекул N, она дает среднее, или вероятное, число изображающих точек в единице объема пространства скоростей.

Найдем теперь *распределение молекул газа по абсолютным значениям их скоростей*. Направления скоростей нас больше не интересуют. Надо найти вероятность того, что модуль скорости молекулы заключен между v и $v+dv$. Эту вероятность будем обозначать через $F(v)\,dv$. Умноженная на N, она дает вероятное число молекул dN с такими скоростями. Новая функция распределения $F(v)$ просто связана с ранее введенной функцией $f(\varepsilon)$. Будем откладывать от одной и той же точки 0 векторы скоростей всех молекул газа. Из них отберем векторы с длинами, заключенными между v и $v+dv$. Соответствующие скоростные точки лежат внутри бесконечно тонкого шарового слоя со средним радиусом v и толщиной dv. Объем этого слоя $d\omega = 4\pi v^2\,dv$. Объемная плотность $f(\varepsilon)$ внутри шарового слоя постоянна, так как она зависит только от модуля скорости v, но не от ее направления. Умножив ее на объем слоя $d\omega$, находим искомую вероятность $f(\varepsilon)\,d\omega = 4\pi v^2 f(\varepsilon)\,dv$. Но для той же вероятности ранее мы писали $F(v)\,dv$. Сравнивая оба выражения получаем

$$F(v) = 4\pi v^2 f(\varepsilon), \tag{73.1}$$

или окончательно

$$F(v) = 4\pi\Big(\frac{m}{2\pi kT}\Big)^{3/2} v^2 \exp\Big(-\frac{\varepsilon}{kT}\Big). \tag{73.2}$$

Значит,

$$dN = 4\pi N\Big(\frac{m}{2\pi kT}\Big)^{3/2} v^2 \exp\Big(-\frac{\varepsilon}{kT}\Big)\,dv. \tag{73.3}$$

Ясно, что функция $F(v)$ удовлетворяет условию нормировки

$$\int_0^\infty F(v)\,dv = 1. \tag{73.4}$$

Этими формулами и решается поставленная задача.

2. График функции $F(v)$ представлен на рис. 53. Кривая $F(v)$ асимметрична и проходит через нуль в начале координат. Напротив, кривая $\varphi(v_x)$, как мы видели, симметрична и в начале координат проходит через максимум. Легко видеть, в чем причина этого различия. Выражение $\varphi(v_x)\,dv_x$ дает вероятность попадания молекулы в бесконечно тонкий плоский слой скоростного пространства между плоскостями $v_x = \text{const}$ и $v_x + dv_x = \text{const}$. Выражение $F(v)\,dv$ есть также вероятность попадания молекулы в бесконечно тонкий слой, но сферический, заключенный между двумя концентрическими сферами $v = \text{const}$ и $v + dv = \text{const}$. Если фиксировать толщину dv_x, то все

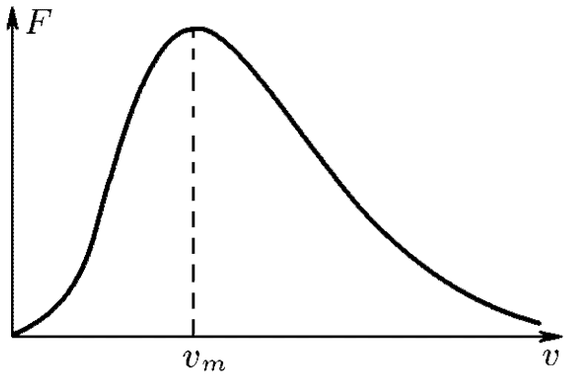


Рис. 53

плоские слои будут одинаковыми, каково бы ни было значение v_x. Естественно, что выражение $\varphi(v_x)\,dv_x$, а с ним и функция $\varphi(v_x)$ максимальны в центре, т. е. при $v_x = 0$. Напротив, при фиксированной толщине сферических слоев dv их объемы возрастают с возрастанием v благодаря наличию множителя $4\pi v^2$ в выражении $d\omega = 4\pi v^2 f(v)\,dv$. Ввиду этого положение максимума на кривой $F(v) = 4\pi v^2 f(v)$ определяется произведением конкурирующих множителей: монотонно убывающего $f(v)$ и монотонно возрастающего v^2.

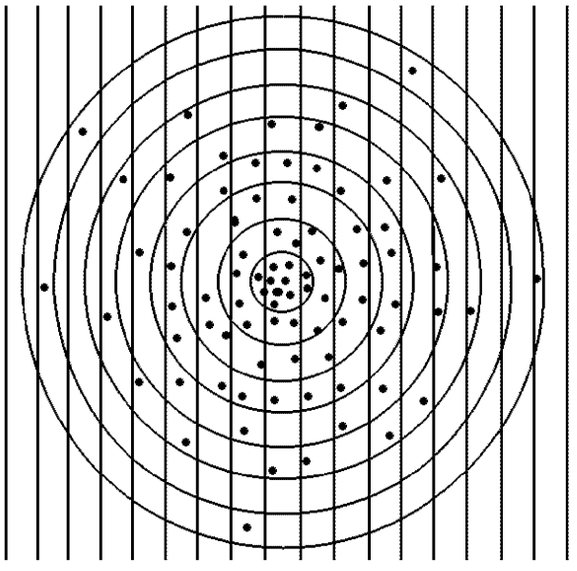


Рис. 54

Следующая аналогия полезна для уяснения указанного различия между функциями $\varphi(v_x)$ и $F(v)$. Допустим, что производится стрельба по мишеням. Попадание пули в то или иное место мишени есть случайное событие, а потому распределение пробоин в мишени подчиняется законам случая. Если мишень разделить на одинаковые вертикальные полоски (рис. 54), то, как и в опыте с доской Гальтона, распределение вероятности попадания в них представится симметричной кривой $\varphi(x)$. Если же мишень разделить на кольца равной толщины, то площади их будут возрастать с возрастанием радиуса, а распределение вероятности попадания в кольца представится асимметричной кривой, напоминающей кривую рис. 53.

3. Скорость, при которой функция $F(v)$ максимальна, называется *наиболее вероятной скоростью*. Будем обозначать ее через v_m. Для нахождения v_m величину F лучше рассматривать как функцию аргумента v^2. Дифференцируя (73.2) по указанному аргументу и приравнивая результат нулю, получим

$$\frac{d}{dv^2}\left[v^2 \exp\left(-\frac{mv^2}{2kT}\right)\right] = \left[1 - \frac{mv^2}{2kT}\right] \exp\left(-\frac{mv^2}{2kT}\right) = 0,$$

откуда

$$v \equiv v_m = \sqrt{2kT/m}. \tag{73.5}$$

Средняя, или средняя арифметическая, скорость молекулы определяется обычной формулой

$$\langle v \rangle = \frac{1}{N}\int v\,dN = \int\limits_0^\infty vF(v)\,dv.$$

Подставляя сюда значение $F(v)$ и интегрируя, получим

$$\langle v \rangle = \sqrt{\frac{8kT}{\pi m}} = v_m\sqrt{\frac{4}{\pi}} = 1{,}13\,v_m. \tag{73.6}$$

Добавим сюда еще среднюю квадратичную скорость, определяющую среднюю кинетическую энергию молекулы $(1/2)m\langle v^2\rangle = (1/2)mv_{\text{кв}}^2 = (3/2)kT$ (см. § 62). Эта скорость

$$v_{\text{кв}} = \sqrt{3kT/m} \approx 1{,}22\, v_m. \tag{73.7}$$

Эти три скорости отличаются друг от друга числовыми множителями порядка единицы, причем $v_{\text{кв}} > \langle v\rangle > v_m$. Поэтому каждая из них может быть использована для общего представления о скоростях теплового движения молекул.

ЗАДАЧИ

1. Написать выражение для среднего числа dN молекул газа, кинетические энергии которых заключены между ε и $\varepsilon + d\varepsilon$.

Ответ. $dN = 2\pi N(\pi kT)^{-3/2}\sqrt{\varepsilon}\exp\left(-\dfrac{\varepsilon}{kT}\right)d\varepsilon$.

2. Найти среднее значение обратной величины скорости молекулы в газе.

Ответ. $\left\langle\dfrac{1}{v}\right\rangle = \sqrt{\dfrac{2m}{\pi kT}} = \dfrac{4}{\pi\langle v\rangle}$.

3. Найти среднее число молекул, компоненты скорости которых, параллельные некоторой оси, лежат в интервале $(v_\parallel, v_\parallel + dv_\parallel)$, а модули перпендикулярной составляющей скорости заключены между $v_\perp$ и $v_\perp + dv_\perp$.

Решение. Искомое число молекул dN равно среднему числу скоростных точек в элементе объема пространств скоростей, заключенном между двумя коаксиальными цилиндрами радиусов $v_\perp$ и $v_\perp + dv_\perp$ и высоты $dv_\parallel$. Объем этого элемента равен $d\omega = 2\pi v_\perp\, dv_\perp\, dv_л$, а среднее число скоростных точек в нем

$$dN = f\,d\omega = 2\pi\left(\frac{m}{2\pi kT}\right)^{3/2}\exp\left(-\frac{\varepsilon}{kT}\right)v_\perp\, dv_\perp\, dv_\parallel.$$

4. В диоде электроны, эмитируемые накаленным катодом, попадают в задерживающее поле анода. До анода доходят лишь достаточно быстрые электроны. Считая, что тепловые скорости эмитируемых электронов (вышедших из катода) распределены по закону Максвелла с температурой $T = 1150$ K, определить долю электронов α, преодолевающих задерживающий потенциал: 1) $V = 0{,}2$ В; 2) $V = 0{,}4$ В. Катодом является тонкая прямолинейная нить, натянутая по оси цилиндрического анода.

Ответ. $\alpha = \exp(-eV/kT)$, где e — заряд электрона (по абсолютной величине);

1) $\alpha = 13{,}4\,\%$; 2) $\alpha = 1{,}8\,\%$.

§ 74. Другое доказательство закона распределения скоростей Максвелла. Принцип детального равновесия

1. Пусть в отсутствие силовых полей газ находится в закрытом сосуде, стенки которого поддерживаются при постоянной температуре. Если в какой-либо момент времени в газе создать какое угодно распределение скоростей между молекулами, то в результате столкновений

молекул между собой и со стенками сосуда установится такое статистической распределение молекул по скоростям, которое уже не будет меняться при дальнейших столкновениях. На этом основано второе доказательство закона распределения скоростей, данное Максвеллом. Оно существенно отличается от первого доказательства, приведенного нами в § 72, в котором представление о столкновениях совсем не используется. Мы приводим второе доказательство Максвелла главным образом потому, что в основе его лежит *принцип детального равновесия* — положение, имеющее самостоятельное значение и играющее большую роль в различных разделах физики. Будем считать, что молекулы при столкновениях ведут себя как идеально твердые и упругие шары. Такое предположение сильно упрощает доказательство, хотя оно и не является обязательным.

2. Рассмотрим две группы молекул, скоростные точки которых лежат в элементах объема $d\omega$ и $d\omega_1$ пространства скоростей с центрами в точках $\mathbf{v}$ и $\mathbf{v}_1$ соответственно. Если N — общее число молекул в сосуде, то средние числа молекул в рассматриваемых группах будут $dN = Nf(\mathbf{v})\,d\omega$ и $dN_1 = Nf(\mathbf{v}_1)\,d\omega_1$. Рассмотрим столкновение молекул первой группы с молекулами второй группы и притом такие, у которых линия центров (т. е. прямая, соединяющая центры шаров в момент столкновения) имеет какое-то фиксированное направление в пространстве, точнее, лежит в пределах бесконечно малого телесного угла $d\Omega$ с фиксированным направлением его оси. Среднее число таких «*прямых*» столкновений будет пропорционально произведению $dN\,N_1$, а с ним и произведению $f(\mathbf{v})f(\mathbf{v}_1)\,d\omega\,d\omega_1$.

В результате каждого «прямого» столкновения скорости $\mathbf{v}$ и $\mathbf{v}_1$ сталкивающихся молекул изменяются и переходят в $\mathbf{v}'$ и $\mathbf{v}_1'$ соответственно. Изображающие точки молекул первой группы перейдут из элемента объема $d\omega$ в элемент объема $d\omega'$ пространства скоростей с центром в точке $\mathbf{v}'$; изображающие точки молекул второй группы перейдут соответственно в элемент объема $d\omega_1'$ с центром в точке $\mathbf{v}_1'$. При столкновении молекулы обмениваются нормальными скоростями, т. е. скоростями, параллельными линии центров; их касательные скорости, т. е. скорости, перпендикулярные к линии центров, остаются без изменения. Если линию центров принять за ось X, то $v_x' = v_{1x}$, $v_{1x}' = v_x$. Отсюда $dv_x' = dv_{1x}$, $dv_{1x}' = dv_x$, а потому $dv_x' dv_{1x}' = dv_x\,dv_{1x}$. Мы видим, что в результате столкновения произведение $dv_x\,dv_{1x}$ остается неизменным, хотя dv_x и dv_{1x} могут изменяться. А так как поперечные размеры объемов $d\omega$ и $d\omega_1$ не меняются, то произведение $d\omega\,d\omega_1$ также остается неизменным:

$$d\omega\,d\omega_1 = d\omega'\,d\omega_1'. \tag{74.1}$$

3. Назовем *обратными* такие столкновения, которым соответствует то же самое направление линии центров и которые переводят скоростные точки сталкивающихся молекул из элементов объема $d\omega'$ и $d\omega_1'$ в исходные элементы $d\omega$ и $d\omega_1$ соответственно. Нетрудно найти по заданным прямым столкновениям обратные. На рис. 55 *а* и *б*

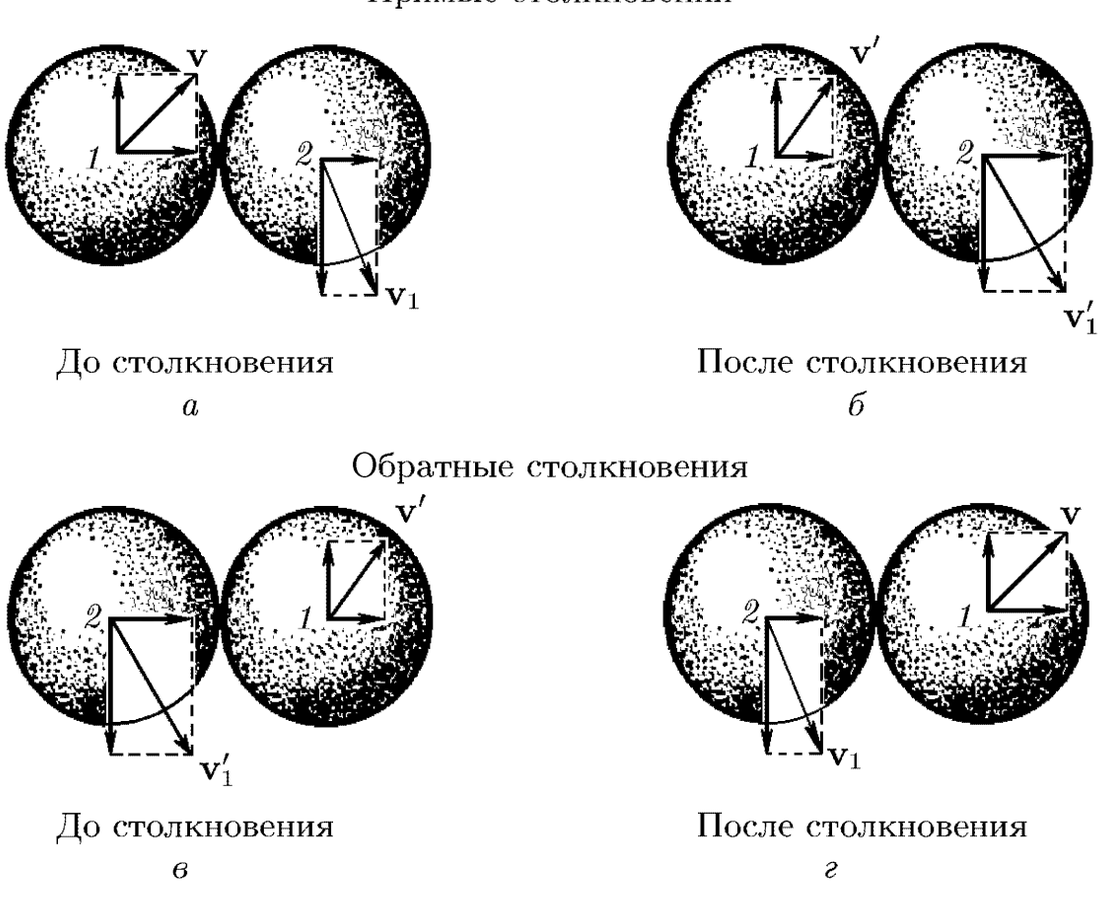


Рис. 55

изображены скорости молекул *1* и *2* до и после столкновения. Эти столкновения будем рассматривать как прямые. Скорости молекул после столкновения легко найти, если все скорости разложить на нормальные и касательные компоненты. Переставим теперь местами молекулы *1* и *2* и допустим, что перед столкновением они имели скорости $\mathbf{v}'$ и $\mathbf{v}_1'$ (рис. 55 *в*). Теперь молекула *2* становится ударяющей, а молекула *1* — ударяемой. Построив скорости после столкновения (рис. 55 *г*), найдем, что они примут исходные значения $\mathbf{v}$ и $\mathbf{v}_1$. Таким образом, столкновения 55 *в* являются обратными по отношению к столкновениям 55 *а* и наоборот.

Число обратных переходов скоростных точек из элементов объема $d\omega'$ и $d\omega_1'$ скоростного пространства в элементы $d\omega$ и $d\omega_1$ того же пространства пропорционально $f(\mathbf{v}')f(\mathbf{v}_1')\,d\omega' d\omega_1'$ или ввиду соотношения (74.1) $f(\mathbf{v}')f(\mathbf{v}_1')\,d\omega\,d\omega_1$. Из соображений симметрии следует, что коэффициент пропорциональности один и тот же для прямых и обратных переходов. Каждое прямое столкновение уводит скоростную точку из элемента объема скоростного пространства $d\omega$, каждое обратное столкновение вводит в тот же объем какую-то другую скоростную точку. Увеличение среднего числа скоростных точек в элементе $d\omega$ в течение какого-то промежутка времени t в результате прямых и обратных столкновений рассматриваемого типа пропорционально

$$dZ \sim [f(\mathbf{v}')f(\mathbf{v}_1') - f(\mathbf{v})f(\mathbf{v}_1)]\,d\omega\,d\omega_1. \qquad (74.2)$$

Прежде чем идти дальше, заметим, что для доказательства соотношения (74.2) можно было бы и не пользоваться моделью твердых упругих шаров. Вместо этого можно было бы построить доказательство на основе общих свойств симметрии, которые свойственны законам механики. Поэтому дальнейшие рассуждения не зависят от специальных предположений относительно формы молекулы и сил, действующих между ними.

4. Потребуем теперь, чтобы среднее число скоростных точек в элементе объема скоростного пространства $d\omega$ не изменялось в течением времени, несмотря на столкновения. В выражении (74.2) скорость $\mathbf{v}$ следует считать фиксированной (точнее, конец вектора $\mathbf{v}$ должен лежать в пределах $d\omega$). Напротив, скорости $\mathbf{v}_1$ и направления линии центров могут быть какими угодно. Скорости $\mathbf{v}'$, $\mathbf{v}'_1$ однозначно определяются заданием $\mathbf{v}_1$ и направления линии центров. Среднее приращение числа скоростных точек в элементе $d\omega$ за рассматриваемый промежуток времени найдется суммированием выражения (74.2) по всем возможным значениям скорости $\mathbf{v}_1$ и всем возможным направлениям линии центров. Чтобы среднее число скоростных точек в $d\omega$ не изменялось в результате столкновений, необходимо и достаточно, чтобы указанная сумма обращалась в нуль. Однако в состоянии хаоса, которым характеризуется тепловое движение молекул, надо потребовать большего. Надо, чтобы обращалась в нуль не только сумма в целом, но и каждое слагаемое (74.2) в отдельности. Смысл этого требования состоит в том, что *в газе в состоянии хаотического движения должны компенсировать друг друга всякие два противоположно направленные процесса. Скорости таких противоположно направленных процессов должны быть одинаковыми.* Это положение называется *принципом детального равновесия.* Если бы оно не выполнялось, то тепловое движение молекул в какой-то мере утратило бы беспорядочный характер и приобрело бы черты, свойственные упорядоченному движению. Принцип детального равновесия, разумеется, справедлив не только для газов, но и *для любых систем в состоянии полного хаоса.*

5. Следующий пример, принадлежащий Я. И. Френкелю (1894–1952), уясняет, почему установившееся хаотическое состояние является не просто состоянием статистического равновесия, а состоянием *детального статистического равновесия.* Пусть население какой-либо страны сосредоточено в городах, попарно связанных между собой железными дорогами. Жители путешествуют по этим дорогам, переезжая из города в город, причем среднее число жителей в каждом городе остается неизменным. Можно ли на основании этого утверждать, что среднее число жителей, переезжающих из одного произвольного города A в другой произвольный город B, равно среднему числу жителей, переезжающих в обратном направлении из B в A? Иными словами, можно ли утверждать, что рассматриваемое статистическое равновесие будет детальным? Нет, этого утверждать нельзя. Например, если число городов три — A, B и C, то постоянство числа жителей в каждом из них можно обеспечить путем движения пассажиров по замкнутому пути: из A в B, из B в C, из C в A и т. д. Однако такое перемещение населения не совместимо с представлением о хаотичности статистического равновесия, исключающей какое бы то ни было упорядоченное движение, в том числе

и круговое. Дополняя пример Френкеля, допустим, что путешествия жителей не являются целенаправленными, а совершенно случайными. Пусть, например, житель города A, отправляясь в путешествие, бросает монету. Выпадение герба или решки решит, в какой город ему ехать — в B или C. Так же поступает каждый житель городов B и C. Тогда неизменность среднего числа жителей в каждом из городов будет поддерживаться через детальное равновесие: число пассажиров, переезжающих в каком-либо направлении, в среднем будет равно числу пассажиров, едущих в обратном направлении.

6. Покажем теперь, как из принципа детального равновесия выводится максвелловский закон распределения скоростей. Надо потребовать, чтобы обращалось в нуль выражение в квадратных скобках формулы (74.2). При этом надо принять во внимание, что функция $f(\mathbf{v})$ может зависеть не от самой скорости $\mathbf{v}$, а от кинетической энергии ε. Это приводит к уравнению

$$f(\varepsilon')f(\varepsilon_1') = f(\varepsilon)f(\varepsilon_1), \tag{74.3}$$

или

$$\frac{f(\varepsilon')}{f(\varepsilon)} = \frac{f(\varepsilon_1)}{f(\varepsilon_1')} \tag{74.3а}$$

при дополнительном условии

$$\varepsilon + \varepsilon_1 = \varepsilon' + \varepsilon_1',$$

которое выражает закон сохранения энергии при столкновении. Для нахождения вида функции $f(\varepsilon)$ рассмотрим такие изменения аргументов, при которых ε_1 и ε_1' постоянны, а потому постоянна и разность $\varepsilon' - \varepsilon$. Тогда из уравнения (74.3) получаем

$$\frac{f(\varepsilon')}{f(\varepsilon)} = \text{const} \tag{74.4}$$

при условии

$$\varepsilon' - \varepsilon = C = \text{const},$$

где постоянная C может иметь любые значения. С аналогичным уравнение мы сталкивались в § 72. Только вместо частного $f(\varepsilon')/f(\varepsilon)$ там стояло произведение $f(\varepsilon')f(\varepsilon)$, а вместо разности $\varepsilon' - \varepsilon$ — сумма $\varepsilon' + \varepsilon$. Но для применимости метода решения это несущественно. Поступая, как и раньше, находим

$$f(\varepsilon) = Ae^{-\alpha\varepsilon},$$

т. е. максвелловский закон распределения скоростей.

7. Небольшое изменение в рассуждении приводит также к теореме о равномерном распределении средней кинетической энергии между различными молекулами газа. Рассмотрим для простоты смесь двух газов. Величины, относящиеся к одному из газов, будем снабжать нижним индексом 1, величины, относящиеся к другому газу — оставлять без индекса. Детальное равновесие должно иметь место по отношению к любым процессам, в том числе и к процессам столкновений между одинаковыми молекулами. Поэтому к каждому газу в отдельности

применимы рассуждения, приведенные выше. Из них следует, что функции распределения для обоих газов должны иметь вид

$$f(\varepsilon) = Ae^{-\alpha\varepsilon}, \quad f_1(\varepsilon_1) = A_1 e^{-\alpha_1\varepsilon_1}. \tag{74.5}$$

Постоянными α и α_1 определяются средние кинетические энергии молекул газа. Поэтому для доказательства теоремы о равномерном распределении кинетической энергии достаточно показать, что $\alpha = \alpha_1$. С этой целью рассмотрим столкновение молекул первого газа с молекулами второго газа и применим к ним принцип детального равновесия. Нетрудно показать (см. п. 8 этого параграфа), что соотношение (74.1) сохраняет силу и для таких столкновений. Далее надо принять во внимание, что относительные скорости в прямых и обратных столкновениях одинаковы по величине. На основании этих двух фактов можно утверждать, что средние числа прямых и обратных столкновений пропорциональны соответственно $f(\mathbf{v})f_1(\mathbf{v}_1)$ и $f(\mathbf{v}')f_1(\mathbf{v}_1')$ с одним и тем же коэффициентом пропорциональности. Поэтому принцип детального равновесия приводит к уравнению

$$f(\varepsilon)f_1(\varepsilon_1) = f(\varepsilon')f_1(\varepsilon_1').$$

Подставляя сюда выражения (74.5), получим

$$\alpha(\varepsilon - \varepsilon') = \alpha_1(\varepsilon_1' - \varepsilon_1).$$

С учетом закона сохранения энергии $\varepsilon + \varepsilon_1 = \varepsilon' + \varepsilon_1'$, или $\varepsilon - \varepsilon' = \varepsilon_1' - \varepsilon_1$, отсюда находим $\alpha = \alpha_1$.

8. В заключении остановимся на смысле и доказательстве соотношения (74.1) для столкновения шаров с различными массами m и m_1. Рассмотрим сначала частный случай, когда шары движутся вдоль линии центров. Тогда состояние движения в каждый момент времени может быть охарактеризовано импульсом первого шара p и импульсом второго шара p_1. Геометрически такое состояние можно представить на плоскости изображающей точкой A, прямоугольными координатами которой являются числа p и p_1 соответственно. После столкновения изображающая точка переместится в новое положение A' с координатами p' и p_1'. На основании законов сохранения импульса и энергии

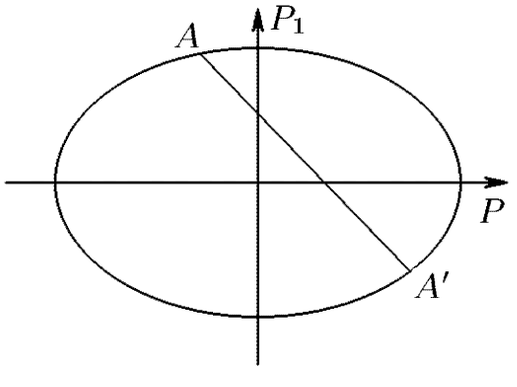


Рис. 56

$$p + p_1 = p' + p_1', \tag{74.6}$$

$$\frac{p^2}{2m} + \frac{p_1^2}{2m_1} = \frac{p'^2}{2m} + \frac{p_1'^2}{2m_1} = E. \tag{74.7}$$

Отсюда видно, что A' лежит в точке пересечения прямой (74.6), наклоненной под углом 135° к оси абсцисс, и эллипса с полуосями $\sqrt{2mE}$ и $\sqrt{2m_1E}$, причем обе эти линии пересекаются также и в точке A (рис. 56). Используя это, легко найти геометрическим построением положение точки A', если известно положение точки A.

Допустим теперь, что импульсы шаров до столкновения изменены, т. е. изменилось положение изображающей точки A, но направление линии цент-

ров осталось неизменным. Тогда изменится и положение изображающей точки A'. Но всегда положение A' однозначно определится, если известно положение A. Пусть точки A могут занимать любое положение в пределах произвольной области D. Тогда точки A' расположатся в пределах некоторой другой области, которую мы обозначим через D'. Докажем, что площади областей D и D' равны между собой. Это утверждение является частным случаем весьма общей теоремы аналитической механики, известной под названием *теоремы Лиувилля* (1809–1882) и играющей важную роль в статистической механике.

Достаточно доказать наше утверждение для бесконечно малой области D произвольной формы, так как из таких бесконечно малых областей можно составить любую область конечного размера. Возьмем два бесконечно близких подобных эллипса типа (74.7), отличающихся друг от друга значением энергии E. Пересечем их двумя бесконечно близкими прямыми, наклоненными под углом 135° к оси абсцисс. В пересечении образуются два бесконечно малых параллелограмма, заштрихованных на рис. 57. Один из этих параллелограммов примем за область D, другой — за область D'. Высоты этих параллелограммов, перпендикулярные к прямой AA', будут одинаковы, как это ясно из построения. Будут равны также и основания AA_1 и $A'A'_1$. В этом легче всего убедиться, заметив, что наши подобные эллипсы могут быть получены из двух концентрических окружностей путем их равномерного растяжения или сжатия вдоль горизонтальной или вертикальной оси в одно и то же число раз. До деформации длины сторон AA_1 и $A'A'_1$ были равны, как это следует из геометрических свойств круга. Это равенство сохранится и после деформации, так как при однородном одностороннем растяжении и сжатии длины параллельных отрезков изменяются в одинаковое число раз. Этим доказано равенство оснований, а следовательно, и площадей параллелограммов D и D'. Заметим, что форма самих параллелограммов D и D', вообще говоря, разная.

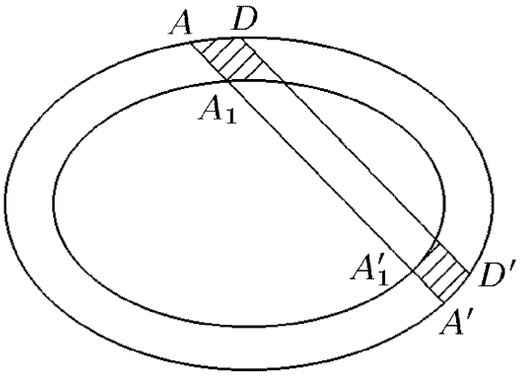


Рис. 57

Не представляет труда обобщить доказанную теорему на случай, когда сталкивающиеся шары имеют не только составляющие скорости вдоль линии центров, но и поперечные скорости, к ней перпендикулярные. Доказательство применимо и в этом случае, так как поперечные скорости при столкновении не изменяются. Только вместо двумерных областей D и D' появятся соответствующие области в *шестимерном пространстве*. Теорема состоит в том, что объемы этих шестимерных областей одинаковы. Наконец, все рассуждения останутся верными, если импульсы p и p_1 заменить соответствующими им скоростями.

§ 75. Среднее число молекул, сталкивающихся со стенкой сосуда

1. Среднее число ударов молекул о стенку сосуда в единицу времени можно оценить следующим простым способом. Пусть n — среднее число молекул в единице объема. Рассмотрим на стенке сосуда элементарную площадку dS и введем прямоугольную координатную

систему XYZ (рис. 58). Ось X направим по нормали к площадке dS, а оси Y и Z расположим в плоскости, перпендикулярной к этой нормали. Введем два упрощающих предположения: 1) скорости всех молекул одинаковы по модулю; 2) молекулы движутся только параллельно координатным осям, а именно так, что одна шестая всех молекул движется в положительном направлении оси X, одна шестая — в отрицательном, и аналогично для осей Y и Z. При таких упрощениях с площадкой dS будут сталкиваться только молекулы, движущиеся к стенке, т. е. в положительном направлении оси X. Число таких молекул в единице объема $n_x = (1/6)n$. За врем dt с площадкой dS столкнутся все молекулы рассматриваемой группы, которые лежат внутри цилиндра с основанием dS и высотой $v\,dt$. Число молекул в этом цилиндре, движущихся к стенке, равно $dz = n_x v S\,dt = (1/6)nvS\,dt$. Среднее число молекул, сталкивающихся в единицу времени с единичной площадкой, будет

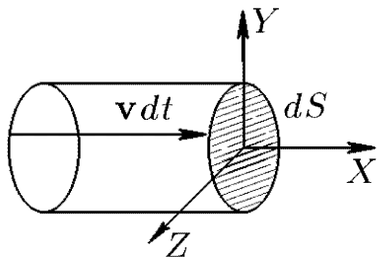


Рис. 58

$$z = \frac{1}{6}\,nv. \tag{75.1}$$

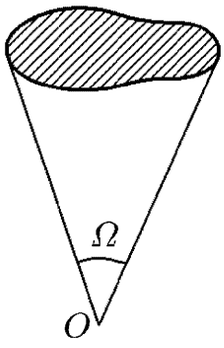


Рис. 59

2. Найдем теперь точное выражение для среднего числа ударов z. В газе в состоянии покоя все направления скоростей молекул равновероятны, т. е. распределены в пространстве изотропно. Найдем среднее число N_Ω молекул, направления скоростей которых лежат в пределах телесного угла Ω (рис. 59). Так как телесный угол, охватывающий все направления в пространстве, равен 4π, то ввиду указанной изотропии

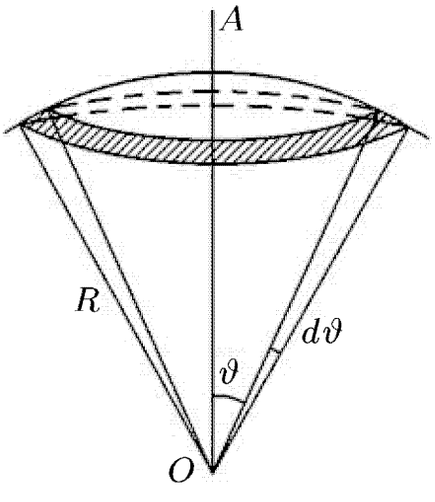


Рис. 60

$$N_\Omega = \frac{N}{4\pi}\,\Omega, \tag{75.2}$$

где N — общее число молекул. В частности, для бесконечно малого телесного угла

$$dN = \frac{N}{4\pi}\,d\Omega. \tag{75.3}$$

Пользуясь формулой (75.3), определим среднее число молекул dN, скорости которых образуют с некоторым фиксированным направлением OA (рис. 60) углы, лежащие между ϑ и $\vartheta + d\vartheta$. Для этого опишем сферу радиуса R с центром в точке O и построим два прямых круговых конуса с общей осью OA, образующие которых составляют с этой осью углы ϑ и $\vartheta + d\vartheta$. Конусы вырежут на сфере кольцевую полоску, заштрихованную на рис. 60. Площадь полоски равна $dS = 2\pi R^2 \sin\vartheta\,d\vartheta$. Она видна из точки O под

телесным углом $d\Omega = dS/R^2 = 2\pi \sin\vartheta\, d\vartheta$. Подставляя это значение в формулу (75.3), получим

$$dN = \frac{N}{2} \sin\vartheta\, d\vartheta. \tag{75.4}$$

Теперь легко найти выражение для z. Внешнюю нормаль к стенке примем за ось X (рис. 61). Угол между осью X и скоростью молекулы обозначим через ϑ. Сначала предположим, что скорости всех молекул одинаковы по величине, а их направления распределены изотропно. Выделим группу молекул с x-компонентами скоростей между v_x и $v_x + dv_x$. Чтобы удариться о стенку, молекулы должны лететь к ней, а потому должно быть $v_x > 0$. Пусть dn — число таких молекул в единице объема. Число ударов о квадратный сантиметр стенки, производимых молекулами выделенной группы в одну секунду, равно $dz = dn\, v_x = v\cos\vartheta\, dn$. В этом легко убедиться, если произвести еще более детальную сортировку молекул по скоростям. Надо выделить молекулы с определенным направлением скорости и на площадке dS построить косой параллелепипед с боковой стороной $\mathbf{v}\, dt$, как указано на рис. 61. Число ударов выделенных молекул о площадку dS за время dt найдется, если их концентрацию умножить на объем параллелепипеда $dS v_x\, dt$. Затем надо просуммировать по всем параллелепипедам, имеющим разные направления боковых сторон, но общую высоту $v_x\, dt$. Разделив результаты на dt, мы и придем к выражению $dz = v\cos\vartheta\, dn$. Поскольку молекулы из числа dn характеризуются общими значениями v и v_x, они движутся под определенным углом к оси X. Точнее, это молекулы, скорости которых образуют с направлением оси X углы между ϑ и $\vartheta + d\vartheta$. По формуле (75.4) $dn = (1/2)n\sin\vartheta\, d\vartheta$, а потому

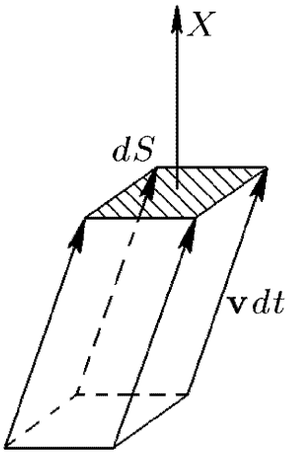


Рис. 61

$$dz = \frac{1}{2} nv \sin\vartheta\cos\vartheta\, d\vartheta.$$

Интегрируя это выражение по всем молекулам, летящим к стенке, т. е. в пределах от $\vartheta = 0$ до $\vartheta = \pi/2$, получим

$$z = \frac{1}{4} nv.$$

Не составляет труда обобщить этот результат на случай, когда скорости молекул не одинаковы по величине, но их направления попрежнему распределены изотропно. Разобьем все молекулы на группы со скоростями $v_1, v_2, \ldots, v_s$. Пусть $n_1, n_2, \ldots, n_s$ означают числа таких молекул в единице объема. Очевидно,

$$z = \frac{1}{4} \sum n_i v_i.$$

По определению среднего

$$\langle v\rangle = \frac{1}{n}\sum n_i v_i,$$

а потому

$$z = \frac{1}{4}\, n\langle v\rangle. \tag{75.5}$$

При максвелловском распределении скоростей для $\langle v\rangle$ следует воспользоваться выражением (73.6). Это дает

$$z = n\sqrt{\frac{kT}{2\pi m}}. \tag{75.6}$$

3. Таким образом, отличие приближенной формулы (75.1) от точной (75.5) состоит в том, что вместо коэффициента $1/6$ надо брать коэффициент $1/4$. На самом деле различие еще меньше. Действительно, чтобы значение кинетической энергии газа получилось правильным, под v в формуле (75.1) следует понимать среднюю квадратичную скорость, связанную с $\langle v\rangle$ соотношением $\langle v_{\text{кв}}\rangle = 1{,}08\,\langle v\rangle$.

В рассмотренной задаче получение точного выражения (75.5) не представляло большого труда. Но, как правило, задачи, которые приходится ставить и решать в кинетической теории газов, очень сложны. Их точные и строгие решения получить очень трудно и часто даже невозможно. Поэтому очень большое значение приобретают соотношения оценочного характера. При их выводе вводятся упрощающие предположения, сильно облегчающие вычисления и в то же время не затрагивающие существенные черты явлений. Обычно такого рода оценочные соотношения отличаются от точных выражений малосущественными числовыми множителями порядка единицы. Метод оценок в дальнейшем нами будет широко применяться.

4. С помощью формулы (75.6) можно оценить среднее число молекул, вылетающих в одну секунду при испарении с единицы поверхности жидкости или твердого тела, граничащей с вакуумом. Для этого сначала рассмотрим случай, когда тело граничит со своим насыщенным паром. При наличии пара идет не только испарение, но и обратный процесс конденсации молекул пара, летящих к поверхности тела. Молекулы, летящие из тела в область, занятую паром, или обратно, могут претерпеть отражение на границе тел. Но при оценках этим эффектом можно пренебречь. Если пар насыщенный, то имеет место детальное равновесие, при котором каждому потоку испаряющихся молекул от тела соответствует равный и противоположно направленный поток молекул пара к телу. Но среднее число молекул пара, ударяющихся о единицу поверхности тела, дается выражением (75.6). Таково же будет и среднее число молекул, вылетающих ежесекундно с единицы поверхности тела. Это число не изменится, если весь пар над телом удалить, так как оно зависит только от состояния самого тела, а не от состояния окружающей среды. Под n в формуле (75.6) следует понимать число молекул насыщенного пара в единице объема. На этом

основан метод измерения давления насыщенных паров тугоплавких металлов (см. задачу 10 к этому параграфу).

ЗАДАЧИ

1. Число ударов молекул о квадратный сантиметр стенки в одну секунду можно представить интегралом $z = n\int v_x\varphi(v_x)\,dv_x$, где интегрирование производится по всем молекулам, летящим к стенке. (Предполагается, что стенка перпендикулярна к оси X.) Убедиться непосредственным расчетом, что при максвелловском распределении скоростей этот интеграл приводится к выражению (75.6).

2. В тонкостенном сосуде объема V, стенки которого поддерживаются при постоянной температуре, находится идеальный газ. Сосуд помещен в вакуум. Как будет меняться с течением времени концентрация молекул n газа внутри сосуда, если в его стенке сделать очень малое отверстие площади S [1])?

Р е ш е н и е. Если отверстие S очень мало, то распределение скоростей исказится очень мало, т. е. останется изотропным и максвелловским. По формуле (75.5) получаем

$$d(Vn) = -\frac{1}{4}\,Sn\langle v\rangle\,dt.$$

Интегрируя это уравнение, получаем

$$n = n_0\,e^{-t/\tau}, \quad \text{где} \quad \tau = 4V/S\langle v\rangle.$$

3. Откачанный тонкостенный сосуд, стенки которого поддерживаются при постоянной температуре, погружен в атмосферу идеального газа с постоянной концентрацией молекул n_0, поддерживаемого при той же температуре. Как будет меняться с течением времени концентрация молекул газа внутри сосуда, если в его стенке сделать очень маленькое отверстие?

О т в е т. $n = n_0(1 - e^{-t/\tau})$. Обозначения такие же, как и в предыдущей задаче.

4. Полностью эвакуированный герметический сосуд помещен в атмосферу, состоящую из смеси двух газов, молекулярные веса которых относятся как 1 : 4, а отношение концентраций (т. е. чисел молекул в единице объема) равно α. Смесь газов вне сосуда поддерживается при постоянных давлении и температуре. В стенке сосуда оказалось малое отверстие, через которое оба газа начали медленно натекать в сосуд. Определить максимальное и минимальное значения отношения концентрации легкого к концентрации тяжелого компонентов газовой смеси в сосуде и моменты времени, когда достигаются эти значения.

Р е ш е н и е. Поступая, как в задачах 1 и 2, для отношения концентраций легкого и тяжелого компонентов внутри сосуда найдем выражение

$$\beta = \alpha\,\frac{1 - e^{-t/\tau_1}}{1 - e^{-t/\tau_2}},$$

где индекс 1 относится к легкому, а индекс 2 — к тяжелому компонентам. Времена τ_1 и τ_2 связаны соотношением $\tau_2/\tau_1 = 2$. Учитывая это, найдем,

[1]) В задачах 2, 3, 4, 5, 7, 8 предполагается, что размеры отверстия и толщина стенки малы по сравнению с длиной свободного пробега (см. § 86 и 95).

что производная $d\beta/dt$ обращается в нуль, когда

$$e^{-t/\tau_2} = \sqrt{2} - 1$$

и, следовательно, когда $\beta = \alpha\sqrt{2}$. Однако этому случаю соответствует не максимум и не минимум на кривой $\beta = \beta(t)$, а точка перегиба. Максимальное и минимальное значения величина β принимает на концах временного интервала $(0, \infty)$. При $t = 0$ получается максимум $\beta_{\text{макс}} = \alpha\tau_2/\tau_1 = 2\alpha$, при $t = \infty$ — минимум $\beta_{\text{мин}} = \alpha$.

5. Полностью эвакуированный тонкостенный герметический сосуд помещен в атмосферу кислорода, поддерживаемого при постоянной температуре и невысоком давлении P. В стенке сосуда оказалось малое отверстие, через которое окружающий кислород стал натекать в сосуд. Через час давление газа в сосуде повысилось от нуля до $P/2$. Какое давление было бы в том же сосуде через то же время, если бы после откачки сосуд был помещен в атмосферу водорода при тех же давлении и температуре?

Ответ. $(15/16)P$.

6. Найти полную кинетическую энергию E молекул идеального одноатомного газа, ударяющихся о квадратный сантиметр стенки в единицу времени.

Ответ. $E = (1/8)mn\langle v^3\rangle$. Для максвелловского распределения

$$E = n\sqrt{\frac{2k^3T^3}{m\pi}} = \frac{1}{16}\,nm\pi\langle v\rangle^3.$$

7. В тонкостенном сосуде, помещенном в вакууме, имеется очень маленькое отверстие, на которое извне направляется параллельный пучок одноатомных молекул, летящих с одной и той же скоростью v_0, перпендикулярной к площади отверстия. Концентрация молекул в пучке равна n_0. Найти в установившемся равновесном состоянии среднюю скорость $\langle v\rangle$, концентрацию молекул n и температуру T газа в сосуде.

Решение. Из-за столкновений молекул со стенками сосуда и между собой внутри сосуда устанавливается максвелловское распределение скоростей. Условия сохранения числа частиц и кинетической энергии газа в сосуде имеют вид

$$n_0 v_0 = \frac{1}{4}\,n\langle v\rangle, \quad \frac{1}{2}\,n_0 m v_0^3 - \frac{1}{16}\,nm\pi\langle v\rangle^3.$$

Используя их, а также формулу (73.6), находим

$$\langle v\rangle = \sqrt{\frac{2}{\pi}}\,v_0, \quad n = n_0\sqrt{8\pi}, \quad T = \frac{mv_0^2}{4k}.$$

8. В тонкостенном сосуде, содержащем одноатомный идеальный газ при температуре T, имеется очень маленькое отверстие, через которое молекулы вылетают в вакуум. Определить среднее значение $\langle\varepsilon\rangle$ кинетической энергии вылетевшей молекулы в предположении, что за время опыта изменения числа молекул и температуры газа в сосуде пренебрежимо малы.

Ответ. $\langle\varepsilon\rangle = 2kT$.

9. Определить, какая часть молекул идеального газа, столкнувшихся со стенкой сосуда за определенное время (например, за одну секунду), имеет кинетическую энергию, превосходящую ε.

Ответ. $\alpha = (1 + \varepsilon/kT)e^{-\varepsilon/kT}$.

10. Вольфрамовая нить, испаряясь в высокий вакуум при температуре $T = 2000$ К, уменьшается в весе, как показали измерения, со скоростью $q = 1{,}14 \cdot 10^{-13}$ г/(с · см). Оценить давление насыщенного пара вольфрама при этой температуре.

Р е ш е н и е. На основании изложенного в п. 4 этого параграфа скорость испарения дается формулой

$$q = \frac{1}{4} nm\langle v \rangle,$$

где n — концентрации атомов насыщенного пара вольфрама. Его давление будет

$$P = \frac{1}{3} nm\langle v^2 \rangle = \frac{4}{3} q \frac{\langle v^2 \rangle}{\langle v \rangle}.$$

При максвелловском распределении

$$\frac{\langle v^2 \rangle}{\langle v \rangle} = 3\sqrt{\frac{\pi kT}{8m}} = 3\sqrt{\frac{\pi RT}{8A}},$$

где A — атомный вес, равный для вольфрама 184. Окончательно получаем

$$P = q\sqrt{2\pi RT/A}.$$

Подставив сюда числовые значения, найдем для давления насыщенных паров вольфрама при $T = 2000$ К, $P = 8{,}6 \cdot 10^{-9}$ дин/см^2 $= 6{,}4 \cdot 10^{-12}$ мм рт. ст. Ввиду малости этой величины ее трудно измерить прямым методом.

§ 76. Опытная проверка закона распределения скоростей Максвелла

1. Одна из первых проверок закона распределения скоростей Максвелла была осуществлена Ричардсоном (1879–1959) в 1921 г. Если раскаленная поверхность металла граничит с вакуумом, то она испускает электроны. Это явление называется *термоэлектронной эмиссией* и имеет разнообразнейшие научно-технические применения. В состоянии статического равновесия над поверхностью металла образуется насыщенный пар из электронов. Ричардсон показал экспериментально, что распределение скоростей электронов пара подчиняется закону Максвелла. Это объясняется тем, что при малой концентрации электронов над поверхностью металла электронный пар может рассматриваться классически, так как температура вырождения, вычисленная по формуле (71.8), значительно ниже температуры электронного газа (равной температуре металла). Обратное соотношение, как было показано в § 71, имеет место для электронов внутри самого металла. Здесь электронный газ вырожден и поэтому должен рассматриваться как квантовая система. Максвелловское распределение для него не справедливо и должно быть заменено *распределением Ферми–Дирака* (см. § 82).

2. Для целей нашего курса основной интерес представляет проверка закона распределения скоростей Максвелла для *молекулярных и атомных газов*. Все опыты, относящиеся к этой области, были

выполнены с атомными пучками. Пионером в этой области был Отто Штерн, впервые измеривший средние скорости атомов в таких пучках. Как уже указывалось в § 60, опыты Штерна показали, что атомы в пучке летят с различными скоростями. В дальнейшем эти опыты были специально приспособлены для изучения распределения атомов по скоростям. Элдридж (1927 г.) и Ламмерт (1926-1929 гг.) сконструировали селекторы скоростей, работавшие на принципе зубчатого колеса, с помощью которого в середине XIX столетия Физо (1819-1896 гг.) измерил скорость света.

Схема опыта показана на рис. 62. Атомы легкоплавкого металла, испаряясь в печи A, выходят наружу через узкую щель S_1. На своем пути они встречают вторую щель S_2, вырезающую узкий атомный

Рис. 62

пучок. Пучок попадает на прорези диска D_1, представляющего собой подобие зубчатого колеса. На оси этого диска насажен второй такой же диск D_2, но прорези в нем смещены относительно прорезей первого диска на малый угол α (около 2°). За диском D_2 помещается третья щель S_3, пройдя через которую пучок попадает на стеклянную пластинку P, охлаждаемую жидким азотом, и конденсируется на ней. Пластинку можно было наблюдать с помощью микроскопа M. Вся система (за исключением, конечно, микроскопа) помещается в высоком вакууме, так что атомы пролетают через нее практически без столкновений друг с другом. Когда диски покоятся, то атомы пучка, пройдя через щели диска D_1, попадают на зубцы диска D_2 и задерживаются ими. В этом случае на пластинку P они попадать не могут. Однако если диски привести во вращение, то атомы с определенными скоростями могут проходить через систему и оседать на пластинке P. Максимум прохождения будет тогда, когда за время пролета от диска D_1 до диска D_2 диски поворачиваются на угол α. Таким образом, при определенной скорости вращения система пропускает атомы только определенных скоростей. По скорости вращения дисков можно вычислить скорость прошедших атомов, осевших на пластинке P. Интенсивность прошедшего пучка, пропорциональную числу атомов в нем, можно оценить по времени, которое требуется на то, чтобы на пластинке появился осадок, видимый в микроскоп. Специально поставленные опыты показали, что видимый осадок появляется при вполне определенном и всегда одинаковом числе осевших атомов. Можно определять относительное число осевших атомов и другими способами, например

по степени прозрачности пластинки P, измеряемой фотометрированием. Меняя угловую скорость вращения, можно выделять пучки с различными скоростями и таким путем изучать распределение атомов по скоростям. Обычно круговые частоты, применявшиеся в опытах описанного типа, менялись в пределах 10–50 $с^{-1}$.

3. Описанный прибор обладает следующим недостатком. Если атомы, имеющие определенную скорость, проходят через щели при угловой скорости вращения ω, то они пройдут через них и при угловых скоростях 2ω, 3ω и т. д. Значит, прибор выделяет пучки не с одной определенной скоростью, а с несколькими (кратными) скоростями. Для устранения этого недостатка Миллер и Куш (1955 г.) заменили диски с зубцами сплошным металлическим цилиндром, вдоль боковой поверхности которого были прорезаны узкие винтовые канавки под малым углом α к образующим цилиндра. При вращении цилиндра через канавки могут пройти только те атомы, скорость которых v удовлетворяет условию $\operatorname{tg}\alpha = \omega R/v$, где R — радиус цилиндра.

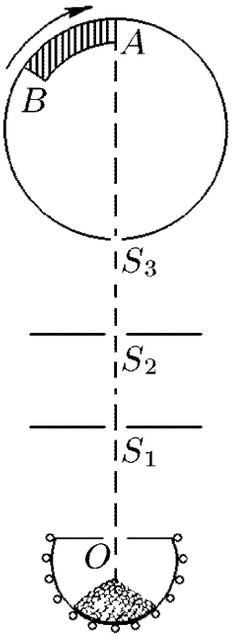


Рис. 63

4. Несколько иной метод для изучения распределения атомов по скоростям применялся в опытах Цартмана, прибор которого изображен на схематическом рис. 63. В электрической печи помещается чистый легкоплавкий металл. При нагревании образуются пары этого металла, которые могут вылетать через узкое отверстие O. Прибор помещен в высоком вакууме. Благодаря малости отверстия O плотность паров металла в печи довольно велика. В ней происходят многочисленные столкновения между атомами. Поэтому можно ожидать, что скорости атомов металла в печи будут распределены по закону Максвелла. Вне печи, где поддерживается высокий вакуум, атомы металла движутся практически без столкновений. Они проходят через узкие щели S_1 и S_2 и попадают внутрь вращающегося цилиндра в те моменты, когда щель S_3 расположена на одной прямой с щелями S_1 и S_2. Если бы атомы летели с бесконечно большими скоростями, то они попадали бы в точку A цилиндра, диаметрально противоположную щели S_3. В действительности на прохождение диаметра цилиндра D атом затрачивает конечное время. Точка попадания смещается по поверхности цилиндра против вращения на расстояние

$$x = \frac{\omega DT}{2} = \frac{D^2\omega}{2v},$$

где v — скорость атома, ω — угловая скорость вращения цилиндра. К внутренней стенке цилиндра прикреплена согнутая в дугу стеклянная пластинка AB. Чтобы лучше задерживать атомы металла, пластинка предварительно покрывается тонким слоем того же металла и тщательно фотометрируется. Затем цилиндр приводят во вращение

и включают печь. Опыт продолжается около 10 часов. После этого пластинка AB с осевшими на ней атомами металла вынимается и снова фотометрируется. По степени почернения пластинки можно определить плотность распределения металла, сконденсировавшегося на ней. Вблизи точки A распределятся атомы металла с наибольшими скоростями, вблизи B — с наименьшими. Предположив, что скорости атомов в парах металла внутри печи распределены по закону Максвелла, можно заранее рассчитать плотность распределения исследуемого металла на пластинке AB.

В опытах Элдриджа исследовались пучки атомов кадмия, в опытах Ламмерта — ртути, в опытах Миллера и Куша — калия и таллия, в опытах Цартмана — висмута. Все эти опыты оказались в превосходном согласии с максвелловским законом распределения скоростей. Этого и следовало ожидать в связи с тем, что было сказано в конце § 71.

ЗАДАЧА

В опыте Штерна (см. рис. 44) на поверхности вращающегося цилиндра C конденсируются молекулы серебра с различными скоростями. Каким скоростям молекул, попадающих на пластинку DD', соответствует ее наибольшее почернение?

Решение. Если установка неподвижна, то молекулы конденсируются в точке D. При вращении всей установки молекулы со скоростями v попадают в D'. Смещение по дуге DD' равно $x = C/v$, где C — постоянная прибора. Число молекул dN со скоростями между v и $v + dv$, ежесекундно испускаемых источником A, пропорционально $vF(v)\,dv$. Выражая dv через dx, представим его в виде $dN = v^3 F(v)\,dx$. Отсюда видно, что линейная плотность распределения молекул, сконденсировавшихся на поверхности цилиндра, пропорциональна $v^3 F(v)$, т. е. $v^5 \exp(-v^2/v_m^2)$. Эта плотность максимальна при $v = \sqrt{5/2}\,v_m$ — наиболее вероятная скорость.

§ 77. Закон распределения Больцмана

1. В отсутствие внешних сил средняя концентрация молекул газа n в состоянии равновесия всюду одинакова. Но этого не будет при наличии силовых полей. Рассмотрим, например, идеальный газ в однородном поле тяжести. В состоянии теплового равновесия температура T должна быть одинаковой по всей толще газа. Иначе в газе возникли бы потоки тепла, направленные в сторону понижения температуры, и состояние газа не было бы равновесным. Для механического равновесия необходимо, сверх того, чтобы концентрация молекул газа убывала с увеличением высоты. Направим ось Z вертикально вверх и найдем закон изменения концентрации n с координатой z в состоянии теплового и механического равновесия. Выделим мысленно бесконечно короткий вертикальный столб газа $ABDC$ (рис. 64) с высотой dz. Пусть площадь основания

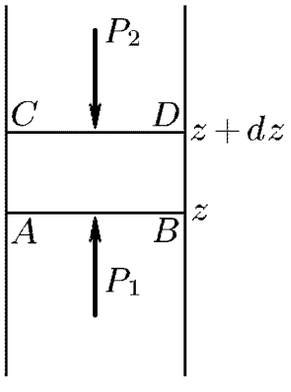


Рис. 64

столба равна единице. Вес столба $nmg\,dz$ должен уравновешиваться разностью давлений

$$P_1 - P_2 = -\frac{dP}{dz}\,dz.$$

Это приводит к соотношению

$$\frac{dP}{dz} = -nmg. \tag{77.1}$$

Подставляя сюда $P = nkT$ и принимая во внимание, что температура T одинакова на всех высотах, получим

$$kT\,\frac{dn}{dz} = -nmg, \quad \text{или} \quad kT\,d\ln n = -mg\,dz.$$

Для справедливости этого соотношения предположение об однородности поля тяжести, использованное при выводе, несущественно. Аналогичное соотношение можно получить и для неоднородного поля. Для этого надо написать условие механического равновесия части газа, заполняющей настолько малую область пространства, что в пределах этой области поле $\mathbf{g}$ может считаться однородным. Условие равновесия в этом случае удобнее писать в векторной форме:

$$kT\,d\ln n = -m(\mathbf{g}\,d\mathbf{r}). \tag{77.2}$$

Физическая природа силового поля $\mathbf{g}$ также не играет роли. Оно не обязательно должно быть гравитационным, а может быть электрическим или каким-либо другим. Важно только, что *поле не должно зависеть от времени и быть консервативным (потенциальным). В неконсервативных полях равновесие невозможно.* В этом нетрудно убедиться, интегрируя по замкнутому контуру обе части соотношения (77.2). Если поле $\mathbf{g}$ не консервативно, то по крайней мере для некоторых контуров интервал $\oint \mathbf{g}\,d\mathbf{r}$ будет отличен от нуля. Интеграл же от левой части (77.2) равен нулю по любому замкнутому контуру ввиду однозначности функции $n(\mathbf{r})$. Получившееся противоречие и доказывает наше утверждение. Следует, однако, отметить, что консервативность (потенциальность) силового поля является только необходимым, но недостаточным условием равновесия газа (см. § 79, п. 1).

Если ε_p — потенциальная энергия молекулы в силовом поле, то $m(\mathbf{g}\,d\mathbf{r}) = d\varepsilon_p$, а потому

$$kT\,d\ln n = -d\varepsilon_p. \tag{77.2а}$$

В этом виде в соотношении (77.2а) уже не осталось никаких признаков однородности и физической природы силового поля. Интегрируя, получаем

$$n = n_0 \exp\Big(-\frac{\varepsilon_p}{kT}\Big). \tag{77.3}$$

Это важное соотношение называется *законом распределения Больцмана* или просто *распределением Больцмана.* Постоянная n_0 имеет смысл значения n при $\varepsilon_p = 0$.

Применительно к однородному полю тяжести, если от концентрации n перейти к давлению газа P, формула (77.3) преобразуется в

$$P = P_0 \exp\Big(-\frac{\mu g z}{RT}\Big), \tag{77.4}$$

где μ — молекулярная масса газа, R — универсальная газовая постоянная, а P_0 — давление на уровне $z = 0$. Это — *барометрическая формула*, с которой мы имели дело в механике (см. т. I, § 92).

2. Приведенный вывод распределения Больцмана является чисто *гидростатическим* — в нем мы по существу отвлекаемся от молекулярной структуры газа, рассматривая его как сплошную среду. Это допустимо лишь для достаточно плотных газов при наличии большого числа столкновений. Требуется, чтобы средний пробег молекулы между двумя последовательными столкновениями был мал по сравнению с толщиной dz слоя $ABDC$ (рис. 64), который рассматривался при выводе распределения Больцмана. Только тогда имеет смысл говорить о давлении, с которым на слой dz действует окружающий газ. В недостаточности гидростатического вывода можно убедиться также с помощью следующих соображений. В гидростатическом выводе величины n и m входят в формулы не независимо, а только через плотность, т. е. в комбинации $\rho = nm$. Если в поле тяжести имеется смесь различных равномерно перемешанных идеальных газов, то согласно гидростатическому выводу такое состояние равномерного перемешивания должно неограниченно долго сохраняться и в дальнейшем, причем давление должно определяться барометрической формулой (77.4), в которой под μ следует понимать среднюю молекулярную массу смеси. Однако это заключение находится в противоречии с тем, что известно о свойствах идеальных газов. *Поведение идеального газа в том или ином объеме совершенно не зависит от того, есть в нем другие идеальные газы или нет.* В состоянии термодинамического равновесия концентрации различных газов в смеси должны убывать с высотой экспоненциально с различными экспонентами, определяющимися молекулярными массами соответствующих компонентов смеси. Концентрация легких газов должна убывать с увеличением высоты медленнее, а тяжелых — быстрее. По мере поднятия относительная концентрация легких газов должна возрастать. В действительности в пределах тропосферы этого не происходит. Но это, конечно, не может служить реабилитацией гидростатического вывода, поскольку все наши рассуждения относятся только к случаю термодинамически равновесной атмосферы. В реальной же тропосфере происходят оживленные движения, приводящие к интенсивному перемешиванию ее нижних и верхних слоев.

3. Можно дать молекулярно-кинетический вывод закона распределения Больцмана, свободный от недостатков, присущих гидростатическому выводу. Приведем вывод, основанный на принципе детального равновесия. Оба доказательства закона распределения скоростей Максвелла, приведенные нами в § 72 и 74, можно без всяких изменений распространить на случай наличия потенциального силового поля.

Поэтому можно считать доказанным, что в состоянии термодинамического равновесия скорости молекул газа в каждой точке пространства распределены по закону Максвелла с температурой T, общей для всего газа. Влияние силового поля сказывается только на изменении концентрации молекул газа от точки к точке. Это значит, что средняя концентрация dn молекул газа, скорости которых лежат в пределах элемента $d\omega$ скоростного пространства, определяется выражением вида

$$dn = nf(v)\,d\omega, \tag{77.5}$$

где функция $f(v)$ определяется законом Максвелла (72.14), а концентрация молекул n зависит только от координат: $n = n(\mathbf{r})$. Наша задача состоит в том, чтобы доказать, что эта зависимость определяется формулой (77.3).

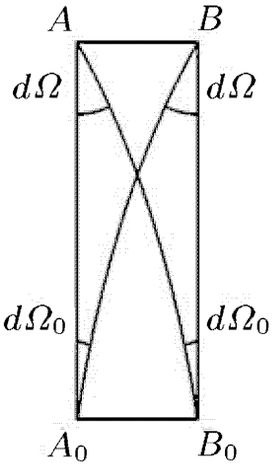


Рис. 65

4. Предположим сначала, что силовое поле во всем пространстве имеет одно и то же направление, хотя его напряженность и может меняться в этом направлении. Рассмотрим две одинаковые бесконечно малые площадки A_0B_0 и AB, перпендикулярные к направлению поля. Пусть прямые A_0A, B_0B и т. д., соединяющие соответствующие точки площадок, параллельны полю. Отвлечемся сначала от столкновений между молекулами. Через каждую точку площадки A_0B_0 проведем траектории молекул, проходящие через контур, ограничивающий вторую площадку AB. (На рис. 65 проведены две такие траектории через точку A_0 и две — через точку B_0.) Направления этих траекторий на площадке A_0B_0 образуют телесный угол, величину которого мы обозначим через $d\Omega_0$. Выделим группу молекул, скорости которых по величине заключены в интервале $(v_0, v_0 + dv_0)$, а направления (на площадке A_0B_0) лежат в пределах телесного угла $d\Omega_0$. Скоростные точки таких молекул заполняют в скоростном пространстве объем $MNN'M'$, величина которого равна $d\omega_0 = v_0^2\,d\Omega_0\,dv_0$ (рис. 66), а их концентрация в обычном пространстве будет $dn_0 = n_0 f(v_0)\,d\Omega_0 v_0^2\,dv_0$. Ясно, что если такие молекулы пересекут площадку A_0B_0, то они пройдут и через площадку AB. Число молекул рассматриваемой группы, проходящих ежесекундно от площадки A_0B_0 через площадку AB, будет равно $dN_0 = dS_0 v_0\,dn_0$, где dS_0 — площадь площадки A_0B_0. Число молекул, летящих по тем же траекториям в обратном направлении от площадки AB и проходящих через площадку A_0B_0, будет $dN = dSvdn$, где dS — площадь площадки AB, равна по

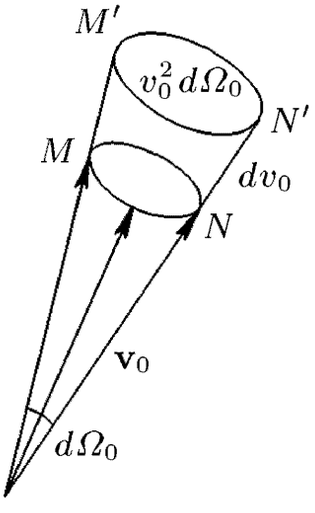


Рис. 66

условию dS_0, а $dn = nf(v)\,d\Omega v^2\,dv$. Если состояние установилось, то принцип детального равновесия требует $dN = dN_0$, т. е.

$$v\,dn = v_0\,dn_0. \tag{77.6}$$

При движении вверх или вниз меняется величина и направление скорости молекул. Вследствие этого меняется и величина телесного угла, в пределах которого направлены касательные к траекториям молекул рассматриваемой группы. Однако, поскольку действующая сила параллельна направлению A_0A, величина скорости молекулы, перпендикулярная к тому же направлению, остается неизменной. Меняется только продольная составляющая скорости, т. е. скорость, параллельная A_0A. Отсюда непосредственно следует, что изменение телесного угла в пучке молекул происходит по закону

$$d\Omega = \frac{\text{const}}{v^2}, \tag{77.7}$$

а потому

$$v^2\,d\Omega = v_0^2\,d\Omega_0. \tag{77.8}$$

Далее по закону сохранения энергии

$$\frac{mv^2}{2} + \varepsilon_p = \frac{mv_0^2}{2}, \tag{77.9}$$

если за нуль принять потенциальную энергию молекул на уровне площадки A_0B_0. Варьируя скорость v при неизменных положениях площадок A_0B_0 и AB (т. е. при постоянстве ε_p), получим

$$v\,dv = v_0\,dv_0. \tag{77.10}$$

Подставим теперь в (77.6) выражения для dn и dn_0. Тогда с учетом (77.8) и (77.10) найдем

$$nf(v) = n_0 f(v_0). \tag{77.11}$$

Если в качестве аргумента ввести кинетическую энергию молекулы, то это соотношение перепишется так:

$$nf(\varepsilon_k) = n_0 f(\varepsilon), \tag{77.12}$$

где $\varepsilon = (1/2)mv_0^2$ — полная энергия молекулы: $\varepsilon = \varepsilon_k + \varepsilon_p$. Но согласно закону Максвелла (77.12)

$$f(\varepsilon_k) \sim \exp\Big(-\frac{\varepsilon_k}{kT}\Big), \quad f(\varepsilon) \sim \exp\Big(-\frac{\varepsilon}{kT}\Big),$$

а потому

$$n = n_0 \exp\Big(-\frac{\varepsilon_p}{kT}\Big). \tag{77.3а}$$

Но это и есть закон распределения Больцмана. Если выражение (77.3) ввести в (77.5), то получится

$$dn = n_0 \exp\Big(-\frac{\varepsilon_{\text{полн}}}{kT}\Big)\,d\omega. \tag{77.13}$$

Это выражение дает среднюю концентрацию молекул dn в рассматриваемом месте пространства, скоростные точки которых лежат в элементе объема $d\omega$ скоростного пространства. Мы видим, что dn определяется только полной энергией молекулы $\varepsilon_{\text{полн}} \equiv \varepsilon = \varepsilon_k + \varepsilon_p$. Чтобы подчеркнуть это обстоятельство, мы снабдили символ ε индексом «полн», который должен напоминать, что речь идет о полной энергии молекулы. Формула (77.13) называется *распределением Максвелла–Больцмана.*

5. Необходимо теперь снять ограничения, введенные при молекулярно-кинетическом выводе распределения Больцмана. Таких ограничений два. Во-первых, мы предполагали, что силовое поле всюду имеет одинаковое направление, а именно — параллельное прямолинейной траектории A_0A (см. рис. 65). От этого ограничения легко освободиться, записав соотношение (77.3) в дифференциальной форме. Действительно, силовое поле в пределах бесконечно малой области газа может считаться однородным. К такому полю применимы наши рассуждения. Но для бесконечно малой области формула (77.3) принимает вид

$$d(\ln n) = -\frac{d\varepsilon_p}{kT}.$$

Она справедлива и для неоднородного поля. Поэтому интегрируя ее, мы снова приходим к формуле (77.3), но уже без всяких ограничений.

Во-вторых, надо принять во внимание столкновения. Из числа dN_0 молекул, прошедших сквозь площадку A_0B_0, не все дойдут до AB из-за столкновений с другими молекулами. Однако наряду с прямыми столкновениями, выводящими молекулы из рассматриваемого пучка, существуют обратные столкновения, за счет которых пучок пополняется другими молекулами. И если состояние статистически равновесно, то по принципу детального равновесия молекулы, выбывшие из пучка, в среднем будут восполнены молекулами, вступившими в него при обратных столкновениях. В результате среднее число молекул рассматриваемого типа, достигших площадки AB, не изменится. То же справедливо и для молекул, прошедших сквозь площадку AB и достигающих площадки A_0B_0. Столкновения, таким образом, приводят только к замене одних молекул другими, но не меняют средние значения dN_0 и dN молекул в пучках. Следовательно, при наличии столкновений доказательство сохраняет силу.

6. Заметим еще, что распределение Больцмана мы получили из соотношения (77.12), считая доказанным закон распределения скоростей Максвелла. Можно было бы поступить наоборот: предполагая доказанным закон распределения Больцмана, например, методами гидростатики, из формулы (77.12) получить закон распределения молекул по скоростям Максвелла. Мы видим, что законы распределения Максвелла и Больцмана тесно связаны и взаимно обусловливают друг друга: из одного с необходимостью следует другой. Оба распределения обусловлены столкновениями между молекулами. В частности, в

законе распределения Больцмана с особой отчетливостью проявляются две противоположные тенденции. Регулярно действующее силовое поле (сила тяжести) стремится сконцентрировать все молекулы на дне сосуда. Беспорядочные толчки, испытываемые молекулами при тепловом движении, препятствуют этому. Они снабжают молекулы газа кинетической энергией, при наличии которой молекулы способны преодолевать силу тяжести и подниматься вверх.

7. В заключение отметим еще один момент. Молекулы воздуха в земной атмосфере движутся вверх с уменьшающимися скоростями, а движущиеся вниз увеличивают свои скорости под действием силы тяжести. Отсюда делали неправильный вывод, что средние скорости молекул наверху, а с ними и температура воздуха должны быть меньше, чем внизу. Но этот парадоксальный вывод находится в противоречии с термодинамикой. Парадокс был разъяснен уже самим Максвеллом. Суть дела заключается в том, что при движении вверх молекулы действительно замедляются, но при этом наиболее медленные молекулы выбывают из пучка. При движении вниз, наоборот, молекулы не только ускоряются, но одновременно пучок пополняется более медленными молекулами. В результате средняя скорость теплового движения молекул остается неизменной. Сила тяжести, как уже отмечалось в п. 3, меняет лишь концентрацию молекул на разных высотах, но не температуру газа. И закон изменения этой концентрации как раз и выводится из требования, чтобы температура оставалась одной и той же на всех высотах. Следующее сравнение, принадлежащее Г. А. Лоренцу (1853–1928), может служить для разъяснения вопроса. Пусть имеются два города A и B, причем числу жителей определенного возраста города A соответствует удвоенное число жителей того же возраста города B. Ясно, что средний возраст жителей в обоих городах будет один и тот же.

ЗАДАЧИ

1. Теплоизолированный сосуд с идеальным газом подвешен на нити в поле тяжести. Из-за действия силы тяжести плотность газа внизу сосуда больше, чем наверху. Нить пережигают, и сосуд свободно падает. Предполагая, что во время падения успевает установиться термодинамическое равновесие, определить равновесную температуру газа, которая в нем установится при падении.

Р е ш е н и е. Температура газа не изменится. При свободном падении газ находится в состоянии невесомости. Начальное состояние его неравновесное — плотность вверху меньше, чем внизу. Однако средняя кинетическая энергия молекул всюду одинакова. При переходе в равновесное состояние плотности выравняются. Но полная кинетическая энергия молекул газа, определяющая его температуру, останется неизменной. Опыт аналогичен известному опыту Гей-Люссака с расширением газа в пустоту (см. § 19).

2. Найти среднюю потенциальную энергию $\langle \varepsilon_p \rangle$ молекулы газа в земной атмосфере, считая последнюю изотермической (с температурой T), а поле тяжести однородным. Вычислить теплоемкость газа c при этих условиях.

О т в е т. $\langle \varepsilon_p \rangle = kT$, $c = c_p$.

3. Теплоизолированный герметический цилиндрический сосуд высоты H, наполненный газом, подвешен в вертикальном положении в однородном поле тяжести. Температура газа в сосуде везде одинакова и равна T. Найти среднюю потенциальную энергию молекулы газа $\langle \varepsilon_p \rangle$.

Ответ. $\langle \varepsilon_p \rangle = \dfrac{1 - \left(1 + \dfrac{mgH}{kT}\right) \exp\left(-\dfrac{mgH}{kT}\right)}{1 - \exp\left(-\dfrac{mgH}{kT}\right)} kT.$

4. В цилиндре предыдущей задачи помещен моль идеального газа с молекулярной массой μ. Найти теплоемкость этого газа, учитывая влияние поля тяжести и предполагая, что $\mu g H \ll RT$.

Ответ. $C = C_V + \dfrac{R}{12}\left(\dfrac{\mu g H}{RT}\right)^2.$

5. Цилиндр радиуса R и длины H, наполненный химически однородным газом, равномерно вращается в однородном поле тяжести вокруг своей геометрической оси с угловой скоростью ω. Найти распределение концентрации молекул газа внутри цилиндра, если его ось направлена вертикально.

Ответ. Число молекул dN с координатами между r и $r + dr$, z и $z + dz$ равно

$$dN = \frac{Ng\left(\dfrac{N\omega}{kT}\right)^2}{\left[1 - \exp\left(-\dfrac{mgH}{kT}\right)\right]\left[\exp\left(\dfrac{m\omega^2 R^2}{2kT}\right) - 1\right]} \times$$

$$\times \exp\left(\frac{m\omega^2 r^2}{2kT}\right) r\, dr \cdot \exp\left(-\frac{mgz}{kT}\right) dz,$$

где N — общее число молекул в сосуде. Ось Z направлена вертикально вверх.

§ 78. Работы Перрена по определению постоянной Авогадро

1. В § 64 уже говорилось о работах Перрена по определению постоянной Авогадро N_A и проверке основных положений молекулярно-кинетической теории вещества путем наблюдения поступательного броуновского движения. Перрену же принадлежит другой метод измерения числа N, по существу также использующий броуновское движение. Этот второй метод основан на законе распределения Больцмана (77.3). В поле тяжести этот закон принимает вид

$$n = n_0 \exp\left(-\frac{mgz}{kT}\right). \tag{78.1}$$

Если бы была известна масса молекулы m, то, измеряя распределение плотности газа по высоте, можно было бы по формуле (78.1) вычислить постоянную Больцмана k, а затем и постоянную Авогадро $N_A = R/k$. Но измерить массу молекулы не менее трудно, чем постоянную k. Эту трудность можно преодолеть, если заметить, что роль молекул могут играть достаточно малые, но макроскопические частицы. В поле тяжести такие гигантские «макромолекулы» не могут все лежать на дне сосуда. Они должны совершать броуновское движение и вести себя подобно идеальному газу с очень большой

молекулярной массой, концентрация которого распределена по высоте согласно формуле Больцмана. «Макромолекулы» должны быть очень велики, чтобы экспериментально можно было измерить их массу. Но в таком случае их концентрация с высотой будет убывать настолько быстро, что получится состояние, практически не отличимое от того, в котором все «макромолекулы» лежат на дне сосуда. Перрен нашел способ преодоления и этой трудности. Надо поместить частицы-макромолекулы в жидкость, плотность которой немного меньше плотности вещества самих частиц. Тогда поле тяжести будет сильно ослаблено архимедовой подъемной силой и возникнет «атмосфера» из макромолекул, распределение концентрации в которой может быть измерено. В установившемся состоянии концентрация частиц определяется законом Больцмана (77.3). При вычислении потенциальной энергии надо учесть, что поднятие частицы сопровождается опусканием такого же объема жидкости. Величина ε_p в рассматриваемом случае равна потенциальной энергии частицы в поле тяжести mgz за вычетом потенциальной энергии жидкости $m_{\text{ж}}gz$, которую она вытесняет, т. е. $\varepsilon_p = (m - m_{\text{ж}})gz$, где m — масса частицы, а $m_{\text{ж}}$ — масса вытесненной ею жидкости. Таким образом,

$$n = n_0 \exp\left[-\frac{(m - m_{\text{ж}})gz}{kT}\right].$$

Если n_1 и n_2 — измеренные концентрации частиц на высотах z_1 и z_2, то по этой формуле получаем

$$k = \frac{(m - m_{\text{ж}})\,(z_2 - z_1)g}{T \ln(n_1/n_2)}. \tag{78.3}$$

2. Одна из трудностей состояла в получении взвешенных частиц совершенно одинакового размера и формы. Перрен пользовался частицами гуммигута и мастики. Растирая гуммигут в воде, Перрен получал эмульсию ярко-желтого цвета, в которой при наблюдении в микроскоп можно было различить множество зернышек сферической формы. Вместо механического растирания Перрен обрабатывал также гуммигут или мастику спиртом, растворяющим эти вещества. При разбавлении такого раствора большим количеством воды получалась эмульсия из таких же сферических зернышек, что и при механическом растирании гуммигута. Для отбора зернышек совершенно одинакового размера Перрен подвергал взвешенные в воде частицы многократному центрофугированию и таким путем получал весьма однородную эмульсию, состоящую из шарообразных частиц с радиусом порядка микрометра. Обработав 1 кг гуммигута, Перрен получил через несколько месяцев фракцию, содержавшую несколько дециграммов зерен желаемого размера. С этой фракцией и были выполнены описываемые здесь опыты, а также опыты по броуновскому движению, о которых говорилось в § 64.

3. При изучении эмульсии надо было производить измерения при ничтожных разностях высот — всего в несколько сотых миллиметра.

Поэтому распределение концентрации частиц по высоте исследовалось с помощью микроскопа. К предметному стеклу микроскопа (рис. 67) приклеивалось очень тонкое стекло с просверленным в нем широким отверстием. Таким путем получалась плоская ванночка (кювета Цейсса (1816–1886)), высота которой была около 100 мкм (0,1 мм). В центре ванночки помещалась капля эмульсии, которая тотчас сплющивалась покровным стеклом. Чтобы избежать испарения, края покровного стекла покрывались парафином или лаком. Тогда препарат можно наблюдать в течение нескольких дней или даже недель. Препарат помещался на столике микроскопа, тщательно установленного в горизонтальном положении. Возможны два способа наблюдения. В первом способе препарат помещается вертикально, а микроскоп — горизонтально; во втором — препарат горизонтально, а микроскоп — вертикально. Перрен пользовался обоими способами, но все количественные измерения были выполнены по второму способу. Объектив был очень сильного увеличения с малой глубиной фокуса, так что одновременно можно было видеть только частицы, находящиеся внутри очень тонкого горизонтального слоя с толщиной порядка микрометра. Частицы совершали интенсивное броуновское движение. Фокусируя микроскоп на определенный горизонтальный слой эмульсии, можно было сосчитать число частиц в этом слое. Затем микроскоп фокусировался на другой слой и снова считалось число видимых броуновских частиц. Таким путем можно было определить отношение концентраций n_1/n_2 на разных высотах. Разность высот измерялась микрометрическим винтом микроскопа.

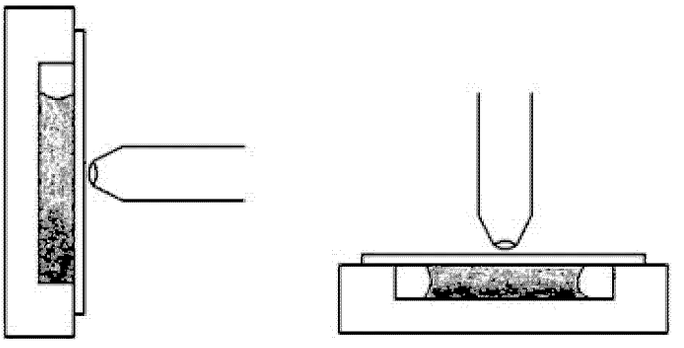


Рис. 67

4. Масса частицы m и масса вытесненной ею воды $m_{\text{ж}}$ вычислялись по размерам частицы и плотности гуммигута (мастики). Размеры частиц нельзя было определить прямым методом, рассматривая отдельную частицу в микроскоп, так как диаметры частиц были порядка или меньше длины световой волны. При этих условиях размеры изображения заметно увеличиваются из-за дифракции света. Однако влияние этого источника ошибок сильно уменьшится, если измерить длину ряда, составленного из известного и достаточно большого числа зернышек, вплотную примыкающих друг к другу. Перрен помещал на предметном столике микроскопа капельку сильно разведенной эмульсии, не покрытой покровным стеклом. Когда испарение почти кончилось, зернышки сбегались под действием сил поверхностного натяжения и собирались местами в довольно правильные ряды, как это видно из рис. 68, заимствованного из работы Перрена. Сосчитав число зернышек, вытянутых вдоль прямой линии известной длины,

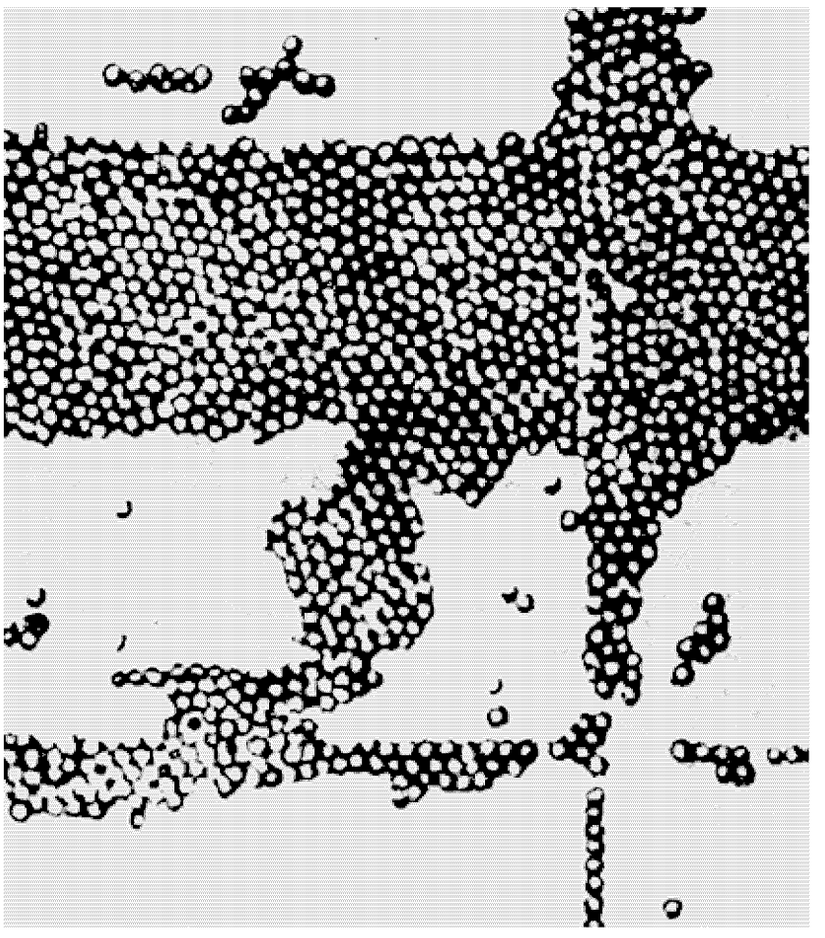


Рис. 68

или число зернышек, соприкасающихся друг с другом на известной площади, легко вычислить диаметр отдельного зернышка. Таким путем с лучшей из своих эмульсий Перрен получил для диаметра зернышка около 0,37 мкм. Кроме этого способа для измерения диаметра частиц, Перрен применил еще два способа, которые мы не описываем. Все они привели к согласованным результатам.

5. Таким образом, все величины, входящие в правую часть формулы (78.3), могут быть измерены экспериментально. После этого можно вычислить постоянную Больцмана k и постоянную Авогадро $N_{\text{А}}$. Результаты Перрена оказались в согласии с другими методами измерения тех же постоянных. Описанные опыты Перрена, как и примыкающие к ним опыты по броуновскому движению, просты в идейном отношении. Но их фактическое осуществление требует громадного труда и большого экспериментального искусства. Эти классические опыты были выполнены в 1908–1911 гг. и имели большое значение для утверждения идей атомистики.

§ 79. Распределение Больцмана и атмосферы планет

1. Применим закон распределения Больцмана к уединенной планете, окруженной газовой атмосферой. Последнюю будем считать изотермической. Кроме того, будем предполагать, что все молекулы одинаковы. Это

предположение не лишает наши рассуждения общности, поскольку каждый газ (если его рассматривать как идеальный), входящий в состав атмосферы, ведет себя независимо от остальных газов. Будем считать, что масса атмосферы пренебрежимо мала по сравнению с массой планеты. Тогда потенциальная энергия молекулы в поле тяготения планеты будет равна GMm/r. Для концентрации молекул n на расстоянии r от центра планеты закон Больцмана (77.3) дает

$$n = n_0 \exp \frac{GMm}{kTr}, \tag{79.1}$$

где M — масса планеты, а G — гравитационная постоянная. Если бы формула (79.1) была применима на всех расстояниях от планеты, то на бесконечности получилось бы конечное значение для концентрации n, а именно $n = n_0$. Но это невозможно, так как общее количество молекул в атмосфере планеты конечно, а объем пространства, окружающего ее, бесконечно велик. Равновесие возможно только при $n_0 = 0$, т. е. при полном отсутствии атмосферы.

2. Невозможность равновесного состояния планетной атмосферы связана с тем, что потенциальная энергия молекулы в поле тяготения планеты в бесконечности остается конечной. Приняв ее за нуль, можно сказать, что молекула при отсутствии столкновений совершала бы *инфинитное* движение, если бы ее полная энергия была положительна (см. т. I, § 57). Такие молекулы (а они всегда появляются в результате столкновений) не могут удерживаться полем тяготения планеты. Поэтому к планетной атмосфере в целом неприменима формула Больцмана (77.3), так как ее вывод предполагал, что газ находится в состоянии термодинамического равновесия. Пусть в некоторый момент скорости молекул в атмосфере распределены по закону Максвелла. Если бы с этого момента молекулы перестали сталкиваться между собой, испытывая лишь упругие столкновения с поверхностью планеты, то все молекулы, скорости которых превышают вторую космическую (см. т. I, § 61), навсегда покинули бы планету. Остались бы только молекулы со скоростями, меньшими второй космической. Они совершали бы финитное движение вокруг планеты, а их скорости были бы распределены по закону Максвелла. Для финитных систем возможно термодинамическое равновесие. Оно будет обязательно больцмановским, если скорости молекул распределены по закону Максвелла. Действительно, к этому случаю полностью применимо рассуждение, изложенное нами в § 77, пп. 4, 5. Для такого распределения в гравитационном поле с напряженностью $\sim 1/r^2$ требуется бесконечное множество молекул, и оно устанавливается бесконечно долго. Если, однако, из всех молекул, совершающих финитное движение, отобрать молекулы с полной энергией ε, удовлетворяющей неравенству $\varepsilon < \varepsilon_0 < 0$, то, каково бы ни было значение ε_0, получится больцмановское распределение с конечным числом частиц и конечным временем установления.

3. Для планеты достаточно большой массы доля молекул со скоростями, превышающими вторую космическую, ничтожна. Вторая космическая скорость в проблеме рассеяния атмосферы называется *скоростью убегания молекулы*, а молекулы со скоростями, превышающими эту скорость, — *убегающими молекулами*. Скорость убегания меняется с расстоянием молекулы от центра планеты. Ввиду ничтожной доли убегающих молекул распределение частиц в атмосфере является *квазиравновесным* и при постоянной температуре может быть описано следующим образом. Подавляющая доля молекул распределена в пространстве по закону Больцмана. На больцмановское распределение накладывается поток убегающих молекул. Вблизи планеты относительная концентрация убегающих молекул в таком потоке

ничтожна. По мере удаления от планеты эта относительная концентрация непрерывно растет. На бесконечности все молекулы являются убегающими. Поток убегающих молекул непрерывно пополняется в результате межмолекулярных столкновений. Это приводит к тому, что *планета в конце концов должна потерять атмосферу*. Почему же Земля, Венера и другие планеты Солнечной системы имеют атмосферы? Потому, что время τ, в течение которого планета теряет атмосферу, очень велико. Время τ, в течение которого масса атмосферы планеты убывает в e раз, называется *временем рассеяния атмосферы*. Его оценкой мы и займемся.

4. Точный расчет времени рассеяния реальной атмосферы потребовал бы знания параметров *верхних слоев атмосферы* и учета происходящих в них процессов. Действительно, рассеяние планетных атмосфер непосредственно определяется только условиями и процессами в верхних слоях атмосферы. Точный расчет, однако, в настоящее время вряд ли возможен, даже с использованием данных о верхней атмосфере Земли, полученных с помощью ракет и искусственных спутников. Оценка времени рассеяния идеализированной изотермической атмосферы, которую мы приводим, может дать результат, отличающийся на порядок и даже больше от действительного времени рассеяния. Однако такая оценка все же дает представление о порядке величины этого времени. Кроме того, эта оценка может служить интересным примером применения кинетической теории газов.

Опишем вокруг планеты сферу σ, концентрическую с поверхностью планеты. Радиус r_σ этой сферы возьмем настолько большим, чтобы столкновениями между молекулами вне сферы σ можно было полностью пренебречь, но этого нельзя делать в пространстве, ограниченном сферой σ. Предположим, что на сфере σ справедливо распределение Максвелла–Больцмана для всех молекул. В отношении молекул, совершающих финитное движение, справедливость этого предположения не вызывает сомнений. Но для убегающих молекул оно верно только приближенно. Введем две скорости убегания: на поверхности планеты и на сфере σ. Обозначим их соответственно через v_0 и v_σ. Если r_0 — радиус планеты, g_0 и $g_\sigma = g_0 r_0^2/r_\sigma^2$ — ускорения свободного падения на поверхности планеты и на сфере σ, то

$$v_0 = \sqrt{2g_0 r_0}, \tag{79.2}$$

$$v_\sigma = \sqrt{2g_\sigma r_\sigma} = r_0\sqrt{2g_0/r_\sigma}. \tag{79.3}$$

Для Земли $v_0 = 11{,}2$ км/с (см. т. I, § 61). Величины v_0 и v_σ связаны между собой уравнением энергии

$$\frac{1}{2}\,mv_0^2 = \frac{1}{2}\,mv_\sigma^2 + \varepsilon_p, \tag{79.4}$$

где ε_p — разность потенциальных энергий на сфере σ и на поверхности планеты.

Дальнейшие вычисления удобно производить, приняв за единицу наиболее вероятную скорость v_m, определяемую соотношением (73.5). Скорость $x = v/v_m$, измеренную в таких единицах, будем называть *безразмерной скоростью*. В частности, безразмерные скорости убегания на поверхности планеты и на сфере σ равны $x_0 = v_0/v_m$ и $x_\sigma = v_\sigma/v_m$. Ввиду (79.2) и (79.3) они связаны соотношением

$$x_\sigma^2 = \frac{r_0}{r_\sigma}\,x_0^2. \tag{79.5}$$

Соотношение (79.4), записанное в безразмерных величинах, будет

$$\frac{\varepsilon_p}{kT} = x_0^2 - x_\sigma^2. \tag{79.6}$$

С учетом соотношения (79.6) из закона распределения Больцмана получим

$$n_0 e^{-x_\sigma^2} = n_0 e^{-x_0^2}, \tag{79.7}$$

где n_σ — концентрация молекул на сфере σ. Наконец, если пользоваться безразмерными скоростями, то максвелловское распределение примет вид

$$dn = \frac{4}{\sqrt{\pi}} x^2 e^{-x^2}\, dx. \tag{79.8}$$

Концентрация убегающих молекул на сфере σ равна

$$\Delta n = \frac{4n_\sigma}{\sqrt{\pi}} J, \tag{79.9}$$

где J означает интеграл

$$J = \int\limits_{x_\sigma}^{\infty} x^2 e^{-x^2}\, dx. \tag{79.10}$$

Средняя безразмерная скорость таких молекул будет

$$c \equiv \langle x \rangle_{x>x_\sigma} = \frac{1}{J} \int\limits_{x_\sigma}^{\infty} x^3 e^{-x^2}\, dx.$$

Интегрируя это выражение по частям, получим

$$c = \frac{1}{2J} (x_\sigma^2 + 1) e^{-x_\sigma^2}. \tag{79.11}$$

Найдем средний поток убегающих частиц Z, исходящий наружу из сферы σ. Поскольку распределение скоростей молекул изотропно, можно воспользоваться формулой (75.5). Средняя скорость рассматриваемых частиц, выраженная в обычных единицах, равна cv_m, а потому эта формула дает $Z = (1/4) S c v_m \Delta n$, где $S = 4\pi r_\sigma^2$ — поверхность сферы σ. Подставив сюда выражения (79.9) и (79.11) и воспользовавшись формулами (79.5) и (79.7), получим

$$Z = 2\sqrt{\pi} \left(\frac{r_0}{r_\sigma}\, x_0^2 + 1 \right) n_0 r_\sigma^2 v_m e^{-x_0^2}. \tag{79.12}$$

Это выражение и дает число молекул, теряемых атмосферой в единицу времени. Его можно представить в виде

$$Z = -dN/dt, \tag{79.13}$$

где N — полное число молекул в атмосфере.

Концентрацию n_0 можно выразить через N. Подавляющая масса атмосферы приходится на тонкий слой, примыкающий к поверхности планеты. В пределах этого слоя можно пренебречь кривизной поверхности планеты, а также изменением ускорения свободного падания с высотой, т. е. положить

$g = g_0$. Тогда распределение Больцмана (77.3) переходит в барометрическую формулу, и мы получаем

$$N = 4\pi r_0^2 n_0 \int\limits_0^\infty \exp\Big(-\frac{mg_0 z}{kT}\Big)\, dz = 4\pi r_0^2 n_0 \frac{kT}{mg_0}.$$

Отсюда и найдется концентрация n_0. Подставляя ее в выражение (79.12) и воспользовавшись уравнением (79.13), придадим последнему вид

$$dN/dt = -N/\tau, \tag{79.14}$$

где введено обозначение

$$\tau = \frac{2\sqrt{\pi} r_0^2 kT}{mg_0 r_\sigma^2 v_m \Big(\dfrac{r_0}{r_\sigma}\, x_0^2 + 1\Big)}\, e^{x_0^2}. \tag{79.15}$$

Интегрирование уравнения (79.14) дает

$$N = N_0 e^{-t/\tau}. \tag{79.16}$$

Из этой формулы видно, что постоянная τ есть введенное выше время рассеяния атмосферы.

5. Формула (79.15) еще не решает задачу, так как она содержит радиус r_σ, который мы еще не определили. В одном предельном случае решение очевидно. Это случай, когда планетная атмосфера *бесконечно разреженная*. В ней полностью отсутствуют столкновения между молекулами, а распределение молекул в пространстве и по скоростям устанавливается в результате столкновений с поверхностью планеты. В рассматриваемом случае следует положить $r_\sigma = r_0$. Используя, кроме того, соотношения $v_m^2 = 2kT/m$ и $x_0 v_m = v_0 = \sqrt{2r_0 g_0}$, получим

$$r = \sqrt{\frac{2\pi r_0}{g_0}}\, \frac{e^{x_0^2}}{x_0(x_0^2+1)}, \tag{79.17}$$

или

$$\tau = \sqrt{\frac{3}{2G_\rho}}\, \frac{e^{x_0^2}}{x_0(x_0^2+1)}, \tag{79.18}$$

где ρ — средняя плотность планеты. Ясно, что величина τ, определяемая этими формулами, имеет смысл *времени рассеяния бесконечно разреженной атмосферы*.

6. Остается определить радиус r_σ, когда атмосфера не является бесконечно разреженной. Для этого надо указать значение какой-то длины l, чтобы при $r > l$ молекулу можно было считать не принадлежащей к атмосфере планеты. Тогда радиус r_σ определится из условия $\lambda = l$, где λ — средняя длина свободного пробега молекулы при $r = r_\sigma$. В частности, атмосфера может считаться бесконечно разреженной, когда условие $\lambda \geqslant l$ соблюдается уже при $r = r_0$.

Для уединенной планеты нельзя заранее (т. е. без точного решения задачи) указать никакой длины, которую можно было бы принять за l. Единственным параметром размерности длины в этом случае является радиус планеты r_0, но он не имеет отношения к рассеянию атмосферы. Однако уединенная планета есть абстракция. Реальная планета вращается вокруг Солнца. А так как величина r_σ, входящая в формулу (79.15), как будет видно из дальнейшего, очень нечувствительна к выбору длины l, то в качестве l

можно взять радиус планетной орбиты. Для обоснования этого заметим, что в системе отсчета, связанной с планетой, на молекулу действуют сила гравитационного притяжения планеты $\mathbf{F}_{\text{пл}}$, сила гравитационного притяжения Солнца $\mathbf{F}_{\text{солн}}$ и сила инерции $\mathbf{F}_{\text{ин}}$, связанная с ускоренным движением центра планеты к Солнцу (от осевого вращения планеты можно отвлечься — оно не играет принципиальной роли). Как было подробно разъяснено в первом томе (см. § 65, 69), сила $\mathbf{F}_{\text{солн}}$ полностью компенсируется силой инерции $\mathbf{F}_{\text{ин}}$, если пренебречь неоднородностью поля тяготения Солнца. С учетом указанной неоднородности компенсация не будет полной, и с этим связано явление приливов в планетной атмосфере. На расстояниях порядка радиуса планетной орбиты ни о какой компенсации речи быть не может. На таких расстояниях сила $\mathbf{F}_{\text{пл}}$ становится меньше результирующей сил $\mathbf{F}_{\text{солн}}$ и $\mathbf{F}_{\text{ин}}$. Если молекула удалилась так далеко, то можно считать, что к планете она уже не вернется.

7. Итак, за l мы примем величину порядка радиуса планетной орбиты. Конкретно, для Земли возьмем $l = 10^8$ км. По формуле (79.1) найдем для различных газов значения радиуса r_σ, при котором $\lambda = 10^8$ км. При этом мы предполагаем, что на земной поверхности атмосферное давление нормальное (средняя длина свободного пробега $\lambda \sim 10^{-5}$ см для всех газов). Полагая $r_0 = 6375$ км, $T = 300$ К, получим табл. 7.

Таблица показывает, что если бы атмосфера состояла из одних только тяжелых газов (тяжелее водорода), то сфера σ практически совпадала бы с

Таблица 7

Газ	Молекулярная масса	r_σ, км	Высота над земной поверхностью $h = r_\sigma - r_0$, км
N_2	28	$6{,}78 \cdot 10^3$	400
He	4	$1{,}09 \cdot 10^4$	4 500
H_2	2	$3{,}72 \cdot 10^4$	28 800

поверхностью Земли. В этом случае можно было бы пользоваться формулой (79.18) для бесконечно разреженной атмосферы. Для водорода радиус r_σ значительно превышает радиус Земли r_0. В случае сложной атмосферы, состоящей из смеси различных газов, величина r_σ определяется наиболее легким компонентом ее (с учетом, конечно, содержания этого компонента в атмосфере). Так, если для земной атмосферы принять, что содержание водорода составляет 10^{-6} от полного числа молекул атмосферы (что примерно вдвое больше действительного числа), то мы получили бы $r_\sigma = 1{,}4 \cdot 10^4$ км. Но даже во всех этих случаях можно пользоваться формулой для бесконечно разряженной атмосферы. Она дает вполне приемлемую оценку. Действительно, во всех случаях, представляющих интерес, значение x_0^2 очень велико, и в формуле (79.15) можно пренебречь единицей по сравнению с $(r_0/r_\sigma)x_0^2$. В этом приближении

$$\frac{\tau_0}{\tau_\sigma} = \frac{r_\sigma}{r_0}, \tag{79.19}$$

где τ_0 и τ_σ означают времена рассеяния, вычисленные по формуле для бесконечно разреженной атмосферы и по формуле (79.15) соответственно. Мы видим, что формула для бесконечно разреженной атмосферы (79.19) завышает время рассеяния всего в несколько раз. Поэтому дальнейшие численные расчеты будут произведены с помощью формулы (79.18).

8. Однако даже с помощью формулы (79.18) не так легко оценить время рассеяния атмосферы. Формула очень чувствительна к температуре атмосферы T, влияние которой сказывается преимущественно через экспоненциальный множитель $e^{x_0^2}$. А так как на разных высотах температура атмосферы разная и, кроме того, подвержена частым и нерегулярным изменениям, то невозможно с достаточной точностью определить значение T, которое следует подставлять в формулу. Поэтому мы воспользуемся формулой (79.18) для решения обратной задачи — по заданному времени τ найдем x_0^2, а затем и температуру планетной атмосферы, при которой она рассеивается в окружающее пространство за время τ. Возраст Земли $\sim 4 \cdot 10^9$ лет $\approx 4 \cdot 10^9 \cdot 3{,}156 \cdot 10^7$ с. Проведем расчет для $\tau = 10^{10}$ лет и $\tau = 10^8$ лет. Для Земли $\rho = 5{,}517$ г/см^3. Подставляя это значение в формулу (79.18) и логарифмируя, приходим к уравнению

$$0{,}4343 x_0^2 = 14{,}19 + \lg x_0 + \lg(x_0^2 + 1), \tag{79.20}$$

в котором 0,4343 есть модуль перехода от натуральных логарифмов к десятичным. Корень уравнения (79.20) легко находится методом последовательных приближений.

Нулевое приближение. В правой части уравнения отбрасываем оба слагаемые, содержащие логарифмы. Получаем $x_0^2 = 32{,}68$.

Первое приближение. В правую часть подставляем значение x_0 из нулевого приближения. Это дает $x_0^2 = 37{,}92$. Таким же путем поступаем в последующих приближениях и получаем:

второе приближение $x_0^2 = 38{,}15$,

третье приближение $x_0^2 = 38{,}15$.

Таким образом, уже второе приближение обеспечивает точность до четырех значащих цифр.

Итак, для $\tau = 10^{10}$ лет $x_0^2 = 38{,}15$. Аналогичные вычисления для $\tau = 10^8$ лет дают $x_0^2 = 33{,}34$. Температура T может быть вычислена по формуле

$$x_0^2 = \frac{2R_0 g_0}{v_m^2} = \frac{R_0 g_0}{kT},$$

или

$$T = \frac{R_0 g_0 m}{k x_0^2}. \tag{79.21}$$

9. Результаты вычислений приведены в табл. 8. Из нее видно, что время τ очень чувствительно к изменениям температуры T. При изменении T на 12–15 % τ меняется на два порядка. Отсюда следует, что рассеяние атмосферы должно сильно возрастать из-за нерегулярных местных колебаний температуры. Рассеяние сильно возрастает также из-за диссоциации двухатомных и многоатомных молекул под действием солнечного излучения. Из таблицы видно, что поле тяготения Земли надежно удерживает в течение геологических эпох все газы земной атмосферы, за исключением водорода и гелия. Формула (79.21) объясняет, почему Луна практически лишена атмосферы, а мощное гравитационное поле Юпитера не позволяет в течение геологических эпох рассеяться сколько-нибудь заметно даже наиболее легкому газу — атомарному и молекулярному водороду. Понятно также, почему Луна лишена атмосферы, а на Титане — шестом спутнике Сатурна — обнаружена атмосфера из мената (CH_4), аммиака (NH_3) и других газов, хотя скорости убегания на обоих спутниках почти одинаковы (2,4 км/с на Луне и 2,6 км/с на Титане). Дело в том, что температура поверхности Титана

(примерно 70–120 K) много ниже температуры лунной поверхности. При такой низкой температуре только наиболее легкие газы — водород и гелий —

Таблица 8

Планеты		x_0^2	Температура T, K				
			H_2	He	H_2O	N_2	O_2
Земля	$\tau = 10^{10}$ лет	38,15	396	792	3 560	5 540	6 340
	$\tau = 10^8$ лет	33,34	454	908	4 090	6 360	7 260
Луна	$\tau = 10^{10}$ лет	37,99	18	36	162	252	288
	$\tau = 10^8$ лет	33,11	20,6	41,2	185	288	330
Марс	$\tau = 10^{10}$ лет	37,97	81	162	729	1 130	1 300
	$\tau = 10^8$ лет	33,17	93	186	837	1 300	1 490
Венера	$\tau = 10^{10}$ лет	38,11	335	670	3 010	4 690	5 360
	$\tau = 10^8$ лет	33,31	384	768	3 460	5 380	6 140
Юпитер	$\tau = 10^{10}$ лет	37,42	12 000	24 000	108 000	168 000	192 000
	$\tau = 10^8$ лет	32,61	13 800	27 600	124 000	193 000	221 000

обладают тепловыми скоростями, достаточными для быстрого улетучивания их в окружающее пространство.

Из планет Солнечной системы наименее благоприятны условия для удержания атмосферы на Меркурии. Скорость убегания с поверхности планеты составляет всего 3,8 км/с. Неблагоприятна также крайне высокая температура на освещенной поверхности планеты. Поэтому Меркурий могут покидать даже молекулы тяжелых газов. Наконец, может иметь значение давление электромагнитного и корпускулярного излучения Солнца, которое на Меркурии довольно значительно и способно заметно «выдувать» молекулы газов из атмосферы Меркурия, если бы таковая существовала.

§ 80. Энтропия и вероятность

1. Согласно феноменологической термодинамике все процессы в замкнутой системе происходят в направлении возрастания энтропии. В конце концов система переходит в равновесное состояние, в котором энтропия достигает максимума, и все процессы в системе прекращаются. Этот вывод, если его понимать буквально, находится в противоречии с основными представлениями молекулярно-кинетической теории. Рассмотрим, например, закрытый сосуд, разделенный перегородкой AB на две одинаковые части I и II (рис. 69). Пусть сначала в части сосуда I находится N молекул идеального газа, а в части II — ни одной. В момент времени $t = 0$ мгновенно удалим перегородку AB. Газ начнет расширяться. Молекулы из части I будут переходить в часть II. Спустя некоторое время возникнет и обратный поток молекул из части II в часть I, после чего начнется и будет продолжаться обмен молекулами между обеими частями. Когда числа молекул N_1 и N_2 в обеих частях

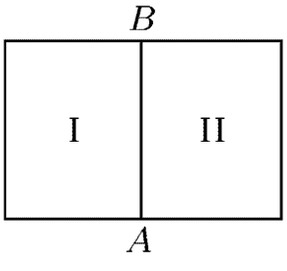


Рис. 69

сосуда, а также их потоки туда и обратно выравняются, наступит состояние равновесия. Но это будет динамическое, а не статическое равновесие. В состоянии динамического равновесия равенство $N_1 = N_2 = N/2$ почти никогда не соблюдается. Равенство относится не к мгновенным значениям N_1 и N_2, а к их *средним значениям* за длительный промежуток времени: $\overline{N}_1 = \overline{N}_2 = N/2$. *Самопроизвольные отклонения чисел N_1 и N_2, а также любых других физических величин от их средних значений, обусловленные тепловым движением, называются флуктуациями.* Броуновское движение и соответствующее ему статистически равновесное распределение броуновских частиц по высоте, описываемое барометрической формулой (78.2), поскольку они связаны с нарушениями статистического равновесия, также относятся к классу флуктуационных явлений. В нашем примере принципиально возможна и такая флуктуация, когда все молекулы газа, первоначально распределенные по всему сосуду, самопроизвольно собираются в одной из равных частей I или II. Чтобы убедиться в этом, предположим, что молекулы газа являются материальными точками, а стенки сосуда абсолютно гладкие. Если в некоторый момент времени t изменить на противоположные скорости всех молекул, то молекулы начнут двигаться в противоположных направлениях, проходя в точности через те же положения, через которые они проходили ранее. Отсюда следует, что если в момент времени 0 все молекулы находились в части сосуда I, то они снова соберутся в той же части в момент $2t$. Почему же такие процессы никогда не наблюдаются? Ответ, даваемый молекулярно-кинетической теорией, состоит в том, что они хотя и принципиально возможны, но при колоссальности числа молекул N крайне маловероятны.

2. Рассчитаем вероятность таких процессов. Пусть в сосуде находится всего одна молекула. Тогда, если нет внешних силовых полей, молекула с равной вероятностью может попасть либо в часть I, либо в часть II. Вероятность попадания ее в эти одинаковые части $P_{\mathrm{I}} = P_{\mathrm{II}} = 1/2$. Введем в сосуд вторую молекулу. Так как молекулы идеального газа не взаимодействуют между собой, то их попадания в ту или иную часть сосуда будут независимыми событиями. Вероятность того, что обе они окажутся в части I, найдется по теореме умножения вероятностей и будет равна $P_{\mathrm{I}} = (1/2) \cdot (1/2) = 1/4$. Если в сосуде N молекул, то, рассуждая аналогично, найдем, что вероятность их попадания в часть I будет $P_{\mathrm{I}} = (1/2)^N$. При $N = 10$ получаем $P_{\mathrm{I}} = (1/2)^{10} = 1/1024 \approx 0{,}001$. Если в течение длительного (в пределе бесконечно длительного) времени фотографировать расположение молекул в сосуде через равные промежутки времени, то на каждый 1000 кадров в среднем придется приблизительно один кадр, на котором будут зафиксированы все 10 молекул только в части сосуда I. То же можно сказать и о части II. По теореме сложения вероятностей получится в среднем 2 кадра на каждую тысячу с молекулами, сосредоточенными либо в части I, либо в части II (безразлично какой). Все это не только

принципиально возможно, но и фактически доступно наблюдению. Однако при $N = 100$ мы получаем $P_{\text{I}} = (1/2)^{100} \approx 10^{-30}$, и практически нет никаких шансов наблюдать соответствующую флуктуацию. При $N = 10^{20}$ для вероятности P_{I} получается слишком малая величина $P_{\text{I}} = (1/2)^{10^{20}} \approx 10^{-3\cdot 10^{19}}$. С такого рода вероятностями и соответствующими им событиями можно совершенно не считаться.

Обобщим наш расчет, так как это обобщение понадобится нам уже в этом параграфе. Пусть V_0 — объем всего сосуда, а V — объем какой-либо его части. Вероятность того, что какая-либо молекула попадает в объем V, равна V/V_0, а вероятность того, что в объеме V окажутся все N молекул идеального газа, представится выражением

$$P = (V/V_0)^N. \tag{80.1}$$

3. Относительно большие флуктуации встречаются только в системах с малым числом частиц (см. § 81, п. 2). Если число частиц в замкнутой системе очень велико, то подавляющее время она проводит в состоянии, в котором все величины лишь незначительно отличаются от их средних значений. В системах с очень большим числом частиц относительно большие флуктуации фактически не встречаются. Все флуктуации малы. Феноменологическая термодинамика не принимает их во внимание. Таким образом, можно сказать, что выводы термодинамики верны, если пренебречь флуктуациями. Вблизи состояния равновесия флуктуации в ту и в другую сторону равновероятны. Но если создать искусственно неравновесное состояние, то в подавляющем большинстве случаев система самопроизвольно будет переходить в состояние с большей вероятностью. С другой стороны, согласно феноменологической термодинамики, все самопроизвольные процессы в замкнутых системах сопровождаются возрастанием энтропии. Поэтому можно ожидать, что между энтропией системы S в каждом состоянии и вероятностью P того же состояния должна существовать *однозначная связь*. Такая гипотеза, введенная Больцманом, оправдалась и оказалась весьма плодотворной. Наша задача состоит теперь в том, чтобы установить эту связь.

4. На первый взгляд кажется, что такая задача неразрешима и даже не имеет смысла, пока не установлено в общем виде, как определять вероятность произвольного состояния любой термодинамической системы. В действительности для решения задачи достаточно знать самые общие свойства, которыми должна обладать вероятность P при любом способе ее определения. Надо только усилить гипотезу о связи между энтропией и вероятностью требованием, чтобы эта связь была *универсальной*, т. е. выражалась формулой $S = f(P)$, где функция $f(P)$ — *одна и та же* для всех тел, в каких бы состояниях они ни находились.

Для нахождения вида функции $f(P)$ рассмотрим две независимые подсистемы в состояниях с вероятностями P_1 и P_2. Их энтропии в этих состояниях будут $S_1 = f(P_1)$ и $S_2 = f(P_2)$. Объединим обе подсистемы в одну систему и обозначим вероятность ее состояния через P_{12}, а

энтропию — через S_{12}. Поскольку подсистемы независимы, $P_{12} = P_1 P_2$, а потому $S_{12} = f(P_{12}) = f(P_1 P_2)$. С другой стороны, согласно термодинамике энтропия сложной системы должна быть равна сумме энтропий составляющих ее независимых подсистем. Следовательно, должно быть

$$f(P_1 P_2) = f(P_1) + f(P_2), \tag{80.2}$$

каковы бы ни были значения вероятностей P_1 и P_2.

Для решения функционального уравнения (80.2) поступаем так же, как мы поступали при решении уравнения (72.2). Предположим, что переменные P_1 и P_2 изменяются так, что их произведение $P_1 P_2$ остается постоянным, причем значение самой постоянной в допустимой области ее изменения может быть каким угодно. Из уравнения (80.2) следует

$$f(P_1) + f(P_2) = \text{const},$$

если

$$P_1 P_2 = \text{const}.$$

Отсюда дифференцированием находим

$$df(P_1) = -df(P_2)$$

при условии, что

$$\frac{dP_1}{P_1} = -\frac{dP_2}{P_2}.$$

Почленное деление приводит к соотношению

$$P_1 \frac{df(P_1)}{dP_1} = P_2 \frac{df(P_2)}{dP_2}.$$

Слева стоит функция только от аргумента P_1, а справа — та же функция, но от аргумента P_2. Значения самих аргументов P_1 и P_2 в допустимых областях изменения их могут быть какими угодно. Это значит, что функция $Pdf(P)/dP$ не меняется при изменении аргумента P, т. е. является постоянной. Постоянная должна быть универсальной, т. е. одной и той же для всех тел, поскольку сама функция $f(P)$ универсальна. Обозначая эту постоянную буквой k, придем к соотношению

$$P \frac{df}{dP} = k, \quad \text{или} \quad df = k \frac{dP}{P}.$$

Отсюда

$$f(P) = k \ln P + C.$$

Постоянная интегрирования C должна равняться нулю. Действительно, подстановка найденного решения в исходное уравнение (80.2) приводит к соотношению

$$k \ln(P_1 P_2) + C = (k \ln P_1 + C) + (k \ln P_2 + C),$$

откуда $C = 0$. Итак,

$$S = k \ln P. \tag{80.3}$$

5. Остается определить числовое значение постоянной k. Для этого достаточно найти независимыми способами, а затем сравнить

две величины: разность энтропий какой-либо системы в двух произвольных состояниях и логарифм отношения ее вероятностей в тех же состояниях. Проще всего воспользоваться идеальным газом. Пусть V_1 и V_2 — объемы моля газа в начальном и конечном состояниях, температуры которых одни и те же. Отношение вероятностей найдется по формуле (80.1), если в ней положить сначала $V = V_1$, а затем $V = V_2$. Таким путем находим

$$S_2 - S_1 = k \ln \frac{P_2}{P_1} = k N_A \ln \frac{V_2}{V_1}.$$

Для той же величины термодинамическая формула (40.7) дает

$$S_2 - S_1 = R \ln \frac{V_2}{V_1}.$$

Сравнением обоих выражений получаем

$$k = R/N_A, \tag{80.4}$$

т. е. k есть постоянная Больцмана. Фундаментальное соотношение (80.3) между энтропией и вероятностью было установлено Больцманом и называется *формулой Больцмана*. Приведенный вывод формулы (80.3) принадлежит Планку (1858–1947). Планком же была введена и фундаментальная постоянная k.

6. При термодинамическом определении энтропии мы встречались с трудностью распространения этого понятия на случай *термодинамически неравновесных состояний* (см. § 42). Формула Больцмана (80.3) дает принципиальный способ преодоления указанной трудности. Надо смотреть на нее как на *определение энтропии*. Правда, для того чтобы это определение получило конкретное содержание, надо дополнить его способами вычисления вероятностей состояний во всех требуемых случаях. Но и без этого видно, что *при таком понимании энтропии закон ее возрастания коренным образом меняет свой характер. Он утрачивает свою абсолютность и превращается в статистический закон. Энтропия замкнутой системы может не только возрастать, но и убывать. И она действительно будет убывать, если только подождать достаточно долго.* Однако процесс убывания снова сменится в дальнейшем процессом возрастания. Что же остается в таком случае от второго начала термодинамики? В чем состоит его физическое содержание? А в том, что *за каким-либо заданным состоянием системы будут следовать состояния еще более вероятные, если и не с необходимостью, то в подавляющем большинстве случаев*. Если система большая, а исходное состояние ее не очень близко к состоянию равновесия, то переходы системы в менее вероятные состояния будут настолько маловероятны, что на практике они совершенно не имеют никакого значения. Тогда закон возрастания энтропии оправдывается практически с абсолютной достоверностью.

В § 42 мы говорили о концепции тепловой смерти Вселенной, выдвинутой Клаузиусом. Здесь следует заметить, что в противоположность

этой концепции Больцманом была высказана так называемая *флуктуационная гипотеза*. Больцман не отрицал применимость второго закона термодинамики к Вселенной в целом. Однако второе начало термодинамики есть статистический закон, согласно которому отступления от термодинамического равновесия — флуктуации — *не только возможны, но и неизбежны*. Больцман считал, что неравновесное состояние Вселенной, в котором она находится сейчас, есть гигантская флуктуация. Эта флуктуация должна исчезнуть. Тогда наступит состояние тепловой смерти Вселенной. Однако это состояние временное. Спустя некоторое время снова возникнет аналогичная гигантская флуктуация, и Вселенная выйдет из состояния тепловой смерти. Затем опять наступит тепловая смерть, и так без конца. Если согласно концепции Клаузиуса тепловая смерть есть окончательное состояние Вселенной, из которого она никогда не может выйти, то по гипотезе Больцмана Вселенная периодически приходит в состояние тепловой смерти и самопроизвольно выходит из него. Однако времена между двумя последовательными гигантскими флуктуациями невообразимо велики по сравнению с временем существования каждой из них. Поэтому можно сказать, что по флуктуационной гипотезе Вселенная должна находиться в состоянии тепловой смерти «почти всегда». Мы видим, что флуктуационная гипотеза, радикально отличаясь от концепции Клаузиуса в принципиальном отношении, приводит практически почти к тому же окончательному результату. Нельзя экстраполировать на Вселенную в целом второе начало термодинамики, если даже смотреть на него как на статистический закон.

7. В термодинамике энтропия определена только с точностью до произвольной аддитивной постоянной. Физический смысл имеют не сами энтропии, а их разности. Между тем формула Больцмана (80.3) выражает энтропию через вероятность состояния вполне однозначно. Это противоречие является только кажущимся. Оно устранится, если заметить, что вероятность не обязательно определять однозначно. *Однозначно должны определяться не сами вероятности, а их отношения в различных состояниях*. Отсюда следует, что вероятность может быть определена только с точностью до какого-то произвольного числового множителя C. Об этом уже говорилось при обсуждении условия нормировки (70.3). Наличие числового множителя скажется в том, что в формуле для S появится произвольная аддитивная постоянная $\ln C$.

Если вероятность нормирована условием (70.3), то она называется *математической вероятностью*. При применениях формулы Больцмана обычно более удобна нормировка, предложенная Планком. Ее выбирают так, чтобы все вероятности, если это возможно, выражались целыми положительными числами. Нормированную так вероятность называют *статистическим весом* или *термодинамической вероятностью состояния*. Статистический вес мы будем обозначать через G и записывать формулу Больцмана в виде

$$S = k \ln G. \tag{80.5}$$

8. В качестве примера рассмотрим снова закрытый сосуд объема V, в котором находится N тождественных молекул идеального газа. Каждую молекулу будем рассматривать как материальную точку, подчиняющуюся законам классической механики. Для того чтобы отличать одну молекулу от других молекул, занумеруем их числами $1, 2, 3, \ldots$. Мгновенное состояние газа будет описано полностью, если указать положение и скорость каждой молекулы в рассматриваемый момент времени. Такое описание состояния газа называется *динамическим*. Однако при статистическом изучении газа динамическое описание его является слишком детальным. Действительно, если мгновенное состояние газа характеризовать указанием точного положения и скорости каждой молекулы, то вероятность такого состояния всегда будет равна нулю (см. § 71). Нужна более грубая характеристика, чтобы каждому состоянию можно было приписать определенную вероятность, отличную от нуля. Отвлечемся сначала от движения молекул и разобьем мысленно объем V на m достаточно малых ячеек с объемами $V_1, V_2, \ldots, V_m$. При грубом описании состояния газа будем считать положение молекулы известным, если указано, в какой объемной ячейке она находится. При такой точности определения положения молекулы состояние газа можно характеризовать указанием *чисел молекул и их номеров* в каждой объемной ячейке. Состояние газа, описанное таким образом, называется *микросостоянием*. Очевидно, что при той точности определения положения молекулы, которая принята при описании микросостояния, всякое перемещение ее в пределах одной и той же объемной ячейки не меняет микросостояния газа. Переход же молекулы из одной ячейки в другую приводит к изменению микросостояния. Пусть p_i есть математическая вероятность того, что какая-либо определенная молекула попадает в i-ю объемную ячейку. Очевидно, $p_1 + p_2 + \ldots + p_m = 1$. Если нет силовых полей, действующих на газ, то $p_i = V_i/V$. Найдем математическую вероятность микросостояния, когда в ячейке 1 находится N_1 молекул с определенными номерами, в ячейке 2 — N_2 также определенных молекул и т. д. По теореме умножения вероятностей она равна $p_1^{N_1} p_2^{N_2} \ldots p_m^{N_m}$.

Если все молекулы газа, как мы предполагаем, тождественны, то с макроскопической точки зрения не имеет значения, какие именно молекулы находятся в той или иной ячейке. Для полного задания состояния газа надо указать только *общие числа молекул* в каждой объемной ячейке, но не их номера. Состояние газа, описанное таким образом, называется *макросостоянием*. Его можно изобразить следующей схемой:

$$\begin{array}{ll} \text{ячейки} & 1 \quad 2 \quad \ldots m \\ \text{число молекул} & N_1 \; N_2 \; \ldots \; N_m. \end{array} \tag{80.6}$$

В отличие от микросостояния всякая перестановка молекул без изменения чисел $N_1, N_2, \ldots, N_m$ не меняет макросостояния, независимо от того, производится ли эта перестановка в пределах одной объемной ячейки или в пределах объема всего газа.

9. Определим математическую вероятность макросостояния (80.6) с заданными числами частиц $N_1, N_2, \ldots, N_m$. Возьмем какое-либо микросостояние с теми же числами частиц в объемных ячейках. Представим себе, что все N частицы в этом микросостоянии закреплены на своих местах. Произведем затем всевозможные перестановки всех N частиц. Поскольку места, в которых они могут находиться, фиксированы, при таких перестановках общее число молекул в каждой объемной ячейке остается неизменным. Мы получим всевозможные комбинации частиц с требуемыми числами $N_1, N_2, \ldots$ в ячейках, которым предписано занимать закрепленные места. Число таких комбинаций равно $N!$. Однако при таком подсчете мы считали различными и такие комбинации, которые получаются друг из друга перестановкой частиц в пределах одной и той же объемной ячейки. Такие перестановки к новым макросостояниям не приводят. Число перестановок в пределах первой ячейки равно $N_1!$, в пределах второй — $N_2!$ и т. д. Поэтому для получения числа G всех возможных макросостояний надо $N!$ разделить на $N_1!N_2!\ldots N_m!$. Итак,

$$G = \frac{N!}{N_1!N_2!\ldots N_m!}. \tag{80.7}$$

Чтобы получить вероятность P макросостояния (80.6), надо эти числа умножить на вероятность одного микросостояния. Это дает

$$P = \frac{N!}{N_1!N_2!\ldots N_m!}\, p_1^{N_1} p_2^{N_2} \ldots p_m^{N_m}. \tag{80.8}$$

До сих пор на объемы ячеек $V_1, V_2, \ldots, V_m$ мы не накладывали никаких ограничений. Наши рассуждения применимы и в тех случаях, когда газ находится в силовом поле. Вероятности $p_1, p_2, \ldots, p_m$ в этих случаях, вообще говоря, различны. Допустим теперь, что силовых полей нет, а объемы ячеек $V_1, V_2, \ldots$ выбраны одинаковыми. Тогда станут одинаковыми и вероятности $p_1, p_2, \ldots$ и притом равными V_1/V. Вместо формулы (80.8) получится

$$P = G\Big(\frac{V_1}{V}\Big)^N.$$

Множитель при G постоянен, т. е. не зависит от $N_1, N_2, \ldots$. На него можно сократить и принять за вероятность макросостояния величину G. Эта величина G и будет статистическим весом рассматриваемого макросостояния. Таким образом, *статистический вес макросостояния можно определить как число равновесных микросостояний, каждое из которых реализует это макросостояние.*

§ 81. Флуктуации

1. Пусть f — любая физическая величина, испытывающая флуктуации. *Флуктуацией* величины f называется отклонение $\Delta f = f - \overline{f}$ мгновенного значения этой величины от ее среднего значения.

Очевидно, что $\Delta\overline{f} = 0$. Поэтому обычно пользуются *средним квадратом* флуктуации, т. е. величиной $\overline{(\Delta f)^2}$. Квадратный корень из этой величины $\sqrt{\overline{(\Delta f)^2}}$ называется *среднеквадратичной флуктуацией*, а ее отношение к среднему значению $\overline{f}$, т. е. $\sqrt{\overline{(\Delta f)^2}}/\overline{f}$, — *среднеквадратичной относительной флуктуацией*.

Усредняя выражение $(\Delta f)^2 = (f - \overline{f})^2 = f^2 - 2f\overline{f} - (\overline{f})^2$, получим

$$\overline{(\Delta f)^2} = \overline{f^2} - \overline{2\overline{f}f} - (\overline{f})^2.$$

Но $\overline{f}$ — величина постоянная, а потому $\overline{\overline{f}f} = \overline{f}\,\overline{f} = (\overline{f})^2$. Следовательно,

$$\overline{(\Delta f)^2} = \overline{f^2} - (\overline{f})^2. \tag{81.1}$$

Усредним теперь произведение двух флуктуирующих величин:

$$fg = (\overline{f} + \Delta f)\,(\overline{g} + \Delta g) = \overline{f}\,\overline{g} - \overline{g}\Delta f + \overline{f}\Delta g + \Delta f\Delta g.$$

Так как $\overline{f}$ и $\overline{g}$ — величины постоянные, а $\overline{\Delta f} = \overline{\Delta g} = 0$, то

$$\overline{fg} = \overline{f}\,\overline{g} + \overline{\Delta f\Delta g}. \tag{81.2}$$

Формула (81.1) содержится здесь как частный случай, который получается при $g = f$.

Величины f и g называются *статистически независимыми*, если $\overline{\Delta f\Delta g} = 0$. Для таких величин

$$\overline{fg} = \overline{f}\,\overline{g}. \tag{81.3}$$

2. Рассмотрим теперь любую физическую систему, состоящую из N независимых одинаковых частей. Примером такой системы может служить идеальный газ, а составных частей — отдельные молекулы. Пусть f_i — произвольная величина, характеризующая i-ю подсистему, например, в приведенном примере — кинетическая энергия i-й молекулы. Тогда в силу предполагаемой аддитивности соответствующая величина для всей системы будет $F = \sum f_i$. Выразим средний квадрат флуктуации величины F через аналогичный квадрат для величины f_i. Очевидно, $\overline{F} = \sum \overline{f}_i = N\overline{f}$, где индекс i опущен, так как предполагается, что все составные части системы тождественны. Далее,

$$F^2 = \Big(\sum f_i\Big)^2 = \sum f_i^2 + \sum_{i\neq j}\sum f_i f_j.$$

А так как эти части независимы, то $\overline{f_i f_j} = \overline{f}_i\overline{f}_j = (\overline{f})^2$. Следовательно,

$$\overline{F^2} = N\overline{f^2} + N(N-1)(\overline{f})^2.$$

Подставляя эти значения в формулу (81.1), получим

$$\overline{(\Delta F)^2} = \overline{F^2} - (\overline{F})^2 = N(\overline{f^2} - \overline{f}^2). \tag{81.4}$$

Отсюда на основании (81.1)

$$\frac{\sqrt{\overline{(\Delta F)^2}}}{\overline{F}} = \frac{\sqrt{N}\sqrt{\overline{\Delta f^2}}}{N\,\overline{f}} = \frac{1}{\sqrt{N}}\,\frac{\sqrt{\overline{(\Delta f)^2}}}{\overline{f}}. \tag{81.5}$$

С увеличением N относительная флуктуация величины F убывает обратно пропорционально $\sqrt{N}$. При больших N относительные флуктуации ничтожны. Этот вывод качественно верен и для неаддитивных величин. С ним связана *достоверность термодинамических результатов для больших макроскопических систем.*

3. Применим формулу (81.4) к вычислению флуктуаций числа молекул в фиксированном объеме V идеального газа. Пусть в большом закрытом сосуде объема V в отсутствие силовых полей находится N молекул идеального газа. Разделим объем V на $z = V/v$ одинаковых объемчиков величиной $v_i = v$ каждый. Если n_i — число молекул в объемчике v_i, а N — в большом объеме V, то $N = \sum n_i$. Среднее число молекул $\overline{n}_i$ в объемчике v_i будет $\overline{n}_i \equiv \overline{n} = Nv/V$, т. е. одно и то же во всех объемчиках. Представим его в виде $\overline{n} = Np$, где $p = v/V$ — вероятность нахождения молекулы в объемчике v. Возьмем теперь в качестве f_i следующие значения: $f_i = 1$, если i-я молекула находится внутри объемчика v, и $f_i = 0$, если она находится в оставшемся объеме $V - v$. Тогда число молекул w в объемчике v можно представить в виде $n = \sum f_i$, предполагая, что суммирование ведется по всем N молекулам объема V. Ясно, что $f_i = f_i^2 = f_i^3 = \ldots$, а потому $\overline{f}_i = \overline{f_i^2} = \overline{f_i^3} = \ldots = p$. Следовательно, по формуле (81.1)

$$\overline{\Delta f_i^2} = \overline{f_i^2} - (\overline{f_i})^2 = p - p^2 = p(1-p).$$

А так как в случае идеального газа величины $f_1, f_2, f_3, \ldots$ статистически независимы, то по формуле (81.4)

$$\overline{\Delta n^2} = Np(1-p) = (1-p)\overline{n}. \tag{81.6}$$

Если $V \to \infty$, то $p \to 0$, а следовательно,

$$\overline{\Delta n^2} = \overline{n}. \tag{81.7}$$

Отсюда получаем для относительной флуктуации плотности газа

$$\sqrt{\overline{\Delta\rho^2}}/\overline{\rho} = 1/\sqrt{\overline{n}}. \tag{81.8}$$

В соответствии со сказанным видим, что *в объемах с большим средним числом частиц* $\overline{n}$ *относительные флуктуации малы и трудно доступны наблюдению.* Наоборот, при малых $\overline{n}$ относительные флуктуации велики.

4. Более общий метод вычислений флуктуации плотности, применимый также к жидкостям и твердым телам, основан на *теореме о равномерном распределении кинетической энергии по степеням свободы.* Рассмотрим малую часть жидкости или газа, окруженную

такой же жидкой или газообразной средой, температура T которой поддерживается постоянной (термостатом). С целью упрощения и наглядности вычислений предположим, что малая часть жидкости или газа заключена в цилиндр с поршнем. Стенки цилиндра идеально проводят тепло, а поршень может ходить в нем без трения. Тогда наличие стенок цилиндра и поршня не будет препятствовать обмену энергий и выравниванию давлений между веществом в цилиндре и термостатом. Тепловое движение молекул вещества вызовет броуновское движение поршня. К этому движению поршня мы и применим теорему о равномерном распределении кинетической энергии по степеням свободы.

Поршень можно рассматривать как гармонический осциллятор, совершающий беспорядочные тепловые колебания. Среднее значение его потенциальной энергии при смещении на x из положения равновесия $x = 0$ равно $(1/2)\varkappa x^2 = (1/2)kT$, где $\varkappa$ — жесткость, соответствующая такому смещению. Если S — площадь поршня, а ΔV — изменение объема системы, то $\Delta V = Sx$. Таким образом, $\overline{(\Delta V)^2} = S^2\overline{x^2} = S^2kT/\varkappa$. Сила, возвращающая поршень в положение равновесия, будет $F = S\dfrac{\partial P}{\partial x}x$, где P — давление газа или жидкости. Поэтому $\varkappa = -S\,\partial P/\partial x = S^2\,\partial P/\partial V$. В результате получим

$$\overline{(\Delta V^2)}_T = \frac{kT}{(\partial P/\partial V)_T} = -kT\left(\frac{\partial V}{\partial P}\right)_T. \tag{81.9}$$

Знак T указывает, что в выводе предполагалось *постоянство температуры окружающей среды* (термостата). Если бы вещество внутри объема V было адиабатически изолировано, то знак T следовало бы заменить на S (постоянство энтропии), т. е.

$$\overline{(\Delta V^2)}_S = -kT\left(\frac{\partial V}{\partial P}\right)_S. \tag{81.10}$$

Формулы (81.9) и (81.10) выражают флуктуации объема *одной и той же массы вещества*, находящейся в термодинамическом равновесии с окружающей средой.

Для идеального газа при постоянной температуре $PV = \text{const}$, так что $(\partial V/\partial P)_T = -V/P$. А так как $PV = nkT$, где n — число молекул в объеме V, то из формулы (81.9) получаем $\overline{(\Delta V)^2} = V^2/n$. Пусть теперь объем V фиксирован, а число молекул идеального газа в нем n меняется из-за тепловых флуктуаций. Если бы объем V увеличился на ΔV, то число молекул в прежнем объеме V уменьшилось бы на $\Delta n = n\Delta V/V$. Отсюда

$$\overline{(\Delta n)^2} = (n/V)^2\,\overline{(\Delta V)^2} = n = \overline{n},$$

что совпадает с прежним результатом (81.7). В окончательном результате различием между n и $\overline{n}$ можно пренебречь, что мы и сделали.

5. Перейдем теперь к вычислению *флуктуаций энергии*. С целью лучшего уяснения метода начнем с вычисления флуктуаций кинетической энергии ε молекулы одноатомного идеального газа в отсутствие

силовых полей. Согласно максвелловскому закону распределения скоростей

$$\overline{\varepsilon} = \frac{1}{Z}\int \varepsilon e^{-\alpha\varepsilon}\, d\Gamma, \tag{81.11}$$

где $\alpha = 1/kT$, $d\Gamma$ — элементы объема пространства скоростей, а Z определяется условием нормировки

$$Z = \int e^{\alpha\varepsilon}\, d\Gamma. \tag{81.12}$$

Дифференцируя это соотношение по параметру α, получим

$$\frac{dZ}{d\alpha} = -\varepsilon \int e^{-\alpha\varepsilon}\, d\Gamma,$$

и формула (81.11) перейдет в

$$\overline{\varepsilon} = -\frac{1}{Z}\frac{dZ}{d\alpha}. \tag{81.13}$$

Отсюда

$$\frac{d\overline{\varepsilon}}{d\alpha} = -\frac{1}{Z}\frac{d^2Z}{d\alpha^2} + \frac{1}{Z^2}\left(\frac{dZ}{d\alpha}\right)^2 = -\frac{1}{Z}\frac{d^2Z}{d\alpha^2} + \varepsilon^2.$$

Аналогично,

$$\overline{\varepsilon^2} = \frac{1}{Z}\int \varepsilon^2 e^{-\alpha\varepsilon}\, d\Gamma = \frac{1}{Z}\frac{d^2Z}{d\alpha^2}. \tag{81.14}$$

Сравнение этой формулы с предыдущей дает

$$\overline{\Delta\varepsilon^2} = \overline{\varepsilon^2} - (\overline{\varepsilon})^2 = -\frac{d\overline{\varepsilon}}{d\alpha},$$

или после подстановки параметра $\alpha = 1/kT$

$$\overline{\Delta\varepsilon^2} = kT^2\frac{d\overline{\varepsilon}}{dT}. \tag{81.15}$$

Так как для идеального одноатомного газа $\overline{\varepsilon} = (3/2)kT$, то в этом случае

$$\overline{\Delta\varepsilon^2} = \frac{3}{2}(kT)^2. \tag{81.16}$$

6. Приведенный метод вычисления $\overline{\Delta\varepsilon^2}$ может быть распространен практически без изменений на случай величин внешнего силового поля. Надо только максвелловское распределение заменить *распределением Больцмана и производить интегрирование не только по скоростям, но и по координатам обычного пространства*. В результате снова получится формула (81.15), в которой, однако, *под ε следует понимать уже полную энергию молекулы, состоящую из кинетической и потенциальной*.

Важно отметить расширение области применимости формулы (81.15) для флуктуации полной энергии ε. Выделим малую часть

(подсистему) изотропной среды (жидкости или газа), находящуюся в статическом равновесии со всей средой, температура T которой поддерживается постоянной. Подсистемой может быть и отдельная молекула. По отношению к выделенной подсистеме окружающая среда играет роль термостата. Из-за обмена энергий между термостатом и подсистемой энергия последней будет непрерывно флуктуировать. *Флуктуации полной энергии ε подсистемы определяются уравнением (81.15). Не имеет значения, меняется ли энергия ε непрерывно (классическая система), или принимает дискретный ряд значений (квантовая система).* Доказательство этого утверждения дается в статистической физике. Оно основано на *каноническом распределении Гиббса*, частными случаями которого являются распределения Максвелла и Больцмана (см. конец следующего параграфа).

В случае макроскопической подсистемы, объем которой поддерживается постоянным, $\overline{\varepsilon}$ имеет смысл *внутренней энергии подсистемы, а $d\overline{\varepsilon}/dT$ — ее теплоемкости C_V при постоянном объеме.* Тогда из формулы (81.15) получается

$$(\overline{\Delta\varepsilon^2})_V = kT^2\, C_V. \tag{81.17}$$

Знак V снова указывает на то, что $(\overline{\Delta\varepsilon^2})_V$ есть средний квадрат флуктуации энергии подсистемы при сохранении ее объема V постоянным.

7. Рассмотрим теперь *флуктуации энтальпии I* подсистемы. Для этого воспользуемся следующим искусственным приемом. Предположим, что подсистема заключена в оболочку с идеально проводящими подвижными стенками, так что объем подсистемы не сохраняется постоянным. Пусть оболочка снаружи подвергается действию постоянных внешних сил, поддерживающих внешнее давление P постоянным. Эти силы увеличивают потенциальную энергию подсистемы на величину PV. Если под ε понимать ту же энергию, что и в предыдущем выводе, то с учетом дополнительной потенциальной энергии PV среднее значение полной энергии подсистемы будет $\overline{\varepsilon} + \overline{PV}$. Но это есть энтальпия подсистемы I. Все предыдущие рассуждения можно повторить без изменений, заменив ε на $\varepsilon + PV$. В результате вместо (81.15) получится

$$(\overline{\Delta I^2})_P = kT^2\, \frac{dI}{dT}. \tag{81.18}$$

Но при $P = \text{const}$ производная dI/dT есть теплоемкость C_P подсистемы при постоянном давлении, а потому

$$(\overline{\Delta I^2})_P = kT^2\, C_P. \tag{81.19}$$

8. Распространим теперь изложенный выше термодинамический метод вычисления флуктуаций на *любые величины*, характеризующие макроскопические свойства подсистемы. Ограничимся при этом *изотропными телами.* Для них любая физическая величина в состоянии термодинамического равновесия есть флуктуация двух других величин, которые могут быть приняты за *независимые переменные.*

Термодинамические величины макроскопических подсистем хотя и испытывают флуктуации, но в случае малости таких подсистем их мгновенные состояния *практически равновесны*. Их состояние также определяется двумя независимыми переменными. Поэтому задача сводится к вычислению тепловых флуктуаций двух независимых переменных. В окончательном результате, определяющем значение среднего квадрата той или иной флуктуации, необходимо указывать, какая из двух величин, выбранных для характеристик состояния подсистемы, *поддерживается постоянной*. Иначе самый результат будет неопределенным, а потому и бессмысленным.

Покажем на примерах, как изложенный метод применяется к вычислению флуктуаций различных физических величин.

Начнем с флуктуаций температуры, предполагая, что рассматриваемая макроскопическая подсистема находится в тепловом контакте с термостатом. Считая температуру T подсистемы функцией независимых переменных V и ε, напишем

$$\Delta T = \left(\frac{\partial T}{\partial V}\right)_\varepsilon \Delta V + \left(\frac{\partial T}{\partial \varepsilon}\right)_V \Delta\varepsilon.$$

В силу независимости V и ε имеем $\overline{\Delta V \cdot \Delta\varepsilon} = 0$, а потому

$$\overline{(\Delta T)^2} = \left(\frac{\partial T}{\partial V}\right)_\varepsilon^2 \overline{(\Delta V)^2} + \left(\frac{\partial T}{\partial \varepsilon}\right)_\varepsilon^2 \overline{(\Delta\varepsilon)^2}.$$

При постоянном объеме производная $(\partial\varepsilon/\partial T)_V$ есть теплоемкость подсистемы C_V. Тогда по формуле (81.17) находим

$$\overline{(\Delta T)^2} = \frac{1}{C_V^2}\,(\overline{\Delta\varepsilon^2})_V = \frac{kT^2}{C_V}. \tag{81.20}$$

Вычислим теперь *флуктуации энтропии* S подсистемы. В качестве независимых переменных выберем V и ε. Рассуждая как в предыдущем примере, напишем

$$\overline{\Delta V \cdot \Delta\varepsilon} = 0, \quad (\overline{\Delta S^2})_V = \left(\frac{\partial S}{\partial \varepsilon}\right)_V^2 (\overline{\Delta\varepsilon^2})_V.$$

Так как $\partial S/\partial\varepsilon = 1/T$, то с учетом (81.17)

$$(\overline{\Delta S^2})_V = kC_V. \tag{81.21}$$

Если бы независимые переменные были приняты P и I, то получилось бы

$$(\overline{\Delta S^2})_P = kC_P. \tag{81.22}$$

Рассмотрим теперь флуктуации давления P. Примем за независимые переменные V и T. Тогда

$$\overline{\Delta V \cdot \Delta T} = 0, \quad (\overline{\Delta P^2})_T = \left(\frac{\partial P}{\partial V}\right)_T^2 (\overline{\Delta V^2})_T$$

или на основании формулы (81.9)

$$(\overline{\Delta P^2})_T = -kT\left(\frac{\partial P}{\partial V}\right)_T. \tag{81.23}$$

Приняв за независимые переменные P и S, мы получили бы

$$(\overline{\Delta P^2})_S = -kT\left(\frac{\partial P}{\partial V}\right)_S = -\gamma kT\left(\frac{\partial P}{\partial V}\right)_T, \tag{81.24}$$

где $\gamma = C_P/C_V$. При этом было использовано термодинамическое соотношение $\left(\frac{\partial P}{\partial V}\right)_S \Big/ \left(\frac{\partial P}{\partial V}\right)_T = \gamma$ (см. § 47).

Вычислим, наконец, *флуктуации плотности* вещества ρ в объеме V. Задача сводится просто к преобразованию формулы (81.9). Прежде всего заметим, что величина $V\partial P/\partial V$ не зависит от величины объема V. Поэтому в таком выражении V можно заменить удельным объемом вещества v. Тогда

$$(\overline{\Delta V^2})_T = -\frac{VkT}{v(\partial P/\partial v)_T}.$$

Так как масса вещества $V\rho$ в объеме V остается постоянной, то $V\Delta\rho + \rho\Delta V = 0$. Следовательно, $(\Delta\rho)^2 = (\rho/V)^2\,(\Delta V)^2$, а потому

$$V(\overline{\Delta\rho^2})_T = -\frac{\rho^2 kT}{v(\partial P/\partial v)_T}. \tag{81.25}$$

Чем меньше объем V, тем больше относительные флуктуации плотности в нем.

В т. IV при рассмотрении явления молекулярного рассеяния света мы воспользуемся большинством формул, выведенных здесь.

§ 82. Метод наиболее вероятного распределения в статистике Больцмана

1. Как определять микро- и макросостояния системы, их вероятности и статистические веса — это наиболее фундаментальные и труднейшие вопросы статистической термодинамики, к тому же еще не исследованные до конца. Мы не имеем возможности осветить здесь эти вопросы с исчерпывающей полнотой, так как для этого потребовались бы основательные знания математики, аналитической механики, электродинамики и квантовой механики. Остановимся только на идейной стороне, опуская те вопросы и доказательства, для которых такое знание необходимо.

Рассмотрим сначала систему N тождественных молекул, помещенных в закрытом сосуде с жесткими теплонепроницаемыми стенками. Система может находиться во внешнем постоянном потенциальном силовом поле. Ее полная энергия остается постоянной. Между молекулами должно происходить взаимодействие. Будем предполагать, что это взаимодействие *слабое*. Это значит, что энергия взаимодействия

пренебрежимо мала, и имеет смысл говорить об энергии каждой молекулы в отдельности, а не только об энергии системы в целом. Такое взаимодействие необходимо, так как только благодаря ему в системе и может установиться определенное статистическое распределение.

2. Подойдем к вопросу сначала с классической точки зрения, принимая молекулы за материальные точки, подчиняющиеся законам классической механики. Динамическое состояние системы определяется заданием координат x, y, z и соответствующих им импульсов p_x, p_y, p_z каждой молекулы. Для упрощения терминологии введем воображаемое шестимерное пространство, каждая точка которого характеризуется шестью координатами x, y, z, p_x, p_y, p_z. Такое пространство называется *фазовым пространством* молекулы, а его точки — *изображающими точками в фазовом пространстве*. Таким образом, мгновенное состояние отдельной молекулы полностью характеризуется положением ее фазовой точки в фазовом пространстве, а динамическое состояние всех N молекул — положением фазовых точек этих молекул в том же фазовом пространстве. Переход к микро- и макросостояниям осуществляется так же, как и в § 80. Только вместо объемных ячеек приходится рассматривать *фазовые ячейки*, т. е. *ячейки в фазовом пространстве*. Необходимо, однако, разъяснить, как определять шестимерные объемы фазовых ячеек и вообще конечных областей фазового пространства. Для этого рассмотрим прежде всего случай, когда фазовая ячейка имеет форму бесконечно малого шестимерного прямоугольного параллелепипеда в фазовом пространстве. Так называется совокупность фазовых точек, координаты которых лежат внутри бесконечно малых интервалов $(x, x + dx), \ldots, (p_z, p_z + dp_z)$. Фазовым объемом такого элементарного шестимерного параллелепипеда называется произведение $dx\, dy \ldots dp_z$. Складывая объемы всех элементарных параллелепипедов, заполняющих какую-либо область фазового пространства, получим *фазовый объем этой области*. Разобьем теперь все фазовое пространство молекулы на достаточно малые области с одинаковыми фазовыми объемами. Такие области и называются *фазовыми ячейками*. Число фазовых ячеек в фазовом пространстве молекулы бесконечно велико, поскольку p_x, p_y, p_z могут принимать все значения от $-\infty$ до $+\infty$. Занумеруем фазовые ячейки числами $1, 2, 3, \ldots$ Энергию молекулы, когда она находится в i-й ячейке, обозначим через ε_i. Для определенности нумерацию условимся производить в порядке возрастания энергии ε ($\varepsilon_1 \leqslant \varepsilon_2 \leqslant \varepsilon_3 \leqslant \ldots$). Для целей статистики, как мы видели, от полного динамического описания состояния молекулы надо перейти к менее полному и грубому описанию. *Мы будем считать состояние отдельной молекулы описанным полностью, если указано, в какой фазовой ячейке она находится.*

3. Принципиальный недостаток такого описания состояния молекулы состоит в том, что объемы фазовых ячеек не фиксированы. И такая неопределенность в последовательно классической теории не может быть устранена, так как классическая теория допускает только непрерывные изменения состояния. Фазовые ячейки, как бы малы

они ни были, можно дробить бесконечно на более мелкие части, и все же в каждой такой части может еще поместиться бесконечное и непрерывное множество фазовых точек. Но классический способ описания состояния частицы заданием ее координат и импульсов имеет принципиальные границы применимости, определяемые соотношением неопределенности Гейзенберга. Координата x и соответствующий ей импульс p_x могут быть заданы только с неопределенностями δx и δp_x, подчиняющимися условию $\delta x \cdot \delta p_x \gtrsim h$. Поэтому естественно выбрать объем фазовой ячейки равным ch^3, где c — постоянная порядка единицы. Уточнение значения этой постоянной несущественно. Так называемая *полуклассическая теория* полагала $c = 1$. Как выяснилось позднее, при таком выборе число квантовых состояний частицы в полуклассической теории совпадает с числом состояний ее в последовательной квантовой механике. Это обстоятельство и делает целесообразным указанный выбор.

Последовательная квантовая механика вообще отказалась от описания динамических состояний координатами и импульсами частиц. Здесь нет необходимости вдаваться в подробности, как в квантовой механике описываются состояния частиц или их систем. Существенно только, что квантовая механика допускает *дискретные состояния*. Система или частица не может перейти из одного состояния в соседнее непрерывно, так как промежуточных состояний не существует. Переход совершается скачком. Говоря о квантовых состояниях, в дальнейшем мы имеем в виду квантовые состояния отдельной молекулы. При этом молекула не обязательно должна рассматриваться как точечная частица, а может иметь и внутреннюю структуру. Для наших целей достаточно ограничиться так называемыми *стационарными состояниями*, т. е. состояниями, не изменяющимися во времени. Они характеризуются определенными значениями или уровнями энергии молекулы $\varepsilon_1, \varepsilon_2, \ldots$ Уровни энергии могут быть *простыми* и *кратными*. Уровень энергии и соответствующее ему квантовое состояние называются *кратными*, или *вырожденными*, если существует несколько состояний молекулы с тем же значением энергии, отличающихся друг от друга значениями других физических величин. В противном случае уровень и квантовое состояние называются *простыми*, или *невырожденными*. Число подуровней, из которых состоит кратный уровень, называется *кратностью уровня* или *кратностью вырождения*. Не теряя общности, мы будем считать все уровни простыми. Если это не так, то достаточно разделить каждый кратный уровень на соответствующие простые подуровни, чтобы свести этот случай к предыдущему.

Описание с помощью квантовых состояний есть *предельно подробное описание*, допускаемое принципами квантовой механики. В этом смысле оно соответствует *динамическому описанию* классической механики. И если квантовые состояния дискретны, то при переходе от квантовой механики к статистике отпадает необходимость в замене этих состояний более грубыми, что необходимо делать в классической статистике. В этом и состоит преимущество квантовой статистики перед классической.

4. Дальнейшие рассуждения не зависят от того, какая принята точка зрения — квантовая или классическая. Чтобы их формально объединить вместе, будем применять квантовую терминологию и в классической теории. Будем говорить, что классическая частица находится в i-м квантовом состоянии с энергией ε_i, если она находится в i-й ячейке своего фазового пространства. Будем предполагать в этом параграфе, что частицы *принципиально различимы*, хотя бы они и были абсолютно тождественны. Это предположение лежит в основе так называемой *статистики Больцмана*. Если его принять, то частицы можно занумеровать, как это делалось в § 80. В последовательно квантовой теории микросостояние системы совпадает с ее квантовым состоянием. Здесь эти понятия тождественны. Микросостояние газа характеризуется числами частиц в каждом квантовом состоянии с указанием их номеров. Для характеристики микросостояния надо указать только числа частиц $N_1, N_2, \ldots$ в каждом квантовом состоянии.

В заданных условиях, в которых находится газ, не все мыслимые микросостояния возможны. Например, если стенки сосуда непроницаемы для частиц, а частицы рождаться и уничтожаться не могут, то будет невозможно микросостояние, в котором одна или несколько частиц находятся вне сосуда. Если система замкнута и ее энергия равна E, то невозможно микросостояние с энергией, отличающейся от E. Все возможные микросостояния системы называются *допустимыми*. Основная гипотеза, принимаемая в статистической механике, состоит в том, что *все допустимые микросостояния замкнутой системы равновероятны*. Если в § 80, когда речь шла только о положениях, но не о скоростях частиц, аналогичная гипотеза не вызывала сомнений и была почти самоочевидной, то здесь ее обоснование потребовало бы весьма сложных и тонких рассуждений. Мы будем считать, что доказательством введенной гипотезы является то, что ее выводы подтверждаются опытом. После всего изложенного ясно, что статистический вес макросостояния и здесь определяется формулой (80.7), т. е.

$$G = \frac{N!}{N_1!N_2!\ldots}. \tag{82.1}$$

5. Найдем теперь такое распределение частиц по квантовым состояниям, которому соответствует максимальное значение статистического веса G, а следовательно, и максимальное значение энтропии S. Это и будет состояние статистического равновесия системы, около которого происходят малые флуктуации. При отыскании максимума надо учесть два дополнительных условия:

$$N_1 + N_2 + \ldots = N = \text{const}, \tag{82.2}$$

$$N_1\varepsilon_1 + N_2\varepsilon_2 + \ldots = E = \text{const}. \tag{82.3}$$

Первое из них выражает постоянство числа частиц в системе, второе — постоянство ее полной энергии. Только микросостояния, удовлетво-

ряющие этим условиям, являются допустимыми, и только они должны приниматься во внимание. Все прочие микросостояния являются невозможными.

Будем предполагать, что очень велико не только общее число молекул N в сосуде, но и велики числа молекул $N_1, N_2, \ldots$ в каждом квантовом состоянии. Правда, это условие не может быть выполнено для всех N_i. Дело в том, что общее число молекул N, хотя и очень велико, но все же конечно. Поэтому целые числа N_i при достаточно больших номерах i неизбежно станут малыми и даже обратятся в нуль при дальнейшем возрастании i. Однако такие молекулы составляют лишь ничтожную долю от общего числа молекул N, и их наличие никак не скажется на статистическом поведении всего газа. Если целые числа N и N_i очень велики, то они могут быть аппроксимированы непрерывно меняющимися аргументами. Для вычисления факториалов можно воспользоваться формулой Стирлинга (1692–1770)

$$N! = \sqrt{2\pi N}\left(\frac{N}{e}\right)^N \exp\frac{\vartheta}{12N}, \tag{82.4}$$

где $0 < \vartheta < 1$ [1]. Эту точную формулу мы заменим приближенной

$$N! = \sqrt{2\pi N}\left(\frac{N}{e}\right)^N. \tag{82.5}$$

Относительная ошибка при вычислениях по этой формуле не превосходит $e^{1/(12N)} - 1 \approx 1/(12N)$. Уже при $N = 10$ она меньше одного процента. Подставляя выражение (81.5) в формулу (81.1) и учитывая соотношение (82.2), получим

$$G = \frac{C(N)}{\prod\limits_i (N_i)^{N_i+1/2}}, \tag{82.6}$$

где постоянная C может зависеть только от N. Для энтропии найдем

$$S = -k\sum_i (N_i + 1/2)\ln N_i + \text{const}.$$

Здесь можно пренебречь $1/2$ по сравнению с N_i. Тогда

$$S = -k\sum_i N_i \ln N_i + \text{const}. \tag{82.7}$$

Вместо того чтобы отыскивать максимум статистического веса, удобнее отыскивать максимум энтропии (82.7) при дополнительных условиях (82.2) и (82.3). Применяя метод Лагранжа, варьируем эти выражения и получаем в максимуме

$$\sum \ln N_i \cdot dN_i = 0, \quad \sum dN_i = 0, \quad \sum \varepsilon_i\, dN_i = 0. \tag{82.8}$$

[1] См., например: *Фихтенгольц Г. М.* Курс дифференциального и интегрального исчисления.— М.: Наука. Т. II (любое издание).

Отсюда

$$\sum(\ln N_i + \beta + \alpha\varepsilon_i)\, dN_i = 0,$$

где α и β — лагранжевы множители, не зависящие от всех переменных N_i. Их выбираем так, чтобы коэффициенты при dN_1 и dN_2 обратились в нуль. Тогда будут равны нулю и коэффициенты при всех остальных dN_i, так как переменные $N_3, N_4, \ldots$ можно принять за независимые. Итак,

$$\ln N_i + \beta + \alpha\varepsilon_i = 0,$$

откуда

$$\overline{N}_i = N_0 e^{-\alpha\varepsilon_i}, \tag{82.9}$$

где $N_0 = e^{-\beta}$ — новая постоянная. Черта над N_i поставлена для того, чтобы подчеркнуть, что речь идет о средних значениях чисел N_i, точнее об их значениях в наиболее вероятном состоянии.

6. Для определения постоянной α заменим адиабатические стенки сосуда теплопроводящими, сохраняя объем сосуда неизменным. Газ в сосуде перестанет быть изолированной системой, но его макроскопическое состояние останется тем же, если только температура окружающей среды равна температуре газа T и поддерживается постоянной. Появятся лишь малые флуктуации энергии, которые при полной изоляции системы были бы невозможны. Но флуктуации не принимаются во внимание при термодинамическом описании систем. Будем теперь бесконечно медленно (квазистатически) изменять температуру окружающей среды. Так как объем газа сохраняется неизменным, то газ не совершает работы, а только обменивается теплотой с окружающей средой. Поэтому $dE = \delta Q = T\, dS$. Энергии квантовых уровней $\varepsilon_1, \varepsilon_2, \ldots$ в таком процессе останутся неизменными. Они зависят лишь от внутренней структуры молекулы и от положения стенок сосуда, которое во время процесса не изменяется. Будет происходить лишь *перераспределение молекул между различными уровнями*, т. е. будут меняться средние числа заполнения $\overline{N}_i$. Для изменения энергии газа получаем $dE = \sum \varepsilon_i\, d\overline{N}_i$, а для изменения энтропии, согласно формуле (82.7), $dS = -k \sum \ln \overline{N}_i\, d\overline{N}_i = k\alpha \sum \varepsilon_i\, d\overline{N}_i$, так как $\sum dN_i = 0$. Подставляя эти значения в соотношение $dE = T\, dS$, получаем

$$\alpha = 1/kT, \tag{82.10}$$

а потому

$$\overline{N}_i = N_0 e^{-\varepsilon_i/kT}. \tag{82.11}$$

Это — *распределение Максвелла–Больцмана.* С изложенной здесь точки зрения оно может быть охарактеризовано как *наиболее вероятное распределение.* Кроме того, показано, что *это распределение верно не только в классической, но и в квантовой статистике.* Постоянная N_0 найдется из условия нормировки

$$\sum \overline{N}_i = N_0 \sum e^{-\varepsilon_i/kT} = N. \tag{82.12}$$

Если квантовые уровни молекулы вырождены, то вместо формулы (82.11), очевидно, следует писать

$$\overline{N}_i = N_0 g_i e^{-\varepsilon_i/kT}, \tag{82.13}$$

где g_i — кратность i-го уровня.

7. Найдем еще выражение для энтропии идеального газа через функцию распределения в классической статистике. С этой целью разделим фазовое пространство молекулы на ячейки с равными фазовыми объемами $\Delta\Omega$. Среднее число частиц в i-й ячейке будет $\overline{N}_i = N f(\varepsilon_i)\,\Delta\Omega$. Подставляя в (82.7) и принимая во внимание, что $\ln \Delta\Omega$ есть величина постоянная, получим

$$S = -kN \sum f(\varepsilon_i) \ln f(\varepsilon_i)\,\Delta\Omega + \text{const},$$

или, заменяя сумму интегралом,

$$S = -kN \int f(\varepsilon) \ln f(\varepsilon)\, d\Omega + \text{const}. \tag{82.14}$$

8. Заметим, наконец, что формула (82.11) применима для статистического описания не только отдельных молекул, но и макроскопических систем. Возьмем большую изолированную макроскопическую систему Σ, которую можно мысленно разделить на одинаковые малые, но макроскопические подсистемы σ, слабо взаимодействующие между собой. Благодаря такому взаимодействию подсистемы могут обмениваться энергией и находиться в различных квантовых состояниях с энергиями ε_i. К подсистемам применимы все рассуждения, приведенные выше для отдельных молекул. Среднее число подсистем, находящихся в i-м квантовом состоянии, будет по-прежнему определяться формулой (82.11). Но равновесное состояние подсистемы не зависит от того, какой средой она окружена, а зависит только от температуры этой среды. Поэтому можно изменить постановку вопроса. Пусть σ — произвольная макроскопическая система, окруженная любой протяженной средой, температура которой поддерживается постоянной. Такую среду называют *термостатом*, а о σ говорят как о «*системе в термостате*». Формула (82.11) применима и к этому случаю. Число $\overline{N}_i$ определяет относительную вероятность того, что система σ при термодинамическом равновесии находится в i-м квантовом состоянии. Понимаемая в таком смысле формула (82.11) называется *каноническим распределением Гиббса*. Это распределение представляет наиболее общую и удобную основу для построения статистической механики.

В частности, на основе распределения Гиббса можно получить и формулу (81.15) для среднего квадрата флуктуации энергии подсистемы (независимо от того, является ли энергия непрерывной, или может принимать только дискретные значения). Действительно, и в этом случае полностью сохраняют силу все рассуждения, приведенные

в § 81, п. 5. Надо только в этих рассуждениях все интегралы заменить суммами. Но это — чисто формальный момент, не затрагивающий существа вопроса.

§ 83. Статистики Ферми–Дирака и Бозе–Эйнштейна

1. Согласно современной квантовой теории все элементарные и сложные частицы разделяются на два класса. К первому классу относятся электроны, протоны, нейтроны и все частицы с так называемым *полуцелым спином*. Эти частицы подчиняются *статистике Ферми–Дирака*. Они называются *фермионами*. Ко второму классу относятся фотоны, π- и K-мезоны и все частицы с *целым спином*. Эти частицы называются *бозонами*. Никаких других возможностей квантовая теория не допускает. Статистика Больцмана, изложенная в предыдущем параграфе, является приближенным предельным случаем, в который переходят при определенных условиях статистики Ферми–Дирака и Бозе–Эйнштейна. В настоящем томе нам почти не придется применять эти квантовые статистики. Но, учитывая их важность в самых различных разделах современной физики, необходимо уже здесь, насколько это возможно, изложить их физические (точнее, статистические) основы.

2. Во всех трех статистиках допустимые микросостояния принимаются равновероятными. Но статистики отличаются друг от друга тем, как они определяют микросостояния и статистические веса макросостояний. Статистика Больцмана стоит на точке зрения *принципиальной различимости частиц*, даже когда частицы абсолютно тождественны. Если частица A находится в квантовом состоянии I, а частица B — в квантовом состоянии II, то получится новое микросостояние, когда эти частицы поменяются местами, т. е. частица A перейдет в состояние II, а частица B — в состояние I. Квантовые статистики Ферми–Дирака и Бозе–Эйнштейна, наоборот, принимают, что при такой перестановке *никаких изменений не произойдет* — получится в точности то же микросостояние. Обе эти статистики стоят на точке зрения *принципиальной неразличимости тождественных частиц*. Различие между статистиками Ферми–Дирака и Бозе–Эйнштейна состоит в следующем. *В статистике Ферми–Дирака принимается, что в каждом квантовом состоянии может находиться не более одной частицы. Статистика Бозе–Эйнштейна подобных ограничений не накладывает. Она допускает, что в каждом квантовом состоянии может находится любое число частиц*. Столь различное поведение бозонов и фермионов обосновывается в квантовой механике теории поля. Здесь об этом говорить преждевременно.

Для пояснения рассмотрим две тождественные частицы A и B, которые требуется распределить по трем квантовым состояниям. Состояния будем изображать клетками. Все равновозможные случаи, допускаемые статистикой Больцмана, представлены на рис. 70, слева.

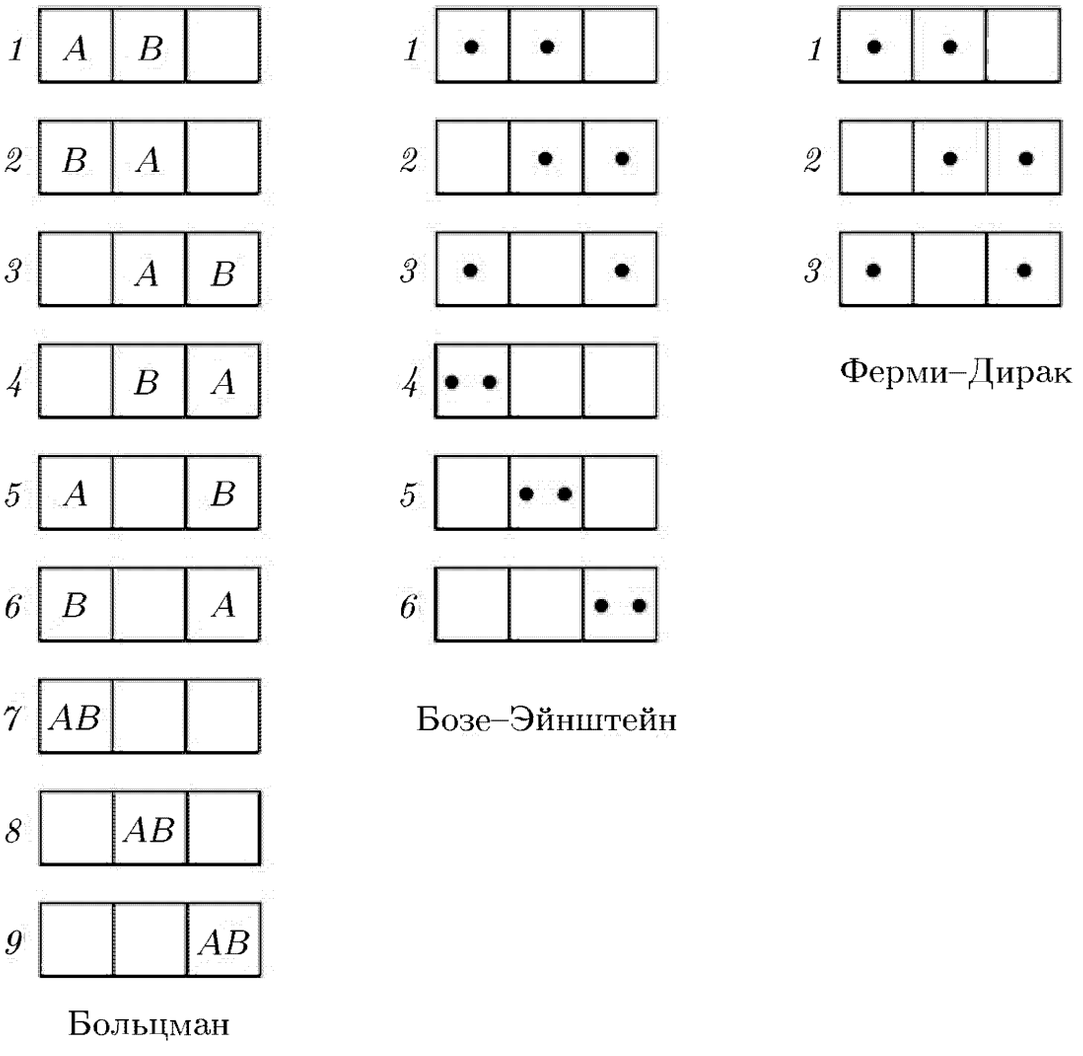


Рис. 70

Всех микросостояний девять, математическая вероятность каждого из них 1/9. По статистикам Бозе–Эйнштейна и Ферми–Дирака (рис. 70, в центре и справа), состояния 1 и 2, 3 и 4, 5 и 6 принципиально неразличимы, и каждая пара таких состояний должна рассматриваться как одно состояние. Частицы «обезличены», их уже нельзя обозначать различными буквами A и B, на нашей схеме они обозначены точками. Если частицы являются бозонами, то число всех возможных микросостояний будет шесть, а вероятность каждого из них 1/6. (Если бы вероятность определялась по Больцману, то каждому из первых трех состояний следовало бы приписать математическую вероятность 2/9, а каждому из последних трех 1/9. Поэтому очевидно, что статистики Больцмана и Бозе–Эйнштейна, а также Ферми–Дирака принципиально различны.) Для фермионов последние три распределения, представленные в среднем столбце, реализоваться не могут. Остаются только три микросостояния, изображенные справа. Вероятность каждого из них равна 1/3 (по Больцману она равна 2/9).

3. Прежде чем идти дальше, решим следующую задачу из комбинаторики. Имеется Z квантовых состояний. Требуется определить число способов, с помощью которых по этим состояниям можно распределить N тождественных частиц. Аналогом этой задачи может служить

следующая. Имеется Z квартир, требуется определить число способов заселения их N людьми. При этом предполагается, что люди «обезличены», так что не имеет значения, какой именно человек поселится в той или иной квартире.

Решим эту задачу сначала для фермионов. В этом случае должно быть $Z \geqslant N$, так как при $N > Z$ фермионы разместить по квантовым состояниям нельзя, поскольку в каждом квантовом состоянии может находиться не более одного фермиона. Изобразим все Z квантовых состояний клетками (рис. 71). В заполненной клетке поставим темный кружок, в свободной — светлый. Произведем затем всевозможные перестановки между Z элементами: N темными и $(Z-N)$ светлыми кружками. В результате получатся всевозможные распределения точек по клеткам. Число таких распределений будет $Z!$. Однако это число надо уменьшить в $N!$ раз, так как перестановки между темными кружками не приводят к новым распределениям. Кроме того, его надо уменьшить еще в $(Z-N)!$ раз, поскольку перестановки между светлыми кружками также не приводят к новым распределениям. В результате для числа «расселений» N фермионов по Z квантовым состояниям получаем

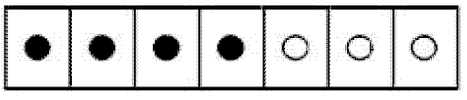


Рис. 71

$$\frac{Z!}{N!(Z-N)!}. \tag{83.1}$$

Допустим теперь, что частицами являются бозоны. В этом случае соотношение между числами Z и N может быть каким угодно. Снова изобразим квантовые состояния Z клетками, разделенными друг от друга $Z-1$ перегородками (рис. 72). Концы крайних клеток оставим свободными. Разместим в этих клетках совершенно произвольно все частицы — точки. Получим $Z+N-1$ элементов: N частиц и $Z-1$ перегородок. Произведем всевозможные перестановки между этими элементами. Получим всевозможные распределения N частиц по Z клеткам. Однако это число надо уменьшить в $N!$ раз, так как перестановки между частицами не приводят к новым распределениям. Кроме того, его надо уменьшить еще в $(Z-1)!$ раз, поскольку перестановки между перегородками также не приводят к новым распределениям. В результате для числа распределений N бозонов по Z квантовым состояниям получим выражение

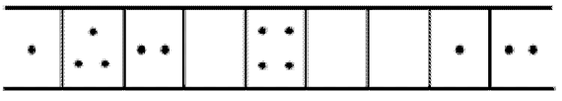


Рис. 72

$$\frac{(Z+N-1)!}{N!(Z-1)!}. \tag{83.2}$$

4. Перейдем теперь к выводу распределений Ферми–Дирака и Бозе–Эйнштейна. Мы имеем в виду идеальный газ, состоящий из фермионов или бозонов, помещенный в сосуде неизменного объема с твердыми, непроницаемыми адиабатическими стенками. Прежде всего

надо решить, как характеризовать макросостояние газа. С этой целью разделим все квантовые состояния частицы на узкие энергетические слои. Каждый слой состоит из квантовых состояний с одинаковыми или очень близкими значениями энергии частицы. Энергии квантовых состояний в i-м слое заключены в интервале $(\varepsilon_i, \varepsilon_i + \delta\varepsilon_i)$. Нет необходимости точно фиксировать ширины слоев $\delta\varepsilon_i$. Достаточно потребовать, чтобы выполнялось условие $\delta\varepsilon_i \ll \varepsilon_i$. Кроме того, число квантовых состояний Z_i в энергетическом слое должно быть очень велико. Макросостояние газа характеризуется заданием чисел частиц N_i в каждом энергетическом слое. Понятно, что любая перестановка частиц в слое не меняет ни микро-, ни макросостояние. Определим теперь число микросостояний, с помощью которых может быть осуществлено рассматриваемое макросостояние газа с фиксированными числами N_i, т. е. статистический вес G этого макросостояния. Число способов, которыми можно распределить N_i частиц по Z_i квантовым состояниям i-го слоя, будет

$$G_i = \frac{Z_i!}{N_i!(Z_i - N_i)!} \quad \text{или} \quad G_i = \frac{(Z_i + N_i - 1)!}{N_i!(Z_i - 1)!}$$

для фермионов и бозонов соответственно. Перемножая все G_i, найдем статистический вес рассматриваемого макросостояния всего газа. Таким образом, для фермионов

$$G = \prod_i \frac{Z_i!}{N_i!(Z_i - N_i)!}, \tag{83.3}$$

а для бозонов

$$G = \prod_i \frac{(Z_i + N_i - 1)!}{N_i!(Z_i - 1)!}. \tag{83.4}$$

Задача состоит в том, чтобы найти наиболее вероятные распределения, обращающие в максимум выражения (83.3) и (83.4) при дополнительных условиях (82.2) и (82.3). Предполагая, что велики не только Z_i, но и все N_i, поступаем так же, как и в статистике Больцмана. Применяя формулу Стирлинга, находим энтропию газа из фермионов и бозонов:

$$S_{\text{ф}} = -k \sum_i [N_i \ln N_i + (Z_i - N_i) \ln(Z_i - N_i)] + \text{const}, \tag{83.5}$$

$$S_{\text{б}} = k \sum_i [(Z_i + N_i - 1) \ln(Z_i + N_i - 1) - N_i \ln N_i] + \text{const}. \tag{83.6}$$

Из условия максимума с учетом (82.2) получаем

$$\sum \ln \frac{N_i}{Z_i - N_i}\, dN_i = 0 \quad \text{(для фермионов)},$$

$$\sum \ln \frac{N_i}{Z_i + N_i - 1}\, dN_i = 0 \quad \text{(для бозонов)}.$$

Эти соотношения отличаются от первого соотношения (82.8) только тем, что вместо $\ln N_i$ в них стоят $\ln \dfrac{N_i}{Z_i - N_i}$ и $\ln \dfrac{N_i}{Z_i + N_i - 1}$. Поэтому по аналогии с (82.9) можно сразу написать

$$\frac{\overline{N}_i}{Z_i - \overline{N}_i} = Ae^{-\alpha\varepsilon_i} \quad \text{(для фермионов)}, \tag{83.7}$$

$$\frac{\overline{N}_i}{Z_i + \overline{N}_i} = Ae^{-\alpha\varepsilon_i} \quad \text{(для бозонов)}, \tag{83.8}$$

причем в последней формуле мы пренебрегли единицей по сравнению с $Z_i + \overline{N}_i$. Постоянная α находится из тех же термодинамических соображений, как и в статистике Больцмана. Она оказывается равной прежнему выражению (82.10). Среднее число частиц $\overline{n}_i$, приходящееся на одно квантовое состояние, равно $\overline{N}_i/Z_i$, т. е.

$$\overline{n}_i = \frac{1}{\exp\dfrac{\varepsilon_i - \mu}{kT} + 1} \quad \text{(для фермионов)}, \tag{83.9}$$

$$\overline{n}_i = \frac{1}{\exp\dfrac{\varepsilon_i - \mu}{kT} - 1} \quad \text{(для бозонов)}. \tag{83.10}$$

Здесь введен новый параметр μ, связанный с A соотношением $A = \exp(\mu/kT)$. Это и есть распределения Ферми–Дирака и Бозе–Эйнштейна.

5. Если $\overline{n}_i \ll 1$, то в знаменателях формул (83.9) и (83.10) можно пренебречь единицами, тогда эти формулы переходят в

$$\overline{n}_i = \exp\frac{\mu - \varepsilon_i}{kT} = \text{const} \cdot \exp\left(-\frac{\varepsilon_i}{kT}\right), \tag{83.11}$$

т. е. в формулу распределения Больцмана. Значит, *распределением Больцмана можно пользоваться лишь тогда, когда малы «числа заполнения» квантовых ячеек, т. е. при условии* $\overline{n}_i \ll 1$. Об этом уже говорилось в конце § 71. Переход статистик Ферми–Дирака и Бозе–Эйнштейна в статистику Больцмана надо понимать в том смысле, что при выполнении условия $\overline{n}_i \ll 1$ формулы (83.9) и (83.10) переходят в больцмановскую формулу (83.11). При этом статистические веса (83.3) и (83.4) отнюдь не переходят в больцмановское выражение (82.1). Однако реальный смысл играют не сами статистические веса, а их логарифмы, точнее, разности логарифмов статистических весов в различных состояниях, определяющие соответствующие приращения энтропии.

6. Для выяснения физического смысла постоянной μ будем пользоваться выражениями для энтропии (83.5) и (83.6). Входящие в них постоянные не существенны, так как они не зависят от числа частиц N. Для конкретности вычислим химический потенциал μ^* ферми-газа. Удобнее всего воспользоваться свободной энергией Ψ. Будем изменять

число частиц в системе N, сохраняя T и V неизменными. Найдем соответствующее приращение свободной энергии $d\Psi$. Поскольку объем V остается неизменным, а частицы между собой не взаимодействуют, то при изменении N энергетические уровни ε_i и соответствующие им числа Z_i меняться не будут. Будут меняться только числа заполнения N_i. Поэтому для приращения энтропии из (82.5) получаем

$$dS = -k \sum_i \ln \frac{N_i}{Z_i - N_i}\, dN_i.$$

В состоянии равновесия имеет место соотношение (83.7), из которого следует $\ln \dfrac{N_i}{Z_i - N_i} = \dfrac{\mu - \varepsilon_i}{kT}$ $\left(\text{так как } A = \exp\left(-\dfrac{\mu}{kT}\right)\right)$. Таким образом,

$$dS = -k \sum_i \frac{\mu - \varepsilon_i}{kT}\, \overline{dN}_i.$$

Сумма $\sum \varepsilon_i\, d\overline{N}_i$ дает приращение внутренней энергии dU. Кроме того, $\sum dN_i = d\sum N_i = dN$, где N — общее число частиц газа. В результате получим

$$T\, dS = -\mu\, dN + dU, \quad \text{или} \quad d\Psi = \mu\, dN.$$

Отсюда

$$\mu^* = \left(\frac{\partial \Psi}{\partial N}\right)_{T,V} = \mu.$$

То же справедливо и в статистике Бозе–Эйнштейна. Мы доказали, таким образом, что в распределениях (83.9) и (83.10) величина μ есть химический потенциал в *термодинамическом смысле*.

7. Химический потенциал μ определяется из условия нормировки

$$\sum Z_i \overline{n}_i = \sum_i \frac{Z_i}{\exp \dfrac{\varepsilon_i - \mu}{kT} \pm 1} = N. \tag{83.12}$$

Очевидно, он зависит от внешних параметров (в нашем случае от объема V), температуры газа T и от числа частиц N. Химический потенциал μ определен с точностью до той же произвольной аддитивной постоянной, что и энергии ε_i. Условимся энергию ε_i наинизшего уровня считать равной нулю. Тогда формулой (83.12) химический потенциал определится однозначно.

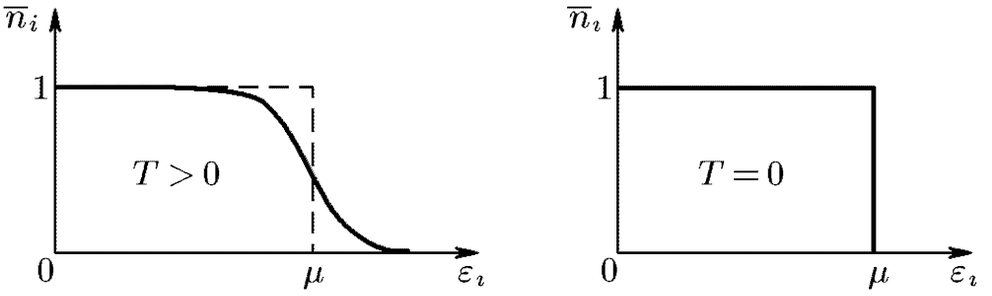
Средние числа заполнения $\overline{n}_i$ по своему смыслу не могут быть отрицательными. Это накладывает определенные ограничения на знак μ в случае бозе-газов (т. е. газов, состоящих из бозонов). Именно, из положительности выражения (82.10) следует $\mu \ll \varepsilon_i$ для всех i. В частности, полагая $i = 1$, получаем $\mu \leqslant 0$. Для бозе-газов, таким образом, химический потенциал отрицателен или равен нулю. Для ферми-газов (т. е. газов, состоящих из фермионов) подобного ограничения не существует. Для газов, подчиняющихся статистике Больцмана, также $\mu < 0$. Действительно, формула (83.11) допускает любой знак для μ.

Однако не надо забывать, что эта формула применима при условии $\overline{n}_i \ll 1$. Из него при $i = 1$ снова получаем $\mu < 0$.

8. На рис. 73 *а* сплошной кривой представлено распределение Ферми–Дирака при $\mu > 0$. Если $T \to 0$, то

$$\overline{n}_i \to \begin{cases} 1 & \text{при} \quad \varepsilon_i < \mu, \\ 1/2 & \text{при} \quad \varepsilon_i = \mu, \\ 0 & \text{при} \quad \varepsilon_i > \mu. \end{cases}$$

Это значит, что при $T = 0$ частицы ферми-газа заполняют все квантовые состояния с энергиями $\varepsilon_i < \mu$. Квантовые состояния с более высокими энергиями не заполнены. Говорят, что при $T = 0$ ферми-



а *б*

Рис. 73

газ находится в состоянии *полного вырождения*. Кривая, изображающая соответствующее распределение, вырождается в прямоугольник (рис. 73 *б*). Изображать на том же рисунке распределения больцмановского и бозе-газов не имеет смысла, так как для этих газов $\mu < 0$. Сравнение распределений Ферми–Дирака, Бозе–Эйнштейна и Больцмана показано на отдельном рис. 74.

Интересный характер вырождения при $T \to 0$ имеет бозе-газ. Химический потенциал бозе-газа при $T = 0$ должен обращаться в нуль. Для доказательства предположим, что при $T = 0$ химический потенциал в нуль не обращается ($\mu < 0$). Тогда при любых i (включая $i = 0$) разности $\varepsilon_i - \mu$ были бы положительными. Значит, при $T \to 0$ знаменатель в формуле (83.10) стремился бы к ∞. При $T = 0$ все $\overline{n}_i$ обратились бы в нуль. И так было бы при любом числе частиц N, что очевидно невозможно. Но этого не произойдет, если при $T = 0$ химический потенциал μ также будет равен нулю. Тогда обратятся в нуль только те $\overline{n}_i$, у которых $i \neq 0$ (так как в этом случае $\varepsilon_i - \mu = \varepsilon_i > 0$). Для числа же частиц на энергетическом уровне $\varepsilon_0 = 0$ при $T = 0$ формула (83.10) приводит к неопределенному выражению $\overline{n}_0 = 1/(e^{0/0} - 1)$. Очевидно, что при

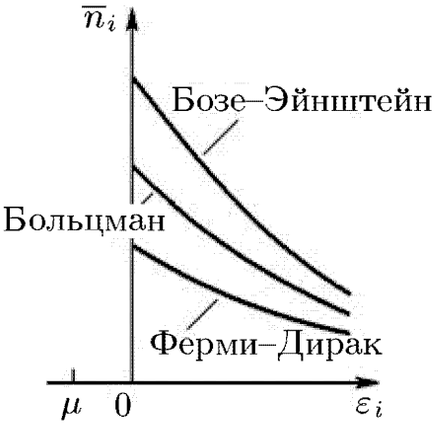


Рис. 74

приближении к абсолютному нулю числа частиц на квантовых уровнях ε_i $(i \neq 0)$ будут стремиться к нулю. Исключение составляют только частицы на нижнем квантовом уровне $\varepsilon_0 = 0$. *При приближении к абсолютному нулю бозе-частицы все более и более будут накапливаться на нижнем квантовом уровне* $\varepsilon_1 = 0$ *и, наконец, все они окажутся на нем при* $T = 0$. Это явление получило название *бозе-эйнштейновской конденсации*. Разумеется, такая «конденсация» не имеет ничего общего с конденсацией пара в жидкость.

Для применения распределений (82.9) и (82.10) к конкретным вопросам надо знать выражения для энергий ε_i и соответствующих им чисел квантовых состояний Z_i. Это будет сделано в надлежащих местах нашего курса при изучении конкретных явлений.

§ 84. Теорема Нернста

1. В 1906 г. термодинамика обогатилась новым фундаментальным законом, открытым Нернстом (1864–1941) эмпирическим путем. Этот закон получил название *тепловой теоремы Нернста*. Теорема Нернста не может быть логически выведена из остальных начал термодинамики, а потому ее часто называют *третьим началом термодинамики*. Мы не будем останавливаться на первоначальной формулировке теоремы, данной самим Нернстом. Она представляет только исторический интерес. Приведем сразу современную расширенную формулировку теоремы, принадлежащую в основном Планку.

Содержание теоремы Нернста сводится к двум утверждениям. Первое утверждение состоит в том, что *при приближении к абсолютному нулю энтропия стремится к определенному конечному пределу*. Поэтому имеет смысл говорить об *энтропии тела при абсолютном нуле температур*. Нетривиальность этого утверждения станет очевидной, если обратиться к термодинамическому определению энтропии

$$S - S_0 = \int_{T_0}^{T} \frac{\delta Q}{T}. \tag{84.1}$$

(Напомним, что интеграл берется по произвольному пути, квазистатически переводящему систему из начального (нулевого) состояния в конечное.) В подынтегральном выражении температура T стоит в знаменателе. Поэтому не очевидно, будет интеграл сходиться или нет при $T \to 0$. Все зависит от поведения δQ вблизи абсолютного нуля. Первая часть теоремы Нернста как раз и состоит в утверждении, что *интеграл сходится*.

Вторая часть теоремы Нернста утверждает, что *все процессы при абсолютном нуле температур, переводящие систему из одного равновесного состояния в другое равновесное состояние, происходят без изменения энтропии*. Из этого утверждения следует, что предел, к которому стремится интеграл (84.1) при $T \to 0$, не зависит от того, в каком конечном состоянии окажется система.

Объединяя обе части вместе, можно дать теореме Нернста следующую формулировку. *При приближении к абсолютному нулю приращение энтропии* $S - S_0$ *стремится к вполне определенному конечному пределу, не зависящему от значений, которые принимают все параметры, характеризующие состояние системы* (например, от объема, давления, агрегатного состояния и пр.).

Теорема Нернста относится только к термодинамически равновесным состояниям систем. К неравновесным и метастабильным состояниям она неприменима. Примером могут служить аморфные твердые тела, например стекла, представляющие собой термодинамически неравновесные системы. Однако они переходят в термодинамически равновесное (кристаллическое) состояние крайне медленно — годами и даже в течение десятков и сотен лет. В этом смысле они и являются метастабильными. Когда говорят о неприменимости к ним теоремы Нернста, то имеют в виду, что эти тела находятся в термостате, охлажденном до абсолютного нуля, и пришли в температурное равновесие с ним. Строго говоря, к неравновесным, в частности, метастабильным состояниям понятие температуры неприменимо. При абсолютном нуле находится термостат, а не само тело. Если бы температура понималась только в таком строго термодинамическом смысле, то указание на неравновесность и метастабильность состояния при формулировке теоремы Нернста было бы излишним.

Если *условиться* энтропию всякой равновесной системы при абсолютном нуле температур считать равной нулю, то всякая неоднозначность в определении энтропии исчезнет. Энтропия, определенная таким образом, называется *абсолютной энтропией*. Теорема Нернста может быть, следовательно, сформулирована следующим образом. *При приближении к абсолютному нулю абсолютная энтропия системы стремится также к нулю независимо от того, какие значения принимают при этом все параметры, характеризующие состояние системы*. Следует, однако, подчеркнуть, что, вопреки распространенному мнению, такой выбор аддитивной постоянной в выражении для энтропии есть не более как *произвольное соглашение*. Энтропия по своей сущности всегда определена с точностью до произвольной аддитивной постоянной. Фактическое содержание теоремы Нернста никак не связано с выбором этой постоянной, а целиком сводится к тем двум утверждениям, которые были сформулированы выше.

2. Абсолютный нуль недостижим, поэтому о справедливости тепловой теоремы Нернста можно судить только по поведению вещества вблизи абсолютного нуля температур. Рассмотрим некоторые следствия теоремы Нернста, подтверждающие эту теорему.

Из первой части теоремы Нернста следует, что при приближении к абсолютному нулю теплоемкости C_P и C_V всех тел должны стремиться к нулю. В самом деле, допустим, например, что давление остается постоянным. Тогда $\delta Q = C_P(T')\,dT'$. Теорема Нернста требует, чтобы

интеграл

$$\int_0^T \frac{\delta Q}{T'} = \int_0^T \frac{C_P(T')}{T'}\, dT'$$

сходился. Но это было бы невозможно, если бы при $T = 0$ теплоемкость C_P не обращалась в нуль. В противном случае можно было бы найти температурный интервал $0 \leqslant T' \leqslant T$, в котором теплоемкость C_P всюду отлична от нуля и, следовательно, сохраняет знак. Пусть C — минимальное значение модуля C_P в этом интервале. Тогда предыдущий интеграл по модулю был бы не меньше интеграла

$$\int_0^T \frac{C}{T'}\, dT' = C \int_0^T \frac{dT'}{T'},$$

который логарифмически расходится. Следовательно, должен был бы расходиться и исходный интеграл $\int \frac{C_P}{T'}\, dT'$, а это противоречит теореме Нернста. Получившееся противоречие и доказывает наше утверждение относительно C_P. Аналогично доказывается, что теплоемкость C_V должна вести себя так же.

Полученные следствия доказывают, что теплоемкости вещества обязательно должны зависеть от температуры. Классическая теория теплоемкостей приводит к противоположному выводу и независимости теплоемкости от температуры (см. § 66 и 68). Поэтому *теорема Нернста не может быть истолкована классически.*

3. Обратимся теперь к следствиям из второй части теоремы Нернста. Для этого воспользуемся условиями, что выражения (45.7) и (45.6) являются полными дифференциалами:

$$\left(\frac{\partial S}{\partial P}\right)_T = -\left(\frac{\partial V}{\partial T}\right)_P, \quad \left(\frac{\partial S}{\partial V}\right)_T = \left(\frac{\partial P}{\partial T}\right)_V.$$

Из теоремы Нернста следует, что при абсолютном нуле левые части этих соотношений обращаются в нуль. Должны быть равны нулю и правые части, а потому

$$\frac{1}{V_0}\left(\frac{\partial V}{\partial T}\right)_P \to 0, \quad \frac{1}{P_0}\left(\frac{\partial P}{\partial T}\right)_V \to 0. \tag{84.2}$$

Это значит, что *при приближении к абсолютному нулю для всех тел должны стремиться к нулю температурный коэффициент объемного расширения и температурный коэффициент давления.*

Однако из уравнения Клапейрона следует, что оба коэффициента (84.2) должны оставаться постоянными вплоть до абсолютного нуля. Это противоречит теореме Нернста. Отсюда следует, что при очень низких температурах уравнение Клапейрона перестает выполняться даже в тех случаях, когда силы взаимодействия между молекулами газа сколь угодно малы. Далее, из второй формулы (84.2) видно, что

вблизи абсолютного нуля давление газа практически не зависит от температуры, а является функцией одной только плотности. Если это имеет место, то говорят, что *газ находится в состоянии вырождения*, а сами газы называют *вырожденными*. Примером вырожденного газа могут служить свободные электроны в металлах уже при обычных температурах (см. § 71). Статистика Больцмана к вырожденным газам неприменима. Эти газы подчиняются либо статистике Ферми–Дирака, либо статистике Бозе–Эйнштейна, в зависимости от того, состоят ли они из фермионов или бозонов.

Внутренняя энергия газа в состоянии вырождения практически не зависит от температуры, а только от его плотности. Действительно, согласно термодинамике

$$\left(\frac{\partial U}{\partial V}\right)_T = T\left(\frac{\partial P}{\partial T}\right)_V - P.$$

Значения T и $(\partial P/\partial T)_V$ стремятся к нулю при $T \to 0$. Давление P вырожденного газа, как мы видели, от температуры практически не зависит. Поэтому производная $(\partial U/\partial V)_T$, а с ней и внутренняя энергия U становятся функциями одной только плотности. С этим и связано то обстоятельство, что электронный газ в металлах при обычных температурах не вносит сколько-нибудь заметного вклада в теплоемкость.

4. Теорему Нернста можно объяснить, если обратиться к вероятностной интерпретации энтропии с помощью формулы Больцмана (80.5). Для этого надо найти статистический вес состояния системы при абсолютном нуле. Классический подход здесь не годится. Он наверняка привел бы к противоречию с теоремой Нернста. Это связано с тем, что классическая механика, даже при абсолютном нуле температур, допускает непрерывное множество динамических состояний системы. Нужен квантовый подход. Будем понимать под квантовым состоянием состояние системы в целом. Саму систему будем предполагать замкнутой. При абсолютном нуле температур энергия системы минимальна. Число допустимых квантовых состояний системы при этом равно либо единице, если уровень минимальной энергии не вырожден, либо какому-то целому числу, равному кратности вырождения, если этот уровень вырожден. Тем же числом выражается и статистический вес состояния. Поэтому для энтропии по формуле Больцмана (80.5) получается *конечное значение*. Таково объяснение первой части теоремы Нернста.

Дадим теперь объяснение второй части этой теоремы. При изменении внешних параметров, например объема или давления системы, квантовое состояние и энергия системы в этом состоянии изменяются. Кратные уровни могут частично или полностью расщепляться на простые уровни.

Простые уровни могут соединяться в один кратный уровень. Однако *общее число простых уровней остается неизменным*. Система, если она находится в термодинамическом равновесии, при абсолютном

нуле температур занимает самый низкий энергетический уровень, все прочие уровни для нее недоступны. Если при изменении внешних параметров кратность этого уровня не изменяется, то остается неизменным статистический вес, а с ним и энтропия системы, как того требует вторая часть теоремы Нернста. Если же кратность нулевого уровня изменится, то произойдет некоторое изменение и энтропии. Однако это изменение ничтожно, и им можно пренебречь.

Например, если вместо простого уровня появится двойной, то статистический вес возрастет в 2 раза, а энтропия получит приращение $\Delta S = k \ln 2$. Оно ничтожно ввиду малости постоянной Больцмана k. Даже если бы кратность нулевого уровня изменилась в 10^{20} раз, то изменение энтропии составляло бы всего $\Delta S = 20k \ln 10 \approx 46k$. Это — также ничтожная величина. По мере возрастания температуры система переходит на высшие энергетические уровни. Статистический вес макросостояния резко возрастает. Начинает возрастать и энтропия.

ЗАДАЧА

Найти статистические веса ферми- и бозе-газов при абсолютном нуле температур. Убедиться, что эти газы удовлетворяют теореме Нернста.

Р е ш е н и е. Пусть $T = 0$. В случае ферми-газов можно, указать энергетический уровень (с номером $i = n$), обладающий следующими свойствами. Энергетические уровни с номерами $i < n$ заполнены целиком, уровни с номерами $i > n$ свободны. Уровень с номером $i = n$ заполнен частично (или в частном случае полностью). Для заполненных уровней $N_i = Z_i$, для незаполненных $N_i = 0$. В обоих случаях все множители в произведении (82.3), за исключением n-го, равны единице. Множитель с номером n отличен от единицы, если соответствующий ему уровень заполнен частично. Итак,

$$G_{\text{Ф–Д}} = \frac{Z_n!}{N_n!(Z_n - N_N)!}. \tag{84.3}$$

Аналогично для бозе-газа

$$G_{\text{Б–Э}} = \frac{(Z_1 + N - 1)!}{N!(Z_1 - 1)!}. \tag{84.4}$$

Если изменяется объем газа, то меняется энергия энергетических уровней. Однако числа Z_1, Z_n, N_n, а также общее число частиц N остаются неизменными. Остаются неизменными статистические веса и энтропии газов. К тому же заключению можно прийти непосредственно на основании формул (83.3) и (83.4), не преобразуя их к виду (84.3) и (84.4). Квантовый (но не классический) больцмановский газ также удовлетворяет теореме Нернста. Однако это замечание имеет чисто формальный характер, так как при абсолютном нуле температур статистка Больцмана неприменима.

§ 85. Квантовая теория теплоемкостей Эйнштейна

1. Квантовая теория в принципе устранила трудности, на которые натолкнулась классическая теория в вопросе о теплоемкости тел. Качественно этот вопрос уже был рассмотрен в § 69. Теперь рассмотрим его количественно. Будем представлять тело как систему N молекул,

слабо взаимодействующих друг с другом. Применим к ней закон распределения Больцмана (81.13), предполагая, что энергетические уровни дискретны. Средняя энергия, приходящаяся на одну молекулу в состоянии термодинамического равновесия, определяется выражением

$$\overline{\varepsilon} = \frac{1}{N} \sum N_i \varepsilon_i.$$

С учетом формулы (81.13) и условия нормировки $\sum N_i = N_0 \times \times \sum g_i e^{-\alpha\varepsilon_i} = N$ получим

$$\overline{\varepsilon} = \frac{\sum \varepsilon_i g_i e^{-\alpha\varepsilon_i}}{\sum g_i e^{-\alpha\varepsilon_i}},$$

или

$$\overline{\varepsilon} = -\frac{1}{Z}\frac{dZ}{d\alpha} = -\frac{d}{d\alpha}(\ln Z), \tag{85.1}$$

где введено обозначение

$$Z = \sum g_i e^{-\alpha\varepsilon_i} = \sum g_i e^{-\varepsilon_i/kT}. \tag{85.2}$$

Выражение (85.2) называется *статистической суммой* или *суммой состояний* и играет важную роль в статистических исследованиях.

2. В качестве примера рассмотрим систему одномерных гармонических осцилляторов. Уровни энергии гармонического осциллятора простые и определяются формулой

$$\varepsilon_i = (i + 1/2)h\nu \tag{85.3}$$

(см. § 69). Для суммы состояний получаем

$$Z = e^{-\alpha h\nu/2} \sum_{i=0}^{\infty} e^{-i\alpha h\nu} = \frac{e^{-\alpha h\nu/2}}{1 - e^{-\alpha h\nu}},$$

а для средней энергии осциллятора

$$\overline{\varepsilon} = -\frac{d}{d\alpha}(\ln Z) = \frac{h\nu}{2} + \frac{h\nu}{e^{h\nu/kT} - 1}. \tag{85.4}$$

В последнем слагаемом мы заменили α на $1/kT$.

Слагаемое $h\nu/2$ есть *нулевая энергия* гармонического осциллятора. Она не зависит от температуры и не имеет отношения к тепловому движению. В теории теплоемкости тел ее можно опустить. Если это сделать, то получится

$$\overline{\varepsilon} = \frac{h\nu}{e^{h\nu/kT} - 1}. \tag{85.5}$$

Эта формула впервые была получена Планком в 1900 г. в его исследованиях по теории теплового излучения. Если $h\nu \ll kT$, что имеет место при высоких температурах, то $e^{h\nu/kT} \approx 1 + \frac{h\nu}{kT}$. В этом

приближении (85.5) переходит в классическую формулу

$$\overline{\varepsilon} = kT. \tag{85.6}$$

Такой результат довольно очевиден, так как при $kT \gg h\nu$ возбуждено очень много энергетических уровней, и их дискретность становится несущественной.

3. Формула (85.5) была положена Эйнштейном в основу квантовой теории теплоемкости твердых тел. Он пользовался той же моделью твердого тела, какая применялась в классической теории. Атомы кристаллической решетки рассматривались как гармонические осцилляторы, совершающие тепловые колебания около положений равновесия с одной и той же частотой ν. Осцилляторы брались трехмерными, т. е. обладали тремя степенями свободы. На каждую степень свободы приходится средняя энергия тепловых колебаний $\overline{\varepsilon}$, а на один атом — $3\overline{\varepsilon}$. Внутренняя энергия одного моля определяется выражением

$$U = 3N\overline{\varepsilon} = \frac{3Nh\nu}{e^{h\nu/kT} - 1}, \tag{85.7}$$

где N_A — постоянная Авогадро. Отсюда получаем для молярной теплоемкости кристаллической решетки твердых тел

$$C_V = \frac{dU}{dT} = \frac{3R(h\nu/kT)^2}{[e^{h\nu/kT} - 1]^2}\, e^{h\nu/kT}. \tag{85.8}$$

Это и есть *формула Эйнштейна.* При высоких температурах, когда $h\nu/kT \ll 1$, она переходит в классическую формулу

$$C_V = 3R.$$

В другом предельном случае низких температур, когда $h\nu/kT \gg 1$, можно пренебречь единицей в знаменателе и получить

$$C_V = 3R\left(\frac{h\nu}{kT}\right)^2 e^{-h\nu/kT}. \tag{85.9}$$

При $T \to 0$ выражения (85.8) и (85.9) стремятся к нулю, как это требует тепловая теорема Нернста.

4. Впрочем, согласие с опытом носит только качественный характер. По формулам (84.8) и (84.9) при $T \to 0$ теплоемкость C_V слишком быстро стремится к нулю — приблизительно экспоненциально. Опыт показывает, что в действительности приближение теплоемкости к нулю идет по степенному закону, т. е. более медленно. При других температурах формула Эйнштейна также находится только в качественном, но не в количественном согласии с опытом. Однако эти расхождения связаны не с существом квантовой теории, а с упрощением расчета, в котором предполагается, что все гармонические осцилляторы колеблются с одной и той же частотой. На самом деле кристаллическую решетку следует рассматривать как связанную систему взаимодействующих частиц. Малые колебания такой системы получаются в результате наложения многих гармонических колебаний

с различными частотами. Число частот очень велико — порядка числа степеней свободы системы. При вычислении теплоемкости тело можно рассматривать как систему гармонических осцилляторов, но с различными частотами. Задача сводится к вычислению этих частот, т. е. отысканию так называемого *спектра частот*. На это было указано уже самим Эйнштейном.

Задача о спектре частот кристаллической решетки твердого тела рассматривалась Дебаем (1884–1966), а затем Борном (1882–1971) и Карманом (1881–1963). Борн и Карман подошли к решению задачи с последовательно атомистической точки зрения. Это очень трудно. Дебай сильно упростил задачу. *При низких температурах основной вклад в теплоемкость вносят низкочастотные колебания, которым соответствуют малые кванты энергии.* Практически только такие низкочастотные тепловые колебания и возбуждены при низких температурах. Колебания с более высокими частотами, которым соответствуют большие кванты энергии, практически не возбуждены. Но низкочастотный спектр тепловых колебаний твердого тела может быть с достаточной точностью рассчитан методами *механики сплошной среды*, отвлекаясь от атомистической структуры тела. Тогда вычисления становятся довольно простыми. Таким путем Дебай построил простую теорию теплоемкости твердых тел, особенно хорошо согласующуюся с опытом при низких температурах. Согласно этой теории *вблизи абсолютного нуля теплоемкость кристаллической решетки твердого тела пропорциональна кубу абсолютной температуры.* Этот результат называется *законом кубов Дебая.* Мы изложим теорию Дебая в т. V (см. § 54).

Теория Эйнштейна, разумеется, применима и к *колебательной теплоемкости* двухатомных или многоатомных газов (см. § 69). Совершенно аналогично может быть построена и теория *вращательной теплоемкости.* Вычисления здесь несколько сложнее из-за более сложной структуры энергетического спектра (см. т. V, § 31, п. 8 и § 51).

5. В металлах, помимо кристаллической решетки, построенной из ионов, имеются еще *свободные электроны.* В простейшей модели их рассматривают как *идеальный электронный газ.* Дело в том, что электрические силы притяжения, действующие на электроны со стороны положительно заряженных ионов, в среднем компенсируются силами отталкивания, действующими со стороны самих электронов. В этой модели вся энергия электронного газа только кинетическая. Поэтому для него применимо уравнение кинетической теории газов (59.8), т. е.

$$PV = \frac{2}{3}\overline{E}. \tag{85.10}$$

Но, как мы уже указывали в § 69, электроны практически не вносят никакого вклада в теплоемкость. Формальное объяснение этого было уже дано в § 84 с помощью теоремы Нернста. Приведем теперь более подробное молекулярно-кинетическое (точнее, статистическое) объяснение. Электронный газ в металлах всегда вырожден, так как

температуры вырождения для всех металлов составляют десятки тысяч градусов (см. § 71). При абсолютном нуле распределение электронов по квантовым состояниям представляется прямоугольником (см. рис. 73 *б*). Электроны совершают весьма интенсивное квантованное движение, но совершенно не участвуют в беспорядочном тепловом движении. Энергия этого квантованного движения E и давление P определяются только концентрацией электронов. Это приближенно справедливо и в том случае, когда температура T отлична от нуля, так как для подавляющего большинства электронов распределение носит тот же характер, что и на рис. 73 *а*. Следовательно, эти электроны по-прежнему не участвуют в тепловом движении и не влияют на теплоемкость. «Прямоугольное» распределение электронов нарушается только внутри очень тонкого энергетического слоя вблизи границы $\varepsilon = \mu$. Толщина этого слоя порядка энергии теплового движения kT. Только эти приграничные электроны и участвуют в тепловом движении. Они-то и вносят дополнительный вклад в энергию E и давление P, зависящий от температуры. За счет этих электронов и появляется теплоемкость электронного газа. Но так как пограничных электронов очень мало, то этот вклад также мал. Расчет показывает, что *теплоемкость электронного газа линейно зависит от температуры*, т. е. имеет вид

$$c_{\text{эл}} = \gamma T,$$

где γ — постоянная.

Аналогично можно показать, что и теплоемкость бозе-газа в состоянии вырождения мала и стремится к нулю при $T \to 0$. Особенность в этом случае состоит в том, что при $T = 0$ все частицы газа накапливаются на самом низком уровне с энергией, равной нулю. Поэтому при $T = 0$ не только кинетическая энергия, но и давление бозе-газа обращаются в нуль.

Глава VII

ЯВЛЕНИЯ ПЕРЕНОСА В ГАЗАХ

§ 86. Средняя длина свободного пробега

1. Средняя скорость теплового движения газовых молекул определяется формулой (73.6). Уже при комнатной температуре она порядка скорости ружейной пули. Например, при 0 °C для молекул водорода, азота и кислорода $\overline{v}$ равна соответственно 1700 м/с, 455 м/с и 425 м/с. На ранней стадии развития кинетической теории газов столь большие значения скоростей молекул некоторым физикам казались невозможными. Если скорости молекул действительно так велики — говорили они, — то запах пахучего вещества должен был бы распространяться от одного конца комнаты к другому практически мгновенно. На самом деле при отсутствии конвективных потоков воздуха время распространения запаха на такие расстояния может составлять многие минуты и даже часы. Распространение запаха осуществляется посредством медленного *процесса диффузии*.

Для демонстрации медленности диффузии газов можно взять стеклянный цилиндр высоты примерно 30 см, закрытый сверху проволочной сеткой. Вдоль оси цилиндра пропущена тонкая стеклянная палочка или трубка, к которой на равных расстояниях прикреплено около 10 полосок фильтровальной бумаги, смоченных в фенолфталеине. На сетку сверху кладется вата, смоченная нашатырным спиртом. Выделяющийся аммиак диффундирует вниз. Диффузия наблюдается по покраснению полосок фильтрованной бумаги. Через 1–2 минуты начинается покраснение верхней полоски. Нижняя полоска начинает краснеть минут через 20. Аммиак легче воздуха, поэтому его проникновение вниз происходит только в результате диффузии. Стеклянный цилиндр служит для предотвращения возникновения потоков воздуха.

Значительно медленнее происходит диффузия в жидкостях, хотя скорости теплового движения здесь такие же, что и в газах и в твердых телах. Если узкий и высокий стеклянный цилиндр наполнить дистиллированной водой, а затем на дно с помощью специальной трубки осторожно опустить кристаллы медного купороса, то последние растворяться, и начнется диффузия. Чтобы ее заметить на глаз, нужны сутки или несколько суток. А для того чтобы получился однородный раствор по всей высоте цилиндра, требуется несколько месяцев. В твердых телах диффузия происходит еще медленнее, и требуются специальные методы, чтобы ее обнаружить.

Так же медленно происходит выравнивание температур между различными частями неравномерно нагретого газа посредством теплопроводности или выравнивание скоростей макроскопического движения газа посредством сил вязкости.

2. Медленность диффузии и аналогичных ей явлений Клаузиус объяснил *столкновениями молекул*. Молекула газа не все время движется свободно, а время от времени испытывает столкновения с

другими молекулами. Свободно она пролетает короткое расстояние от одного столкновения до следующего. В момент столкновения скорость молекулы испытывает резкое изменение как по модулю, так и по направлению. В результате траектория молекулы получается не прямой, а ломаной линией с большим количеством звеньев. Молекула беспорядочно мечется туда и сюда, и ее общее продвижение вперед происходит сравнительно медленно. Для количественного описания явления Клаузиус ввел понятие *средней длины свободного пробега*, т. е. среднего расстояния, которое пролетает молекула от одного столкновения до следующего.

Для вычисления средней длины свободного пробега будем пользоваться *моделью твердых шаров*. Между столкновениями молекулы шары движутся по инерции прямолинейно и равномерно. В моменты столкновений между молекулами развиваются очень большие силы отталкивания, изменяющие их скорости по величине и направлению. Разумеется, такая грубая модель передает далеко не все черты явлений, которые происходят при столкновениях. Молекулы могут распадаться и соединяться. Атомы могут ионизоваться, переходить в возбужденные состояния и т. д. Все это оставим сейчас без внимания. Модель твердых шаров может приблизительно правильно описать только *процессы рассеяния* молекул, в которых происходят изменения скорости и направления движения этих частиц в результате столкновений их между собой и со стенками сосуда, в котором заключен газ.

Для упрощения расчета предположим, что движется только одна молекула с постоянной скоростью v, а все остальные молекулы неподвижны. Будем называть движущуюся молекулу молекулой A. Вообразим, что с молекулой A жестко связана концентрическая с ней твердая сфера S вдвое большего диаметра. Назовем эту сферу *сферой ограждения* молекулы A. В момент столкновения расстояние между центрами сталкивающихся молекул равно диаметру молекулы d. Следовательно, в этот момент центр неподвижной молекулы, с которой столкнулась молекула A, окажется на поверхности сферы ограждения последней. Очевидно, он не может проникнуть внутрь этой сферы. Между двумя последовательными столкновениями молекулы A ее сфера ограждения описывает цилиндр, длина которого и есть свободный пробег молекулы A. Из таких цилиндров складывается поверхность, описываемая с течением времени сферой ограждения (рис. 75). Для краткости будем называть эту поверхность ломаным цилиндром. Если центр другой молекулы лежит внутри или на боковой поверхности этого цилиндра, то она столкнется с молекулой A. В противном случае столкновения не произойдет.

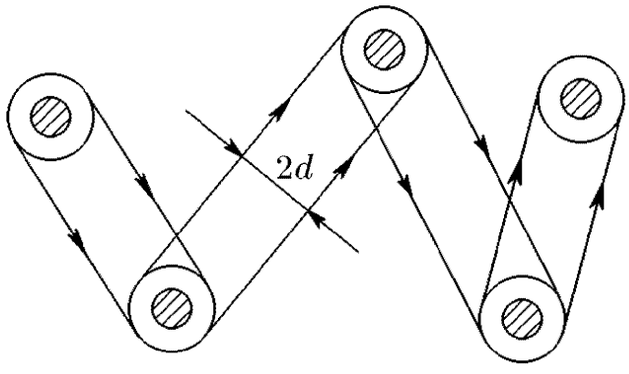


Рис. 75

Пусть V — объем ломаного цилиндра, описываемого сферой S в единицу времени. Среднее число z столкновений движущейся молекулы с остальными молекулами в единицу времени равно среднему числу последних в объеме V, т. е. $z = Vn$, где n — число молекул в единице объема. Мы предполагаем, что средняя длина свободного пробега λ очень велика по сравнению с диаметром сферы ограждения $2d$. Тогда можно пренебречь теми частями объема V, которые приходятся на изломы цилиндра, т. е. при вычислении V цилиндр можно считать прямым, а его высоту равной скорости молекулы v. В этом приближении $V = \sigma v$, где $\sigma = \pi d^2$ — площадь поперечного сечения цилиндра. Следовательно,

$$z = n\sigma v. \tag{86.1}$$

Путь, пройденный молекулой A за единицу времени, равен v. Разделив его на среднее число столкновений z, найдем среднюю длину свободного пробега молекулы:

$$\lambda = 1/n\sigma. \tag{86.2}$$

Из вывода следует, что при получении формул (86.1) и (86.2) можно рассуждать так, как если бы все молекулы, с которыми сталкивается молекула A, были точечными, а радиус молекулы A увеличен вдвое, т. е. молекула A заменена ее сферой ограждения. Такая замена может рассматриваться как вычисленный прием для учета конечных размеров молекул, сталкивающихся с молекулой A. Этот прием будет использован в следующем параграфе при введении понятия эффективного сечения.

Конечно, формулы (86.1) и (86.2) не точны, поскольку в основу их вывода положено предположение, что движется только одна молекула, а все остальные неподвижны. Математически строгий расчет был дан Максвеллом с учетом максвелловского распределения молекул по скоростям. Но в физике уточнение Максвелла мало существенно, так как его расчет выполнен не для реальных молекул, а для модели твердых упругих шаров. Максвелл получил:

$$z = \sqrt{2}\, n\sigma\overline{v} = 1{,}41\, n\sigma\overline{v}, \tag{86.3}$$

$$\lambda = \frac{1}{\sqrt{2}n\sigma} = \frac{0{,}707}{n\sigma}. \tag{86.4}$$

Выражения (86.3) и (86.4) отличаются от приближенных формул (86.1) и (86.2) только числовыми коэффициентами, близкими к единице. Это несущественно во всех расчетах, которые сами проводятся с точностью до числовых коэффициентов. Такое положение имеет место в излагаемой ниже теории *явлений переноса — диффузии, внутреннего трения* и *теплопроводности*. Ввиду сложности точной теории этих явлений приходится довольствоваться приближенными расчетами, часто довольно грубыми. В таких расчетах несущественно сохранять числовые множители $\sqrt{2}$ и $1/\sqrt{2}$, что обычно и делается. Упрощенные формулы (86.1) и (86.2) дают не только правильные порядки величин, но, что особенно важно, приводят к верной зависи-

мости числа столкновений и длины свободного пробега от концентрации и размеров молекул.

3. Сам Максвелл получил выражения (86.3) и (86.4) в результате довольно кропотливых и сложных вычислений. Между тем их можно получить из формул (86.1) и (86.2) путем весьма простых рассуждений почти без вычислений. Появление множителя $\sqrt{2}$ становится при этом особенно ясным. Приведем соответствующий вывод.

При рассмотрении процесса столкновения играет роль не абсолютная скорость выделенной молекулы A, а ее скорость *относительно молекулы, с которой она сталкивается*. Выделим мысленно группу молекул, которые движутся относительно молекулы A с одной и той же относительной скоростью $v_{i\,\text{отн}}$. Пусть n_i — число таких молекул в единице объема. Число столкновений z_i молекулы A с молекулами выделенной группы в единицу времени можно найти по формуле (86.1), которая дает $z_i = n_i \sigma v_{i\,\text{отн}}$. Полное число столкновений молекулы A со всеми остальными молекулами найдется суммированием этого выражения по всем скоростным группам, т. е. по всем возможным значениям индекса i:

$$z = \sum_i n_i \sigma v_{i\,\text{отн}}.$$

Введя среднюю относительную скорость

$$\overline{v}_{\text{отн}} = \frac{1}{n} \sum n_i v_{i\,\text{отн}},$$

получим

$$z = n \sigma \overline{v}_{\text{отн}}, \tag{86.5}$$

и, следовательно,

$$\lambda = \frac{\overline{v}}{\overline{v}_{\text{отн}}} \frac{1}{n\sigma}. \tag{86.6}$$

Задача свелась к вычислению средней относительной скорости $\overline{v}_{\text{отн}}$ какой-либо молекулы относительно всех остальных молекул газа.

4. Для решения этой задачи дадим другую интерпретацию максвелловского закона распределения скоростей. В прежней интерпретации *закон Максвелла давал распределение скоростей всех молекул газа в один и тот же момент времени*. Но на него можно смотреть как на *закон распределения скоростей одной и той же молекулы* (например, молекулы A), которые она последовательно принимает в различные моменты времени. Воспользуемся следующей интерпретацией. Пусть $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N$ — скорости, принимаемые молекулой A непосредственно после первого, второго и последующих столкновений. Если число N стремится к бесконечности, то эти скорости распределятся по закону Максвелла. Это непосредственно следует из равноправия всех молекул и хаотичности молекулярного теплового движения. В моменты столкновений на молекулу A действуют беспорядочно меняющиеся силы $\mathbf{F}_1, \mathbf{F}_2, \ldots$ Они-то и приводят к установлению максвелловского распределения скоростей молекул A в рассматриваемые моменты времени. Под $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N$ мы понимаем скорости относительно системы отсчета, в которой газ как целое покоится. Введем теперь скорости молекулы A относительно остальных молекул, которыми она обладала в промежутках между последовательными столкновениями. Пусть $\mathbf{v}_{1\,\text{отн}}$ означает скорость молекулы A после первого столкновения *относительно молекулы, с которой произошло это столкновение*, $\mathbf{v}_{2\,\text{отн}}$ — скорость после второго столкновения *относительно молекулы, с которой произошло это*

второе столкновение, и т. д. Как известно из механики, при рассмотрении относительного движения двух частиц одну из них можно считать неподвижной. Относительное движение второй частицы (например, частицы A) формально описывается уравнением Ньютона, как в неподвижной системе. При этом силы $\mathbf{F}_1$, $\mathbf{F}_2$, ... остаются прежними, но масса частицы A должна быть заменена приведенной массой. Если молекулы одинаковы ($m_1 = m_2 = m$), то приведенная масса равна $m_1 m_2/(m_1 + m_2) = m/2$. Таким образом, при относительном движении все происходит так, как если бы масса молекулы уменьшилась в два раза. Так как силы $\mathbf{F}_1, \mathbf{F}_2, \ldots$ и моменты времени их действия остались прежними, то в эти моменты относительные ускорения молекулы A будут вдвое превосходить ее же ускорение в неподвижной системе отсчета. Отсюда непосредственно следует, что *распределение относительных скоростей молекулы будет максвелловским*. А так как эффективная масса молекулы в относительном движении вдвое меньше m, то все средние относительные скорости окажутся больше соответствующих абсолютных скоростей в $\sqrt{2}$ раз. В частности, $\overline{v}_{\text{отн}} = \sqrt{2}\,\overline{v}$, и формулы (86.5) и (86.6) переходят в (86.3) и (86.4).

5. Рассмотрим теперь более важный случай, когда сталкивающиеся молекулы различны. Пусть одна молекула сорта 1 с массой m_1 и радиусом r_1 движется в среде молекул сорта 2 с массами m_2, радиусами r_2 и концентрацией n_2. Если бы молекулы сорта 2 были неподвижны, то остались бы справедливыми прежние формулы (86.1) и (86.2). В них надо было бы только заменить z на $\overline{z_{12}}$, n — на n_2, v — на $\overline{v}_1$, σ на $\sigma_{12} = \pi(r_1 + r_2)^2$. Сферой ограждения молекулы 1 теперь является концентрическая с ней сфера радиуса $d = r_1 + r_2$. Учтем теперь максвелловское распределение скоростей, используя формулы (86.5) и (86.6). На основании изложенного выше средняя относительная скорость $\overline{v}_{\text{отн}}$ и средняя скорость молекулы 1 $\overline{v}_1$ обратно пропорциональны квадратным корням $\sqrt{\mu}$ и $\sqrt{m_1}$, т. е.

$$\overline{v}_{\text{отн}} = \overline{v}_1 \frac{\sqrt{m_1}}{\mu} = \overline{v}_1 \sqrt{\frac{m_1 + m_2}{m_2}}, \quad \text{где} \quad \mu = \frac{m_1 m_2}{m_1 + m_2}.$$

С учетом соотношения $m_1 \overline{v}_1^{\,2} = m_2 \overline{v}_2^{\,2}$ этот результат можно представить в более симметричной форме:

$$\overline{v}_{\text{отн}} = \sqrt{\overline{v}_1^2 + \overline{v}_2^2}. \tag{86.7}$$

Для среднего числа столкновений $\overline{z_{12}}$, претерпеваемых молекулой сорта 1 с молекулами сорта 2 в единицу времени, получаем

$$\overline{z}_{12} = n_2 \sigma_{12} \sqrt{\overline{v}_1^2 + \overline{v}_2^2}, \tag{86.8}$$

а для средней длины свободного пробега молекулы сорта 1

$$\lambda_1 = \frac{1}{n_2 \sigma_{12} \sqrt{1 + (\overline{v}_2/\overline{v}_1)^2}} = \frac{1}{n_2 \sigma_{12} \sqrt{1 + m_1/m_2}}. \tag{86.9}$$

При $m_1 = m_2$ эти выражения переходят в максвелловские формулы (86.3) и (86.4).

6. Наконец, рассмотрим случай смеси двух различных газов. Пусть r_1 и r_2 означают радиусы молекул этих газов, а n_1 и n_2 — их концентрации. Теперь движущаяся молекула может сталкиваться не только с молекулами, себе подобными, но и с молекулами другого типа. В соответствии с этим с ней следует связать две сферы ограждения

в зависимости от того, с молекулами какого типа она сталкивается. Введем четыре величины

$$\sigma_{11} = \pi(2r_1)^2 = 4\pi r_1^2, \quad \sigma_{12} \equiv \sigma_{21} = \pi(r_1 + r_2)^2,$$
$$\sigma_{22} = \pi(2r_2)^2 = 4\pi r_2^2.$$

Для чисел столкновений в единицу времени молекулы первого и молекулы второго газов получаем соответственно (без учета максвелловского распределения скоростей)

$$z_1 = (n_1\sigma_{11} + n_2\sigma_{12})\overline{v}_1, \quad z_2 = (n_1\sigma_{21} + n_2\sigma_{22})\overline{v}_2, \tag{86.10}$$

а для средних длин свободного пробега

$$\lambda_1 = \frac{1}{n_1\sigma_{11} + n_2\sigma_{12}}, \quad \lambda_2 = \frac{1}{n_1\sigma_{21} + n_2\sigma_{22}}. \tag{86.11}$$

ЗАДАЧИ

1. Газ состоит из молекул с массами m_1 и m_2, концентрации которых равны соответственно n_1 и n_2. Найти выражение для средней длины свободного пробега молекулы каждого газа с учетом максвелловского распределения скоростей.

Ответ.

$$\lambda_1 = \frac{1}{n_1\sigma_{11}\sqrt{2} + n_2\sigma_{12}\sqrt{1 + m_1/m_2}},$$
$$\lambda_2 = \frac{1}{n_1\sigma_{21}\sqrt{1 + m_2/m_1} + n_2\sigma_{22}\sqrt{2}}. \tag{86.12}$$

2. Для приближенного вычисления средней длины свободного пробега молекулы Клаузиус предположил, что все молекулы газа движутся с одинаковыми скоростями, направления которых распределены в пространстве изотропно. Получить выражение для λ в этом предположении.

Решение. Найдем среднюю скорость молекул относительно одной из них (например, правой). Относительная скорость молекулы, движущейся под углом ϑ к скорости первой молекулы v_1, определяется выражением

$$v_{\text{отн}} = 2v \sin\frac{\vartheta}{2}$$

(рис. 76). Число молекул, скорости которых образуют с направлением v_1 углы между ϑ и $\vartheta + d\vartheta$, дается формулой (75.4). Используя ее, получаем

$$v_{\text{отн}} = v\int_0^\pi \sin\frac{\vartheta}{2}\,\sin\vartheta\,d\vartheta = \frac{4}{3}\,v.$$

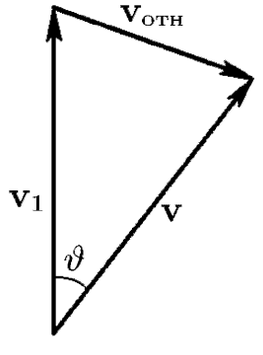


Рис. 76

После этого по формуле (86.6) находим

$$\lambda = \frac{3}{4}\,\frac{1}{n\sigma}. \tag{86.13}$$

3. Найти выражение для среднего полного числа столкновений ν молекул газа в единице объема в единицу времени.

Решение. Число столкновений одной молекулы с остальными дается выражением (86.5). Для n молекул его надо умножить на n и разделить на два. Деление на два необходимо потому, что при нашем подсчете каждая молекула учитывается дважды: один раз как ударяющая, другой — как ударяемая. В результате получаем

$$\nu = \frac{zn}{2} = \frac{1}{2}\, n^2 \sigma \overline{v}_{\text{отн}} = \frac{1}{\sqrt{2}}\, n^2 \sigma \overline{v}. \tag{86.14}$$

4. Газ состоит из смеси двух газов с концентрациями n_1 и n_2. Найти выражение для среднего полного числа столкновений ν_{12} молекул одного газа с молекулами другого газа в единице объема в единицу времени.

Ответ.

$$\nu_{12} = n_1 n_2 \sigma_{12} \overline{v}_{\text{отн}} = n_1 n_2 \sigma_{12} \overline{v}_1 \sqrt{1 + m_1/2} = n_1 n_2 \sigma_{12} \overline{v}_2 \sqrt{1 + m_2/m_1}. \tag{86.15}$$

§ 87. Эффективное сечение

1. *Площадь сечения сферы ограждения молекулы по большому кругу называется эффективным сечением молекулы, точнее, газокинетическим эффективным сечением молекулы при рассеянии ее на других молекулах.* Если рассеяние происходит на таких же молекулах, то эффективное сечение $\sigma = \pi d^2$, где d — диаметр молекулы. Если же молекула радиуса r_1 рассеивается на молекулах с радиусом r_2, то эффективное сечение будет $\sigma = \sigma_{12} = \pi(r_1 + r_2)^2$.

Понятие эффективного сечения широко используется при рассмотрении различных явлений, возникающих при столкновении частиц. При столкновении частица может изменить направление своего движения, т. е. рассеяться. Она может поглотиться, диссоциировать молекулу или ионизовать атом, с которыми она сталкивается, и т. д. В соответствии с этим говорят об *эффективных сечениях рассеяния*, *поглощения*, *диссоциации*, *ионизации* и пр. Во всех таких случаях при вычислении среднего числа столкновений, приводящих к требуемому результату, можно для наглядности представить, что рассматриваемая частица (будем продолжать называть ее частицей A) окружена некоторой непроницаемой «сферой ограждения», а частицы, с которыми она сталкивается, являются точечными. Если частица A движется, а прочие частицы неподвижны, то они называются *полевыми частицами*, а частица A — *пробной частицей*. Среднее число столкновений пробной частицы с полевыми определяется формулой (86.1), где n — концентрация полевых частиц, а v — скорость пробной частицы относительно полевых. На формулу (86.1) следует смотреть как на определение понятия эффективного сечения σ соответствующего процесса.

В экспериментах по столкновениям ядерных и элементарных частиц более удобна другая интерпретация формулы (86.1) и эффективного сечения. Здесь обычно бывает неподвижна частица A, а прочие частицы налетают и бомбардируют ее. В связи с этим эти частицы, если они движутся параллельно, называются *пучком*, а частица A — *мишенью*, которую они бомбардируют. Величина $I = nv$ есть *интенсивность пучка*, т. е. число частиц, проходящих через единичную

площадку, перпендикулярную к пучку, в единицу времени. Применительно к рассматриваемому случаю формулу (86.1) следует переписать в виде

$$\sigma = \frac{z}{I} = \frac{\Delta N}{I}. \tag{87.1}$$

Величина $\Delta N = z$ означает среднее число частиц, выбывших из пучка в единицу времени в результате столкновений с частицей-мишенью A. Таким образом, можно дать следующее определение эффективного сечения частицы по отношению к какому-либо процессу. *Эффективным сечением σ называется отношение среднего числа частиц, выбывших из пучка в единицу времени при столкновениях, приводящих к требуемому результату* (рассеянию пучка, поглощению, ионизации, прилипанию и пр.), *к интенсивности самого пучка*. Преимущество такой интерпретации эффективного сечения состоит в том, что она не связана ни с какими модельными представлениями.

2. Эффективное сечение тех или иных процессов, вообще говоря, сильно зависит от относительной скорости сталкивающихся частиц. Рассмотрим, например, *ионизацию атомов* при столкновениях. Если кинетическая энергия $\varepsilon_{\text{отн}}$ относительного движения сталкивающихся атомов меньше *энергии ионизации* атома, то последняя невозможна. Эффективное сечение ионизации равно в этом случае нулю. Когда $\varepsilon_{\text{отн}}$ равна энергии ионизации или превосходит ее, ионизация становится невозможной. Ясно поэтому, что эффективное сечение ионизации должно зависеть от относительной скорости $v_{\text{отн}}$ сталкивающихся атомов. Сильная зависимость эффективного сечение от $v_{\text{отн}}$ имеет место для таких процессов, как поглощение нейтронов атомными ядрами, деление тяжелых атомных ядер под действием нейтронов, химические и термоядерные реакции и пр. Расчет эффективных сечений подобных процессов производится с помощью законов и вычислительных методов квантовой механики. В настоящей главе речь будет идти только о процессах *упругого рассеяния* молекул и атомов на других молекулах и атомах. В этих случаях внутреннее состояние сталкивающихся частиц не изменяется. Эффективное же сечение таких процессов очень слабо зависит от относительной скорости частиц. Вот почему для их изучения можно пользоваться моделью твердых шариков, в которой поперечное сечение σ совсем не зависит от относительной скорости.

3. В действительности наблюдается некоторое уменьшение эффективного сечения рассеяния молекул с увеличением относительной скорости. Объяснение этому было дано Сёзерлендом (1859–1912) в 1893 г. Он использовал модель твердых упругих шаров, но учел силы притяжения, с которыми молекулы действуют друг на друга в промежутках между столкновениями. Силы притяжения несколько сближают молекулы, пролетающие мимо друг друга, и делают возможными некоторые столкновения, которые при отсутствии этих сил не могли бы произойти. Это ведет к увеличению эффективного сечения рассеяния σ.

Исследуем вопрос с количественной стороны. Рассматривая относительное движение, молекулу A будем считать неподвижной, а молекулу B — движущейся (рис. 77). Относительную скорость молекулы B на бесконечности обозначим через $\mathbf{v}_0$. Пусть b — *прицельное расстояние между молекулами*,

т. е. длина перпендикуляра, опущенного из центра молекулы A на направление прямой, вдоль которой направлен вектор $\mathbf{v}_0$. Если молекулы не взаимодействуют и $b > d$, то столкновение между ними невозможно. При наличии

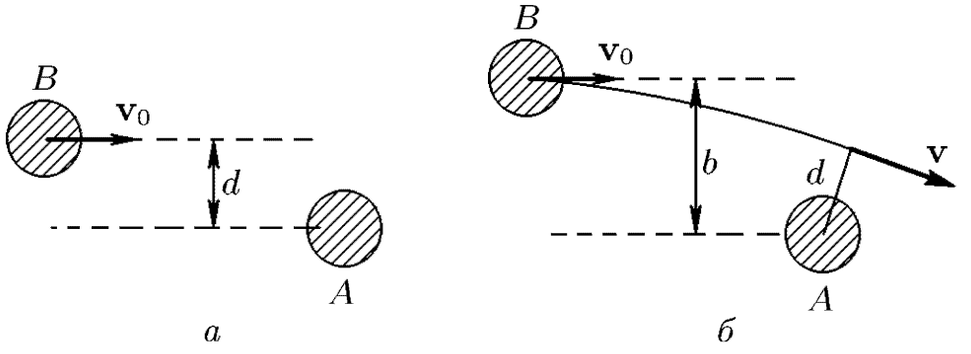


Рис. 77

же сил притяжения столкновение может произойти и в этом случае, как это видно из рис. 77 *б*. Если b — максимальное значение прицельного расстояния, при котором столкновение еще произойдет, то эффективное сечение будет $\sigma = \pi b^2$. Эффективное сечение при отсутствии сил притяжения $\sigma_0 = \pi d^2$. Так как силы — центральные, то по закону площадей $v_0 b = v d$, где v — скорость молекулы B в момент максимального сближения с A. Возведя в квадрат, получим $\sigma v_0^2 = \sigma_0 v^2$. Скорость v найдется из уравнения энергии $(1/2)\mu v^2 = (1/2)\mu v_0^2 + A$, где A — работа центральных сил притяжения при перемещении молекулы B из «бесконечности» в положение максимального сближения с молекулой A, а μ — приведенная масса. Введем кинетическую энергию относительного движения $\varepsilon_{\text{отн}} = (1/2)\mu v_0^2$. Тогда

$$\sigma \varepsilon_{\text{отн}} = \sigma_0 (\varepsilon_{\text{отн}} + A),$$

или

$$\sigma = \sigma_0 (1 + A/\varepsilon_{\text{отн}}).$$

Эта формула, как ясно из ее вывода, применима в тех случаях, когда имеется всего одна пара взаимодействующих молекул A и B. Применять ее к молекулам газа можно только при условии, когда взаимодействия в газе могут рассматриваться как *парные*. В этом приближении учитываются силы притяжения, действующие на молекулу только со стороны одной молекулы, достаточно сблизившейся с рассматриваемой. Действия всех остальных молекул не учитываются. Так можно поступать, когда газ достаточно разряжен, а молекулярные силы притяжения убывают с увеличением расстояния между взаимодействующими молекулами достаточно быстро. Случаи, когда в разряженном газе сближаются и начинают взаимодействовать три и больше молекулы, маловероятны и по этой причине не принимаются во внимание.

В случае газа $\varepsilon_{\text{отн}}$ может принимать всевозможные значения при переходе от одной пары взаимодействующих молекул к другим. Поэтому целесообразно ввести некоторое *среднее эффективное сечение молекулы*, усреднив $1/\varepsilon_{\text{отн}}$ по всем относительным скоростям. Выполнив это и обозначая усредненное эффективное сечение прежней буквой σ, очевидно, придем к формуле вида

$$\sigma = \sigma_0 \left(1 + \frac{S}{T}\right), \tag{87.2}$$

где S — новая постоянная, называемая *постоянной Сёзерленда*. Она имеет размерность температуры. Формула (87.2) также называется *формулой Сёзерленда*.

4. Если эффективное сечение σ существенно зависит от относительной скорости сталкивающихся частиц, то формулы (86.14) и (86.15) следует писать в виде

$$\nu = \frac{1}{2}\, n^2 \langle \sigma v_{\text{отн}} \rangle, \tag{87.3}$$

$$\nu_{12} = n_1 n_2 \, \langle \sigma v_{\text{отн}} \rangle, \tag{87.4}$$

причем усреднение должно быть произведено по *всему спектру относительных скоростей частиц газа*, в частности по максвелловскому распределению, если таковое установилось. Формулы (87.3) и (87.4) применяются при расчете скоростей химических и термоядерных реакций. В этих случаях σ очень сильно зависит от $v_{\text{отн}}$. (Для воспламенения горючей смеси надо нагреть ее до некоторой минимальной температуры!) Формула (87.3) дает среднее число реакций в единице объема в единицу времени, когда все реагирующие частицы одинаковы, а формула (87.4) — когда реакции происходят между разными частицами. Если в каждой реакции выделяется энергия E, то, очевидно, энергетическая мощность «реактора», отнесенная к единице объема, будет

$$P = \nu E. \tag{87.5}$$

ЗАДАЧА

Показать, что в системе центра масс рассеяние шаров при упругих столкновениях сферически симметрично.

Р е ш е н и е. Пусть $\mathbf{v}_1$ — скорость первого, а $\mathbf{v}_2$ — скорость второго шаров (рис. 78). В системе центра масс

$$m_1 \mathbf{v}_1 + m_2 \mathbf{v}_2 = 0.$$

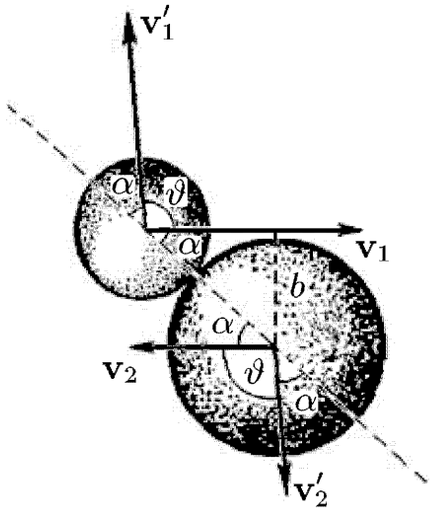


Рис. 78

При столкновении шары в этой системе обмениваются нормальными компонентами импульсов, тогда как касательные компоненты их остаются неизменными. Отсюда следует, что скорости шаров в системе центра масс не изменяются по величине, а только поворачиваются на один и тот же угол ϑ (*угол рассеяния*). Пусть α — угол, который образовывали в момент удара начальные скорости шаров с линией центров. Тогда, как видно из рисунка, $2\alpha + \vartheta = \pi$, т. е. $\alpha = \pi/2 - \vartheta/2$. прицельное расстояние $b = d \sin\alpha = d\, \cos(\vartheta/2)$.

Пусть $\sigma = \pi d^2$, а $d\sigma = 2\pi b\, db$ — площадь кольца с радиусами b и $b + db$. Тогда

$$d\sigma = (1/2)\pi d^2 \sin\vartheta \, d\vartheta.$$

Вероятность рассеяния в телесный угол $d\Omega = 2\pi \sin\vartheta \, d\vartheta$ равна

$$\frac{d\sigma}{\sigma} = \frac{d\Omega}{4\pi},$$

т. е. пропорциональна $d\Omega$. Коэффициент пропорциональности $1/4\pi$ не зависит от ϑ. А это значит, что рассеяние сферически симметрично.

§ 88. Ослабление пучка молекул в газе

1. Допустим, что в газе распространяется *параллельный пучок молекул*. Это может быть *внешний* пучок, состоящий из молекул другого газа. Но пучок может состоять и из молекул *того же газа*. Можно представить себе, например, что в какой-то момент времени в газе отмечены молекулы с определенным направлением скорости.

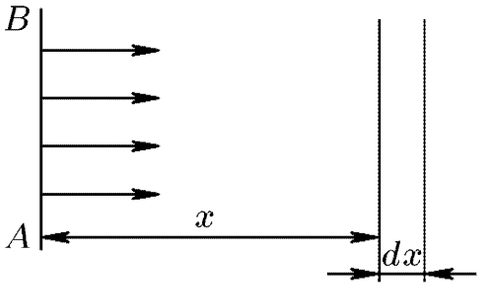


Рис. 79

Пусть J_0 — интенсивность пучка, когда он пересекает плоскость AB, перпендикулярную к нему (рис. 79). Найдем интенсивность J того же пучка на расстоянии x от плоскости AB. Возьмем бесконечно тонкий слой газа с толщиной dx и площадью поперечного сечения $S = 1$. Число молекул газа в нем равно $nS\,dx = n\,dx$. Согласно определению эффективного сечения (87.1) среднее число частиц, выбывающих из пучка из-за столкновений с одной молекулой газа, равно $J\sigma$, а из-за столкновений с $n\,dx$ молекулами $dN = J\sigma n\,dx = (J/\lambda)\,dx$. На такую величину уменьшится интенсивность пучка после прохождения слоя dx, а потому

$$dJ = -\frac{J}{\lambda}\,dx. \tag{88.1}$$

Интегрирование этого выражения дает

$$J = J_0 e^{-x/\lambda}. \tag{88.2}$$

Из-за рассеяния интенсивность пучка убывает экспоненциально. В связи с этим величину $1/\lambda$ называются *коэффициентом рассеяния*. Согласно формуле (88.1) величина dx/λ определяет вероятность рассеяния на пути dx, а $1/\lambda$ — *вероятность рассеяния на единице длины*.

Формуле (88.2) можно также дать следующее толкование. Если N_0 — число частиц, прошедших через площадку AB, то число частиц, прошедших без столкновения расстояние x, определяется выражением

$$N = N_0 e^{-x/\lambda}. \tag{88.3}$$

Число частиц, претерпевших столкновение в слое $(x, x + dx)$, равно $|dN| = (1/\lambda)\,N_0 e^{-x/\lambda}\,dx$. Средний путь, пройденный частицами без столкновений,

$$\overline{x} = \frac{1}{N_0}\int x|dN| = \frac{1}{\lambda}\int\limits_0^\infty x e^{-x/\lambda}\,dx = \lambda.$$

Он, как и следовало ожидать, совпадает с длиной свободного пробега λ.

2. Формула (88.3) применима также к другому случаю, существенному для излагаемой ниже элементарной теории явлений переноса.

Представим себе, что в некоторый момент времени скорости всех молекул газа изменили направления на противоположные, но сохранили свои модули. Такая замена не скажется на хаотичности движения молекул газа. Однако каждая молекула в точности повторит предшествующее движение, но в обратном направлении. Отсюда получается следующий результат. Пусть площадку AB пересекло N_0 молекул, двигавшихся перпендикулярно к ней. Отметим все эти молекулы. На расстоянии x перед площадкой AB число отмеченных молекул будет $N = N_0 e^{-x/\lambda}$. По мере приближения к площадке AB оно возрастает из-за столкновений. Число отмеченных молекул, претерпевших столкновения между x и $x + dx$, определяется прежним выражением

$$|dN| = \frac{1}{\lambda} N_0 e^{-x/\lambda}\, dx = \frac{N}{\lambda}\, dx.$$

3. На формуле (88.2) или (88.3) основан прямой метод измерения длины свободного пробега. Этот метод был предложен и осуществлен М. Борном и Е. Борман в 1921 г.

Идея опыта состояла в следующем. Путем испарения получался пучок атомов серебра, резко ограниченный диафрагмами. На пути пучка помещались 4 коаксильных диска на расстоянии 1 см друг от друга (рис. 80), в которых были вырезаны одинаковые круглые отверстия с центрами на оси системы. Через них и проходили атомы серебра. На каждом диске прикреплялся стеклянный квадрант, вершина которого лежала на оси системы. (На рис. 80 один из квадрантов изображен сбоку.) Квадранты были повернуты друг относительно друга на 90°, так что на каждый из них направлялась четвертая часть атомов пучка. Вся система помещалась в кварцевой трубке, давление воздуха в которой менялось с помощью насоса и измерялось манометром. Диски охлаждались жидким азотом. На пути от источника атомы серебра частично рассеивались молекулами воздуха, а затем конденсировались на квадрантах, стоящих на пути пучка. Пусть N_1, N_2, N_3, N_4 — числа атомов серебра, осевших на квадрантах. Тогда, согласно формуле (88.3), должно быть

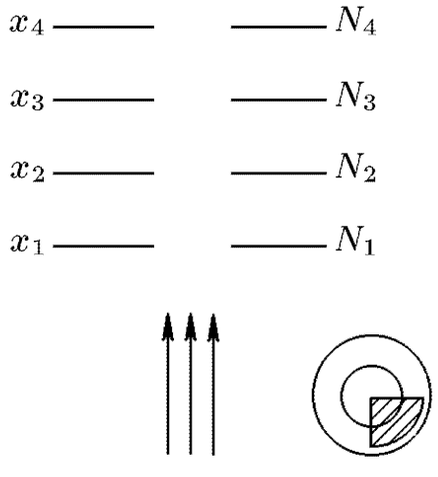


Рис. 80

$$\frac{N_1}{N_2} = \exp\frac{x_2 - x_1}{\lambda}, \qquad \lambda = \frac{x_2 - x_1}{\ln(N_1/N_2)}.$$

Аналогичные соотношения можно написать для каждой пары квадрантов. Отношение N_1/N_2 можно было измерить по степени почернения стеклянных квадрантов, которая находилась путем фотометрирования. Затем, зная расстояние между дисками $(x_2 - x_1)$, можно было найти и среднюю длину свободного пробега λ. Найденные таким путем значения λ удовлетворительно согласуются с результатами других

методов, которые будут изложены ниже. Было показано также, что величина λ обратно пропорциональна давлению воздуха P в трубке, как этого требует формула (86.2).

4. Опыты Борна и Борман дают эффективное сечение рассеяния атомов серебра на молекулах воздуха. Более важным является случай, когда и пучок, и газ состоят из *одинаковых* молекул. Тогда средняя длина свободного пробега λ является *характеристикой самого газа*. Зная ее, можно вычислить эффективное поперечное сечение σ по формуле (86.3), а затем найти и газокинетический диаметр молекулы, связанный с σ соотношением $\sigma = \pi d^2$. Именно таким путем впервые Лошмидтом (1821–1895) были определены геометрические размеры молекул.

Основные сведения о длинах свободного пробега были получены косвенными методами. Они основаны на изучении так называемых *явлений переноса*: вязкости, теплопроводности и диффузии. С феноменологической точки зрения явление вязкости мы рассматривали в томе I в разделе механики жидкостей и газов, а теплопроводности — в гл. 4 этого тома. В следующих параграфах мы изложим эти явления, а также явление диффузии с молекулярно-кинетической точки зрения. Строгая молекулярно-кинетическая теория перечисленных явлений очень сложна. Она сводится к приближенным решениям так называемого *кинетического уравнения Больцмана*. Последнее является *основным* в кинетической теории газов. В принципе оно позволяет найти функцию распределения молекул газа по координатам и скоростям не только в состоянии равновесия, но и тогда, когда в газе происходят различные процессы. Однако уравнением Больцмана мы пользоваться не будем. При изложении упрощенной теории вязкости и теплопроводности газов мы изберем значительно более простой путь, использующий понятие средней длины свободного пробега. В теории диффузии метод средней длины свободного пробега не всегда удобен. Поэтому при изложении теории диффузии мы дополним его другим методом, основанным на *соотношении Эйнштейна* (91.3). Упрощенные теории охватывают все существенные черты явлений переноса. Только значения числовых коэффициентов в формулах получаются не совсем точными.

§ 89. Вязкость и теплопроводность газов

1. Наличие *вязкости* в газах обычно иллюстрируют на следующем примере. Между двумя параллельными пластинками AB и CD (рис. 81 *а*) находится воздух или другой газ. При движении пластинки CD появляется сила, действующая на пластинку AB и направленная в сторону движения. Эта сила и есть *сила вязкости*. Впрочем, о вязкости можно говорить лишь тогда, когда расстояние между пластинками AB и CD очень велико по сравнению со средней длиной свободного пробега молекулы газа. Тогда от наличия пластин можно отвлечься и

говорить о силах, действующих внутри самого газа. Будем представлять себе газ неограниченным и движущимся стационарно плоскопараллельными слоями в горизонтальном направлении. Скорость этого макроскопического движения u меняется в направлении, перпендикулярном к слоям. Это направление мы примем за ось X (рис. 81 *б*).

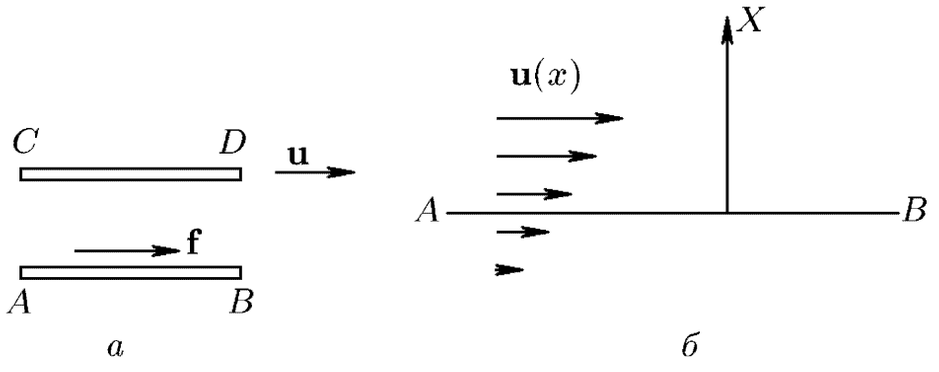


Рис. 81

Таким образом, мы предполагаем, что $u = u(x)$. Рассечем мысленно газ на две половины плоскостью MN, параллельной слоям. Допустим для определенности, что скорость $u(x)$ возрастет с возрастанием x. Тогда верхняя половина газа будет действовать на нижнюю с силой, направленной вправо, а нижняя на верхнюю — с силой, направленной влево. Это и есть силы вязкости.

С молекулярной точки зрения происхождение сил вязкости объясняется следующим образом. Если бы газ покоился, то все направления скоростей его молекул были бы равновероятны. Средняя скорость и среднее количество движения каждой молекулы были бы равны нулю. При наличии *упорядоченного движения* газа средняя скорость молекулы не нуль, а равна $\mathbf{u} = \mathbf{u}(x)$. С этой скоростью связано количество движения $\mathbf{g} = m\mathbf{u}$, которым обладает рассматриваемая молекула. Такое количество движения условимся называть *упорядоченным*. Молекулы, лежащие над плоскостью AB, обладают большим упорядоченным количеством движения, чем молекулы, расположенные под ней. Переходя из пространства над плоскостью MN в пространство под ней, молекулы передают часть своего упорядоченного количества движения молекулам, с которыми они сталкиваются в пространстве ниже плоскости MN. Это проявляется в том, что газ, расположенный ниже этой плоскости, подвергается действию силы, направленной в сторону скорости $\mathbf{u}$. Аналогично, более медленные молекулы, попадая из «нижнего» пространства в «верхнее», при столкновениях отнимают часть упорядоченного количества движения у молекул, расположенных над плоскостью MN. В результате газ в верхнем пространстве испытывает тормозящую силу, направленную против скорости $\mathbf{u}$. Эти силы и являются силами вязкости.

Для пояснения возникновения вязкости полезна следующая аналогия. Две железнодорожные платформы движутся по параллельным рельсам с несколько отличающимися скоростями. Грузчики, находящиеся на платформе, перебрасывают мешки с песком со своей

платформы на соседнюю. Ясно, что в результате этого быстрее движущаяся платформа будет тормозиться, а медленнее движущаяся — ускоряться.

2. Исследуем теперь явления вязкости газов количественно. Для лучшего уяснения существа дела рассмотрим сначала предельно упрощенный расчет. Будем считать скорости теплового движения всех молекул одинаковыми и равными v. Кроме того, при рассмотрении теплового движения будем рассуждать так, как если бы все молекулы были разделены на шесть одинаковых потоков, параллельных координатным осям. Таким образом, одна шестая всех молекул будет двигаться сверху вниз, одна шестая — снизу вверх. Только молекулы этих двух потоков и участвуют в передаче количества движения. Молекулы остальных четырех потоков движутся параллельно плоскости MN. Они к передаче количества движения прямого отношения не имеют. На тепловое движение молекул сверху вниз и снизу вверх накладывается упорядоченное движение вправо, причем скорость этого движения u однозначно определяется положением молекулы, точнее, ее координатой x. Будем предполагать, что изменения упорядоченной скорости u на длине свободного пробега очень малы по сравнению с тепловой скоростью v. Тогда можно выбрать такую систему отсчета, в которой упорядоченная скорость u в интересующей нас части газа будет также очень мала по сравнению с тепловой скоростью v. В дальнейшем движение рассматривается именно в такой системе отсчета.

Возьмем на плоскости MN (рис. 82) единичную площадку S. Начало координат поместим в той же плоскости. Подсчитаем количество движения, ежесекундно переносимое молекулами газа через площадку S. Число молекул, пересекающих площадку S сверху вниз в единицу времени, определяется формулой (75.1), т. е.

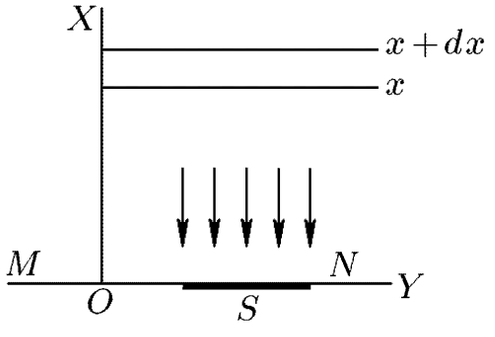


Рис. 82

$$N_0 = \frac{1}{6}\, nv.$$

Пусть N из этих молекул прошли перед площадкой путь x без столкновений. Число N определяется формулой (88.3). Из нее находим, что число молекул, претерпевших последнее столкновение в слое между x и $x + dx$, равно

$$dN = \frac{1}{\lambda}\, N_0 e^{-x/\lambda}\, dx = \frac{nv}{6\lambda}\, e^{x/\lambda}\, dx.$$

При столкновении в этом слое молекула получает количество движения $g(x)$ и, двигаясь далее без столкновений, переносит его через площадку S. Количество движения, переносимое в единицу времени через площадку S всеми N_0 молекулами, определяется интегралом

$$G_+ = \int g(x)\, dN.$$

Так как на длине свободного пробега скорость u меняется мало, то функцию $g(x)$ можно разложить по степеням x, оборвав это разложение на линейном члене, т. е. $g(x) = g_0 + x(dg/dx)_0$. В этом приближении

$$G_+ = g_0 \int dN + \frac{nv}{6\lambda}\frac{dg}{dx}\int\limits_0^\infty e^{x/\lambda} x\,dx.$$

Вычислив интегралы, найдем

$$G_+ = \frac{1}{6}nvg_0 + \frac{1}{6}nv\lambda\frac{dg}{dx}. \tag{89.1}$$

Заметим, что это выражение можно записать в виде $G_+ = (1/6) \times \times nvg(\lambda)$. Отсюда видно, что при вычислении G_+ можно рассуждать так, как *если бы все молекулы, летящие к площадке S, претерпевали последние столкновения на расстоянии λ от этой площадки и далее двигались к ней без столкновений.* Можно пользоваться доказанным положением для сокращения изложения при изучении и других явлений переноса. Отметим только, что концентрация n может меняться в пространстве. Однако это обстоятельство никак не отразится на справедливости формулы (89.1). Действительно, из вывода ясно, что под n следует понимать значение концентрации на самой площадке S, независимо от того, рассматривается ли верхний или нижний пучки молекул, участвующих в переносе величины g. Было бы грубой ошибкой считать, что концентрацию n надо брать на расстоянии $\pm\lambda$ от площадки S, где молекулы претерпели «последние столкновения». Такой способ расчета должен применяться *только к переносимой величине g, но не к концентрации n.*

По аналогии с (89.1) можно утверждать, что молекулы, летящие снизу вверх, переносят в том же направлении количество движения

$$G_- = \frac{1}{6}nvg_0 - \frac{1}{6}nv\lambda\frac{dg}{dx}. \tag{89.1а}$$

Полное количество движения, ежесекундно переносимое через площадку S в положительном направлении оси X (снизу вверх), найдется вычитанием (89.1) из (89.1а). Оно равно

$$G = -\frac{1}{3}nv\lambda\frac{dg}{dx} = -\frac{1}{3}nmv\lambda\frac{du}{dx}. \tag{89.2}$$

Этот перенос проявляется в том, что вдоль плоскости MN действует вязкое касательное напряжение

$$\tau_{xy} = \eta\ \frac{du}{dx}, \tag{89.3}$$

где

$$\eta = \frac{1}{3}nmv\lambda. \tag{89.4}$$

Мы получили не только *ньютоновский закон вязкости* (89.3), но и нашли *выражение для самой вязкости η.*

3. Но всякий тензор напряжений должен быть *симметричным*. В противном случае нарушался бы закон сохранения момента количества движения (см. т. I, § 74). Поэтому вязкие напряжения должны действовать не только в плоскостях течения газа, но и в плоскостях, перпендикулярных к ним. Необходимо поэтому выяснить, как возникают эти напряжения, и убедиться, что они удовлетворяют условию симметрии $\tau_{xy} = \tau_{yx}$. Ориентируем с этой целью бесконечно малую площадку dS перпендикулярно к направлению течения газа (рис. 83). Таким образом, по-прежнему, предполагается, что газ течет параллельно оси Y, и рассматриваются группы молекул, тепловые скорости которых параллельны оси X. Из-за наличия молекулярного течения молекулы имеют боковую составляющую скорости $u(x)$. Благодаря этому в рассматриваемом случае и появляется поток молекул, пронизывающий площадку dS. Рассмотрим пучок молекул, приходящих сверху. На основании доказанного выше можно рассуждать так, как если бы все молекулы пучка совершили последние столкновения на расстоянии λ от площадки dS (измеренном вдоль оси X). Количество таких молекул, пронизывающих площадку dS в единицу времени, равно $(1/6)\,nu\lambda\,dS$. Переносимое ими количество движения будет $dG_+ = (1/6)\,mvnu\lambda\,dS$ и направлено вниз. Поток молекул, приходящих снизу, будет меньше, а именно $(1/6)\,nu(-\lambda)\,dS$. Связанный с ними поток импульса равен $dG_- = (1/6)\,mvnu(-\lambda)\,dS$ и направлен вверх. Разность этих двух потоков

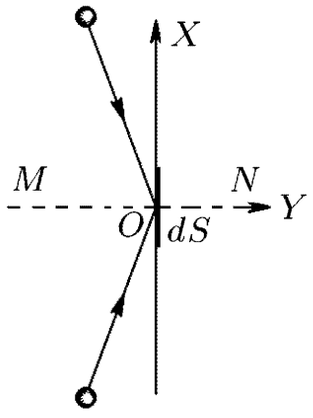


Рис. 83

$$dG = dG_- - dG_+ = -\frac{1}{3}\,mvn\lambda\,\frac{du}{dx}$$

даст полный поток количества движения, переносимый через площадку dS вверх в единицу времени. Он проявляется в появлении касательного напряжения

$$\tau_{yx} = \frac{1}{3}\,mnv\lambda\,\frac{dv}{dx},$$

которое действует в плоскостях, перпендикулярных к направлению течения газа. Таким образом, мы выяснили происхождение «поперечных» касательных напряжений и доказали, что $\tau_{xy} = \tau_{yx}$.

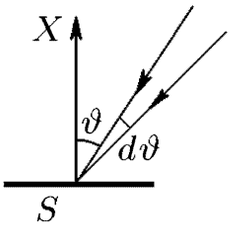


Рис. 84

4. Можно было бы усовершенствовать рассуждения, не прибегая к искусственному разделению молекул на шесть взаимно перпендикулярных потоков. Будем считать сначала, что скорости молекул одинаковы по модулю, но распределены по направлениям изотропно. Рассчитаем касательное напряжение τ_{xy}, действующее в плоскостях, параллельных слоям текущего газа. Расчет касательных напряжений в перпендикулярных плоскостях τ_{yx} производится аналогично. Число молекул

в единице объема, скорости которых направлены под углами между ϑ и $\vartheta + d\vartheta$ к нормали к площадке S (рис. 84), дается выражением (74.5), т. е. равно $(1/2)\, n \sin\vartheta\, d\vartheta$. При подсчете переносимого ими количества движения можно рассуждать так, как если бы все они претерпели последнее столкновение на расстоянии λ от площадки, если это расстояние измерять в направлении движения молекулы, или на расстоянии $\lambda\cos\vartheta$, если его измерять вдоль оси X. Рассматриваемые молекулы переносят сверху вниз количество движения

$$dG_+ = \frac{nv}{2}\sin\vartheta\, d\vartheta\, g(\lambda\cos\vartheta)\cos\vartheta = \\ = \frac{nv}{2}\, g_0 \sin\vartheta\,\cos\vartheta\, d\vartheta + \frac{nv\lambda}{2}\,\frac{dg}{dx}\,\cos^2\vartheta\sin\vartheta\, d\vartheta.$$

Полное количество движения, переносимое сверху вниз, найдется интегрированием этого выражения по ϑ в пределах от $\vartheta = 0$ до $\vartheta = \pi/2$. Оно равно

$$G_+ = \frac{nv}{4}\, g_0 + \frac{nv\lambda}{6}\,\frac{dg}{dx}.$$

Аналогично находится количество движения G_-, переносимое снизу вверх. Для полного потока количества движения

$$G = G_- - G_+ = -\frac{nv\lambda}{3}\,\frac{dg}{dx}$$

получается такое же выражение, как и в более элементарном расчете, приведенном выше, а для η — прежнее выражение (89.4).

Теперь не представляет труда учесть разброс скоростей. Для этого надо только усреднить по всем скоростям произведения $u\lambda$, т. е. вместо (89.4) написать

$$\eta = \frac{1}{3}\, nm\langle v\lambda\rangle. \tag{89.5}$$

Если же пренебречь зависимостью эффективного сечения σ, а с ним и λ от скорости, то надо усреднять только v, т. е.

$$\eta = \frac{1}{3}\, nm\lambda\overline{v}.$$

5. Совершенно так же может быть рассмотрено явление *теплопроводности*. Здесь вместо переноса количества движения речь идет о *переносе энергии*. В той области температур, где справедлива *классическая теория теплоемкостей*, энергия молекулы пропорциональна температуре и может быть представлена в виде $\varepsilon = mc_vT$, где c_v — удельная теплоемкость газа при постоянном объеме. Пусть газ находится между двумя бесконечными проводящими плоскостями, перпендикулярными к оси X, причем температуры их T_1 и T_2 поддерживаются постоянными. Тогда будет происходить передача теплоты совершенно так же, как в аналогичном случае происходит передача количества движения. Для вычисления потока теплоты q можно воспользоваться формулой (89.2), заменив в ней импульс g на энергию $\varepsilon = mc_vT$. Тогда получится

$$q = -\frac{1}{3}\, nm\overline{v}\lambda c_v\,\frac{dT}{dx}.$$

Сравнивая эту формулу с (52.3), получаем для теплопроводности газа

$$\varkappa = \frac{1}{3}\, nm\bar{v}c_v\lambda. \tag{89.7}$$

Необходимо отметить, что концентрация в неравномерно нагретом покоящемся газе не может оставаться всюду одинаковой. Для предотвращения возникновения макроскопических движений должно оставаться постоянным по всему объему газа давление P, а следовательно, и произведение nT. В неравномерно нагретом газе это возможно только тогда, когда концентрация n меняется от точки к точке. Но это обстоятельство, как было показано в п. 2, совершенно несущественно для применимости формулы (89.2).

6. Сравнением формул (89.6) и (89.7) получаем интересное соотношение

$$\varkappa/\eta = c_v. \tag{89.8}$$

Опыт в основном подтвердил эту зависимость, но в несколько более общей форме

$$\varkappa/\eta = Ac_v, \tag{89.9}$$

где A — числовой коэффициент порядка единицы. Для разных газов он имеет разные значения и слабо зависит от температуры. В табл. 9 приведены значения коэффициента A для некоторых газов при 0 °C. Расхожде-

Таблица 9

Газ	A	Газ	A
Гелий	2,51	Хлор	1,79
Неон	2,47	Окись углерода	1,91
Аргон	2,53	Окись азота	1,86
Криптон	2,54	Двуокись углерода	1,67
Ксенон	2,57	Закись азота	1,74
Водород	2,02	Аммиак	1,41
Азот	1,97	Метан	1,73
Кислород	1,91	Этилен	1,44

ние между опытом и теорией, ввиду приближенного характера последней, вполне естественно. Строгая, но очень сложная теория, развитая Чепменом, показала, что для всех сферически симметричных невращающихся молекул $A \geqslant 5/2$. Численные расчеты на специальных моделях молекул (например, твердых шарах) показали, что A лишь немногим больше $5/2$. Значения A для инертных газов согласуются с этим результатом.

При изложении теории теплоемкостей (см. § 69) было показано, что тепловые вращательные движения молекул прекращаются, если температура газа становится заметно ниже *вращательной характеристической температуры*. Газ ведет себя как одноатомный. Поэтому следует ожидать, что для всех газов коэффициент A при низких температурах должен стремиться к предельному значению $5/2$, какое он имеет для одноатомных газов. Этот вывод был подтвержден Эйкеном (1884–1950) для водорода. Эйкен нашел, что для водорода $A = 2{,}25$ при $T = 81$ K и $A = 2{,}37$ при $T = 21$ K, тогда как при 0 °C $A = 2{,}02$. (Напомним, что вращательная характеристическая температура для водорода $T_r \sim 175$ K).

7. Введя плотность газа $\rho = nm$, формулы (89.6) и (89.7) можно переписать в виде

$$\eta = \frac{1}{3}\rho\overline{v}\lambda, \quad \varkappa = \frac{1}{3}\rho\overline{v}\lambda c_v. \tag{89.10}$$

Так как λ обратно пропорциональна ρ, то отсюда следует, что *вязкость и теплопроводность не зависят от плотности газа*. К такому выводу впервые пришел Максвелл, и этот вывод показался ему парадоксальным. Однако опыты, поставленные самим Максвеллом и другими физиками, подтвердили указанный вывод.

Независимость вязкости и теплопроводности от плотности газа имеет простое объяснение. Если плотность газа велика, то в переносе импульса и энергии участвует *много молекул*. Однако передача импульса и энергии за время между двумя последовательными столкновениями производится *малыми порциями* и на *малые расстояния*. Если же плотность мала, то уменьшается и число молекул, участвующих в переносе. Но это уменьшение полностью компенсируется тем, что теперь молекулы переносят импульс и энергию *более крупными порциями* и на *большие расстояния*.

8. Однако так будет происходить только до тех пор, пока плотность газа ρ не очень мала. Допустим, по-прежнему, что речь идет о передаче импульса или тепла от одной пластинки к другой, ей параллельной. *Независимость коэффициентов η и $\varkappa$ от плотности ρ справедлива до тех пор, пока расстояние между пластинками d велико по сравнению с длиной свободного пробега λ*. В противоположном случае, когда $\lambda \gg d$, столкновения между молекулами самого газа в явлениях переноса перестают играть роль. Существенными становятся только столкновения молекул со стенками сосуда, в котором заключен газ. Молекула, столкнувшись с горячей стенкой, получает от нее кинетическую энергию и переносит ее к более холодной стенке, не испытывая промежуточных столкновений. То же можно сказать и о переносе импульса. Ясно, что в таком случае перенос тепла или импульса будет происходить тем медленнее, чем сильнее разрежен газ. Такой перенос формально может быть описан прежними формулами (89.3) и (52.3). Только теперь η и $\varkappa$ лишь условно можно назвать вязкостью и теплопроводностью газа. Они зависят не только от свойств самого газа, но и от расстояния между пластинками. Их можно оценить прежними формулами (89.6) и (89.7), заменив величину λ на расстояние d между пластинками. Таким образом, *независимость вязкости и теплопроводности от плотности газа имеет место лишь до тех пор, пока длина свободного пробега λ мала по сравнению с какими-то характерными размерами сосуда d* (например, расстоянием между противоположными стенками, когда они параллельны). *Когда λ и d становятся сравнимыми друг с другом, при уменьшении плотности коэффициенты η и $\varkappa$ начинают убывать. В дальнейшем это убывание идет по линейному закону: коэффициенты η и $\varkappa$ становятся пропорциональными плотности ρ*. На этом основано устройство *сосуда Дьюара* (см. § 1). Пока расстояние d между двойными стенками сосуда

Дьюара велико по сравнению с λ, откачка газа из пространства между этими стенками ничего не дает для теплоизоляции. Но она становится эффективной после перехода через промежуточную область, в которой $\lambda \sim d$.

9. Закон Максвелла о независимости вязкости и теплопроводности от плотности должен нарушаться и для плотных газов, когда размеры молекул уже не могут считаться пренебрежимо малыми по сравнению с длиной свободного пробега. Опыт показывает, что в таких случаях вязкость газов с повышением плотности возрастает, особенно быстро при низких температурах. Для паров, близких к конденсации, может играть роль и то обстоятельство, что молекулы группируются в агрегаты, состоящие из нескольких молекул. При обычных плотностях для большинства газов наблюдающаяся зависимость η от плотности едва ли выходит за пределы вероятных ошибок измерений. Слабая же зависимость от плотности в этих случаях может быть объяснена действием молекулярных сил, которые в элементарной теории переноса не учитываются.

10. В заключение кратко остановимся на вопросе о зависимости коэффициентов переноса η и $\varkappa$ от температуры. С этой целью подставим в формулы (89.6) и (89.7) значение λ из (86.2). Получим

$$\eta = \frac{m}{3\sigma}\,\overline{v}, \quad \varkappa = \frac{mc_v}{3\sigma}\,\overline{v}. \tag{89.11}$$

Если считать σ не зависящим от температуры, то оба коэффициента η и $\varkappa$ будут пропорциональны $\overline{v}$, т. е. $\sqrt{T}$. На самом деле они растут с увеличением температуры быстрее. Отсюда следует, что эффективное сечение σ убывает с увеличением температуры. Это можно объяснить действием молекулярных сил. Например, если воспользоваться формулой Сёзерленда (87.2), то для вязкости получится

$$\eta = \frac{A\sqrt{T}}{1 + S/T}, \tag{89.12}$$

где A — постоянная. Для некоторых газов эта формула хорошо передает изменение коэффициента η с изменением температуры в довольно широких диапазонах температур. Так, например, это установлено на опыте для N_2 (температурный интервал от -76 до $250\,^\circ$C) и CO_2 (от -21 до $302\,^\circ$C). Постоянные Сёзерленда для этих газов в указанных температурных интервалах равны 103 K (для N_2) и 240 K (для CO_2). При низких температурах формула (89.12) дает слишком малые значения η для всех газов. К водороду и гелию формула (89.12) применима плохо. Так, для гелия значения S, обеспечивающие совпадение формулы (89.12) с опытом, возрастают от 80 при обычных температурах примерно до 200 для температур, близких к $800\,^\circ$C.

ЗАДАЧИ

1. Определить расход массы газа Q при изотермическом стационарном пуазейлевом течении его вдоль прямолинейной трубы длины l с постоянным поперечным сечением S.

Р е ш е н и е. Возьмем бесконечно малый участок трубы длины dx. При течении на этом участке газ можно считать несжимаемым и применить к нему формулу Пуазейля (1799–1869)

$$Q = -C\,\frac{\rho S^2}{\eta}\,\frac{dP}{dx},$$

где C — постоянная, для круглой трубы $C = 1/8\pi$ (см. т. I, § 97). Исключая плотность ρ с помощью уравнения Клапейрона, получим

$$Q = -C\,\frac{\mu S^2}{\eta RT}\,P\,\frac{dP}{dx}.$$

Расход Q при стационарном течении одинаков вдоль всей трубы. Вязкость η зависит только от T и при изотермическом течении остается постоянной. Поэтому интегрирование дает

$$Q = C\,\frac{\mu S^2}{2\eta RT}\,\frac{P_1^2 - P_2^2}{l}. \tag{89.13}$$

2. Найти стационарное распределение температур в плоскопараллельном слое газа, на границах которого поддерживаются постоянные температуры. Считать, что $k \sim \sqrt{T}$. То же для сферического и цилиндрического слоев (см. § 53).

Ответ. В формулах § 53 надо сделать замену $T \to T^{3/2}$, $T_1 \to T_1^{3/2}$, $T_2 \to T_2^{3/2}$.

§ 90. Самодиффузия в газах

1. Допустим, что закрытая горизонтальная труба разделена на две части перегородкой. По одну сторону перегородки находится какой-то газ 1, а по другую — газ 2. Пусть давления и температуры обоих газов одинаковы. Если удалить перегородку, то газы начнут перемешиваться. Причиной этого является хаотическое тепловое движение молекул. Спустя некоторое время концентрации компонентов смеси станут одинаковыми по обе стороны перегородки. Такое проникновение молекул одного газа в среду молекул другого газа называется *взаимной, или концентрационной, диффузией газов*. Для наблюдения диффузии трубу можно расположить и вертикально. Только в этом случае сверху перегородки должен находиться более легкий, а снизу — более тяжелый газ.

Если газы по обе стороны перегородки тождественны, то диффузия также будет происходить. В этом случае она называется *самодиффузией*. Макроскопически самодиффузию наблюдать нельзя, так как из-за тождественности молекул она не может проявиться ни в каком макроскопическом явлении. Для наблюдения самодиффузии надо как-то «пометить» часть молекул газа. Практически это можно сделать, если вместо однородного газа взять смесь двух изотопов его, один из которых является радиоактивным. Можно также взять смесь двух различных газов (например, CO и N_2), молекулы которых одинаковы по массе и практически имеют одинаковые размеры.

2. Допустим, что концентрации «меченых» молекул n_1 и «немеченых» n_2 меняются в направлении оси трубы. Примем эту ось за координатную ось X. Если общая концентрация частиц $n = n_1(x) + n_2(x)$ и температура газа остаются постоянными, то будет постоянно и давление во всем газе. Поэтому макроскопическое движение в газе

возникнуть не может, и перемешивание молекул возможно за счет одной только диффузии. Молекулы газа 1 и молекулы газа 2 будут дифференцировать в противоположных направлениях — от мест большей концентрации к местам меньшей концентрации. Для количественного описания явления введем понятие *диффузионного потока. Диффузионным потоком Γ называется количество молекул рассматриваемого типа, проходящих при диффузии через единичную площадку, перпендикулярную к градиенту концентрации, в единицу времени.* Задача теории диффузии сводится к вычислению диффузионных потоков.

Для вычисления потоков Γ_1 и Γ_2 обоих газов при самодиффузии можно воспользоваться формулой (89.2). Роль «переносимой величины» g в этом случае играют относительные концентрации газов $c_1 = n_1/n$ и $c_2 = n_2/n$. Подставляя их в (89.2), находим для диффузионного потока первого газа

$$\Gamma_1 = -\frac{1}{3}\, n\overline{v}\lambda\, \frac{dc_1}{dx} = -\frac{1}{3}\,\overline{v}\lambda\, \frac{dn_1}{dx}.$$

Аналогичное выражение напишем для газа 2. Таким образом,

$$\Gamma_1 = -Dn\,\frac{dc_1}{dx}, \quad \Gamma_2 = -Dn\,\frac{dc_2}{dx}, \quad \text{где} \quad D = \frac{1}{3}\,\overline{v}\lambda. \tag{90.1}$$

Так как полная концентрация $n = n_1 + n_2$ остается постоянной, то формулы (90.1) можно также записать в виде

$$\Gamma_1 = -D\,\frac{dn_1}{dx}, \quad \Gamma_2 = -D\,\frac{dn_2}{dx}. \tag{90.2}$$

Но если величина n не постоянна, а меняется в пространстве под действием какой-либо причины (например, силового поля или градиента температуры), то формулы (90.2) становятся непригодными. В таких случаях надо пользоваться более общими формулами (90.1).

Из постоянства полной концентрации $n = n_1 + n_2$ следует $dn_1/dx + dn_2/dx = 0$, а потому $\Gamma_1 = -\Gamma_2$. Диффузионные потоки обоих газов одинаковы по модулю, но направлены противоположно.

Формулы (90.1) показывают, что *диффузионный поток пропорционален градиенту концентрации.* Это — *закон Фика.* Величина D называется *коэффициентом диффузии.* В случае самодиффузии она определяется выражением

$$D = \frac{1}{3}\,\overline{v}\lambda. \tag{90.3}$$

3. Закон Фика справедлив и для взаимной диффузии различных газов. Однако здесь формула (90.3), вообще говоря, неприменима. Ею можно пользоваться только в тех случаях, когда концентрация рассматриваемого газа очень мала по сравнению с концентрацией другого газа смеси. При выполнении этого условия средняя длина свободного пробега молекул рассматриваемого газа λ определяется столкновениями их только с молекулами другого газа смеси. Взаимные же столкновения между молекулами рассматриваемого газа, ввиду

малости его относительной концентрации, роли не играют. Аналогичная ситуация встречается и при *диффузии нейтронов* в графите. Это явление используется для *замедления нейтронов* в ядерных реакторах (см. т. V, § 95). Ввиду малости концентрации нейтронов последние ведут себя как молекулы идеального газа, сталкивающиеся время от времени с ядрами атомов графита. Диффузия нейтронов хорошо описывается формулой (90.3). Длина свободного пробега нейтронов в графите порядка сантиметра.

Когда концентрации обоих газов в смеси одного порядка, то при построении теории концентрационной диффузии по методу средней длины свободного пробега надо вводить две средние длины свободного пробега: для молекул первого и второго газов. Эти длины свободного пробега являются характеристиками рассматриваемой смеси газов. Теория получается довольно громоздкой и плохо согласуется с экспериментом. Лучшие результаты дают теории, в которых понятие длины свободного пробега не используется. Мы изложим в § 92 один из вариантов приближенных теорий такого типа, в основе которого лежит *формула Эйнштейна*, выводимая в следующем параграфе.

§ 91. Связь диффузии с подвижностью частицы

1. Понятием *подвижности* мы уже пользовались в § 64 при рассмотрении теории броуновского движения. Там это понятие применялось к броуновским частицам, движущимся в жидкости под действием постоянных сил. Но оно применимо и к частицам молекулярных и атомных размеров, например к ионам, движущимся в электролите. На заряженный ион в постоянном электрическом поле действует постоянная сила $\mathbf{F}$, пропорциональная напряженности электрического поля. Кроме того, при движении ион испытывает столкновения с окружающими молекулами и ионами. При отсутствии внешней силы $\mathbf{F}$ движение иона было бы *полностью беспорядочным*. При наличии постоянной силы $\mathbf{F}$ появляется *преимущественное направление скорости иона*. На беспорядочное тепловое движение накладывается *регулярное движение* в направлении силы $\mathbf{F}$. Скорость этого регулярного движения обозначим через $\mathbf{u}$. Вектор $\mathbf{u}$ есть не что иное, как средняя скорость, с которой перемещается ион: $\mathbf{u} = \langle \mathbf{v} \rangle$. Если электрическое поле не очень сильное, то установившаяся скорость иона пропорциональна приложенной к нему силе $\mathbf{F}$. К этому случаю и относится понятие подвижности. *Подвижностью частицы называется коэффициент пропорциональности B между регулярной скоростью* $\mathbf{u}$ *и силой* $\mathbf{F}$:

$$\mathbf{u} = B\mathbf{F}. \tag{91.1}$$

Несущественно, что частицами являются ионы. Все сказанное справедливо и для молекул, и для любых других частиц.

2. Допустим теперь, что имеется «газ» каких-то частиц в постоянном и однородном силовом поле. «Газ» настолько разрежен, что силам взаимодействия между его частицами можно полностью пренебречь.

Примером такого «газа» может служить совокупность броуновских частиц, взвешенных в жидкости. Другим примером является обычный идеальный газ в силовом поле. Если $\mathbf{F}$ — сила, действующая на частицу «газа» в силовом поле, то потенциальная энергия ее в этом поле будет $\varepsilon_p = -Fx$. (Предполагается, что ось X направлена в сторону действующей силы.) Если состояние стационарно, а температура постоянна, то концентрация частиц «газа» меняется в пространстве в соответствии с формулой Больцмана

$$n = n_0 \exp\Big(-\frac{\varepsilon_p}{kT}\Big) = n_0 \exp\frac{Fx}{kT}. \tag{91.2}$$

Но микропроцессы не прекращаются даже тогда, когда состояние стационарно. Поскольку есть градиент концентрации, в газе происходит диффузия. Диффузионный поток в положительном направлении оси X определяется выражением $\Gamma_{\text{диф}} = -D\,dn/dx$. А так как «газ» находится в силовом поле, то существует также и «силовой поток» молекул с плотностью $\Gamma_{\text{сил}} = B\,Fn$. В состоянии равновесия должно быть

$$-D\frac{dn}{dx} + B\,Fn = 0.$$

Подставляя сюда выражение (91.2), получаем после сокращения на n

$$D = kT\,B. \tag{91.3}$$

Это соотношение между диффузией и подвижностью частицы было установлено Эйнштейном и носит его имя.

§ 92. Концентрационная диффузия в газах

1. Рассмотрим теперь смесь двух различных газов с концентрациями n_1 и n_2, изменяющимися в направлении оси X. Давление и температура смеси предполагаются постоянными, так что общая концентрация $n = n_1(x) + n_2(x)$ — одна и та же во всем газе. Диффузионные потоки газов определяются выражениями

$$\Gamma_1 = -D_{12}\frac{dn_1}{dx}, \quad \Gamma_2 = -D_{21}\frac{dn_2}{dx},$$

где D_{12} — *коэффициент диффузии газа 1 в газ 2*, а D_{21} — *коэффициент диффузии газа 2 в газ 1*. Благодаря наличию диффузионных потоков на тепловое движение газов накладывается упорядоченное движение их в направлении оси X. Скорости такого упорядоченного движения обозначим через $\mathbf{u}_1$ и $\mathbf{u}_2$. Согласно соотношению Эйнштейна (91.3) вычисление коэффициентов диффузии D_{12} и D_{21} сводится к вычислению подвижностей молекул газов. Этим методом мы и воспользуемся.

2. Вычислим подвижность B_1 молекул первого газа. Для этого рассмотрим какую-либо одну молекулу этого газа, которую назовем молекулой 1. Решим сначала следующую задачу. Какая постоянная сила $\mathbf{F}_1$ должна действовать на молекулу 1, чтобы поддерживать ее регулярное движение с постоянной скоростью $\mathbf{u}_1$? Если эта сила будет найдена, то подвижность B_1 найдется из соотношения $\mathbf{u}_1 = B_1\mathbf{F}_1$. Ясно, что сила $\mathbf{F}_1$ в среднем должна

уравновешиваться силами ударов, действующими на молекулу 1 при столкновениях. При ее вычислении можно отвлечься от изменения концентраций n_1 и n_2 в пространстве и считать эти концентрации постоянными. Тогда столкновения молекулы 1 с молекулами того же (первого) газа можно не принимать во внимание. Они движутся с той же упорядоченной скоростью $\mathbf{u}_1$, а потому столкновения с ними не вносят никакого вклада в величину интересующей нас силы. Надо учесть столкновения только с молекулами второго газа.

Пусть z_{12} — число столкновений, претерпеваемых молекулой 1 в одну секунду с молекулами второго газа, а $\Delta\mathbf{p}_1$ — изменение ее импульса при одном столкновении. Полное изменение импульса молекулы 1 в одну секунду в результате столкновений с молекулами второго газа будет $z_{12}\,\Delta\mathbf{p}_1$. Если у этой величины изменить знак и усреднить ее по всем столкновениям, то мы и получим интересующую нас силу $\mathbf{F}_1$. Среднее значение произведения двух величин, вообще говоря, нельзя заменять произведением средних значений этих величин. Однако если сделать такую замену, то это может сказаться только на несущественном числовом коэффициенте порядка единицы. Поэтому в целях упрощения вычислений примем $\mathbf{F}_1 = -\overline{z}_{12}\,\langle\mathbf{p}_1\rangle$. Среднее число столкновений $\overline{z}_{12}$ дается выражением (86.8). Что касается величины Δp_1, то при усреднении выпадет та часть ее, которая связана с тепловым движением. Поэтому от теплового движения молекулы 1 можно отвлечься и написать $\langle\Delta\mathbf{p}_1\rangle = m_1\,\langle\Delta\mathbf{u}_1\rangle$. Точно так же можно отвлечься от теплового движения молекулы 2, с которой сталкивается молекула 1. Таким образом, задача свелась к рассмотрению столкновения двух молекул с массами m_1 и m_2, которые в лабораторной системе отсчета движутся со скоростями $\mathbf{u}_1$ и $\mathbf{u}_2$. Центр масс этих молекул движется со скоростью $V = \dfrac{m_1\mathbf{u}_1 + m_2\mathbf{u}_2}{m_1 + m_2}$. Если молекулы считать идеально упругими шарами, то в системе центра масс они рассеиваются сферически симметрично (см. задачу к § 87). Это означает, что в системе центра масс средние скорости упорядоченного движения молекул после столкновения равны нулю. Значит, в лабораторной системе обе эти скорости будут равны $\mathbf{V}$. Поэтому для среднего изменения скорости $\mathbf{u}_1$ при столкновении получим

$$\langle\Delta\mathbf{u}_1\rangle = \mathbf{V} - \mathbf{u}_1 = \frac{m_2}{m_1 + m_2}\,(\mathbf{u}_2 - \mathbf{u}_1),$$

а потому

$$\mathbf{F}_1 = \overline{z}_{12}\,\frac{m_1 m_2}{m_1 + m_2}\,(\mathbf{u}_1 - \mathbf{u}_2) = \overline{z}_{12}\,\mu(\mathbf{u}_1 - \mathbf{u}_2), \tag{92.1}$$

где μ — приведенная масса.

Заметим, между прочим, что если бы мы рассматривали столкновения молекулы 1 с молекулами первого газа, то мы пришли бы к такому выражению, в котором вместо массы m_2 стояла бы масса m_1, а вместо разности $\mathbf{u}_1 - \mathbf{u}_2$ — разность $\mathbf{u}_1 - \mathbf{u}_1$, т. е. нуль. Это лишний раз доказывает, что столкновения с молекулами первого газа можно не принимать во внимание.

Учтем теперь, что $n_1\mathbf{u}_1 + n_2\mathbf{u}_2 = 0$. Если бы это было не так, то нарушилось бы постоянство полной концентрации частиц n, что привело бы к макроскопическому движению всего газа. С учетом этого

$$\mathbf{F}_1 = \frac{n}{n_2}\,\mu\overline{z}_{12}\,\mathbf{u}_1,$$

а потому

$$B_1 = \frac{n_2}{n\mu\overline{z}_{12}}.$$

Подставляя сюда значение $\overline{z}_{12}$ из (86.8), получим

$$B_1 = \frac{1}{\mu n \sigma_{12} \sqrt{\overline{v}_1^2 + \overline{v}_2^2}}. \tag{92.2}$$

Формула Эйнштейна (91.3) окончательно дает

$$D_{12} = D_{21} = \frac{kT}{\mu n \sigma_{12} \sqrt{\overline{v}_1^2 + \overline{v}_2^2}}. \tag{92.3}$$

В формулу (92.3) не входят ни n_1, ни n_2. Отсюда следует, что *коэффициент концентрационной диффузии газов не зависит от концентрации компонентов газовой смеси*. Этот вывод поддерживается опытом.

§ 93. Броуновское движение как процесс диффузии

1. В § 64 теория броуновского движения строилась на основе рассмотрения движения *одной частицы*. Но к этому явлению можно подойти с другой точки зрения. Можно рассматривать совокупность одинаковых броуновских частиц в жидкости как некоторый «газ», заполняющий пространство. При наличии градиентов концентрации в таком газе из-за броуновского движения будет происходить диффузия. Можно выразить коэффициент диффузии D через средний квадрат смещения броуновской частицы, которое она претерпевает за определенное время τ. Если в полученное соотношение вместо D подставить его выражение по формуле Эйнштейна (91.3), то в результате должна получиться основная формула теории броуновского движения (64.4). Целью настоящего параграфа и является рассмотрение теории броуновского движения с этой точки зрения.

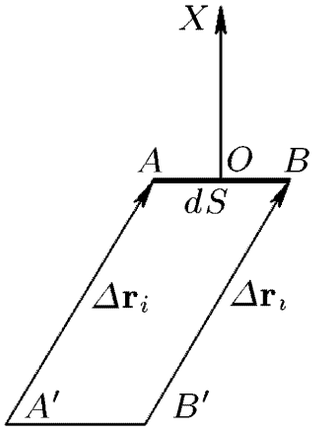


Рис. 85

Пусть в однородной жидкости в отсутствие внешних силовых полей распределены тождественные броуновские частицы с концентрацией $n(x)$, меняющейся только в направлении оси X. Вычислим диффузионный поток Γ таких частиц через произвольное сечение, перпендикулярное к оси X. Возьмем в этом сечении бесконечно малую площадку dS (рис. 85). Выделим группу броуновских частиц, которые за время τ смещаются на один и тот же вектор $\Delta\mathbf{r}_i$. Пусть будет велика не только полная концентрация броуновских частиц n, но и концентрация их $n_i(x)$ в каждой группе. Число частиц i-й группы dN_i, проходящих через площадку dS за время τ, будет равно числу их в косом цилиндре $ABB'A'$ с основанием AB и образующей $\Delta\mathbf{r}_i$, т. е.

$$dN_i = \int n_i(x)\, dV.$$

Линейные размеры площадки dS можно выбрать малыми по сравнению с $\Delta\mathbf{r}_i$. Тогда элемент объема dV можно представить в виде $dV = dS\, dx$ и написать

$$dN_i = dS \int n_i(x)\, dx = dS \int\limits_{-\Delta x_i}^{0} n_i(x_0 + \xi)\, d\xi,$$

где x_0 — координата центра O площадки dS. Выбрав τ, а следовательно, и Δx_i достаточно малыми, разложим функцию $n_i(x_0+\xi)$ по степеням ξ и оборвем это разложение на линейном члене. Тогда

$$dN_i = n_i(0)\,dS \int\limits_{-\Delta x_i}^{0} d\xi + \Big(\frac{dn_i}{dx}\Big)_{x=x_0} dS \int\limits_{-\Delta x_i}^{0} \xi\,d\xi,$$

или после интегрирования

$$dN_i = dS\Big[n_i\Delta x_i - \frac{1}{2}\,\frac{dn_i}{dx}\,(\Delta x_i)^2\Big].$$

Аргумент x_0 мы опустили, предполагая, что концентрация n_i и ее производная dn_i/dx берутся в центре площадки dS.

Избыток dN броуновских частиц, проходящих через площадку dS в положительном направлении оси X, над числом частиц, проходящих в противоположном направлении, найдется суммированием предыдущего выражения по всем группам частиц:

$$dN = dS\sum_i n_i\Delta x_i - \frac{dS}{2}\sum_i \frac{dn_i}{dx}\,(\Delta x_i)^2.$$

Среднее значение первой суммы равно нулю. Действительно, концентрации n_i относятся к центру площадки dS, а смещения броуновских частиц в положительном и отрицательном направлениях равновероятны. Для вычисления второй суммы заметим, что по определению среднего

$$n\langle\Delta x^2\rangle = \sum n_i(\Delta x_i)^2.$$

Величины Δx_i, как независимые параметры, не зависят от x. Средний квадрат смещения $\langle\Delta x^2\rangle$ также не может зависеть от x ввиду однородности жидкости и отсутствия силовых полей. Поэтому дифференцирование предыдущего соотношения по x дает

$$\langle\Delta x^2\rangle\,\frac{dn}{dx} = \sum \frac{dn_i}{dx}\,(\Delta x_i)^2.$$

В результате для среднего значения $d\overline{N}$ получаем

$$d\overline{N} = -dS\,\frac{\langle\Delta x^2\rangle}{2}\,\frac{dn}{dx}.$$

Чтобы найти средний диффузионный поток броуновских частиц Γ, надо эту величину разделить на dS и τ. Таким путем получаем

$$\Gamma = -\frac{\langle\Delta x^2\rangle}{2\tau}\,\frac{dn}{dx}. \tag{93.1}$$

Отсюда видно, что выравнивание концентраций броуновских частиц можно рассматривать как *процесс диффузии* с коэффициентом диффузии

$$D = \frac{1}{2}\,\frac{\langle\Delta x^2\rangle}{\tau}. \tag{93.2}$$

2. Так как по своему смыслу коэффициент диффузии D не может зависеть от произвольно выбранного времени τ, то уже из формулы (93.2) видно, что $\langle\Delta x^2\rangle \sim \tau$. Эта зависимость была проверена Перреном в опытах, описанных в § 64. Далее, используя формулу (91.3), получаем

$$\langle\Delta x^2\rangle = 2kTB\tau. \tag{93.3}$$

Это — основная формула теории броуновского движения. Она отличается от формулы (64.4) только обозначениями. Правда, в формуле (64.4) речь шла о смещениях *одной и той же частицы* в последовательные и равные друг другу промежутки времени. Формула же (93.3) предполагает, что происходят одновременные смещения *множества броуновских частиц*. Однако ввиду тождественности броуновских частиц и однородности жидкости между этими смещениями по существу нет разницы.

Если ввести средний квадрат полного смещения броуновской частицы, то получится

$$D = \frac{\langle \Delta r^2 \rangle}{6\tau}. \tag{93.4}$$

3. Формулы (93.2) и (93.4), конечно, справедливы и для диффузии молекул газа или жидкости. Однако точное вычисление коэффициента диффузии по этим формулам здесь практически невозможно. Для этого надо было бы точно знать средний квадрат смещения молекулы за время τ. Практически возможны лишь приближенные оценки. Оценим, например, с помощью формулы (93.4) коэффициент самодиффузии молекул газа. Упрощая расчет, будем считать, что время между двумя последовательными столкновениями молекулы одно и то же. Возьмем его в качестве τ. Кроме того, будем считать, что и смещение молекулы между двумя последовательными столкновениями одно и то же и равно $\lambda = \overline{v}\tau$. Подставляя в формулу (93.4) $\langle \Delta r^2 \rangle = \lambda^2$, получим

$$D = \frac{\lambda^2}{6\tau} = \frac{\lambda \overline{v} \tau}{6\tau} = \frac{1}{6}\lambda \overline{v}.$$

Этот результат отличается от ранее выведенной формулы (90.3) числовым множителем $1/2$.

§ 94. Термическая диффузия в газах

1. Если концы прямолинейной трубы, заполненной однородной смесью двух различных газов, имеют разные температуры, то вдоль тубы возникают диффузионные потоки этих газов, направленные в противоположные стороны. Направления потоков определяются следующим правилом. Если массы молекул m_1 и m_2 не слишком близки, то более массивные молекулы стремятся перейти в более холодные области, а менее массивные — в более нагретые. Если же m_1 и m_2 почти одинаковы, но молекулы имеют разные размеры, то в холодные области устремляются более крупные молекулы, а в теплые — менее крупные. Описанное явление называется *термической диффузией* или, короче, *термодиффузией*.

С феноменологической точки зрения возникновение потока частиц под влиянием градиента температуры совершенно естественно. Термодиффузия при наличии температурного градиента возникает вследствие столкновений между разнородными молекулами. Однако более конкретное наглядное толкование этого явления весьма затруднительно. Это, по-видимому, имеет глубокие основания. Направление термодиффузионного потока существенно зависит от характера взаимодействия между молекулами. Например, если молекулы считать силовыми центрами, отталкивающимися друг от друга по закону $F \sim 1/r^{\nu}$,

то, как показывает строгая теория, термодиффузионные потоки меняют знак при переходе через $\nu = 5$. При $\nu = 5$ термодиффузия исчезает. Интересно отметить, что закон $F \sim 1/r^5$ был принят Максвеллом, так как в этом случае вычисления сильно упрощаются и становятся выполнимыми. Поэтому-то термодиффузия и осталась им незамеченной. Это явление теоретически было предсказано независимо друг от друга Энскогом и Чепменом в 1917 г. На опыте термодиффузия в газах наблюдалась Чепменом и Дутсоном в том же году. Строгие теории термодиффузии Энскога и Чепмена очень сложны и громоздки. Излагать их мы не имеем возможности. Попытки же элементарной трактовки явления термодиффузии малоутвердительны, и нет смысла на них останавливаться. Термическая диффузия в жидкостях была обнаружена экспериментально в середине прошлого века.

2. В результате термодиффузии происходит частичное распределение газовой смеси и связанное с ним возникновение градиентов концентрации каждого из компонентов. Это вызывает концентрационную диффузию, стремящуюся устранить указанное разделение. Если концы трубы поддерживаются при постоянных температурах, то в трубе в конце концов установится стационарное состояние с вполне определенным распределением температуры и концентрации. Допустим, например, что на концах трубы, заполненной однородной смесью, состоящей из 50 % молекул O_2 и 50 % молекул N_2, поддерживаются температуры 0 и 500 °C. Тогда, как показывает опыт, в установившемся состоянии у более нагретого конца будет 50,27 % N_2 и 49,73 % O_2, а у более холодного 49,73 % N_2 и 50,27 % O_2.

Пусть газовая смесь состоит из двух компонентов 1 и 2. При наличии и градиента концентрации, и градиента температуры поток частиц компонента 1 записывается в виде

$$\Gamma_1 = -nD_{12}\frac{dc_1}{dx} + n\frac{D_T}{T}\frac{dT}{dx}, \tag{94.1}$$

где $c_1 = n_1/n = n_1/(n_1 + n_2)$ — относительная концентрация этого компонента, D_{12} — коэффициент диффузии, D_T — *коэффициент термодиффузии*. Поток Γ_1 можно также записать в виде

$$\Gamma_1 = -nD_{12}\left(\frac{dc_1}{dx} - \frac{k_T}{T}\frac{dT}{dx}\right). \tag{94.2}$$

Величина $k_T = D_T/D_{12}$ называется *термодиффузионным отношением*. В стационарном состоянии в закрытой трубе $\Gamma_1 = \Gamma_2 = 0$, а потому

$$\frac{dc_1}{dx} = \frac{k_T}{T}\frac{dT}{dx}.$$

Если k_T считать постоянным, то интегрирование дает

$$c_1(T) - c_1(T_0) = k_T \ln\frac{T}{T_0}. \tag{94.3}$$

Разность концентраций $c_1(T) - c_1(T_0)$ при температурах T и T_0 называется *разделением*. Обычно k_T почти не зависит от температуры,

однако заметно изменяется при изменении состава, а потому формула (94.3) сохраняет смысл лишь при условии, что неоднородность состава в сосуде незначительна.

3. Термическая диффузия (как и обычная концентрационная диффузия) в сочетании с тепловой конвекцией используется на практике для *разделения изотопов*. Метод этот был предложен и осуществлен Клузиусом и Дикелем в 1938 г. Принцип метода весьма прост. Представим себе закрытый высокий прямоугольный ящик, наполненный смесью двух газов и поставленный вертикально (рис. 86). Пусть противоположные стенки ящика поддерживаются при разных температурах. Вследствие термодиффузии более легкий газ будет собираться у горячей стенки, тяжелый — у холодной. Разность равновесных концентраций каждого компонента у горячей и холодной стенок определяется формулой (94.3). Обогащенный легким компонентом газ у горячей стенки будет подниматься вверх, а обогащенный тяжелым компонентом — опускаться вниз у холодной стенки (тепловая конвекция). В результате этого вверху будет преимущественно концентрироваться легкий компонент смеси, а внизу — тяжелый. На практике применяют вертикальные цилиндрические трубки длиной 5–10 м и диаметром до 10 мм. По оси трубки натягивается платиновая проволока, нагреваемая электрическим током до 1000–1700 °С. Платиновая проволока играет роль горячей, а внутренняя поверхности трубки — холодной стенок. Из таких трубок, поставленных друг над другом, образуется *разделительная колонка*. Обогащенный легким компонентом газ из каждой трубки поступает в вышестоящую трубку и там обогащается дальше.

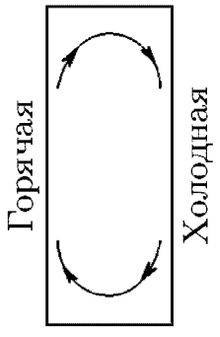


Рис. 86

§ 95. Явления в разреженных газах

1. *Если средняя длина свободного пробега λ того же порядка, что и характерный линейный размер сосуда d, в котором заключен газ, или больше, то состояние газа называют вакуумом.* Воздух в комнате, например, при атмосферном давлении в состоянии вакуума не находится, так как в этом случае $\lambda \sim 10^{-5}$ см. Однако в сосуде, линейные размеры которого меньше 10^{-5} см (поры дерева и многих других пористых тел), тот же воздух уже находится в условиях вакуума.

Различают три вида вакуума: 1) *низкий*, когда λ меньше характерного размера сосуда d, но приближается к нему; 2) *средний*, когда λ сравнима с d; 3) *высокий* (или *глубокий*), когда λ значительно больше d. Газ в состоянии высокого вакуума называется *ультраразреженным*.

В плотных газах $\lambda \ll d$. В этих случаях столкновения между молекулами самого газа играют основную роль в его поведении. Только такие случаи и имелись в виду во всем предшествующем изложении (за исключением § 89, п. 8). В другом предельном случае, когда газ

становится ультраразреженным, столкновения между самими молекулами относительно редки и перестают играть заметную роль. Основную роль в этом случае играют столкновения молекул со стенками сосуда. Это уже было показано на примере зависимости вязкости и теплопроводности газа от его плотности (см. § 89, п. 8).

Одной из особенностей высокого вакуума является невозможность возникновения в нем конвекционных потоков. Это связано с тем, что в высоком вакууме молекулы практически не сталкиваются между собой, а движутся от стенки к стенке совершенно независимо. Наиболее трудным для теории является случай среднего вакуума, когда $\lambda \sim d$.

2. Эффузия разреженного газа. Пусть сосуд разделен перегородкой на две части A и B. Часть A заполнена газом, в части B газа нет. Выделим мысленно на поверхности перегородки площадку s. Число молекул, ежесекундно ударяющихся об эту площадку, определяется формулой (75.5), т. е.

$$N = \frac{1}{4}\, n\overline{v} s = ns\sqrt{\frac{kT}{2\pi m}} = C\,\frac{Ps}{\sqrt{mT}}, \tag{95.1}$$

где C — постоянная, равная $\sqrt{1/2\pi k}$. Проделаем теперь в перегородке отверстие, площадь которого равна s. Чему равно число молекул, пролетающих ежесекундно через это отверстие из A и B? Ответ зависит от размеров отверстия, толщины перегородки и средней длины свободного пробега λ.

При обычных давлениях и не слишком малом отверстии средняя длина свободного пробега очень мала по сравнению с размерами отверстия. В этом случае вблизи отверстия возникает упорядоченное коллективное движение газа, направленное к отверстию. Его можно рассматривать как гидродинамическое течение, обусловленное разностью давлений в газе. Распределения концентрации и скоростей молекул газа вблизи отверстия претерпят существенные изменения по сравнению с теми, какими они были бы при отсутствии отверстия. Формула (95.1) к рассматриваемому случаю не применима, так как она выведена в предположении, что молекулы газа движутся хаотически. Но если размеры отверстия, а также толщина перегородки малы по сравнению с λ, то столкновения между молекулами перестают играть роль. Все определяется столкновениями молекул со стенками сосуда. Если в перегородке проделать малое отверстие, то площадь стенок, с которыми сталкиваются молекулы, изменится пренебрежимо мало. Это никак не скажется на распределении концентрации и скоростей молекул во всем сосуде, в частности и вблизи отверстия. В этом случае формула (95.1) применима.

Поток молекул газа через отверстие в стенке называется *эффузионным потоком*, если размеры отверстия и толщина стенки малы по сравнению с длиной свободного пробега λ.

Допустим теперь, что по разные стороны перегородки находится один и тот же газ, но при разных давлениях и температурах. Если газ

находится в состоянии высокого вакуума, то возникнут два эффузионных потока: из A в B и из B в A. Ввиду отсутствия столкновений между молекулами эти два потока совершенно независимы друг от друга. Поэтому количество молекул, ежесекундно проходящих через отверстие s из A в B, определится выражением

$$N = \frac{Cs}{\sqrt{m}}\left(\frac{P_A}{\sqrt{T_A}} - \frac{P_B}{\sqrt{T_B}}\right), \tag{95.2}$$

где P_A, P_B, T_A, T_B — давления и температуры газа в A и B. В состоянии равновесия, когда средние числа молекул в A и B остаются неизменными, должно быть $N = 0$, т. е.

$$\frac{P_A}{\sqrt{T_A}} = \frac{P_B}{\sqrt{T_B}}. \tag{95.3}$$

Наконец, рассмотрим случай, когда по разные стороны перегородки находятся разные газы: в части A — газ 1 с молекулами массы m_1, в части B — газ 2 с молекулами массы m_2. В результате эффузии газ 1 проникнет в B, а газ 2 — в A. Пусть $P_{1,A}$ и $P_{1,B}$ — парциальные давления газа 1 по разные стороны перегородки. Аналогичные обозначения введем для газа 2. Поток газа 1 из A в B будет

$$N_1 = \frac{Cs}{\sqrt{m_1}}\left(\frac{P_{1,A}}{\sqrt{T_A}} - \frac{P_{1,B}}{\sqrt{T_B}}\right).$$

Обратный поток газа 2 из B в A:

$$N_2 = \frac{Cs}{\sqrt{m_2}}\left(\frac{P_{2,B}}{\sqrt{T_B}} - \frac{P_{2,A}}{\sqrt{T_A}}\right).$$

В начальный момент, когда $P_{2,A} = P_{1,B}$,

$$\frac{N_1}{N_2} = \sqrt{\frac{m_2}{m_1}}\sqrt{\frac{T_B}{T_A}}\,\frac{P_1}{P_2}. \tag{95.4}$$

В частности, когда температуры и начальные давления одинаковы,

$$\frac{N_1}{N_2} = \sqrt{\frac{m_2}{m_1}}.$$

Эффузионные потоки при прочих равных условиях обратно пропорциональны квадратным корням из масс молекул. На этом основан один из методов разделения изотопов. В нем используется эффузия через мембрану со множеством малых отверстий.

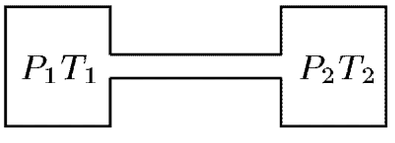


Рис. 87

3. Тепловая эффузия. Пусть два сосуда 1 и 2 соединены между собой трубкой (рис. 87) и поддерживаются при разных температурах T_1 и T_2. Когда поперечное сечение трубки очень велико по сравнению с длиной свободного пробега, газ можно рассматривать как сплошную среду. Условие равновесия в этом случае носит гидродинамический характер: должны быть равны

давления P_1 и P_2 в обоих сосудах. В противоположном случае, когда длина свободного пробега очень велика по сравнению с поперечными размерами трубки, гидродинамический подход неприменим. Условие равновесия требует, чтобы среднее число частиц газа, проходящих через трубку в одном направлении, было равно среднему числу частиц, проходящих в противоположном направлении. Это условие приводит к соотношению (95.3), или в новых обозначениях

$$\frac{P_1}{\sqrt{T_1}} = \frac{P_2}{\sqrt{T_2}}. \tag{95.5}$$

Следовательно, если температуры T_1 и T_2 различны, то при равновесии будут различны и давления P_1 и P_2.

Допустим теперь, что сосуд разделен пористой перегородкой на две части, поддерживаемые при разных температурах T_1 и T_2. Пусть размеры пор малы по сравнению с длиной свободного пробега. (Обычно пористые перегородки удовлетворяют этому условию уже при атмосферном давлении.) Тогда к рассматриваемому случаю применимо соотношение (95.5). Если первоначальные давления P_1 и P_2 были равны, то газ начнет перетекать в направлении от более низкой к более высокой температуре. Это явление называется *тепловой диффузией* или *эффектом Кнудсена* (1871–1949).

Поль (1884–1976) предложил следующую демонстрацию этого явления. Берется пористый стакан из необожженной глины, внутри которого находится электрический нагреватель (рис. 88). Воздух из сосуда может выходить наружу через стеклянную трубку, нижний конец которой погружен в воду. Так как температура внутри сосуда выше температуры окружающего воздуха, то наружный воздух непрерывно всасывается внутрь сосуда. Давление в сосуде повышается, и излишек воздуха непрерывно выходит через стеклянную трубку в виде пузырьков.

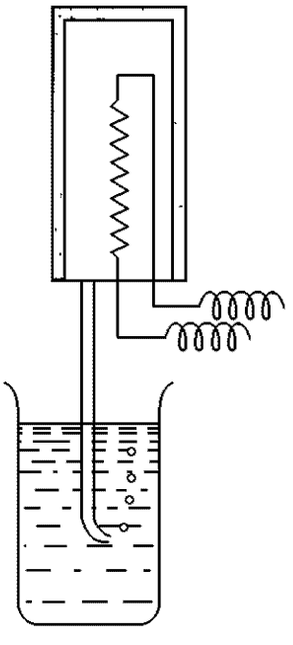


Рис. 88

Тепловая диффузия играет важную роль в явлениях природы. Днем поверхность земли нагревается солнечными лучами. Воздух из более глубоких слоев почвы выходит по капиллярам на поверхность и рассеивается ветром. Ночью наружный слой почвы охлаждается и возникает обратный поток воздуха с поверхности в более глубокие слои почвы. Так возникает обмен воздуха в почве, необходимый для нормальной жизни растений.

4. Условие (95.5) не так просто обосновать. Оно было бы очевидным, если бы соединительная трубка была бесконечно короткой. Тогда ее можно было бы рассматривать как малое отверстие в стенке между сосудами 1 и 2. Но если соединительная трубка длинная, то дело обстоит сложнее. Допустим, что соединительная трубка имеет цилиндрическую форму. С левого конца в нее ежесекундно входит $N_1 = (1/4)\, n_1 \overline{v}_1 s$ частиц. Часть из этих частиц отражается обратно

в сосуд 1, часть проходит в сосуд 2. Число прошедших частиц можно представить в виде $N_{12} = (1/4)\, n_1 \overline{v}_1 s \alpha_{12}$, где α_{12} — «коэффициент прохождения» в направлении от сосуда 1 к сосуду 2. В обратном направлении из 2 в 1 проходит $N_{21} = (1/4)\, n_2 \overline{v}_2 \alpha_{21}$ частиц, где α_{21} — коэффициент прохождения в этом направлении. В установившемся состоянии $N_{12} = N_{21}$, т. е.

$$\alpha_{12} n_1 \overline{v}_1 = \alpha_{21} n_2 \overline{v}_2. \tag{95.6}$$

Трудность вопроса состоит в доказательстве соотношения $\alpha_{12} = \alpha_{21}$. Что это соотношение, по-видимому, справедливо, показывают следующие соображения. Коэффициент прохождения ультраразреженного газа через трубку не может зависеть от его давления, так как молекулы такого газа между собой практически не сталкиваются, а претерпевают столкновения только со стенками трубки. Значительно труднее выяснить влияние температуры. Значения коэффициентов α_{12} и α_{21} зависят от характера взаимодействия молекул со стенкой при столкновениях. Допустим, что молекулы газа приходят в тепловое равновесие со стенкой в результате уже одного или немногих столкновений, причем отражение их является изотропным. Если эта гипотеза справедлива, то относительная доля молекул, выбывающих из пучка при отражении, зависит только от температуры точки, в которой произошло столкновение, но не будет зависеть от направления распространения пучка. Один пучок распространяется в сторону повышения, другой — в сторону понижения температуры. Точки на поверхности трубки, в которых молекулы отражаются и выбывают из пучков, проходятся пучками в обратной последовательности. Но это обстоятельство не может сказаться на потере частиц в результате всех отражений, а потому $\alpha_{12} = \alpha_{21}$. Тогда (95.6) сводится к

$$n_1 \overline{v}_1 = n_2 \overline{v}_2, \tag{95.7}$$

а это соотношение уже легко приводится к виду (95.5). То обстоятельство, что закон (95.5) подтверждается на опыте, может рассматриваться как экспериментальное доказательство соотношения $\alpha_{12} = \alpha_{21}$.

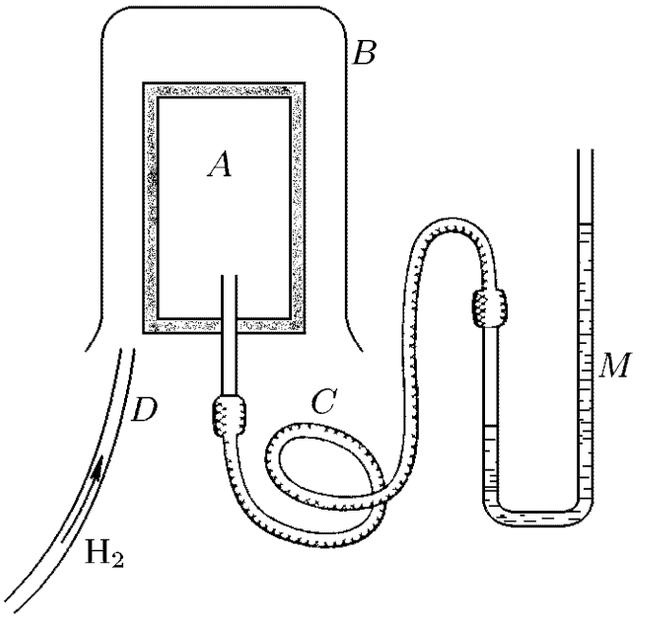


Рис. 89

5. Изотермическая эффузия через пористую перегородку. Допустим, что сосуд разделен на две части пористой перегородкой. Пусть по разные стороны перегородки находятся разные газы, а размеры пор малы по сравнению с длиной свободного пробега. Предположим, что давление и температуры газов одинаковы. Тогда будет справедливо соотношение (95.5). Если $m_1 > m_2$, то $N_1 < N_2$. Это значит, что более легкий газ будет быстрее проходить через пористую перегородку, чем более тяжелый. Явление называется *изотермической эффузией через пористую перегородку*.

Возьмем стакан A из пористой глины, соединенный резиновой трубкой C с U-образным водяным манометром M (рис. 89). Сверху

наденем на него более широкий стеклянный стакан B. Подведем под стеклянный стакан резиновую трубку D, по которой пропускается водород. Вследствие изотермической эффузии водород быстрее поступает внутрь пористого стакана, чем воздух выходит из него. Манометр покажет увеличение давления газа внутри пористого стакана. Манометр покажет увеличение давления газа внутри пористого стакана. Если снять стеклянный стакан, то наоборот, давление газа в пористом стакане начнет падать и сделается меньше давления наружного воздуха.

6. Тепловое скольжение. Допустим, что поверхность тела нагрета неравномерно. Для простоты предположим, что эта поверхность плоская, а температура возрастает в направлении оси X (рис. 90). Примыкающий к поверхности тела газ становится также неравномерно нагретым. Молекулы газа при отражении от тела передают ему не только нормальный, но и тангенциальный импульс. Но так как молекулы приходят справа с большими тепловыми скоростями, то они передают телу больший тангенциальный импульс, чем молекулы, приходящие слева. В результате возникает тангенциальная составляющая силы, действующая на тело справа налево. По третьему закону Ньютона на пристеночный слой газа должна действовать равная и противоположно направленная сила. Газ придет в движение в направлении оси X, т. е. в сторону возрастания температуры. Это явление называется *тепловым скольжением*.

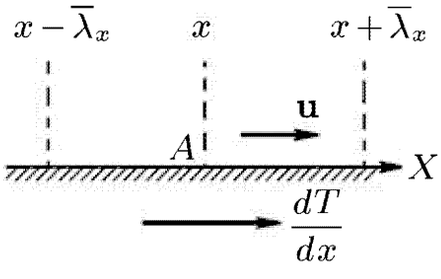


Рис. 90

Нетрудно оценить скорости газа u, когда процесс теплового скольжения станет стационарным. Пусть $\overline{v}_x$ означает среднее значение модуля x-составляющей тепловой скорости молекулы газа, а $\overline{\lambda}_x$ — среднее значение модуля проекции длины свободного пробега на ось X. Рассмотрим какую-либо точку A на поверхности тела с координатой x. При рассмотрении передачи импульса в точке A можно рассуждать так, как если бы все молекулы, попадающие в эту точку, испытали последние столкновения в плоскостях $x+\overline{\lambda}_x$ и $x-\overline{\lambda}_x$ (см. § 89, п. 2). Если газ скользит со скоростью u, то средние значения скорости молекулы вдоль оси X в этих плоскостях будут соответственно $\overline{v}_x(x+\overline{\lambda}_x)-u$ и $\overline{v}_x(x-\overline{\lambda}_x)+u$. При стационарном скольжении передача тангенциального импульса от газа к телу и обратно прекратится. Это будет при выполнении условия

$$\overline{v}_x(x+\overline{\lambda}_x)-u=\overline{v}_x(x-\overline{\lambda}_x)+u, \quad \text{откуда} \quad u=\overline{\lambda}_x\frac{d\overline{v}_x}{dx}.$$

Очевидно, $\overline{\lambda}^2=\overline{\lambda}_x^2+\overline{\lambda}_y^2+\overline{\lambda}_z^2=3\overline{\lambda}_x^2$. Не внося существенной ошибки, положим $\overline{\lambda}_x=\lambda/\sqrt{3}$. Далее $m\overline{v}_x^2\approx kT$. Используя эти соотношения, получим

$$u\approx\frac{\lambda}{2}\sqrt{\frac{k}{3mT}}\frac{dT}{dx}\approx\frac{\lambda}{3}\sqrt{\frac{k}{mT}}\frac{dT}{dx}. \tag{95.8}$$

Отсюда видно, что тепловое скольжение может быть заметным лишь в разреженных газах, так как $\lambda\sim 1/P$.

7. Радиометрический эффект. *Радиометрический эффект состоит в том, что неравномерно нагретые тела, помещенные в разреженных газах, самопроизвольно приходят в движение в направлении от более нагретой стороны к менее нагретой.* Неравномерное нагревание обычно осуществляется односторонним освещением тела, с чем и связано название эффекта. Силы, приводящие тело в движение, называются *радиометрическими*. Они имеют двоякое происхождение.

Первая сила возникает из-за теплового скольжения газа от менее нагретых участков поверхности тела к более нагретым. Благодаря вязкости в движение вовлекается и основная часть газа в окрестности тела (см. рис. 91, где изображено в разрезе течение газа вокруг пластинки. Более нагретая поверхность ее зачернена). Благодаря закону сохранения импульса тело должно прийти в движение в обратном направлении, т. е. холодной стороной вперед. Значит, появляется сила, действующая на него в том же направлении. Такой силой объясняется, между прочим, оседание пыли на холодных стенках вблизи батарей центрального отопления, она перемещает взвешенные в воздухе пылинки в направлении от нагретых тел к холодным.

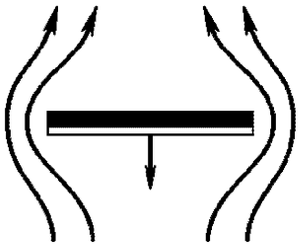


Рис. 91

Вторая сила имеет следующее происхождение. Молекулы газа при отражении от более нагретой стороны тела сообщают ему больший импульс, чем молекулы, отражающиеся от менее нагретой стороны. Поэтому и возникает радиометрическая сила, направленная от более нагретой к менее нагретой стороне тела.

Первая сила является преобладающей в слабо разреженных газах. Она обратно пропорциональна давлению, как в этом можно убедиться с помощью формулы (95.8). Вторая сила играет основную роль в сильно разреженных газах. Она пропорциональна давлению. В промежуточной области существенны обе силы.

Радиометрический эффект при низких давлениях удобно наблюдать с помощью *радиометра Крукса* (1832–1919). Основной частью этого прибора являются слюдяные крылышки, прикрепленные к колпачку, надетому на острие иглы. Таким образом, крылышки могут вращаться вокруг вертикальной оси практически без трения. Крылышки зачернены с одной стороны и помещены в стеклянном баллоне с высоким вакуумом. При освещении крылышки приходят во вращение светлой стороной вперед.

ЗАДАЧА

Изотермическая эффузия газа через пористую перегородку (поры которой малы по сравнению с длиной свободного пробега) используется для разделения изотопов. Естественная смесь изотопов помещается в сосуд с пористыми стенками. Газ, прошедший через поры сосуда в результате эффузии, откачивается и собирается в специальном резервуаре. С ним производится второй цикл эффузии, затем третий и т. д., пока не будет достигнута требуемая степень разделения изотопов. Сколько циклов эффузии

необходимо произвести, чтобы отношение концентраций частиц легкого и тяжелого изотопов увеличить в 10 раз, если молекулярные массы их равны соответственно μ_1 и μ_2?

Ответ. $N \geqslant \dfrac{2}{\lg(\mu_2/\mu_1)}$.

§ 96. Молекулярное течение ультраразреженного газа через прямолинейную трубу

1. Течение ультраразреженного газа через трубу существенно отличается от течения Пуазейля (см. т. I, § 97) вязкой жидкости или плотного газа. Это различие обусловлено тем, что течение ультраразреженного газа определяется исключительно столкновениями его молекул со стенками трубы. Столкновения молекулы между собой никакой роли не играют. Результатом этого является следующая особенность течения ультраразреженного газа. Движение молекул газа, входящих в трубу с одного конца, совершенно не зависит от движения молекул, поступающих на нее с другого конца. Полный поток молекул через трубу можно представить как разность двух независимых потоков, проходящих в противоположных направлениях. Если это условие выполняется, то течение газа называют *молекулярным течением*, или *течением Кнудсена*.

Рассмотрим стационарное молекулярное течение через трубу, длина которой l очень велика по сравнению с ее поперечным размером a. (В случае цилиндрической трубы под a будем понимать ее радиус.)

Допустим сначала, что через отверстие на одном конце трубы поступает ежесекундно N_1 молекул, а на другом конце поддерживается полный вакуум. Определим число молекул N, которые проходят через трубу и выходят их второго конца ее. Число N существенно зависит от характера отражения молекул от стенок трубы. Если бы, например, стенки трубы были абсолютно гладкими, а молекулы отражались от них зеркально, то все молекулы, вошедшие в трубу с одного конца, вышли бы из другого конца, т. е. было бы $N = N_1$. На самом деле такой идеализированный случай никогда не осуществляется. В реальном опыте значительная доля молекул, ударившихся о стенку трубы, летит обратно. Определить вид зависимости N от N_1 и от параметров трубы можно из соображений размерности. Из механизма явления следует, что должна существовать функциональная связь между величинами N, N_1, a, l. Из этих величин можно составить две независимые безразмерные комбинации, а именно N/N_1 и a/l. При течении по трубе одна из них должна быть функцией другой: $N/N_1 = f(a/l)$, так что

$$N = N_1 f(a/l).$$

Функция $f(a/l)$ зависит от формы поперечного сечения трубы, а также от характера отражения молекул от ее стенок. Очевидно $f(0) = 0$, так как при $a = 0$ выходящий поток N обращается в нуль, каково бы ни было значение N_1. Предполагая, что функция $f(a/l)$ разлагается в степенной ряд, произведем это разложение и оборвем его на линейном

члене. Тогда получим

$$N = C N_1 \frac{a}{l},$$

где C — постоянная, зависящая от формы поперечного сечения трубы и от характера отражения молекул от ее стенок. В частности, она может зависеть от того, как меняется температура стенки вдоль трубы.

Пусть теперь через один конец в трубу поступает ежесекундно N_1 молекул, а через другой — N_2. Ввиду независимости обоих потоков, через поперечное сечение трубы будет проходить число молекул, равное

$$N = C \frac{a}{l} (N_1 - N_2). \tag{96.1}$$

Можно представить себе, что труба соединяет два сосуда. В одном поддерживается давление P_1 и температура T_1, в другом — давление P_2 и температура T_2. Если концентрации молекул в сосудах равны n_1 и n_2 соответственно, то

$$N_1 = \frac{1}{4} S n_1 \overline{v}_1, \quad N_2 = \frac{1}{4} S n_2 \overline{v}_2,$$

где S — площадь поперечного сечения трубы.

Используя соотношения $\overline{v} = \sqrt{8kT/\pi m}$ и $P = nkT$, преобразуем выражение для N к виду

$$N = A \frac{a^3}{l\sqrt{m}} \Big(\frac{P_1}{\sqrt{T_1}} - \frac{P_2}{\sqrt{T_2}} \Big), \tag{96.2}$$

где A — новая постоянная:

$$A = \frac{C}{\sqrt{2\pi k}} \frac{S}{a^2}. \tag{96.3}$$

Для массы газа, ежесекундно протекающего через поперечное сечение трубы, получаем

$$Q = A\sqrt{m} \frac{a^3}{l} \Big(\frac{P_1}{\sqrt{T_1}} - \frac{P_2}{\sqrt{T_2}} \Big). \tag{96.4}$$

3. Числовые коэффициенты C и A можно оценить с помощью следующих элементарных соображений. Рассмотрим круглую трубу и будем предполагать, что температура газа одна и та же по всей трубе. Протекание газа через трубу можно рассматривать как процесс диффузии. Следовательно, $N = -DS\, dn/dx$, где $D = (1/3)\lambda\overline{v}$ — коэффициент диффузии (ось X направлена вдоль оси трубы). Для стационарного процесса $N = \text{const}$, а потому $dn/dx = \text{const}$. Значит, $dn/dx = (n_2 - n_1)/l$, и далее

$$N = \frac{1}{3} \lambda \overline{v} S \frac{n_1 - n_2}{l}.$$

При кнудсеновском течении столкновениями между молекулами можно полностью пренебречь. Длина свободного пробега полностью определяется столкновениями молекул со стенками трубы. По порядку

величины она равна диаметру трубы $2a$. Принимая это значение, получим

$$N = \frac{2}{3}\, a\overline{v} S\, \frac{n_1 - n_2}{l},$$

или

$$N = \frac{8}{3}\, \frac{a}{l}\, (N_1 - N_2). \tag{96.5}$$

Сравнение этой формулы с (96.1) показывает, что для круглой трубы

$$C = \frac{8}{3}, \quad A = \frac{8}{3}\sqrt{\frac{\pi}{2k}}. \tag{96.6}$$

Приведенный элементарный вывод показывает также, что в случае неизменности температуры вдоль трубы N и Q строго пропорциональны разности давлений $P_1 - P_2$. Напротив, при изменении температуры вдоль трубы пропорциональность между теми же величинами и разностью $(P_1/\sqrt{T_1} - P_2/\sqrt{T_2})$ — только приближенная. Она справедлива лишь до тех пор, пока в разложении функции $f(a/l)$ в степенной ряд можно ограничиться только линейным членом.

Формулы (96.2) и (96.3) при числовых значениях постоянных C и A (96.6) называются *формулами Кнудсена.*

4. Формула (96.4) показывает, что при прочих равных условиях *расход газа Q пропорционален кубу радиуса трубы.* Это должно учитываться при конструировании вакуумных установок. Допустим, что мощность высоковакуумного насоса позволяет откачивать в секунду V литров газа, а труба, соединяющая насос с откачиваемым баллоном, способна пропускать за то же время v литров. Если $v \ll V$, то применять мощный насос бессмысленно. Для правильного использования насоса размеры соединительной трубы надо выбирать так, чтобы было $v \sim V$.

5. Приведем теперь более строгий молекулярно-кинетический вывод полученных формул. Наиболее существенным моментом нашего вывода будет предположение относительно характера взаимодействия ударяющихся молекул со стенками трубы. Предположим, что после удара о стенку молекулы отражаются обратно так, что их скорости становятся распределенными по закону Максвелла при температуре, равной температуре стенки. Это предположение означает, что молекулы газа воспринимают температуру стенки, а их скорости становятся распределенными изотропно уже в результате однократных ударов о стенку. Хотя это и не совсем правильно, но такое предположение является простейшим и в рассматриваемом вопросе приводит в основном к правильным результатам.

Строго говоря, мы выражались не совсем точно. Отраженные молекулы движутся только от стенки, среди них нет молекул, движущихся к стенке. Поэтому о максвелловском распределении скоростей отраженных молекул можно говорить лишь условно. Смысл нашего предположения состоит в том, что если отраженные молекулы пополнить

таким же числом молекул, летящих с теми же, но противоположно направленными скоростями, то получится максвелловское распределение.

Допустим теперь, что с единичной площадкой S в одну секунду сталкивается $N_{\text{ст}}$ молекул. Найдем долю этих молекул $dN_{\text{ст}}$, отражающихся в телесный угол $d\Omega$, ось которого составляет угол ϑ с нормалью к площадке S (рис. 92). Так как по нашему предположению распределение отраженных молекул по углам и скоростям не зависит от скоростей и направления движения падающих молекул, то можно предположить, что падающие молекулы вместе с отраженными распределены по закону Максвелла. Пусть n — число всех молекул в единице объема. Тогда число молекул в телесном угле $d\Omega$, падающих в единицу времени на площадку S под углом ϑ к нормали, будет $n\overline{v}\,S\cos\vartheta\,\dfrac{d\Omega}{4\pi}$. Таково же будет и число молекул $dN_{\text{ст}}$, отразившихся в симметрично расположенный телесный угол по другую сторону нормали. Полное число падающих молекул дается формулой (75.5), т. е. $N_{\text{ст}} = (1/4)n\overline{v}$. Вводя его, получим,

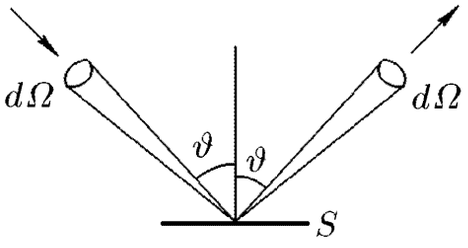


Рис. 92

$$dN_{\text{ст}} = \frac{N_{\text{ст}}}{\pi}\cos\vartheta\,d\Omega. \tag{96.7}$$

6. Вернемся к задаче о молекулярном течении газа через трубу. Трубу будем считать цилиндрической, радиуса a. Так как расход

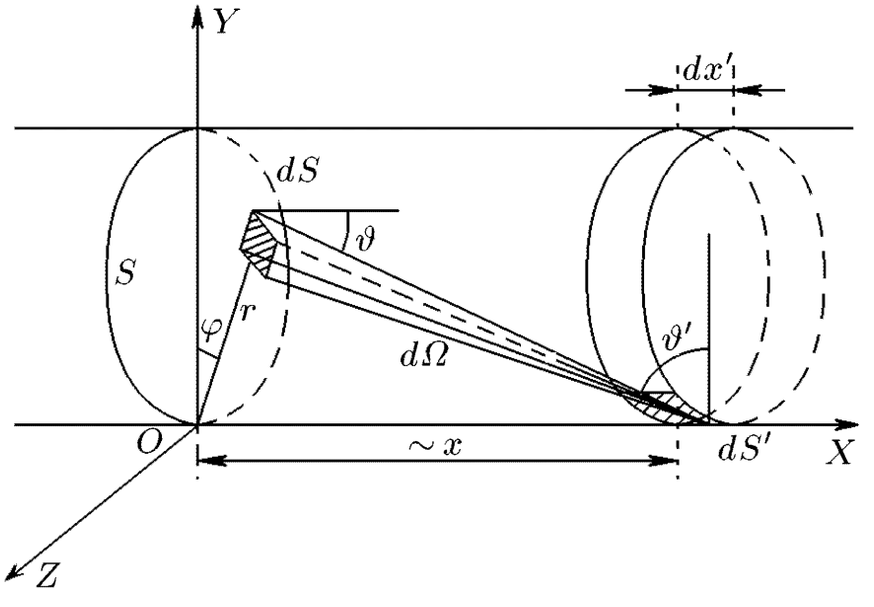


Рис. 93

газа Q один и тот же через все сечения трубы, то для его вычисления можно взять сечение S, проходящее через середину трубы (рис. 93). Плоскость сечения S примем за координатную плоскость YZ, ось X

направим по одной из образующих цилиндра. Пусть dS — элементарная площадка в сечении S. Возьмем на боковой поверхности цилиндра бесконечно короткий поясок ширины dx' и на нем элементарную площадку dS'. Из середины площадки dS' площадка dS видна под телесным углом $d\Omega = \dfrac{dS\cos\vartheta}{R^2}$. Число молекул dN, летящих от площадки dS' и проходящих через dS в одну секунду, определяется выражением

$$dN = \frac{N_{\text{ст}}}{\pi}\cos\vartheta'\,d\Omega = \frac{N_{\text{ст}}\cos\vartheta\cos\vartheta'}{\pi R^2}\,dS\,dS',$$

где R — расстояние между площадками, ϑ и ϑ' — углы между нормалями к ним и линией, соединяющей центры площадок. Величина $N_{\text{ст}}$ относится к месту нахождения площадки dS' и является функцией ее координаты x: $N_{\text{ст}} = N_{\text{ст}}(x)$.

Для определения полного числа молекул N, проходящих через сечение S в единицу времени, надо выражение для dN проинтегрировать по сечению S и боковой поверхности цилиндра. Но так как все площадки dS' на пояске расположены совершенно одинаково относительно сечения S, то dS' можно сразу заменить на площадь пояска $2\pi a\,dx$. Для упрощения вычислений вершину телесного угла $d\Omega$ можно поместить на оси Y, как это сделано на рис. 93. Пусть y и z — координаты центра площадки dS. Тогда

$$\cos\vartheta = \frac{x}{R},\quad \cos\vartheta' = \frac{y}{R},\quad N = 2a\int y\,dS\int\limits_{-l/2}^{+l/2} N_{\text{ст}}(x)\,\frac{x}{R^4}\,dx.$$

Если бы число ударов $N_{\text{ст}}$ было одинаково по всей длине трубы, т. е. не зависело от x, то подынтегральное выражение $N_{\text{ст}}(x)x/R^4$ было бы нечетной функцией x, и интеграл по x обратился бы в нуль. Заметив это и предполагая, что функция $N_{\text{ст}}(x)$ не слишком быстро меняется вдоль трубы, разложим ее в ряд по степеням x и оборвем разложение на квадратичном члене:

$$N_{\text{ст}} = N_{\text{ст}}(0) + \left(\frac{dN_{\text{ст}}}{dx}\right)_{x=0} x + \frac{1}{2}\left(\frac{d^2N_{\text{ст}}}{dx^2}\right)_{x=0} x^2.$$

При интегрировании по x первое и последнее слагаемое не внесут никакого вклада в интеграл, и мы получим

$$N = 2a\,\frac{dN_{\text{ст}}}{dx}\int y\,dS\int\limits_{-l/2}^{+l/2}\frac{x^2}{R^4}\,dx.$$

Считая трубу длинной, заменим в последнем интеграле конечные пределы бесконечными. Для вычисления всего интеграла введем в плоскости сечения S полярные координаты r и φ, поместив начало полярной

системы координат в точку O. Тогда $y = r\cos\varphi$, $R^2 = r^2 + x^2$, $dS = r\,dr\,d\varphi$, а потому

$$N = 2a\frac{dN_{\text{ст}}}{dx}\int\limits_{-\pi/2}^{+\pi/2}\cos\varphi\,d\varphi\int\limits_{0}^{2a\cos\varphi}r^2\,dr\int\limits_{-\infty}^{+\infty}\frac{x^2}{r^2+x^2}\,dx.$$

Для интеграла по x получаем

$$\int\limits_{-\infty}^{+\infty}\frac{x^2}{r^2+x^2}\,dx = \left(\frac{1}{2r}\operatorname{arctg}\frac{x}{r} - \frac{1}{2}\frac{x}{r^2+x^2}\right)_{-\infty}^{+\infty} = \frac{\pi}{2r}.$$

Выполнив остальные интегрирования, найдем

$$N = \frac{8\pi a^3}{3}\frac{dN_{\text{ст}}}{dx}. \tag{96.8}$$

При стационарном течении величина N, а с ней и производная $dN_{\text{ст}}/dx$ остаются постоянными вдоль трубы. Используя это, а также выражение $N_{\text{ст}} = (1/4)n\overline{v}$, без труда получим

$$\frac{dN_{\text{ст}}}{dx} = \frac{N_1 - N_2}{\pi a^2 l}.$$

После этого формула (96.8) приводится к виду (96.5). Совпадение числовых значений коэффициента C в строгом и оценочном выводах, конечно, является случайным.

ЗАДАЧИ

1. Стеклянный сосуд с толщиной стенок $l = 5$ мм и емкостью $V = 1$ л наполнен азотом и окружен вакуумом. В стенке сосуда образовался узкий цилиндрический канал радиуса $a = 0{,}1$ мм. Начальное давление газа в сосуде настолько мало, что радиус канала пренебрежимо мал по сравнению с длиной свободного пробега молекул газа. Как меняется во времени концентрация молекул газа в сосуде? Определить время τ, по истечении которого давление газа в сосуде уменьшится в e раз, если температура поддерживается постоянной и равна $T = 300$ К.

Ответ. $n = n_0 e^{-t/\tau}$, $\tau = \dfrac{3Vl}{2\pi a^3\overline{v}} = 5\cdot 10^3$ с $= 83{,}4$ мин.

2. Полностью эвакуированный стеклянный сосуд с толщиной стенок $l = 3$ мм и емкостью $V = 1$ л погружен в атмосферу углекислого газа (CO_2). В стенке сосуда образовался узкий цилиндрический канал диаметра $D = 0{,}1$ мм. Давление окружающего газа настолько мало, что диаметр канала пренебрежимо мал по сравнению с длиной свободного пробега молекул газа. Как меняется во времени концентрация молекул газа в сосуде? Определить время τ, по истечении которого давление газа в сосуде будет составлять $(e-1)/e = 0{,}628$ от давления окружающего газа при условии, что температура поддерживается постоянной и равна $T = 300$ К.

Ответ. $n = n_0(1 - e^{-t/\tau})$, $\tau = \dfrac{12Vl}{\pi D^3\overline{v}} = 3\cdot 10^4$ с $= 5\cdot 10^2$ мин $= 8{,}33$ ч.

3. Сосуды с объемами V_1 и V_2 соединены между собой цилиндрическим капилляром радиуса a и длины l, по которому происходит изотермическое кнудсеновское перетекание газа из одного сосуда в другой. Как будут меняться во времени концентрации молекул газа в сосудах n_1 и n_2, если их начальные значения были равны n_{10} и n_{20}?

Р е ш е н и е. Перетекание описывается системой дифференциальных уравнений:

$$\frac{dn_1}{dt} = \frac{n_2 - n_1}{\tau_1}, \quad \frac{dn_2}{dt} = \frac{n_1 - n_2}{\tau_2}, \quad \tau_1 = \frac{3V_1 l}{2\pi a^3 \overline{v}}, \quad \tau_2 = \frac{3V_2 l}{2\pi a^3 \overline{v}}.$$

Вычитая из одного уравнения другое, получим

$$\frac{d}{dt}(n_1 - n_2) = \frac{n_1 - n_2}{\tau}, \quad \text{где} \quad \frac{1}{\tau} = \frac{1}{\tau_1} + \frac{1}{\tau_2}.$$

Интегрируя и используя начальные условия, найдем

$$n_1 - n_2 = (n_{10} - n_{20})e^{-t/\tau}.$$

Подставляем это значение в исходную систему уравнений и интегрируем. В результате получим

$$n_1 = \frac{n_{10}\tau_1 + n_{20}\tau_2}{\tau_1 + \tau_2} + \frac{\tau_2}{\tau_1 + \tau_2}(n_{10} - n_{20})e^{-t/\tau},$$

$$n_2 = \frac{n_{20}\tau_2 + n_{10}\tau_1}{\tau_1 + \tau_2} + \frac{\tau_1}{\tau_1 + \tau_2}(n_{20} - n_{10})e^{-t/\tau}.$$

Глава VIII

РЕАЛЬНЫЕ ГАЗЫ

§ 97. Молекулярные силы и отступления от законов идеальных газов

1. Законы идеальных газов — приближенные законы. Отступления от них носят как количественный, так и качественный характер. Количественные отступления проявляются в том, что уравнение Клапейрона–Менделеева $PV = RT$ соблюдается для реальных газов только приближенно. Качественные отступления носят более глубокий характер. Реальные газы могут быть переведены в жидкое и твердое состояния. Это было бы невозможно, если бы газы строго следовали уравнению Клапейрона.

Отступления от законов идеальных газов связаны с тем, что между молекулами газа действуют силы, которые в теории идеальных газов во внимание не принимаются. Эти силы могут приводить к образованию химических соединений. Тогда они называются *химическими*, или *валентными*, *силами*. Если же химические соединения не образуются, то силы взаимодействия между молекулами и атомами называются *молекулярными силами*. Мы будем предполагать, что в газе химические реакции не происходят или закончились, так что химические силы нас интересовать не будут. Если газ ионизован, то появляются силы *кулоновского притяжения* и *отталкивания* между ионами, имеющимися в газе. Такими силами определяется поведение *плазмы*, т. е. *квазинейтрального ионизованного газа*. (Квазинейтральным называется такой газ в котором с большой точностью заряды положительных ионов компенсируются зарядами отрицательных ионов.) Физика плазмы представляет обширный, бурно развивающийся раздел физики, имеющий широкие применения в астрофизике, теории распространения радиоволн, в проблеме управляемых термоядерных реакций и пр. Однако о свойствах плазмы здесь говорить преждевременно. О ней пойдет речь после изучения других разделов физики (электричества, атомной физики). Здесь мы будем рассматривать только газы, состоящие из *электрически нейтральных* молекул или атомов. Химические силы по своей природе не отличаются от молекулярных сил. И те и другие сводятся к силам *электрических взаимодействий* между элементарными зарядами, из которых построены атомы и молекулы. Однако полное понимание природы молекулярных и химических сил стало возможным только после возникновения квантовой механики. Не имея возможности останавливаться на этом вопросе, ограничимся некоторыми значениями, оставаясь в рамках классических представлений.

2. На далеких расстояниях молекулярные силы являются силами притяжения. Эти силы притяжения называют также *силами Ван-дер-Ваальса* (1837–1923) по имени голландского ученого, заложившего основы молекулярной теории реальных газов. Но каким образом могут возникать силы притяжения между сложными частицами, построенными из элементарных зарядов, если эти частицы в целом электрически нейтральны? Для того чтобы это понять, необходимо принять во внимание два обстоятельства. Во-первых, заряды противоположного знака внутри частицы не совмещены в одной точке. Благодаря этому их внешнее действие не вполне уничтожается — вокруг всякого атома или молекулы имеется электрическое поле, более или менее быстро убывающее с увеличением расстояния. Во-вторых, под действием внешнего поля положение или движение зарядов внутри молекулы слегка изменяется таким образом, что положительные заряды смещаются в направлении электрического поля, а отрицательные — в противоположном направлении. Это явление называется *электрической поляризацией*. Представим себе теперь две сложные нейтральные частицы 1 и 2 на столь малом расстоянии друг от друга, что электрическое поле $\mathbf{E}_1$, возбуждаемое первой частицей в месте нахождения второй, имеет еще заметную величину. Под влиянием этого поля частица 2 *поляризуется*, и возбуждаемое ею электрическое поле $\mathbf{E}_2$ в месте нахождения первой частицы усилится. Под действием поля $\mathbf{E}_2$ первая частица также поляризуется, что поведет к усилению поля $\mathbf{E}_1$, и т. д. В результате частицы будут обращены друг к другу противоположно заряженными сторонами. Они будут притягиваться подобно магнитам, обращенным друг к другу противоположными полюсами. Рассмотренные силы называются *дисперсионными*. Происхождение этого термина связано с тем, что поляризуемостью молекул определяются также *диэлектрическая проницаемость* и *показатель преломления* газа (оптическая дисперсия).

Помимо дисперсионных сил между молекулами газа могут действовать еще так называемые *дипольно-ориентационные силы*. Это тоже силы притяжения, обычно меньшие дисперсионных сил. Кроме того, дипольно-ориентационные силы зависят от температуры газа. Они возникают тогда, когда молекулы газа поляризованы уже в отсутствие внешнего электрического поля. Такие молекулы называются *полярными*. Полярные молекулы во внешнем электрическом поле поворачиваются подобно тому, как поворачивается магнитная стрелка в магнитном поле. Представим опять две полярные молекулы, взаимодействующие между собой. В электрическом поле одной молекулы другая будет поворачиваться. Молекулы будут стремится повернуться так, чтобы их противоположные концы были обращены друг к другу. В результате возникает сила притяжения. Так как ориентация постоянно сбивается тепловым движением, то ясно, что дипольно-ориентационные силы должны зависеть от температуры газа.

Если расстояние между молекулами велико по сравнению с их размерами, то легко показать, что дисперсионные и дипольно-ориентационные силы должны убывать обратно пропорционально седьмой

степени расстояния. Этим результатом нам нигде не придется пользоваться, а потому мы не приводим его доказательства. Читатель сам легко может провести соответствующие расчеты, после того как он приступит к изучению электричества.

Конечно, изложенные соображения недостаточны для понимания природы молекулярных сил. К тому же классическая теория вообще не способна последовательно решить проблему молекулярных и химических сил. Это видно уже из того, что она не может объяснить существование самих атомов и молекул как систем, построенных из элементарных электрически заряженных частиц. Такие системы с ее точки зрения не могли бы быть устойчивыми.

На близких расстояниях, когда электронные оболочки взаимодействующих частиц взаимно проникают друг в друга, силы молекулярного притяжения переходят в силы отталкивания. Теория, основанная на квантовой механике, показывает, что силы отталкивания очень велики, когда расстояние между взаимодействующими частицами мало. Они очень быстро убывают с увеличением расстояния. Когда расстояние превосходит «диаметр» взаимодействующих частиц, силы отталкивания экспоненциально убывают с возрастанием расстояния.

3. Взаимодействие молекул удобно характеризовать потенциальной энергией взаимодействия $U(x)$ как функцией расстояния x между центрами сблизившихся молекул. Сами молекулы для простоты можно считать сферическими. На основании изложенного функция $U(x)$ должна иметь вид, графически представленный на схематическом рис. 94 жирной кривой. Она имеет минимум, в котором силы притяжения уравновешиваются силами отталкивания.

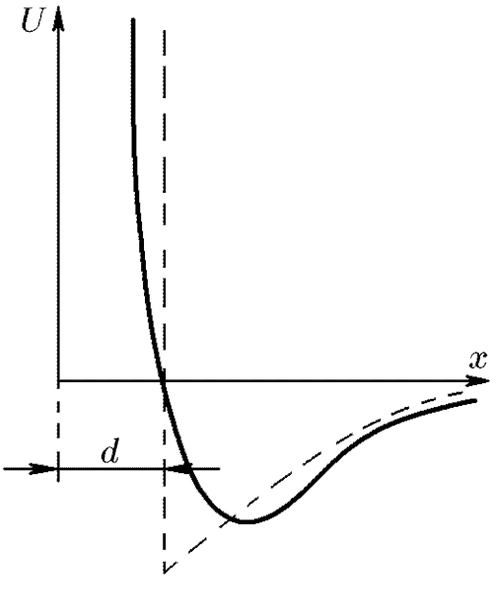


Рис. 94

Во многих вопросах теории газов к хорошим результатам приводит следующая аппроксимация функции $U(x)$:

$$U(x) = \frac{a_1}{x^{12}} - \frac{a_2}{x^6}, \qquad (97.1)$$

где a_1 и a_2 — постоянные. Она называется *потенциалом Леннарда–Джонса* (1894–1954). Первый член соответствует силам отталкивания, второй — силами притяжения Ван-дер-Ваальса. Сила притяжения убывает обратно пропорционально седьмой степени расстояния (так как $F = -dU/dx$). Поэтому член $-a_2/x^6$ может считаться обоснованным теоретически (при больших x). Что касается первого члена, то на него следует смотреть как на простую аппроксимацию.

В теории уравнения состояния Ван-дер-Ваальса применяется еще более грубая аппроксимация. Так как кривая $U(x)$ слева круто под-

нимается вверх, то этот участок кривой заменяется вертикальной прямой, как это изображено на рис. 94 штриховой линией. Если d — расстояние этой прямой от начала координат, то центры взаимодействующих частиц не могут сблизиться на расстояние, меньше d. Рассматриваемая аппроксимация соответствует поэтому *модели твердых упругих* шаров, между которыми действуют силы притяжения. Этой моделью мы и будем пользоваться в дальнейшем. Силы отталкивания учитываются тем, что размеры шаров считаются конечными. Эти силы проявляются только в моменты столкновений. Расстояние d играет роль *диаметра молекулы*. Конечно, диаметр молекулы относится к числу не вполне четко определенных величин.

Еще менее четко определенным является понятие *сферы молекулярного действия*, которое иногда вводится. Оно определяется тем, что силы притяжения молекулы проявляются только внутри сферы молекулярного действия, а вне этой сферы считаются равными нулю. Ясно, что радиус сферы молекулярного действия не может быть указан совершенно точно, так как это зависит от степени точности, предъявляемой к расчету. Ориентировочно он порядка 10^{-7} см.

§ 98. Уравнение Ван-дер-Ваальса

1. Учтем теперь влияние молекулярных сил на вид уравнения состояния газов, пользуясь моделью твердых упругих шаров. Начнем сначала с влияния *сил отталкивания* или, что то же самое, с влияния конечных размеров молекул. Будем предполагать, что силы притяжения между молекулами не действуют. Влияние конечных размеров молекул качественно понять легко. При одних и тех же температурах и концентрациях число ударов о стенку больше в случае молекул конечного размера, чем в случае точечных молекул. Это объясняется тем, что передача импульса в газе по пространству, не занятому молекулами, происходит с *тепловыми скоростями*, а по пространству, заполненному абсолютно твердыми молекулами, — с *бесконечной скоростью*. В результате давление газа возрастает.

Исследуем теперь вопрос количественно. Будем предполагать, что плотность газа не очень велика. Тогда случаи, когда одновременно сталкиваются и приходят во взаимодействие три молекулы или больше, будут относительно очень редки. Много чаще будут встречаться такие случаи, когда сталкиваются между собой только две молекулы, а остальные молекулы в момент столкновения на них не действуют. Такие столкновения называются *парными*. Мы учтем только парные столкновения и полностью пренебрежем влиянием тройных, четверных и прочих столкновений. Ясно, что таким путем нельзя получить уравнение состояния газа, пригодное при больших плотностях. Можно рассчитывать лишь на получение поправок к уравнению Клапейрона.

Допустим сначала, что в сосуде объема V с гладкими стенками находятся две одинаковые молекулы *1* и *2*, совершающие тепловое

движение (рис. 95). Молекулы сталкиваются со стенками и между собой. Из-за этого возникает давление на стенки. Величина давления определяется *суммарной кинетической энергией* обеих молекул и не зависит от того, как эта энергия распределена между молекулами (см. § 59). При вычислении давления можно считать, что одна молекула, например *2*, все время остается неподвижной, а другая (молекула *1*) движется с удвоенной кинетической энергией. Результат расчета от этого не изменится. Центры молекул не могут сблизиться на расстояние, меньшее d. Окружим молекулу *2 сферой ограждения* радиуса d, как это мы делали в § 86 при вычислении средней длины свободного пробега. Движущуюся молекулу *1* можно считать точечной. Очевидно, она не может проникнуть внутрь сферы ограждения неподвижной молекулы. Это значит, что объем, доступный молекуле *1*, благодаря присутствию молекулы *2* уменьшается на объем сферы ограждения, т. е. на величину $(4/3)\pi d^3$. Эта величина равна учетверенной сумме объемов обеих молекул.

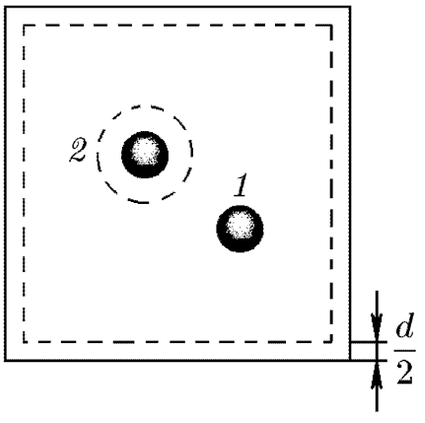


Рис. 95

Пусть теперь в сосуде имеется N одинаковых молекул. При вычислении давления на стенку сосуда можно рассуждать так, как если бы половина из них $(1/2)N$ покоилась и была заменена соответствующими сферами ограждения, а молекулы другой половины были точечными и двигались с удвоенной кинетической энергией. Тогда мы имели бы идеальный газ из $N' = N/2$ точечных молекул с температурой $T' = 2T$. Этим молекулам был бы доступен объем сосуда V за исключением объема, занимаемого $N/2$ сферами ограждения других молекул. Обозначим этот последний объем символом b. Тогда объем, доступный движущимся молекулам, будет равен $V - b$. Давление, оказываемое этими молекулами на стенки сосуда, равно

$$P = n'kT' = \frac{N'}{V-b}\,kT' = \frac{NkT}{V-b}.$$

Если в сосуде находится моль газа, то

$$P(V-b) = RT. \tag{98.1}$$

Очевидно,

$$b = \frac{N}{2}\,\frac{4\pi}{3}\,d^3 = \frac{2\pi}{3},$$

т. е. объем b равен учетверенному объему всех N молекул газа.

2. При выводе мы пренебрегли следующим обстоятельством. Центр движущейся молекулы не может подойти к стенке на расстояние, меньшее $d/2$. Для него недоступен объем пристеночного слоя толщины $d/2$. Граница такого слоя на рис. 95 изображена штриховой линией. Его объем равен $S \cdot d/2$, где S — площадь внутренней поверхности

сосуда. Такой объем учтен не был. Это можно делать, когда $S \cdot d/2 \ll \ll b$, т. е. $S \ll (4/3)\pi N d^2$. Если сосуд имеет форму шара радиуса R, то это условие сводится к

$$R \ll d\sqrt{N}. \tag{98.2}$$

Для одного моля при $d \sim 10^{-8}$ см это дает $R \ll 10^4$ см $= 100$ м. Давление газа в шаре такого радиуса было бы порядка 10^{-9} атм.

3. Так как давление газа P не может быть отрицательным, то из формулы (98.1) следует, что при всех давлениях $V > b$. Значение $V = b$ достигается только при $P = \infty$. Это заключение неправильно. Постоянная b равна учетверенному объему всех молекул газа. Между тем твердые шары, каковыми в нашей модели являются молекулы, можно разместить в меньшем объеме. Их, конечно, невозможно разместить в объеме $b/4$, в точности равном сумме объемов всех шаров. Даже при наиболее плотной упаковке останутся промежутки между шарами. При наиболее плотной упаковке (см. рис. 175 *а*) общий объем всех шаров $b/4$ составляет $\pi/(3\sqrt{2}) = 0{,}74$ объема всего тела (см. § 134). Отсюда следует, что минимальный объем, до которого может быть сжат газ, по нашей модели равен $(3\sqrt{2}/4\pi)b = 0{,}337b$, и не b, как это было получено из уравнения (98.1). Почему же уравнение (98.1) привело нас к неправильному заключению? Потому, что мы вывели его в предположении парных столкновений, а использовали тогда, когда это условие не выполняется. При всякой более плотной упаковке каждый шар одновременно касается нескольких соседних шаров, а такие случаи при выводе были исключены из рассмотрения.

4. Рассмотрим теперь влияние *сил молекулярного притяжения*. Предполагая, что сил отталкивания нет, изменим модель газа. Молекулы будем считать точками, между которыми действуют силы притяжения. В отличие от сил отталкивания, действующих на близких расстояниях, силы молекулярного притяжения являются *силами дальнодействующими*. Во взаимодействии участвует сразу много молекул, и схема парных столкновений становится непригодной. Окружим каждую молекулу сферой молекулярного действия. Если эта сфера целиком находится внутри газа, то силы, действующие на рассматриваемую молекулу со стороны окружающих молекул, в среднем уравновешиваются (рис. 96). Но этого не будет, когда молекула находится вблизи границы газа со стенкой. Здесь сфера молекулярного действия лишь частично проходит в газе. Появляется избыток молекул, тянущих рассматриваемую молекулу внутрь газа, над молекулами, тянущими ее наружу. Таким образом, вблизи стенки возникает *пристеночный слой газа*, толщина которого равна радиусу сферы молекулярного действия.

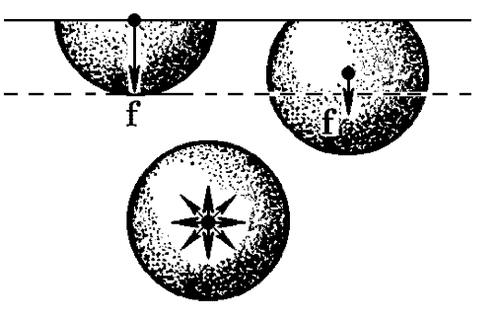


Рис. 96

Каждая молекула этого слоя в среднем подвергается действию силы f, направленной в сторону газа. Величина силы f максимальна, когда молекула находится у самой стенки, и убывает при удалении от нее.

Когда молекула газа летит к стенке, а затем отражается от нее, то меняется ее импульс. Изменение импульса подсчитывается совершенно так же, как это делалось в § 59 при вычислении давления идеального газа. Ежесекундное изменение импульса всех молекул, падающих на единицу площади стенки и отражающих от нее, равно $(1/3)\,nm\langle v^2\rangle$. Однако, в отличие от идеальных газов, импульс налетающих молекул изменяется не только под действием сил давления со стороны стенки, но и под действием сил, с которыми их тянут внутрь газа молекулы пристеночного слоя. В частности, под действием этих последних сил молекула может отразиться внутри пристеночного слоя, не долетев до стенки. Вместо сил, действующих на налетающие молекулы, можно по третьему закону Ньютона ввести равные им и противоположно направленные силы, действующие на стенку и на молекулы пристеночного слоя. Пусть P — средняя сила давления газа на стенку, а P_i — средняя сила (отнесенная к единице площади), с которой молекулы пристеночного слоя втягиваются внутрь газа. Тогда

$$P + P_i = \frac{1}{3}\, nm\langle v^2\rangle,$$

или

$$P + P_i = nkT. \tag{98.3}$$

Из изложенного видно, что давление на стенку P не зависит от материала стенки. Роль стенки может выполнять сам газ. Проведем мысленно произвольное сечение, разделяющее газ на две части. Давление одной части на другую будет таким же, как если бы эта другая часть была твердой стенкой. Оно равно P, а не $P + P_i$ или какой-либо другой комбинации этих величин. Именно это давление P входит в уравнение гидродинамики и газодинамики.

Сила P_i называется *внутренним*, или *молекулярным*, *давлением*. Ее можно представить в виде $P_i = \langle N_{\text{сл}} f\rangle$, где f — сила, действующая на молекулу пристеночного слоя, а $N_{\text{сл}}$ — число молекул в нем, отнесенное к единице площади. Можно также написать $P_i \sim \langle N_{\text{сл}}\rangle\langle f\rangle$. Обе величины $\langle N_{\text{сл}}\rangle$ и $\langle f\rangle$ пропорциональны плотности или обратно пропорциональны объему газа. Предполагая опять, что газ взят в количестве одного моля, можно положить

$$P_i = \frac{a}{V^2}, \tag{98.4}$$

где a — постоянная, характерная для рассматриваемого газа. Тогда (98.3) переходит в

$$\left(P + \frac{a}{V^2}\right) V = RT. \tag{98.5}$$

5. Теперь надо учесть совместное действие сил притяжения и сил отталкивания. Для неплотных газов, к которым относятся наши

рассуждения, поправки на силы притяжения и отталкивания можно вводить *независимо*. Тогда в результате комбинации формул (98.1) и (98.5) получится

$$\left(P + \frac{a}{V^2}\right)(V - b) = RT. \tag{98.6}$$

Это уравнение называется *уравнением Ван-дер-Ваальса*. Ван-дер-Ваальс получил его теоретически в 1873 г. Теоретический вывод уравнения (98.6) применим при выполнении условий

$$b \ll V, \quad a/V^2 \ll P. \tag{98.7}$$

Кроме того, вывод предполагает, что молекулы газа сферически симметричны, поскольку он относится к модели твердых упругих шаров. С этим связано то обстоятельство, что в действительности даже для неплотных газов величины a и b зависят от температуры.

Для плотных газов уравнение Ван-дер-Ваальса, как количественное соотношение, не годится. Однако качественно оно правильно передает поведение и таких газов. Поэтому для изучения качественного поведения вещества мы будем в дальнейшем применять уравнение Ван-дер-Ваальса во всей области изменения давлений и температур. Но тогда на него надо смотреть, как на *приближенное полуэмпирическое уравнение*. Газы, точно подчиняющиеся уравнению Ван-дер-Ваальса, называются *газами Ван-дер-Ваальса*. Ясно, что они являются идеализациями.

Нетрудно записать уравнение Ван-дер-Ваальса для произвольного числа ν молей газа. Если газ занимает объем V, то молярный объем будет V/ν. Этой величиной надо заменить V в уравнении (98.6). В результате получится

$$\left(P + \frac{a\nu^2}{V^2}\right)\left(\frac{V}{\nu} - b\right) = RT, \tag{98.8}$$

или

$$\left(P + \frac{a\nu^2}{V^2}\right)(V - b\nu) = \nu RT. \tag{98.9}$$

6. Согласно закону Бойля–Мариотта произведение PV для идеальных газов остается постоянным, если только температура поддерживается постоянной. Посмотрим, как ведет себя это произведение для газа Ван-дер-Ваальса. Экспериментально удобно исследовать зависимость PV от P, оставляя температуру постоянной. Теоретически, однако, проще получить зависимость PV от плотности $\rho = \mu/V$. Перепишем уравнение (98.6) в виде

$$P = \frac{RT}{V - b} - \frac{a}{V^2}. \tag{98.10}$$

После умножения на V и замены в правой части V на μ/ρ получим

$$PV = \frac{\mu RT}{\mu - b\rho} - \frac{a}{\mu}\rho,$$

где μ — молярная масса газа. Для отыскания экстремума этого выражения находим его производную по ρ. Приравнивая ее нулю и учитывая, что $\mu > b\rho$, найдем, что экстремум имеет место при

$$\rho = \frac{\mu}{b}\left(1 - \sqrt{\frac{RTb}{a}}\right). \tag{98.11}$$

Исследование второй производной показывает, что это есть минимум. Таким образом, если изотермически сжимать газ из состояния бесконечного разрежения, то сначала он будет сжиматься сильнее идеального газа, а после прохождения через минимум PV — слабее. Это и понятно. Для разреженного газа бóльшую роль играют силы притяжения, способствующие сжатию, чем силы отталкивания, препятствующие сжатию. Напротив, для плотных газов более существенное влияние на их поведение оказывают силы отталкивания. С повышением температуры минимум PV смещается в сторону меньших ρ. Это связано с тем, что с повышением температуры увеличивается число столкновений, а потому возрастает роль сил отталкивания. Если $RTb/a = 1$, то минимум смещается в начало координат $\rho = 0$. Температура, при которой это происходит, называется *температурой Бойля*. Она равна

$$T_{\text{B}} = \frac{a}{Rb}. \tag{98.12}$$

Ниже температуры Бойля при сжатии газа произведение PV проходит через минимум. Так ведет себя при комнатных температурах большинство газов. Выше температуры Бойля силы отталкивания всегда перевешивают силы притяжения: с возрастанием ρ произведение PV растет монотонно. При комнатных температурах так ведут себя водород и гелий.

§ 99. Другой метод введения поправки на силы притяжения между молекулами. Уравнение Дитеричи

1. Влияние пристеночного молекулярного слоя на уравнение состояния можно учесть другим способом, который приводится ниже. Отвлечемся сначала от молекулярных сил притяжения, с которыми стенка может действовать на молекулы газа. Будем предполагать, что стенка действует на молекулы газа только при столкновениях с ними. Как выяснено в предыдущем параграфе, молекулы пристеночного слоя подвергаются действию результирующей силы, направленной внутрь газа. Вследствие этого концентрация молекул в пристеночном слое должна убывать при приближении к стенке в соответствии с формулой Больцмана

$$n = n_\infty e^{-U/kT}, \tag{99.1}$$

где U — потенциальная энергия молекулы. Этим и объясняется уменьшение давления газа на стенку. Энергия U является функцией расстояния x молекулы от стенки. Она максимальна у самой стенки и быстро убывает с возрастанием x. Ее значение на бесконечности U_∞ условимся считать

равным нулю. (Бесконечными считаются расстояния больше радиуса сферы молекулярного действия.) Давление на стенку определяется пристеночной концентрацией n_0, а не внутренней концентрацией n_∞. Считая молекулы точечными, можем написать, как для идеального газа:

$$P = n_0 kT = n_\infty kT e^{-U_0/kT}, \tag{99.2}$$

где U_0 — значение потенциальной энергии молекулы у стенки сосуда. Если выполнено условие $U_0 \ll kT$, то $e^{U_0/kT} \approx 1 + U_0/kT$, а потому

$$P + n_\infty U_0 = n_\infty kT. \tag{99.3}$$

Сила притяжения, действующая на пристеночную молекулу, а с ней и потенциальная энергия U_0, пропорциональны концентрации молекул газа $n_\infty = N/V$. Поэтому можно написать $U_0 = \alpha n_\infty$, где α — постоянная для рассматриваемого газа величина. Предыдущая формула преобразуется к виду $(P + \alpha n_\infty^2) = n_\infty kT$, или

$$\left(P + \frac{a}{V^2}\right) V = RT, \tag{99.4}$$

если ввести новую постоянную $a = \alpha N^2$. Этот результат совпадает с (98.5). Заметим еще, что через постоянную a потенциальная энергия U_0 выражается формулой

$$U_0 = \frac{a}{NV}. \tag{99.5}$$

2. Молекулы пристеночного слоя притягиваются не только молекулами газа, но и молекулами стенки. Это должно привести к уплотнению пристеночного слоя и к увеличению давления газа на стенку. Получается парадоксальный результат: давление газа на стенку и постоянная a в уравнении Ван-дер-Ваальса должны зависеть от материала стенки. Этот результат не согласуется с опытом, а вывод его неверен. Разберемся в этом вопросе.

Сила F, действующая на единицу площади стенки из-за ударов налетающих молекул, действительно возрастает. Но это не единственная сила, действующая на стенку. Стенка притягивает газ. По третьему закону Ньютона газ притягивает стенку с равной, но противоположно направленной силой. Обозначим модуль через F_1. Чтобы получить давление газа на стенку, силу F_1 надо вычесть из F. Покажем, что в пределах точности расчета, принятой при выводе уравнения Ван-дер-Ваальса, результат не будет зависеть от материала стенки.

Пусть $U'(x)$ — потенциальная энергия молекулы в поле сил притяжения стенки. Формулу Больцмана теперь надо писать в виде

$$n = n_\infty \exp\left(-\frac{U + U'}{kT}\right).$$

Поэтому вместо (99.3) мы придем к соотношению

$$F = n_\infty kT - n_\infty (U_0 + U_0').$$

Вычислим теперь силу F_1. Стенка притягивает молекулу с силой f_1, абсолютная величина которой равна $f_1 = dU'/dx$. число молекул в слое единичной площади и толщины dx есть $n\,dx$. Следовательно,

$$F_1 = \int_0^\infty n f_1\,dx = \int_0^\infty n\,\frac{dU'}{dx}\,dx = \int_{U_0'}^0 n\,dU'.$$

В принятом приближении зависимость концентрации n от координаты x учитывать не надо, а потому $F_1 = -n_\infty U_0'$. Вычитая F_1 из F, найдем

$$P = F - F_1 = n_\infty kT - n_\infty U_0,$$

что совпадает с (99.3). Таким образом, давление P не зависит от материала стенки.

Обобщить это доказательство на случай произвольно больших плотностей газа, положив в основу рассмотрения механизм явления, довольно затруднительно. Однако в этом и нет необходимости. Независимость давления газа на стенку сосуда от материала стенки можно доказать на основе общих соображений. Рассмотрим закрытый цилиндрический сосуд, наполненный газом, противоположные основания которого AB и CD сделаны из различных материалов (рис. 97). Допустим, что установилось состояние равновесия. Если бы давления на стенки AB и CD были разными, то сосуд пришел бы в движение, и равновесие, вопреки предположению, было бы невозможно.

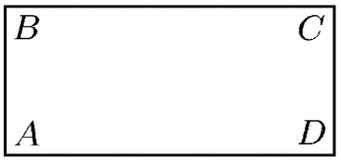


Рис. 97

3. Вернемся к формуле (99.2), но не будем аппроксимировать показательную функцию линейной. Тогда с учетом соотношения (99.5) получим

$$P = n_\infty kT \exp\Big(-\frac{a}{RTV}\Big) = \frac{RT}{V} \exp\Big(-\frac{a}{RTV}\Big).$$

Введем сюда поправку на конечный объем молекул совершенно так же, как это делалось в предыдущем параграфе. Тогда

$$P(V - b) = RT \exp\Big(-\frac{a}{RTV}\Big). \tag{99.6}$$

Это *уравнение Дитеричи*. В пределе, когда $b \ll V$, а $a \ll RTV$, уравнение (99.6) переходит в уравнение Ван-дер-Ваальса. Чтобы выполнить такой предельный переход, аппроксимируем в формуле (99.6) показательную функцию линейной. Получим

$$P = \frac{RT}{V - b} - \frac{a}{V(V - b)}.$$

В последнем слагаемом величиной b следует пренебречь. Это приводит к ошибке второго порядка по малым поправкам a и b. Величинам же такого порядка при выводе уравнения Ван-дер-Ваальса мы пренебрегли. Таким образом, мы снова возвращаемся к уравнению (98.10).

Уравнение Дитеричи является таким же полуэмпирическим уравнением, что и уравнение Ван-дер-Ваальса. Оба уравнения можно считать теоретически обоснованными только при выполнении условий (98.7). Для умеренных давлений уравнение Дитеричи значительно лучше уравнения Ван-дер-Ваальса, но зато совершенно непригодно для высоких давлений.

4. Кроме уравнения Ван-дер-Ваальса, было предложено много эмпирических или полуэмпирических уравнений состояния реальных газов. За счет увеличения эмпирических постоянных, входящих в эти уравнения, удается достигнуть лучшего согласия с опытом по сравнению с тем, что дает уравнение Ван-дер-Ваальса. Однако уравнение Ван-дер-Ваальса благодаря своей простоте и ясному физическому смыслу входящих в него постоянных до сих пор является наиболее распространенным уравнением для анализа качественного поведения реальных газов и жидкостей. Приводим некоторые из наиболее известных уравнений состояния.

Уравнение Бертло:

$$\left(P + \frac{a}{TV^2}\right)(V - b) = RT. \tag{99.7}$$

Уравнение Клаузиуса:

$$\left(P + \frac{a}{T(V + c)^2}\right)(V - b) = RT. \tag{99.8}$$

Уравнение Камерлинга-Оннеса:

$$PV = RT\left(1 + \frac{B_2}{V} + \frac{B_3}{V^2} + \ldots\right), \tag{99.9}$$

где $B_2, B_3, \ldots$ называются вторым, третьим и последующими *вириальными коэффициентами*. Они являются функциями температуры. Таким уравнением мы уже пользовались в § 33 при приведении шкалы газового термометра к термодинамической шкале. Там же было показано, что уравнение состояния всякого газа может быть приведено к виду (99.9). Однако уравнение (99.9) получает конкретное содержание только после того, как будут найдены выражения для входящих в него вириальных коэффициентов, как функций температуры.

Уравнения Бертло и Клаузиуса отличаются от уравнения Ван-дер-Ваальса поправками, вводимыми чисто эмпирически. Уравнение Бертло вблизи критической точки (см. следующий параграф) не имеет никаких преимуществ по сравнению с уравнением Ван-дер-Ваальса. Зато при умеренных давлениях оно лучше согласуется с опытом. Уравнение Клаузиуса точнее уравнения Ван-дер-Ваальса, поскольку оно содержит третью эмпирическую постоянную c. Благодаря этому с помощью уравнения Клаузиуса можно учесть отклонения от закона соответствующих состояний (см. следующий параграф).

ЗАДАЧА

Вычислить температуру Бойля для газа, подчиняющегося уравнению Дитеричи.

Ответ.

$$T_{\text{B}} = a/Rb. \tag{99.10}$$

§ 100. Изотермы газа Ван-дер-Ваальса

1. Наиболее содержательные результаты получаются из уравнения Ван-дер-Ваальса путем анализа изотерм. Уравнение изотермы можно представить в виде (98.6) или (98.10), если считать температуру T постоянной. При высоких температурах последний член в формуле (98.10) можно отбросить. Тогда изотерма будет гиперболой, асимптотами которой являются изобара $P = 0$ и изохора $V = b$. Всякая другая изобара $P = \text{const}$ будет пересекать такую изотерму в одной точке.

Для исследования изотерм при любых значениях T умножим уравнение (98.6) на V^2.

После раскрытия скобок уравнение изотермы примет вид

$$PV^3 - (RT + Pb)V^2 + aV - ab = 0. \tag{100.1}$$

Это уравнение третьей степени по V, в которое давление P входит в качестве параметра. Поскольку его коэффициенты вещественны, уравнение имеет либо один вещественный корень, либо три корня. Каждому корню на плоскости VP соответствует точка, в которой изобара $P = \text{const}$ пересекает изотерму. В первом случае, когда корень один, и точка пересечения будет одна. Так будет, как мы видели, при любых давлениях, если температура достаточно высока. Изотерма имеет вид монотонно опускающейся кривой MN (рис. 98). При более низких температурах и надлежащих значениях давления P уравнение (100.1) имеет три корня V_1, V_2, V_3. В таких случаях изобара $P = \text{const}$ пересекает изотерму в трех точках L, C, G (рис. 98). Изотерма содержит волнообразный участок $LBCAG$. Она сначала монотонно опускается вниз (участок DB), затем на участке BA монотонно поднимается вверх, а за точкой A снова монотонно опускается. При некоторой промежуточной температуре три корня V_1, V_2, V_3 становятся равными. Такая температура и соответствующая ей изотерма называются *критическими*. Критическая изотерма FKH всюду монотонно опускается вниз, за исключением одной точки K, являющейся точкой перегиба изотермы. В ней касательная к изотерме горизонтальна. Точка K называется *критической точкой*. Соответствующие ей давление P_k, объем V_k и температура T_k называются также *критическими*. Говорят, что вещество находится в *критическом состоянии*, если его объем и давление (а следовательно, и температура) равны критическим.

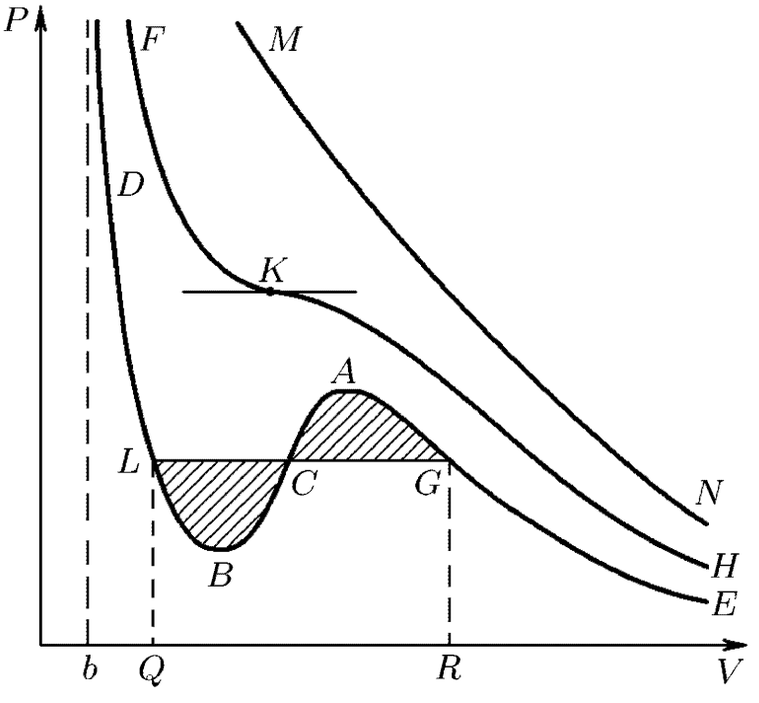


Рис. 98

Критическая температура и критическое состояние никоим образом не являются понятиями, связанными исключительно с уравнением Ван-дер-Ваальса. Такие понятия можно было бы ввести для любого уравнения состояния, изотермы которого имеют волнообразный характер, а при повышении температуры переходят в монотонные. Для всего последующего изложения существен только такой ход изотерм, а не конкретная форма уравнения состояния. Более того, возможность существования вещества в критическом состоянии не может быть обоснована с помощью уравнения Ван-дер-Ваальса. Это видно уже из того, что уравнением Ван-дер-Ваальса приходится пользоваться вне области его применимости. К понятию критического состояния

можно прийти путем анализа экспериментальных изотерм вещества без использования какого бы то ни было «теоретического» уравнения состояния, как это и было на самом деле. Однако мы не встанем на такой путь чистого эмпиризма, а постараемся показать, какая существует связь между критическими явлениями, уравнением состояния и общими условиями термодинамического равновесия. Ради простоты проведем все рассуждения почти исключительно на основе уравнения Ван-дер-Ваальса. Но его следует рассматривать не как истинное, а как «модельное» уравнение состояния вещества. После этого разъяснения можно не опасаться, что у читателя возникнет неправильное представление, что критические явления нельзя понять без уравнения Ван-дер-Ваальса.

2. Для нахождения критических параметров P_k, V_k, T_k учтем, что в критической точке уравнение (100.1) переходит в

$$P_k V^3 - (RT_k + P_k b)V^2 + aV - ab = 0.$$

Поскольку в этом случае все три корня совпадают и равны V_k, уравнение должно приводиться к виду

$$P_k(V - V_k)^3 = 0.$$

Возводя в куб и сравнивая коэффициенты, получим три уравнения:

$$P_k V_k^3 = ab, \quad 3P_k V_k^2 = a, \quad 3P_k V_k = RT_k + P_k b.$$

Решая их, найдем

$$V_k = 3b, \quad P_k = \frac{a}{27b^2}, \quad T_k = \frac{8a}{27Rb}. \tag{100.2}$$

К тем же результатам можно прийти, заметив, что критическая точка K является точкой перегиба изотермы, касательная в которой горизонтальна, а потому в точке K должны соблюдаться соотношения

$$\left(\frac{\partial P}{\partial V}\right)_T = 0, \quad \left(\frac{\partial^2 P}{\partial V^2}\right)_T = 0. \tag{100.3}$$

Решая эти уравнения совместно с уравнением изотермы (98.6), снова приведем к формулам (100.2).

Отношение $K_k = RT_k/P_k V_k$ называется *критическим коэффициентом*. Согласно уравнению Ван-дер-Ваальса критический коэффициент равен

$$K_k \equiv \frac{RT_k}{P_k V_k} = \frac{8}{3} = 2{,}67. \tag{100.4}$$

В действительности критические коэффициенты для реальных газов имеют различные значения, и все они больше $8/3$, как показывает табл. 10.

Критическая температура, по Ван-дер-Ваальсу, связана с температурой Бойля (98.12) соотношением

$$T_{\text{В}} = \frac{27}{8} T_k = 3{,}375 T_k. \tag{100.5}$$

Качественно это соотношение согласуется с опытом. Числовой коэффициент в соотношении между T_B и T_k, измеренный на опыте, оказался равным для гелия 3,5, водорода 3,3, азота 2,4.

Таблица 10

Вещество	K_k	Вещество	K_k
H_2	3,03	SO_2	3,60
He	3,13	C_6H_6	3,76
N_2	3,42	H_2O	4,46
O_2	3,42	CO_2	4,49

3. Примем в качестве единиц объема, давления и температуры критические значения этих величин. Объем, давление и температура, измеренные в таких единицах, называются *приведенными*. Они определяются выражениями:

$$\varphi = \frac{V}{V_k}, \quad \pi = \frac{P}{P_k}, \quad \tau = \frac{T}{T_k}. \tag{100.6}$$

Уравнение состояния, записанное в этих безразмерных переменных, называется *приведенным уравнением состояния*. Для газов Ван-дер-Ваальса из (100.2) находим

$$V = 3b\varphi, \quad P = \frac{a\pi}{27b^2}, \quad T = \frac{8a}{27Rb}\,\tau.$$

После подстановки этих выражений в уравнение Ван-дер-Ваальса оно принимает вид

$$\left(\pi + \frac{3}{\varphi^2}\right)\left(\varphi - \frac{1}{3}\right) = \frac{8}{3}\,\tau. \tag{100.7}$$

В этом виде уравнение состояния не содержит никаких индивидуальных констант вещества. *Приведенные уравнения состояния одинаковы для всех веществ*. Этот вывод обладает большей общностью, чем уравнение Ван-дер-Ваальса, из которого он был получен. Для справедливости вывода конкретный вид уравнения состояния не играет роли. Существенно только, что оно содержит *три* параметра: a, b и R. Всякое уравнение состояния, обладающее этим свойством, записанное в безразмерных величинах φ, π, τ, должно быть также одинаковым для всех веществ. Это положение называется *законом соответственных состояний*. Соответственными называются такие состояния разных веществ, которые имеют одинаковые значения приведенных параметров φ, π и τ. Из закона соответственных состояний следует, что *если для различных веществ из трех параметров φ, π, τ совпадают значения каких-либо двух, то будут совпадать и значения третьего параметра*, т. е. эти вещества находятся в соответственных состояниях.

ЗАДАЧИ

1. Исследовать вид изотерм и получить выражения для критических параметров для газа, подчиняющегося уравнению Дитеричи.

Р е ш е н и е. Уравнением изотермы является (99.6), где температура T должна считаться постоянной. Найдя производную $(\partial P/\partial V)_T$, легко убедиться, что ее знак совпадает со знаком квадратного трехчлена $f(V) = -RTV^2 + a(V-b)$. В области $V \leqslant b$ трехчлен заведомо отрицателен, т. е. в этой области он не имеет корней. Если $a - 4RTb > 0$, то трехчлен имеет два корня V_1 и V_2, причем $b \leqslant V_1 < V_2$. В точках $V = V_1$ и $V = V_2$ трехчлен, а с ним и производная $(\partial P/\partial V)_T$ меняют знаки. Отсюда следует, что изотерма будет типа волнистой кривой $DBAE$ (см. рис. 98). При $a - 4RTb < 0$ трехчлен $f(V)$ не имеет вещественных корней, производная $(\partial P/\partial V)_T$ всюду отрицательна, изотерма представляется монотонно опускающейся кривой типа MN. Условие равенства корней $a - 4RTb = 0$ определяет критическую температуру. Критический объем найдется решением уравнения $-RT_kV^2 + a(V-b) = 0$. Выполнив вычисления, найдем

$$T_k = \frac{a}{4Rb}, \quad V_k = 2b, \quad P_k = \frac{a}{4b^2e^2}. \tag{100.8}$$

Критический коэффициент

$$K_k \equiv \frac{RT_k}{P_kV_k} = \frac{e^2}{2} = 3{,}7, \tag{100.9}$$

а температура Бойля

$$T_{\text{В}} = 4T_k. \tag{100.10}$$

2. Записать приведенное уравнение состояния Дитеричи.
О т в е т.

$$\pi\Big(\varphi - \frac{1}{2}\Big) = \frac{e^2}{2}\,\tau e^{-2/\varphi\tau}. \tag{100.11}$$

§ 101. Изотермы реального газа. Правило Максвелла. Непрерывность газообразного и жидкого состояний вещества

1. Не все состояния вещества, совместимые с уравнением Ван-дер-Ваальса, могут быть реализованы в действительности. Для этого необходимо еще, чтобы они были *термодинамически устойчивы*. Одно из необходимых условий термодинамической устойчивости физически однородного вещества состоит в выполнении неравенства (51.14), т. е.

$$\Big(\frac{\partial P}{\partial V}\Big)_T < 0. \tag{101.1}$$

Физически оно означает, что при изотермическом уравнении давления объем тела должен уменьшаться. Иными словами, при возрастании V все изотермы должны монотонно опускаться. Между тем, ниже критической температуры на изотермах Ван-дер-Ваальса имеются поднимающиеся участки типа BCA (см. рис. 98). Точки, лежащие на таких участках, соответствуют неустойчивым состояниям вещества, которые практически реализованы быть не могут. При переходе к практическим изотермам эти участки должны быть выброшены, как это и сделано на рис. 99.

Таким образом, реальная изотерма распадается на две ветви EGA и BLD, отделенные друг от друга. Естественно предположить, что этим двум ветвям соответствуют различные агрегатные состояния вещества. Ветвь EA характеризуется относительно большими значениями объема или малыми значениями плотности; она соответствует *газообразному состоянию вещества*. Напротив, ветвь BD характеризуется относительно малыми объемами, а следовательно, большими плотностями; она соответствует *жидкому состоянию вещества*.

Мы распространяем, следовательно, уравнение Ван-дер-Ваальса и на область жидкого состояния. Таким путем удается получить удовлетворительное качественное описание явления перехода газа в жидкость и обратно.

2. Возьмем достаточно разреженный газ при температуре ниже критической. Исходное состояние его на диаграмме VP изображается точкой E (рис. 99). Будем сжимать газ квазистатически, поддерживая температуру T по изотерме вверх. Можно было бы думать, что она достигает крайнего положения A, где изотерма обрывается. В действительности, однако, начиная с некоторой точки G, давление в системе перестает повышаться, и она распадается на две *физически однородные части*, или *фазы*: газообразную и жидкую.

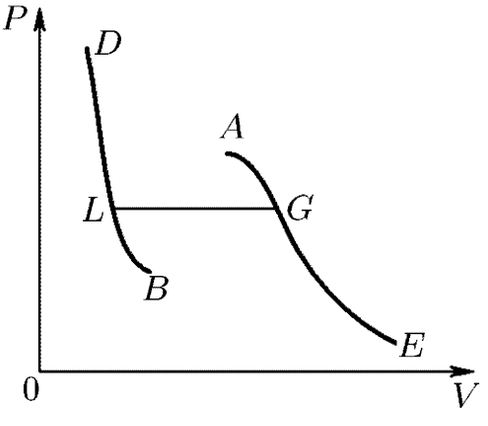


Рис. 99

Процесс изотермического сжатия такой двухфазной системы изображается участком GL горизонтальной прямой. При этом во время сжатия плотности жидкости и газа остаются неизменными и равными их значениям в точках L и G соответственно. По мере сжатия количество вещества в газообразной фазе непрерывно уменьшается, а в жидкой фазе — увеличивается, пока не будет достигнута точка L, в которой все вещество перейдет в жидкое состояние.

Такой ход изотерм впервые был обнаружен на опыте в конце XVIII века голландским ученым ван Марумом (1750–1837). Ван Марум провел серию опытов с целью проверки закона Бойля–Мариотта. Одним из исследуемых газов был аммиак. С увеличением давления объем газа сначала уменьшался обратно пропорционально давлению, как того требовал закон Бойля–Мариотта. Однако когда давление достигло 7 атм, оно неожиданно переставало расти при дальнейшем сжатии, хотя объем и продолжал уменьшаться. При давлении 7 атм происходило сжижение газообразного аммиака. Начиная с этого момента, сжатие вещества приводило лишь к увеличению количества жидкого и уменьшению газообразного аммиака.

После ван Марума последовали многочисленные попытки путем сжатия газов перевести их в жидкое состояние. Особенно больших успехов в этом направлении достиг Фарадей (1791–1867). Он применял

комбинированный метод, сочетая сжатие газа с его охлаждением. Однако все эти исследования производились вслепую. Оставалось неясно, почему при сжатии одних газов происходит сжижение, а других — нет. Влияние охлаждения и сама возможность сжижения газов не были ясны. Положение разъяснилось после классических работ английского физика Томаса Эндрюса (1813–1885), выполненных в течение 1861–1869 гг. Эндрюс систематически исследовал ход изотерм углекислоты (CO_2) при различных температурах и на основе этих исследований ввел понятие критической температуры. Углекислота им была выбрана сознательно, так как она обладает критической температурой (31 °C), лишь незначительно превышающей комнатную, и сравнительно невысоким критическим давлением (72,9 атм). Оказалось, что при температуре выше 31 °C изотермы углекислоты монотонно опускаются вниз, т. е. имеют гиперболический вид. Ниже этой температуры на изотермах углекислоты появляются горизонтальные участки, на которых изотермическое сжатие газа приводит к его конденсации, но не к увеличению давления. Таким путем было установлено, что *сжатием газ можно превратить в жидкость только тогда, когда его температура ниже критической. Но это невозможно сделать никаким повышением давления, если температура газа выше критической.*

3. Положение горизонтального участка изотермы GL легко определить, пользуясь термодинамическим равенством Клаузиуса

$$\oint \frac{\delta Q}{T} = 0. \tag{101.2}$$

Для этого заметим, что из состояния G (см. рис. 98) вещество можно перевести в состояние L двумя изотермическими процессами: по изотерме GCL двухфазного состояния вещества и по теоретической изотерме физически однородного вещества $GACBL$, содержащей неустойчивый участок ACB. Применим равенство Клаузиуса к квазистатическому круговому процессу $GCLBCAG$. Это — изотермический процесс, а потому равенство Клаузиуса принимает вид $\oint \delta Q = 0$. Кроме того, $\delta Q = dU + P\,V$, $\oint dU = 0$, так что $\oint P\,dV = 0$, или

$$\int\limits_{GCL} P\,dV + \int\limits_{LBCAG} P\,dV = 0.$$

Это соотношение можно переписать так:

$$\int\limits_{LCG} P\,dV = \int\limits_{LBCAG} P\,dV.$$

Отсюда следует, что площадь прямоугольника $QLGR$ должна быть равна площади криволинейной фигуры $QLBCAGR$. Значит, прямую GCL надо провести так, чтобы равнялись площади $GACG$ и $CBLC$, заштрихованные на рис. 98. Это правило называется *правилом Максвелла*. В § 112, п. 2 мы приведем другое доказательство этого правила.

4. Если за исходное состояние вещества взять жидкое (точка D на изотерме), то при изотермическом расширении точка, изображающая его состояние, будет перемещаться вниз по изотерме, пока она не достигнет положения L, начиная с которого появится новая — газообразная — фаза. Дальнейший изотермический процесс расширения двухфазной системы пойдет вдоль горизонтального участка изотермы LCG. В точке G все вещество перейдет в газообразное состояние. Затем изотермическое расширение идет по участку изотермы GE, причем вещество все время остается в газообразном состоянии.

При специальных условиях могут быть реализованы состояния, изображаемые участками изотермы GA и BL. Эти состояния называются *метастабильными*. Участок GA изображает так называемый *пересыщенный пар*, участок BL — *перегретую жидкость*. Обе фазы обладают ограниченной устойчивостью. Каждая из них может существовать до тех пор, пока она не граничит с другой, более устойчивой фазой. Например, пересыщенный пар переходит в насыщенный, если в него ввести каплю жидкости. Перегретая жидкость закипает, если в нее попадают пузырьки воздуха или пара [1]. Подробнее метастабильные состояния будут рассмотрены в гл. X.

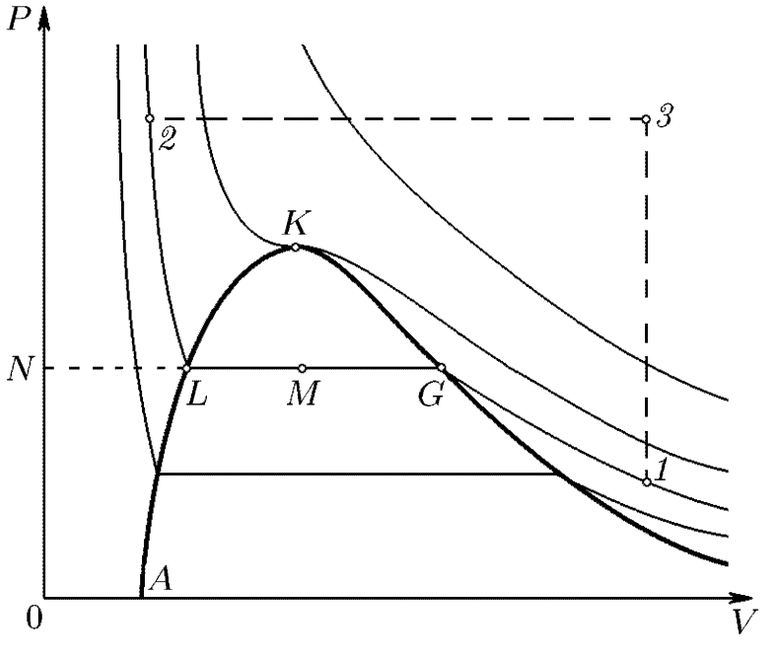


Рис. 100

5. На рис. 100 схематически изображено семейство реальных изотерм. Жирная кривая $ALKG$, соединяющая концы горизонтальных участков изотерм, делит плоскость VP на две области. Область между кривой $ALKG$ и изобарой $P = 0$ соответствует *двухфазным состояниям вещества*. Это значит, что каждая точка этой области изображает такое состояние вещества, в котором оно не является физически однородным, а состоит из жидкости и ее насыщенного пара. Ниже изобары

[1] Наряду с термином «газ» распространен также термин «пар». Эта двойственная терминология возникла в то время, когда считалось, что пар может быть превращен в жидкость, а газ нет. После того как все газы были сконденсированы, для такой двойственности не осталось оправдания. Газ и пар — это одно и то же. Когда говорят о паре какой-либо жидкости, то этим желают подчеркнуть, что речь идет о газе, получающемся испарением этой жидкости. Только по привычке мы говорим о водяном паре, а не о водяном газе, о насыщенном паре, а не о насыщенном газе и пр.

$P = 0$ двухфазное состояние невозможно, так как давление пара не может быть меньше нуля. Область над кривой $ALKG$ и ниже изобары $P = 0$, напротив, соответствует *однофазным состояниям*. Каждая точка этой области изображает состояние физически однородного вещества. Впрочем, точки под граничной кривой $ALKG$, наряду с двухфазными состояниями вещества, могут изображать и однофазные состояния, которым соответствуют теоретические изотермы, имеющие волнообразный характер. Это, во-первых, метастабильные состояния — *перегретая жидкость* и *пересыщенный пар*. Во-вторых, это абсолютно неустойчивые состояния, которым соответствуют поднимающиеся участки теоретических изотерм. Такие состояния могут играть роль только в теоретических рассуждениях, подобных тем, которые применялись при доказательстве правила Максвелла.

6. Возьмем в двухфазной области какую-либо точку M (рис. 100). Вещество, состояние которого изображается этой точкой, состоит из жидкости и газа. Определим относительное содержание этих фаз. Для простоты будем предполагать, что масса взятого вещества равна единице. Тогда удельные объемы жидкости и газа изобразятся длинами отрезков NL и NG соответственно, а объем всего вещества V_M — длиной отрезка NM. Если масса газа равна $m_{\text{г}}$, а жидкости — $m_{\text{ж}}$, то

$$V_M = NM = m_{\text{ж}} NL + m_{\text{г}} NG.$$

Так как $m_{\text{ж}} + m_{\text{г}} = 1$, то это соотношение можно переписать в виде

$$(m_{\text{ж}} + m_{\text{г}})NM = m_{\text{ж}} NL + m_{\text{г}} NG.$$

Отсюда

$$\frac{m_{\text{ж}}}{m_{\text{г}}} = \frac{NG - NM}{NM - NL},$$

или

$$\frac{m_{\text{ж}}}{m_{\text{г}}} = \frac{MG}{LM}. \tag{101.3}$$

Таким образом, точка M делит отрезок LG на части LM и MG, обратно пропорциональные массам жидкости и газа. Этот результат называется *правилом рычага*.

7. Рассматривая рис. 100, можно прийти к важному принципиальному заключению. Пусть начальное состояние газа изображается точкой *1* при температуре ниже критической. Будем изотермически сжимать газ. В точке G начнется превращение его в жидкость. Этот процесс закончится в точке L. При дальнейшем сжатии по изотерме $L2$ вещество будет оставаться все время жидким. Мы перевели вещество из газообразного состояния *1* в жидкое *2* и имели возможность наблюдать, *где начинается и как происходит это превращение*. Но можно тот же переход произвести иначе. Нагреем газ при постоянном объеме, чтобы его температура и давление стали выше критических (прямая

13). Затем, оставляя давление постоянным, охладим газ ниже критической температуры (прямая *32*). Мы придем в то же состояние *2*, т. е. сконденсируем газ в жидкость. Однако при этом *вещество все время будет оставаться физически однородным, свойства его будут меняться непрерывно, и не имеет смысла говорить, что в какой-то момент произошла конденсация.*

Начальное и конечное состояния *1* и *2* физически однородного вещества в однофазной области можно выбрать каким угодно. Переход из одного состояния в другое может быть произведен вдоль произвольной кривой, соединяющей точки *1* и *2*. Если эта кривая пересекает двухфазную область, то будет нарушаться физическая однородность вещества — вещество разделится на жидкость и газ. При выходе из двухфазной области вещество снова станет физически однородным и останется таковым в конечном состоянии *2*. Если же кривая перехода обходит двухфазную область, то никаких скачкообразных превращений вещества из газа в жидкость или обратно наблюдаться не будет — вещество все время будет оставаться физически однородным. Такое поведение вещества было установлено Эндрюсом путем анализа экспериментальных изотерм углекислоты. Когда состояние вещества изображается точкой над граничной кривой $ALKG$, то, согласно Эндрюсу, различие между жидкостью и газом будет только *количественным, но не качественным*. Вещество из одного состояния можно перевести в другое непрерывно. Именно это имеют в виду, когда, следуя Ван-дер-Ваальсу, говорят о *непрерывности жидкого и газообразного состояний вещества*. Только в области под граничной кривой $ALKG$, где вещество распадается на две фазы, отделенные одна от другой границей раздела и обладающие различными физическими свойствами, физически оправдано различать жидкое и газообразное состояния вещества.

8. Теория Ван-дер-Ваальса, которую мы излагали выше, при всей ее простоте дает стройную, качественно верную картину поведения вещества при его превращениях из жидкого состояния в газообразное и обратно. Она впервые привела в систему запутанные явления, связанные с такими превращениями. Невольно испытываешь чувство восхищения перед простотой и красотой теории, охватывающей столь громадный круг явлений. И тем не менее мы еще очень далеки от полного физического понимания этих явлений. В наиболее интересной области сильно сжатых газов и жидкостей, включая область двухфазных состояний вещества, уравнение Ван-дер-Ваальса не только не обосновано теоретически, но даже приводит к значительным количественным расхождениям с опытом. Теория Ван-дер-Ваальса может быть охарактеризована как термодинамическая теория, основанная на уравнении состояния, которое является результатом экстраполяции теоретических и экспериментальных данных. Между тем, всякая последовательная молекулярная теория должна без привлечения каких бы то ни было добавочных соображений дать полную количественную и качественную картину поведения вещества и, в частности, его фазовых превращений. Такой физической теории пока еще нет.

ЗАДАЧИ

1. Применим к круговым процессам $LBCL$ и $CAGC$ (см. рис. 98) равенство Клаузиуса (101.2). Тогда найдем, что площади циклов $LBCL$ и $CAGC$ равны нулю, т. е. кривая $LBCAG$ должна совпадать с прямой LCG. В чем причина этого противоречия?

Р е ш е н и е. Дело в том, что в двухфазной области изображающая точка определяет состояние вещества не однозначно. Она может изображать равновесное состояние либо двухфазной системы, либо физически однородного, но абсолютно неустойчивого вещества. Рассмотрим, например, цикл $LBCL$. В точке C при переходе с кривой BC на прямую CL физически однородное вещество распадается на жидкость и газ. Это необратимый процесс. К нему равенство Клаузиуса неприменимо. Надо пользоваться неравенством (38.5). В рассматриваемом случае оно сводится к неравенству $\oint\limits_{LBCL} P\,dV \leqslant 0$ и означает, что площадь цикла $LCBL$ отрицательна. Это действительно так. Для обратного цикла $LCBL$ площадь положительна, и неравенство Клаузиуса выполняться не может. Это означает, что такой цикл невозможен, т. е. в точке C вещество из двухфазного состояния не может превратиться в однофазное. Аналогично обстоит дело с циклом $GACG$. Прямой цикл $GACG$ термодинамически возможен, обратный — невозможен.

2. Чему равна теплоемкость C_P вещества в двухфазном состоянии, изображаемом точкой под кривой $ALKG$ (см. рис. 100)?

О т в е т. $C_P = \infty$. Достаточно заметить, что в указанной области изобары совпадают с изотермами.

§ 102. Свойства вещества в критическом состоянии. Определение критических параметров

1. Понятие критической температуры и критического состояния, как уже указывалось выше, было введено Эндрюсом на основе исследования экспериментальных изотерм углекислоты (CO_2). Предшественниками Эндрюса были Каньяр де ля Тур (1776–1859) и Д. И. Менделеев (1834–1907). Каньяр де ля Тур в 1822 г. заметил, что кварцевый шарик, помещенный в запаянную с обоих концов толстую металлическую трубу, заполненную спиртом, перекатывался в ней, почти не встречая сопротивления, когда труба была нагрета выше определенной температуры. С целью сделать явление видимым Каньяр де ля Тур повторил опыт, заменив металлическую трубу запаянными толстостенными стеклянными трубками, заполненными различными жидкостями. Оказалось, что указанное явление наступает при температуре исчезновения видимой границы раздела между паром и жидкостью. На основе таких опытов Каньяр де ля Тур пришел к заключению, что для каждой жидкости существует такая температура, выше которой вещество может существовать только в газообразном состоянии. К аналогичному заключению пришел и Д. И. Менделеев в 1861 г. Он заметил, что при определенной температуре прекращалось поднятие жидкости в капиллярах, т. е. поверхностное натяжение обращалось в нуль. При той же температуре обращалась в нуль скрытая теплота парообразования. Такую температуру Менделеев назвал *температурой*

абсолютного кипения. Выше этой температуры, согласно Менделееву, газ не может быть сконденсирован в жидкость никаким увеличением давления.

2. Критическую точку K мы определим как точку перегиба критической изотермы, в которой касательная к изотерме горизонтальна (см. § 100, п. 1). Ее можно определить также как точку, в которую в пределе переходят горизонтальные участки изотерм при повышении температуры до критической (см. рис. 100). На этом основан способ определения критических параметров P_k, V_k, T_k, принадлежащий Эндрюсу. Строится система изотерм при различных температурах. Предельная изотерма, у которой горизонтальный участок LG переходит в точку, будет критической изотермой, а указанная точка — критической точкой. Недостаток способа Эндрюса заключается в его громоздкости.

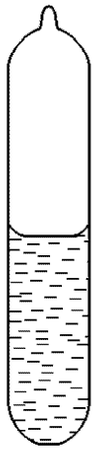


Рис. 101

3. Значительно более простым является метод исчезновения мениска, по существу принадлежащий Каньяру де ля Туру. Берется стеклянная или кварцевая ампула, частично заполненная жидкостью. Воздух из ампулы удаляется кипячением, а затем ампула запаивается. После этого ампула будет заполнена физически неоднородным веществом, состоящим из жидкости и ее насыщенного пара, отделенными друг от друга резкой видимой границей, называемой мениском (рис. 101). Для демонстрации подходящими веществами могут служить жидкая углекислота ($t_k = 31\,^\circ\mathrm{C}$, $P_k = 72{,}9$ атм) или эфир ($t_k = 194\,^\circ\mathrm{C}$, $P_k = 35$ атм). Заполненная ампула помещается в печь и нагревается. Для равномерности нагрева воздух в печи интенсивно перемешивается. Тепловым расширением стенок ампулы можно пренебречь, так что процесс нагревания происходит практически при постоянном объеме. Допустим, что количество вещества подобрано так, что точка *1*, изображающая начальное состояние его, лежит на вертикальной прямой $1K$, проходящей через критическую точку K (рис. 102). По мере нагревания изображающая точка перемещается вверх, оставаясь на прямой $1K$. Пока температура ниже критической, мениск виден отчетливо. При нагревании мениск остается практически на одной и той же высоте. При приближении к критической температуре кривизна мениска непрерывно уменьшается из-за уменьшения поверхностного натяжения на границе между жидкостью и ее насыщенным паром. Мениск становится плоским и, наконец, совсем исчезает, когда температура достигнет критического значения T_k. Начиная с этого момента,

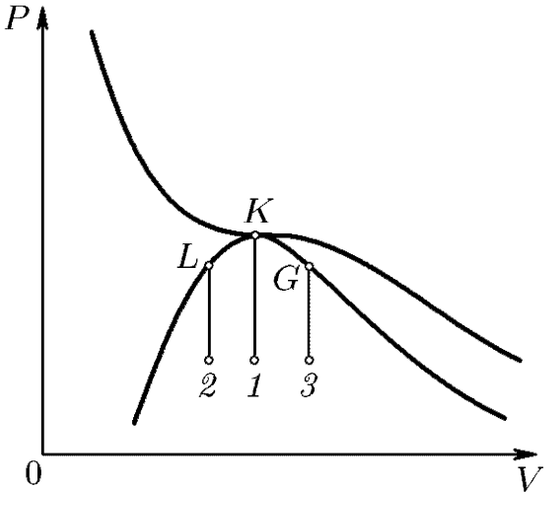


Рис. 102

вещество становится физически однородным и продолжает оставаться таковым при дальнейшем нагревании. При обратном охлаждении вещество продолжает оставаться физически однородным, пока его температура выше критической. При прохождении через критическую точку K снова происходит разделение вещества на две фазы. Образуются мелкие капельки тумана, вещество становится мутным. Капельки быстро оседают, и снова появляется мениск, разделяющий жидкость и пар.

4. Допустим теперь, что жидкости в ампулу налито больше, «чем нужно». Тогда точка, изображающая состояние вещества, будет при нагревании перемещаться из начального положения *2* вдоль вертикальной прямой $2L$ (рис. 102). В положении L, когда температура еще ниже критической, все вещество превратится в жидкость. Поэтому казалось бы, что никакого исчезновения мениска при нагревании происходить не должно. Мениск будет просто монотонно перемещаться вверх ампулы. Когда он достигнет верхнего конца ампулы, последняя окажется целиком заполненной однородной жидкостью. Аналогичное явление должно было бы происходить и в том случае, когда жидкости налито меньше, «чем нужно». Из исходного положения *3* изображающая точка при нагревании должна перемещаться вверх вдоль вертикальной прямой $3G$. В точке G, т. е. еще до достижения критической температуры, все вещество перейдет в газообразное состояние. Опять, казалось бы, никакого исчезновения мениска на границе между жидкостью и паром наблюдаться не должно. Мениск просто должен был бы перемещаться вниз и исчезать на дне ампулы, когда температура еще не достигла критической.

Таким образом, казалось бы, что для наблюдения исчезновения мениска в ампулу должно быть налито строго определенное количество жидкости. Если бы это было так, то метод исчезновения мениска практически нельзя было бы осуществить. В действительности дело обстоит совсем не так. Для исчезновения мениска вовсе не обязательно наполнять ампулу строго определенным количеством жидкости. Исчезновение мениска наблюдается при любом количестве жидкости в ампуле, если только оно не очень сильно отклоняется от того количества, «которое требуется». Дело в том, что согласно первому уравнению (100.3) сжимаемость вещества в критической точке бесконечно велика:

$$\gamma \equiv \frac{1}{V}\left(\frac{\partial V}{\partial P}\right)_T = \infty. \tag{102.1}$$

Поэтому в критической точке и в ее окрестности вещество внизу ампулы должно заметно уплотняться под действием собственного веса, а вверху становиться менее плотным. На такой «гравитационный эффект» обратил внимание бельгийский физик Гуи еще в 1892 г. Однако экспериментальное обнаружение эффекта сделано сравнительно недавно. Для того чтобы получить представление о величине эффекта, мы приводим в табл. 11 распределение по высоте плотности гептана

(C_7H_{16}), измеренное А. З. Голиком и Е. Т. Шиманской в Киевском университете (см. также задачу к этому параграфу). В таблице t_m означает температуру исчезновения мениска, которая принимается равной

Таблица 11

Расстояние от дна ампулы, см	0,2	3,8	4,6	5,8	6,8
Плотность при $t = t_m$, г/см^3	0,273	0,235	0,205	0,195	0,185
Плотность при $t - t_m = 1\,^\circ$C, г/см^3	0,252	0,240	0,230	0,215	0,200

критической. Критическая плотность для гептана равна 0,2355 г/см^3, критическая температура 266,8 °C, критическое давление 26,8 атм. Из таблицы видно, что при изменении высоты всего на 6,6 см плотность гептана при критической температуре изменяется на 37 %, а при температуре градусом выше — на 22 %, если критическую плотность принять за 100 %. Для сравнения вычислим по барометрической формуле относительное изменение плотности воздуха с той же высотой $h = 6{,}6$ см при температуре $T = 273$ К. Оно равно

$$\frac{\delta\rho}{\rho} = \frac{\mu g h}{RT} = 0{,}82 \cdot 10^{-5} = 0{,}82 \cdot 10^{-3}\,\%,$$

т. е. примерно в пятьдесят тысяч раз меньше, чем для гептана в окрестности критической точки. При интерпретации многих опытных данных гравитационный эффект не принимался во внимание, что иногда приводило к ошибочным выводам. Теперь становится ясным, что в описанных выше опытах вовсе не обязательно наполнять ампулу строго определенным количеством жидкости. Благодаря сильному изменению плотности вещества с высотой при критической температуре в ампуле может находиться в критическом состоянии (т. е. иметь еще и критическую плотность) только бесконечно тонкий слой вещества. В этом слое и происходит исчезновение мениска. Таким образом, получается совсем не жесткое требование, которое следует предъявлять к количеству налитой жидкости. Для применимости метода исчезновения мениска требуется только, чтобы упомянутый выше бесконечно тонкий слой вещества не выходил за пределы ампулы.

5. Метод исчезновения мениска удобен для точных измерений критической температуры. Определив критическую температуру, можно затем сравнительно просто измерить критическое давление. Для этого можно, например, подвергать вещество изотермическому сжатию и наблюдать, когда на критической изотерме появится точка перегиба. Наибольшие трудности представляет измерение критической плотности, поскольку в критической точке сжимаемость вещества бесконечна. Критическую плотность можно измерить на той же установке, какая применяется в методе исчезновения мениска. Для этого надо взять много ампул и налить в них жидкости либо «слишком много», либо «слишком мало», чтобы при нагревании мениск достигал вершины или дна ампулы. В моменты, когда это происходит, ампула

будет целиком заполнена либо только жидкостью, либо только насыщенным паром. Разделив в эти моменты массу вещества на объем ампулы, найдем плотность жидкости $\rho_{\text{ж}}$ и насыщенного пара $\rho_{\text{п}}$ при соответствующей температуре. Построим теперь кривые зависимости плотности от температуры для насыщенного пара (нижняя кривая на рис. 103) и жидкости, находящейся с ним в равновесии (верхняя кривая на рис. 103). Кроме того, построим кривую AK для средней арифметической плотности жидкости и пара, т. е. для величины $(1/2)(\rho_{\text{ж}} + \rho_{\text{п}})$. Опыт показывает, что на самом деле она с большой точностью является прямой. Точка пересечения этой прямой с вертикальной изотермой $T = T_k$ и будет критической точкой.

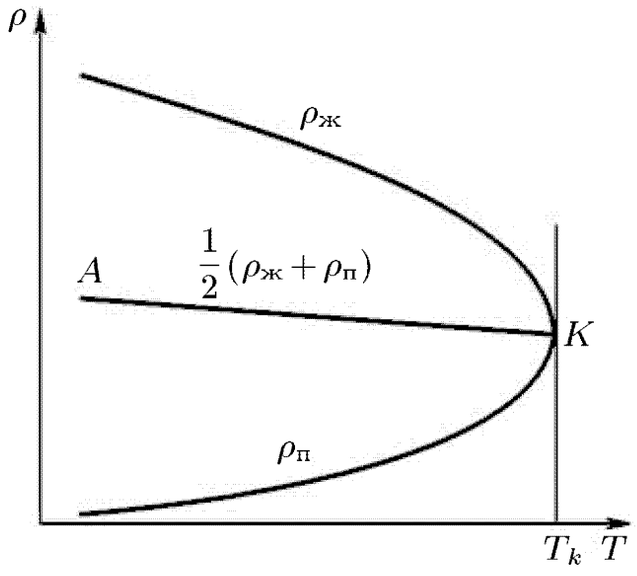


Рис. 103

6. А. И. Надеждин (1858–1886), один из учеников профессора М. А. Авенариуса (1835–1895), применял для определения критической температуры несколько иной метод. Он изобрел прибор, названный им *дифференциальным денсиметром*. Устройство прибора понятно

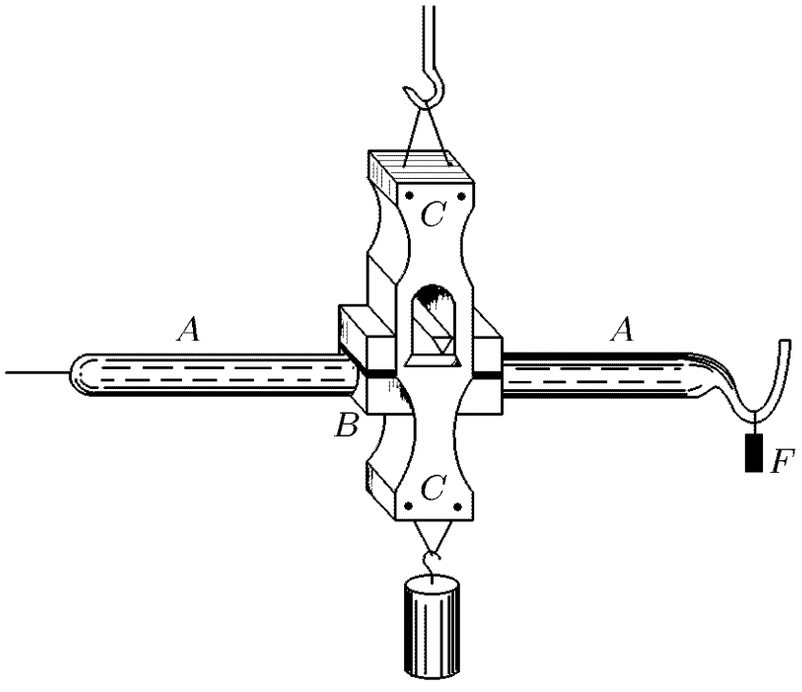


Рис. 104

из рис. 104. Пустая трубка AA устанавливалась в горизонтальном положении с помощью грузика F. Если трубку частично заполнить жидкостью, то жидкость собирается в одном из ее концов, и равновесие

нарушается. Однако при нагревании до критической температуры, когда плотности жидкости и пара выравниваются, равновесие восстанавливается вновь. Этот остроумный метод обладает тем достоинством, что он применим и для непрозрачных веществ (бром, йод), когда мениск невидим, а также в тех случаях, когда вещество разъедает стекло (вода при высоких температурах). Однако он менее точен, чем метод исчезновения мениска, и из-за гравитационного эффекта всегда дает завышенное значение критической температуры.

7. Отметим в заключение еще некоторые свойства вещества в критической точке и ее окрестности. Из формулы (35.3), т. е.

$$C_P - C_V = -T\left(\frac{\partial P}{\partial T}\right)_V^2 \left(\frac{\partial V}{\partial P}\right)_T, \tag{102.2}$$

следует, что разность теплоемкостей $C_P - C_V$ в критической точке обращается в бесконечность. Действительно, изотермическая сжимаемость в этой точке бесконечно велика, а производная $(\partial P/\partial T)_V$ никаких особенностей не имеет. Поэтому должна обращаться в бесконечность и теплоемкость C_P, так как для устойчивости вещества необходимо, чтобы C_V была больше нуля. Такой результат представляется совершенно естественным, если принять во внимание, что в двухфазном состоянии вещества теплоемкость C_P бесконечно велика (см. задачу 2 к предыдущему параграфу). Она остается таковой и в пределе, когда двухфазное состояние переходит в критическое. Что касается теплоемкости C_V, то теоретические соображения, на которых мы останавливаться не можем, тоже приводят к заключению, что в критической точке $C_V = \infty$. Этот вывод, по-видимому, подтверждается опытом.

Более интересной особенностью вещества в критическом состоянии является *медленность установления равновесного состояния.* Например, для гептана равновесное распределение плотности вещества в поле тяжести при сохранении температуры строго постоянной устанавливается в течение не менее суток. Если после установления равновесия температура изменится хотя бы на сотые доли градуса, то также требуется время не менее суток, чтобы установилось новое равновесное состояние. С этим связано явление *гистерезиса плотности.* Оно состоит в следующем. Если плотность вещества в ампуле измерять в одной и той же точке, постепенно повышая температуру, а затем постепенно понижая ее, то при одной и той же температуре получатся разные значения плотности в зависимости от того, как достигнута эта температура — при нагревании или при охлаждении. Гистерезис плотности (для гептана) исчезает при температуре на 10–15 °C выше критической.

ЗАДАЧА

Найти распределение плотности в поле тяжести физически однородного вещества, подчиняющегося уравнению Ван-дер-Ваальса, в окрестности критической точки.

Р е ш е н и е. В критической точке $(\partial P/\partial V)_T = (\partial^2 P/\partial V^2)_T = 0$. Поэтому первый член разложения Тейлора в окрестности этой точки имеет вид

$$P - P_k = \frac{1}{6}\Big(\frac{\partial^3 P}{\partial V^3}\Big)_{\text{кр}}(V - V_k)^3 + \Big(\frac{\partial P}{\partial T}\Big)_{\text{кр}}(T - T_k).$$

Вычислив производные из уравнения Ван-дер-Ваальса и воспользовавшись формулами (100.2), получим

$$P - P_k = -\frac{9}{16}\,\frac{RT_k}{V_k^4}\,(V - V_k)^3 + \frac{3}{2}\,\frac{R}{V_k}(T - T_k).$$

Вместо V введем плотность $\rho = \mu/V$, где μ — молярная масса. Использовав еще уравнение гидростатики $P - P_k = -\rho_k g h$, получим в рассматриваемом приближении

$$\frac{\rho - \rho_k}{\rho_k} = -\frac{2}{3}\sqrt[3]{6\,\frac{\mu g h + (3/2)R(T - T_k)}{RT_k}}. \tag{102.3}$$

Высота h отсчитывается от того уровня, где плотность вещества равна критической, причем положительным считается направление вверх. В частности, при $T = T_k$

$$\frac{\rho - \rho_k}{\rho_k} = -\frac{2}{3}\sqrt[3]{6\,\frac{\mu g h}{RT_k}}. \tag{102.4}$$

Вдали от критической точки газ можно считать идеальным. В этом случае для относительного изменения плотности с высотой мы имели бы

$$\frac{\delta\rho}{\rho} = -\frac{\mu g h}{RT}. \tag{102.5}$$

При одинаковых температурах и молярных массах эта величина меньше предыдущей в $\alpha = \dfrac{2\sqrt{6}}{3}\Big(\dfrac{RT}{\mu g h}\Big)^{2/3}$ раз. Для воздуха ($\mu = 28{,}8$ г/моль, $T_k = 132{,}5$ К) при высоте $h = 1$ см коэффициент $\alpha \approx 8700$, $(\rho - \rho_k)\rho_k \approx -1/100$.

§ 103. Внутренняя энергия газа Ван-дер-Ваальса

1. Если известно уравнение состояния физически однородного вещества, то с помощью общих методов термодинамически можно найти внутреннюю энергию его как функцию объема V при постоянной температуре. Для этого надо воспользоваться уравнением (34.2). Применим этот метод к газу Ван-дер-Ваальса, предполагая, что газ взят в количестве одного моля. Перепишем еще раз уравнение (34.2):

$$\Big(\frac{\partial U}{\partial V}\Big)_T = T\Big(\frac{\partial P}{\partial T}\Big)_V - P. \tag{103.1}$$

Из уравнения Ван-дер-Ваальса находим

$$T\Big(\frac{\partial P}{\partial T}\Big)_V = \frac{RT}{V - b} = P + \frac{a}{V^2}.$$

Таким образом,

$$\Big(\frac{\partial U}{\partial V}\Big)_T = \frac{a}{V^2}. \tag{103.2}$$

Интегрирование вдоль изотермы дает

$$U = -\frac{a}{V} + f(T), \tag{103.3}$$

где $f(T)$ — «постоянная интегрирования», которая может зависеть от температуры T. Ее можно выразить через теплоемкость газа C_V. Согласно (18.3)

$$C_V = \left(\frac{\partial U}{\partial T}\right)_V = \frac{df}{dT}. \tag{103.4}$$

Отсюда следует, что теплоемкость C_V газа Ван-дер-Ваальса может зависеть только от температуры T. Поэтому из последнего соотношения получаем

$$f(T) = \int C_V(T)\, dT,$$

и, следовательно,

$$U = \int C_V(T)\, dT - \frac{a}{V}. \tag{103.5}$$

Если можно пренебречь зависимостью теплоемкости от температуры, то эта формула переходит в более простую:

$$U = C_V T - \frac{a}{V}. \tag{103.6}$$

При выводе предполагалось, что вещество является физически однородным. Поэтому полученные результаты справедливы только для однофазных состояний вещества. Для двухфазных состояний они несправедливы. В частности, если изотерма, вдоль которой производилось интегрирование при получении формулы (103.3), пересекает двухфазную область, то интегрирование надо производить не по реальной, а по теоретической изотерме, содержащей поднимающийся участок, соответствующий абсолютно неустойчивым состояниям вещества.

2. Формулу (103.5) можно получить также из простых молекулярно-кинетических соображений. Внутренняя энергия газа складывается из кинетической энергии теплового движения его молекул и потенциальной энергии их взаимодействия. Ввиду теоремы о равномерном распределении кинетической энергии по степеням свободы кинетическая энергия теплового движения однозначно определяется температурой газа и не зависит от его объема. Если температура поддерживается постоянной, то остается постоянной и кинетическая энергия. Если пользоваться моделью идеально твердых шаров, то силам отталкивания (развивающимся при столкновениях шаров) не будет соответствовать, никакая потенциальная энергия. Вся потенциальная энергия обусловлена только силами притяжения между молекулами. Ее можно вычислить макроскопически, так как силы притяжения между молекулами сводятся к молекулярному давлению $P_i = a/V^2$, с которым поверхностный слой газа давит на остальную массу газа. Работа, которую надо затратить против этого внутреннего давления, и дает

приращение потенциальной энергии при изотермическом расширении газа. При расширении от объема V_1 до объема V это приращение равно

$$\int\limits_{V_1}^{V} P_i\,dV = \int\limits_{V_1}^{V} \frac{a}{V^2}\,dV = -\frac{a}{V} + \frac{a}{V_1} = -\frac{a}{V} + \text{const}.$$

Прибавив его к кинетической энергии теплового движения $f(T)$, мы снова придем к формуле (103.3).

ЗАДАЧА

Найти выражение для энтропии газа Ван-дер-Ваальса.

Решение. $dS = (\partial S/\partial V)_T\,dV + (\partial S/\partial T)_V\,dT$. Используя формулу (45.17) и определение теплоемкости, получим

$$dS = \left(\frac{\partial P}{\partial T}\right)_V dV + \frac{C_V}{T}\,dT.$$

Далее, пользуемся уравнением Ван-дер-Ваальса и интегрируем. Для энтропии одного моля газа получаем

$$S = R\ln(V-b) + \int \frac{C_V(T)}{T}\,dT + \text{const}. \tag{103.7}$$

Для газа с переменным числом частиц

$$S = \nu\left[R\ln\frac{V-\nu b}{\nu} + \int \frac{C_V(T)}{T}\,dT + \text{const}\right], \tag{103.8}$$

где ν — число молей в газе, а аддитивная постоянная в квадратных скобках от числа частиц не зависит. Если теплоемкость C_V не зависит от температуры, то

$$S = \nu\left[R\ln\frac{V-\nu b}{\nu} + C_V\ln T + \text{const}\right]. \tag{103.9}$$

§ 104. Эффект Джоуля–Томсона для газа Ван-дер-Ваальса

1. В § 46 была изложена термодинамическая теория эффекта Джоуля–Томсона. Для получения из нее конкретных результатов надо знать уравнение состояния газа. Примем, что таковым является уравнение Ван-дер-Ваальса. Начнем с дифференциального эффекта Джоуля–Томсона. В § 46 было показано, что если по разные стороны пробки в опыте Джоуля–Томсона поддерживается малая разность давлений $P_2 - P_1 = \Delta P$, то соответствующая ей разность температур, возникающая при стационарном течении газа, определяется формулой

$$\frac{\Delta T}{\Delta P} = \left(\frac{\partial T}{\partial P}\right)_I = \frac{T(\partial V/\partial T)_P - V}{C_P}. \tag{104.1}$$

Так как газ течет от большего давления к меньшему, то $P_2 - P_1$ есть величина отрицательная. Если при протекании через пробку температура газа понижается ($\Delta T/\Delta P > 0$), то эффект Джоуля–Томсона называется *положительным*. В противоположном случае,

когда температура повышается ($\Delta T/\Delta P < 0$), эффект называется *отрицательным*.

Вычислять производную $(\partial V/\partial T)_P$ из уравнения Ван-дер-Ваальса неудобно, поскольку оно является уравнением третьей степени относительно V. Удобнее воспользоваться тождеством

$$\left(\frac{\partial V}{\partial T}\right)_P = -\left(\frac{\partial V}{\partial P}\right)_T \left(\frac{\partial P}{\partial T}\right)_V = -\left(\frac{\partial P}{\partial T}\right)_V \Big/ \left(\frac{\partial P}{\partial V}\right)_T$$

и преобразовать формулу (104.1) к виду

$$\frac{\Delta T}{\Delta P} = -\frac{T(\partial P/\partial T)_V + V(\partial P/\partial V)_T}{C_P(\partial P/\partial V)_T}. \tag{104.2}$$

Вычисляя производные $(\partial P/\partial T)_V$ и $(\partial P/\partial V)_T$ из уравнения Ван-дер-Ваальса и подставляя их значения в эту формулу, получим

$$\frac{\Delta T}{\Delta P} = \frac{bRT/(V-b)^2 - 2a/V^2}{C_P(\partial P/\partial V)_T}. \tag{104.3}$$

2. Рассмотрим простейший случай разреженного газа. В этом случае члены, содержащие a и b, являются малыми поправками. Можно ограничиться линейным приближением, отбросив все члены с высшими степенями a и b. В этом приближении следует $V-b$ заменить на V, а производную $(\partial P/\partial V)_T$ вычислить из уравнения Клапейрона. Тогда

$$\frac{\Delta T}{\Delta P} = \frac{2a/RT - b}{C_P}. \tag{104.4}$$

Если $2a/RT - b > 0$, то $\Delta T/\Delta P > 0$, и газ охлаждается. Если, напротив, $2a/RT - b < 0$, то $\Delta T/\Delta P < 0$, и газ нагревается. В частности, при $b = 0$, $a \neq 0$ газ всегда охлаждается; при $a = 0$, $b \neq 0$ — всегда нагревается. Первый результат понятен, так как при расширении кинетическая энергия теплового движения газа расходуется на работу против сил молекулярного притяжения. Второй результат не столь очевиден, так как здесь при расширении молекулярные силы не производят никакой работы. Однако результат процесса Джоуля–Томсона зависит не только от изменения внутренней энергии газа, но и от внешней работы, которую он совершит при расширении. Процесс Джоуля–Томсона есть процесс, происходящий при постоянной энтальпии. Если $a = 0$, то уравнение Ван-дер-Ваальса переходит в $P(V-b) = RT$. Поэтому энтальпия газа [1])

$$I = U + PV = (C_V + R)T + Pb = C_P T + Pb.$$

Условие постоянства энтальпии дает $C_P\Delta T + b\Delta P = 0$, т. е.

$$\frac{\Delta T}{\Delta P} = -\frac{b}{C_P} < 0,$$

[1]) С помощью формулы (35.2) легко убедиться, что соотношение Р. Майера $C_P = C_V + R$ в рассматриваемом случае справедливо.

что находится в согласии с формулой (104.4).

При $T = 2a/Rb$ изменение температуры равно нулю. Температура

$$T_i = \frac{2a}{Rb} = \frac{27}{4} T_k = 2T_B \tag{104.5}$$

называется *температурой инверсии дифференциального эффекта Джоуля–Томсона.* Ниже этой температуры в опыте Джоуля–Томсона газ охлаждается, выше — нагревается. Для большинства газов температура инверсии лежит значительно выше комнатной температуры. Такие газы в опыте Джоуля–Томсона охлаждаются. Для водорода и гелия температура инверсии лежит значительно ниже комнатной температуры. Для них эффект Джоуля–Томсона отрицателен, т. е. эти газы нагреваются.

3. Перейдем теперь к случаю плотных газов. Применяя к ним уравнение Ван-дер-Ваальса, мы получим результаты, верные лишь качественно. Но излагаемые методы применимы не только к уравнению Ван-дер-Ваальса, но и к любому уравнению состояния. Положив в основу расчета какое-либо более точное эмпирическое уравнение состояния, можно получить и более точные результаты.

Так как величина C_P положительна, а производная $(\partial P/\partial V)_T$ отрицательна, то знак эффекта Джоуля–Томсона будет определяться знаком числителя в формуле (104.3). Эффект отрицателен, т. е. имеет место нагревание, если

$$\frac{bRT}{(V-b)^2} - \frac{2a}{V^2} > 0,$$

или

$$T > \frac{2a}{Rb}\left(\frac{V-b}{V}\right)^2. \tag{104.6}$$

Положительный эффект, т. е. охлаждение, получится в случае

$$T < \frac{2a}{Rb}\left(\frac{V-b}{V}\right)^2. \tag{104.7}$$

При температуре

$$T = \frac{2a}{Rb}\left(\frac{V-b}{V}\right)^2 \tag{104.8}$$

поправки Ван-дер-Ваальса a и b взаимно компенсируют друг друга, и дифференциальный эффект Джоуля–Томсона становится нулевым. Уравнение (104.8) изображает на плоскости VT кривую, называемую *кривой инверсии дифференциального эффекта Джоуля–Томсона.* В приведенных параметрах τ и φ оно имеет вид

$$\tau = \frac{27}{4}\left(\frac{\varphi - 1/3}{\varphi}\right)^2. \tag{104.9}$$

Соответствующая кривая инверсии изображена на рис. 105 сплошной линией. В области, ограниченной этой кривой и осью абсцисс, дифференциальный эффект Джоуля–Томсона положителен ($\Delta T < 0$). В

области над кривой инверсии эффект отрицателен ($\Delta T > 0$). Область $\varphi < 1/3$ не имеет физического смысла, так как для газа Ван-дер-Ваальса объем V не может быть меньше b, а следовательно, $\varphi > 1/3$.

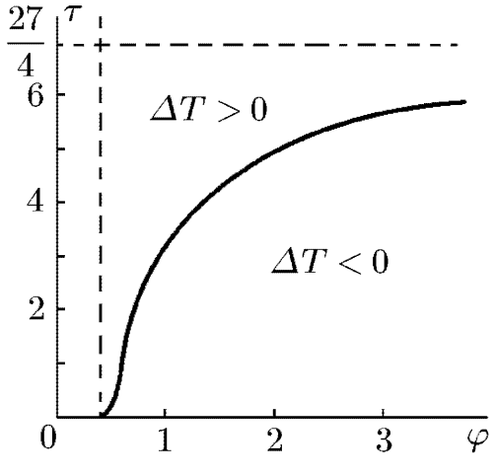


Рис. 105

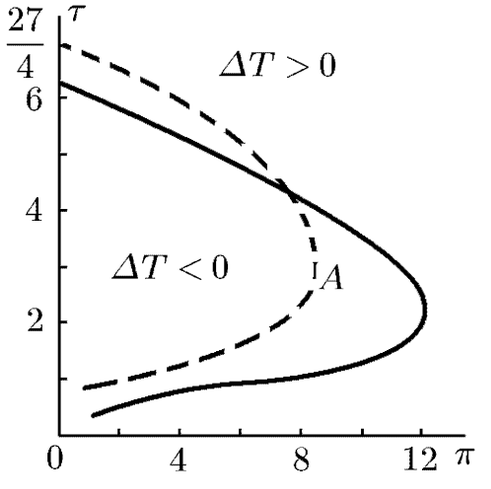


Рис. 106

Для практических применений уравнение кривой инверсии (104.9) удобнее записать в координатах τ, π. С этой целью воспользуемся приведенным уравнением Ван-дер-Ваальса (100.7). Исключая из него и из уравнения (104.9) объем φ, после несложных вычислений получим следующее уравнение кривой инверсии:

$$\pi = 24\sqrt{3\tau} - 12\tau - 27. \tag{104.10}$$

Решая это уравнение, найдем

$$\tau = 3\left(1 \pm \frac{1}{6}\sqrt{9-\pi}\right)^2. \tag{104.11}$$

График этой двузначной функции приведен на рис. 106 (штриховая линия). Он показывает, что при каждом давлении, пока оно не превосходит определенного предела ($\pi < 9$), существуют две точки инверсии дифференциального эффекта Джоуля–Томсона. Эффект Джоуля–Томсона положителен ($\Delta T < 0$) ниже верхней и выше нижней точек инверсии. Положение точек инверсии зависит от давления. По мере повышения давления нижняя точка инверсии поднимается, а верхняя опускается, пока они не сольются в точке A (рис. 106). Эти предсказания теории Ван-дер-Ваальса качественно согласуются с опытом, хотя и существуют значительные расхождения. В качестве примера, наряду с ван-дер-ваальсовой кривой инверсии, мы приводим экспериментальную кривую инверсии для водорода. Она изображена на рис. 106 сплошной линией. (Для водорода $T_k = 33{,}3$ К, $P_k = 12{,}8$ атм.)

4. Перейдем теперь к рассмотрению интегрального эффекта Джоуля–Томсона. В § 46 уже говорилось, что интегральный эффект Джоуля–Томсона получают, заставляя газ перетекать через вентиль, по разные стороны которого поддерживается большая разность давле-

ний. Изменение температуры газа $T_2 - T_1$ при интегральном эффекте определяется формулой (46.2), т. е.

$$T_2 - T_1 = \int_{P_1}^{P_2} \left(\frac{\partial T}{\partial P}\right)_I dP = \int_{P_1}^{P_2} \frac{T(\partial V/\partial T)_P - V}{C_P}\, dP. \qquad (104.12)$$

В зависимости от знака подынтегрального выражения изменение температуры $T_2 - T_1$ может быть как положительным, так и отрицательным. В частности, если во всем диапазоне давлений дифференциальный эффект Джоуля–Томсона положителен, то таким же будет и интегральный эффект, т. е. в результате дросселирования газ должен охлаждаться. При комнатных температурах это имеет место для большинства газов, в частности для воздуха и углекислоты. Если взять баллон с газообразной углекислотой под давлением 100–200 атм и заставить ее вытекать в атмосферу через вентиль, то получится настолько значительное охлаждение, что углекислота перейдет в твердое состояние. Совсем иначе ведет себя водород. Для него при комнатных температурах дифференциальный, а следовательно, и интегральный эффект Джоуля–Томсона отрицателен; при дросселировании водород нагревается. Такое нагревание иногда приводило к катастрофам, в которых сильно сжатый водород самопроизвольно воспламенялся при истечении из поврежденных труб. При внезапном расширении водород может охлаждаться лишь тогда, когда его температура ниже $-80\,^\circ$C. При более высоких температурах водород всегда нагревается.

В связи с использованием интегрального эффекта Джоуля–Томсона для получения низких температур представляет интерес следующий вопрос. Пусть газ при начальной температуре T_1 расширяется в процессе Джоуля–Томсона до некоторого постоянного давления P_2 (например, атмосферного). Какое следует взять начальное давление P_1, чтобы получилось максимальное охлаждение? Разность температур $T_2 - T_1$ определяется формулой (104.12). При фиксированной температуре T_1 она зависит только от начального давления P_1 (поскольку конечное давление P_2 также фиксировано). Из формулы (104.12) получаем

$$\frac{d}{dP_1}(T_2 - T_1) = \frac{dT_2}{dP_1} = -\left(\frac{\partial T}{\partial P}\right)_I \Bigg|_{P=P_1, T=T_1}.$$

Приравнивая эту производную нулю, находим условие максимума:

$$\left(\frac{\partial T}{\partial P}\right)_I = 0 \quad \text{при} \quad P = P_1, \quad T = T_1. \qquad (104.13)$$

Но уравнение $(\partial T/\partial P)_I = 0$ есть уравнение кривой инверсии дифференциального эффекта. Таким образом, чтобы получить максимальное охлаждение, необходимо взять начальную точку T_1, P_1 на кривой инверсии дифференциального эффекта Джоуля–Томсона.

В одном, и притом наиболее важном, случае интегрирование в формуле (104.12) просто доводится до конца. Это случай, когда газ в исходном состоянии находится под высоким давлением, а после дросселирования его давление падает настолько низко, что в конечном состоянии он может рассматриваться как идеальный. Расчет проще выполнить не по формуле (104.12), а непосредственно с помощью равенства энтальпий газа в начальном и

конечном состояниях. Используя выражение (103.5), равенство энтальпий для газа Ван-дер-Ваальса можно записать в виде

$$\int_{T_0}^{T_1} C_V(T)\,dT - \frac{a}{V_1} + P_1V_1 = \int_{T_0}^{T_2} C_V(T) - \frac{a}{V_2} + P_2V_2.$$

Оба интеграла, входящие сюда, мы объединим в один. В правой части можно пренебречь членом a/V_2 и воспользоваться уравнением Клапейрона $P_2V_2 = RT_2$. В этом приближении

$$\int_{T_2}^{T_1} C_V(T)\,dT - \frac{a}{V_1} + P_1V_1 = RT_2.$$

Давление P_1 находим из уравнения Ван-дер-Ваальса. В результате получаем

$$\overline{C}_V(T_1 - T_2) - \frac{2a}{V_1} + \frac{RT_1V_1}{V_1 - b} = RT_2,$$

где $\overline{C}_V$ — средняя теплоемкость газа при постоянном объеме в температурном интервале от T_1 до T_2. Из последней формулы находим

$$T_2 - T_1 = \frac{1}{R + \overline{C}_V}\Big(\frac{RbT_1}{V_1 - b} - \frac{2a}{V_1}\Big). \tag{104.14}$$

Так как знаменатель — величина существенно положительная, то знак эффекта определяется только знаком числителя. Эффект положителен (охлаждение) при $T_1 < \dfrac{2b}{Rb}\,\dfrac{V_1 - b}{V_1}$ и отрицателен (нагревание) при $T_1 > \dfrac{2a}{Rb}\,\dfrac{V_1 - b}{V_1}$. В дальнейшем в целях упрощения написания формул начальные параметры будем обозначать просто T, V, P, опуская индекс 1. При температуре

$$T = \frac{2a}{Rb}\,\frac{V - b}{V} \tag{104.15}$$

эффект исчезает. Эта температура называется *температурой инверсии интегрального эффекта Джоуля–Томсона.* Она всегда выше соответствующей температуры для дифференциального эффекта. Уравнение (104.15) определяет на плоскости кривую, называемую *кривой инверсии интегрального эффекта Джоуля–Томсона.* Записанное в приведенных параметрах, уравнение (104.15) имеет вид

$$\tau = \frac{27}{4}\,\frac{\varphi - 1/3}{\varphi}. \tag{104.16}$$

Кривая, изображаемая уравнением (104.16), приведена на рис. 107 (сплошная линия). Если исходная точка с приведенным объемом φ и приведенной температурой τ, изображающая начальное состояние газа, лежит между кривой инверсии (104.16) и осью абсцисс, то при дросселировании газ будет охлаждаться. Если же она лежит над кривой инверсии интегрального эффекта, то газ будет нагреваться. На рис. 107 штрихом изображена также кривая инверсии дифференциального эффекта. Она всегда лежит под кривой инверсии интегрального эффекта. Это и понятно. Действительно, выше было отмечено, что если во всем интервале изменения давления дифференциальный эффект положителен (отрицателен), то тем более будет положителен (отрицателен) интегральный эффект. Обратное

несправедливо, так как для положительности интегрального эффекта не обязательно, чтобы подынтегральное выражение в формуле (104.12) было всюду положительно. На отдельных участках оно может быть отрицатель-

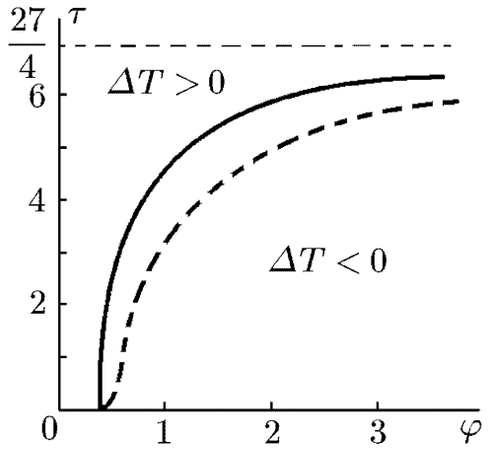


Рис. 107

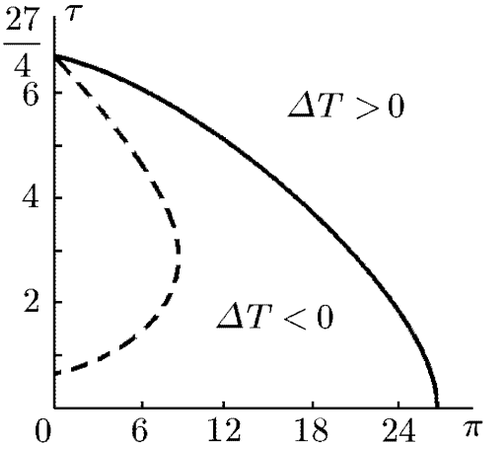


Рис. 108

ным, и тем не менее интегральный эффект Джоуля–Томсона может остаться положительным. Таким образом, область плоскости VT, в которой дифференциальный эффект положителен, всегда лежит внутри области, в которой положителен интегральный эффект.

Уравнение (104.16), записанное в переменных π и τ, имеет вид

$$\tau = \frac{3}{4}\sqrt{81 - 3\pi}. \tag{104.17}$$

Кривая, изображенная этим уравнением, представлена на рис. 108 сплошной линией. Соответствующая кривая инверсии для дифференциального эффекта изображена штриховой линией. В отличие от дифференциального эффекта интегральный эффект Джоуля–Томсона имеет для каждого давления ($\pi < 27$) лишь одну точку инверсии.

ЗАДАЧА

Найти кривую инверсии дифференциального эффекта Джоуля–Томсона для газа, подчиняющегося уравнению Дитеричи.

Ответ. $T = \dfrac{2a}{Rb}\left(1 - \dfrac{b}{V}\right)$. В приведенных параметрах $\tau = 8\left(1 - \dfrac{1}{2\varphi}\right)$ или $\pi = (8 - \tau)e^{5/2 - 4/\tau}$.

§ 105. Методы получения низких температур и сжижения газов

1. В технике применяются три основных метода для получения низких температур: 1) *испарение жидкостей*, 2) *использование эффекта Джоуля–Томсона*, 3) *обратимое адиабатическое расширение газа с совершением внешней работы*. Применяются также различные *охлаждающие смеси*. Так, путем смешения поваренной соли со снегом можно получить температуру ниже минус 20 °C, а смешением хлороформа или эфира с твердой углекислотой — минус 77 °C. По принципу испарения жидкостей работают домашние холодильники. Методы 2) и

3) основаны на газовых законах. Термодинамическая теория метода 2) содержится в формуле (104.12). Идея метода 3) по существу заключена в уравнении адиабаты Пуассона в форме (21.4). Однако это уравнение справедливо только для идеальных газов. При низких температурах, в особенности вблизи температуры сжижения газа, оно несправедливо. Поэтому необходимо подробнее остановится на теории метода 3), не вводя предположения об идеальности газа.

При обратном адиабатическом расширении остается постоянной энтропия газа S. Рассматривая ее как функцию температуры и давления, можно написать для элементарного обратимого процесса расширения:

$$\Delta S = \left(\frac{\partial S}{\partial T}\right)_P \Delta T + \left(\frac{\partial S}{\partial P}\right)_T \Delta P = 0.$$

Очевидно,

$$\left(\frac{\partial S}{\partial T}\right)_P = \frac{1}{T}\left(\frac{T\partial S}{\partial T}\right)_P = \frac{1}{T}\left(\frac{\delta Q}{\partial T}\right)_T = \frac{C_P}{T}.$$

Кроме того, согласно формуле (45.18),

$$\left(\frac{\partial S}{\partial P}\right)_T = -\left(\frac{\partial V}{\partial T}\right)_P.$$

Поэтому

$$\frac{C_P}{T}\,\Delta T - \left(\frac{\partial V}{\partial T}\right)_P \Delta P = 0.$$

Отсюда

$$\frac{\Delta T}{\Delta P} = \left(\frac{\partial T}{\partial P}\right)_S = \frac{T(\partial V/\partial T)_P}{C_P}. \tag{105.1}$$

Если температурный коэффициент объемного расширения положителен (как у газов), то при адиабатическом расширении получается охлаждение. Соответствующий интегральный эффект описывается формулой

$$T_2 - T_1 = \int_{P_1}^{P_2} \frac{T}{C_P}\left(\frac{\partial V}{\partial T}\right)_P dP. \tag{105.2}$$

Сравнивая ее с формулой (104.12), видим, что при прочих равных условиях охлаждение при обратимом адиабатическом расширении будет больше, чем при дросселировании. Охлаждение будет происходить не только в случае реальных, но и в случае идеальных газов. Это объясняется тем, что работа производится за счет убыли внутренней энергии газа, а последняя для идеальных газов является монотонно возрастающей функцией одной только температуры. С молекулярно-кинетической точки зрения явление было подробно рассмотрено в § 67. Охлаждение связано с тем, что молекулы газа вынуждают отражаться от движущегося поршня. Давление газа может быть недостаточным, чтобы преодолеть вредные сопротивления и привести поршень в движение. Тогда его можно привести в движение мотором. Все равно

охлаждение будет происходить. Используя мотор, мы вынуждаем молекулы газа отражаться от движущейся стенки и совершать работу. Эффект Джоуля–Томсона в идеальных газах не имеет места (см. § 46), а потому с идеальными газами с его помощью нельзя получить никакого охлаждения.

2. На рисунках 109 и 110 изображены принципиальные схемы холодильных циклов, работающих по принципу адиабатического расширения и дросселирования. Цикл с применением дросселирования газа (рис. 110) был независимо друг от друга предложен в 1895 г.

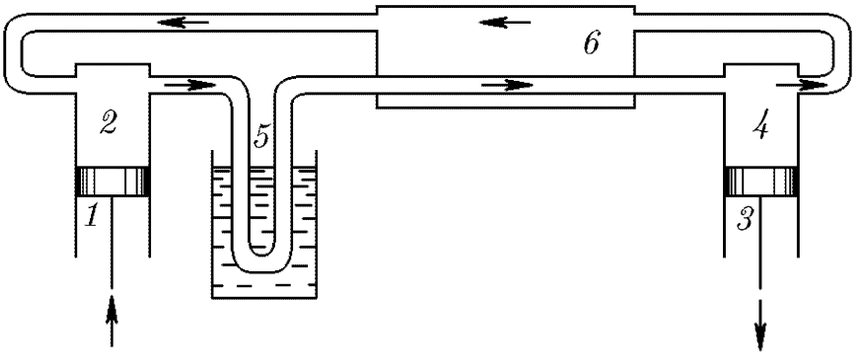


Рис. 109

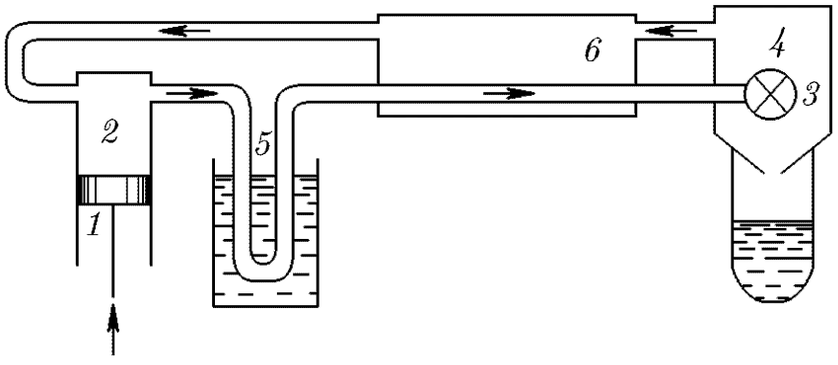


Рис. 110

Хемпсоном (1840–1900) в Англии и Линде (1842–1934) в Германии. Цикл с адиабатическим расширением газа (рис. 109) был фактически запатентован Сименсом (1816–1892) еще в 1857 г. Однако только французскому инженеру Клоду удалось в 1902 г. с помощью такого цикла получить жидкий воздух. Поэтому метод получения низких температур и сжижения газов с использованием этого цикла получил название *метода Клода.*

Мы совершенно не будем касаться конструктивных деталей реальных холодильных установок. Ограничимся только схематическим рассмотрением принципов их действия. Поршень компрессора *1* сжимает газ в цилиндре *2*. При этом газ нагревается. Сжатый и нагретый газ, покидая цилиндр *2*, охлаждается водой *5*. Далее в установках типа Клода газ поступает в цилиндр *детандера 4*, т. е. поршневой машины, в которой осуществляется адиабатическое расширение газа —

газ производит работу над поршнем *3*. В установках типа Хемпсона–Линде вместо детандера применяется *вентиль 3*, через который производится дросселирование газа. В обоих случаях газ охлаждается. Охлажденный газ, прежде чем вернуться в цилиндр *2*, проходит через вспомогательное устройство *6*, называемое *теплообменником*. Обычно теплообменник представляет собой двойную трубу: наружную и вставленную в нее внутреннюю. Трубы свернуты в винтовую спираль (змеевик). По внутренней трубе поступает газ в детандор, или в вентиль, по наружной возвращается газ, охлажденный в детандоре или после дросселирования через вентиль. Охлажденный поток газа в наружной трубе дополнительно охлаждает новую порцию газа, текущего ему навстречу по внутренней трубе к детандору или вентилю. Это охлаждение и является назначением теплообменника. Таким образом, после каждого хода поршня в детандор или к вентилю будет поступать все более холодный газ. Наконец, наступает такой момент, когда поступающий газ охлаждается настолько сильно, что после очередного адиабатического расширения или дросселирования он начинает сжижаться.

3. По сравнению с методом адиабатического охлаждения метод, основанный на эффекте Джоуля–Томсона, обладает большей простотой. В нем не возникает проблемы смазки, поскольку используемая аппаратура не содержит никаких подвижных частей, работающих при низких температурах. Однако за эту простоту приходится расплачиваться дорогой ценой — огромной потерей эффективности охлаждения и необходимостью работать при высоких давлениях с использованием больших количеств газа. Охлаждение, которое можно получить адиабатическим расширением, обычно много больше того, что дает эффект Джоуля–Томсона. Но при этом встречаются существенные трудности, связанные со смазкой подвижных узлов: при низких температурах масло замерзает. Клод применял прокладки из сухой обезжиренной кожи. Роль смазки играл сам воздух, просачивающийся в небольшом количестве между уплотнением поршня и стенками цилиндра. П. Л. Капица (1894–1984) в 1934 г. предложил применять в поршневых детандорах поршень без уплотнительных колец из кожи. В построенном им детандоре между боковой поверхностью цилиндра и поршнем имелся зазор $\sim$ 0,035 мм (при диаметре поршня 30 мм), заполненный газом, который играл роль смазки. Чтобы избежать «заклинивания поршня», на нем были сделаны кольцевые канавки глубиной и шириной по 0,25 мм, расположенные на расстоянии $\sim$ 5 мм друг от друга, обеспечивающие выравнивание давления газа на боковую поверхность поршня. Свой детандор Капица успешно применил для сжижения гелия. Предварительное охлаждение газообразного гелия производилось жидким азотом. Другое решение, снимающее проблему смазки, состоит в замене поршневого детандора турбиной. Это было предложено еще Рэлеем (1842–1919) в 1898 г. Однако первые работающие *турбодетандеры*, по-видимому, были изготовлены только в начале 30-х годов в Германии. Впервые анализ работы турбодетандора был

произведен П. Л. Капицей. Он выявил преимущества турбодетандоров перед ранее применявшимися поршневыми детандорами. Современные мощные ожижители работают по принципу турбодетандора.

4. Циклы, описанные выше, предназначены для *непрерывного сжижения газов*. Однако значительно проще производить сжижение *периодически* методом *однократного адиабатического расширения сжатого газа*. Такой метод впервые был применен Кальете (1832–1913) в 1877 г. для сжижения воздуха, а затем Ольшевским (1846–1915) в 1887 г. для сжижения водорода. Сжатый газ подвергался сначала предельно возможному предварительному охлаждению, а затем адиабатически расширился. Попытки названных ученых ожижить указанные газы лишь частично увенчались успехом. Кальете не удалось получить жидкий воздух, а Ольшевскому — жидкий водород. Они наблюдали только кратковременное появление тумана, состоящего из мельчайших капелек этих жидкостей. Успешное применение метода однократного адиабатического расширения относится к 1932 г., когда Симону удалось получить таким путем жидкий гелий. С тех пор этот метод широко применяется для получения небольших количеств жидкого гелия. Он получил название *экспансионного метода*. В 1935 г. Симон, Кук и Пирсон получили тем же методом жидкий водород. Экспансионный метод пригоден только в том случае, когда теплоемкость сосуда меньше теплоемкости находящегося в нем газа. Это условие выполняется лишь при очень низких температурах, когда теплоемкость твердых тел сильно падает. Поэтому экспансионный метод применяется практически только для сжижения водорода и гелия. Этим же объясняется неудача опытов Кальете по сжижению кислорода.

5. Имея в распоряжении сжиженный газ, можно добиться дальнейшего понижения температуры, заставляя жидкость кипеть под пониженным давлением. Этим воспользовались польские ученые Вроблевский (1845–1888) и Ольшевский (1820–1883), впервые получившие жидкий кислород. Они использовали тот же прием для сжижения газов, которым пользовался Фарадей. Была взята толстенная изогнутая стеклянная трубка с запаянным концом. Запаянный конец трубки погружался в жидкий этилен, кипящий под пониженным давлением при температуре $-130\,^\circ$C (т. е. ниже критической температуры кислорода $-118{,}8\,^\circ$C). Другой конец трубки соединялся с баллоном, в котором содержался сильно сжатый кислород. При открывании соединительного крана на стенках трубки появлялись капельки жидкого кислорода, которые, скатываясь, собирались на дне трубки.

6. Жидкий водород был впервые получен Дьюаром (1842–1923) в 1898 г. в Лондонском Королевском институте. Им был использован эффект Джоуля–Томсона. Сжатый до высокого давления водород предварительно охлаждался ниже температуры инверсии в змеевике, погруженном в жидкий воздух, кипящий под пониженным давлением, а затем подвергался дросселированию. Понадобилось еще десять лет, прежде чем Камерлинг-Оннесу (1853–1926) и его сотрудникам в Лейденском университете удалось перевести в жидкое состояние гелий.

Спустя еще 15 лет жидкий гелий стали производить и в других местах. В настоящее время жидкий гелий получают в больших количествах не только в научных лабораториях, но и на заводах.

Температуры кипения обычного гелия (^{4}He) при нормальном давлении равна 4,2 К. Заставляя ^{4}He кипеть под пониженным давлением, можно достигнуть температуры 0,7–1 К. Для получения температур ниже 1 К употребляют ванны с жидким ^{3}He, который имеет более низкую температуру кипения (3,2 К). Откачивая ванну с жидким ^{3}He, удается понизить температуру до 0,3 К. Для получения еще более низких температур применяется метод *адиабатического размагничивания* парамагнитных солей и эффект охлаждения при растворении ^{3}He в сверхтекучем ^{4}He. О методе адиабатического размагничивания мы будем говорить в третьем томе нашего курса. Наинизшая температура, достигнутая таким методом, составляет 0,0014 К.

Глава IX

ПОВЕРХНОСТНОЕ НАТЯЖЕНИЕ

§ 106. Поверхностное натяжение и некоторые явления, с ним связанные

1. На молекулу жидкости действуют силы притяжения со стороны окружающих молекул. Если молекула находится внутри жидкости и удалена от ее поверхности на расстояние, превышающее радиус сферы молекулярного действия, то, как было выяснено в § 98 (см. рис. 96), эти силы в среднем уравновешиваются. Если же молекула находится в приграничном слое, толщина которого равна радиусу сферы молекулярного действия, то появляется результирующая сила, направленная внутрь жидкости. Поэтому для извлечения молекулы из внутренних частей жидкости на ее поверхность требуется затрата работы. *Работа, которую надо затратить, чтобы изотермически и квазистатически увеличить поверхность жидкости на единицу при сохранении ее объема неизменным, называется поверхностным натяжением жидкости.*

Изотермическая работа, как известно, равна убыли свободной энергии системы (см. § 45). Свободную энергию жидкости, на которую не действуют внешние силы, можно представить в виде

$$\Psi = \Psi_{\text{об}} + \Psi_{\text{пов}}, \tag{106.1}$$

где $\Psi_{\text{об}}$ — объемная составляющая свободной энергии, а $\Psi_{\text{пов}}$ — поверхностная. Первая составляющая пропорциональна объему, вторая — поверхности жидкости (при условии, что плотность жидкости и ее температура поддерживаются постоянными). Разделить полную свободную энергию Ψ на указанные две части принципиально можно следующим образом. При увеличении объема жидкости (когда форма ее остается неизменной) отношение поверхности к объему убывает и в пределе обращается в нуль. Если же объем жидкости V достаточно велик, то поверхностной частью свободной энергии можно полностью пренебречь и вычислить объемную плотность свободной энергии по формуле $\psi_{\text{об}} = \Psi_{\text{об}}/V = \Psi/V$. Величина $\psi_{\text{об}}$ не зависит от объема жидкости, а только от ее плоскости и температуры. Если же объем жидкости не велик, то величина $\Psi_{\text{об}}V$ уже не будет равна всей свободной энергии жидкости Ψ. Оставшаяся часть $\Psi - \psi_{\text{об}}V$ и даст величину поверхностной свободной энергии.

На основании определения, введенного выше, можно написать

$$\Psi_{\text{пов}} = \sigma F,$$

где F — площадь поверхности жидкости, а σ — поверхностное натяжение. Таким образом, *поверхностное натяжение можно также определить как свободную поверхностную энергию жидкости, приходящуюся на единицу ее поверхности.*

В одном предельном случае можно полностью пренебречь объемной частью свободной энергии. Это случай, когда жидкость существует в форме *тонких пленок*. Примером могут служить мыльные пленки. В этих случаях говорят, что жидкость находится в *пластинчатом состоянии*. В таких состояниях явления, связанные с поверхностным натяжением, выступают в наиболее простом и чистом виде, поскольку на них не накладываются эффекты, обусловленные объемными свойствами тел. Мы будем проводить многие рассуждения применительно к жидкостям в пластинчатом состоянии, так как это облегчает рассуждения. Полученные результаты будут справедливы не только для жидких пленок, но и для жидкостей вообще. Надо только брать пленки, толщина которых не меньше диаметра сферы молекулярного действия.

2. Почему же σ называют поверхностным натяжением? Для ответа на этот вопрос рассмотрим следующий опыт. Возьмем проволочный каркас, имеющий форму прямоугольника (рис. 111). Сторона CD каркаса может свободно скользить вдоль направляющих проволок AC и BD. Затянем площадь $ABDC$ мыльной пленкой. Пленка эта двойная. Она подобно листу бумаги имеет две стороны — переднюю и заднюю — и состоит из двух одинаковых простых пленок, между которыми находится жидкость. Опыт показывает, что пленка стремится сократиться, и перемычка CD приходит в движение вверх. Для удержания в равновесии перемычки CD к ней надо приложить определенную силу, например подвесить грузик. Так как пленка состоит их двух простых пленок, то модуль этой силы удобно обозначить через $2f$, считая, что на каждую простую пленку действует сила f. Найдем f. Поддерживая температуру постоянной, увеличим силу f на бесконечно малую величину. Тогда перемычка CD начнет бесконечно медленно перемещаться вниз. При смещении перемычки на расстояние Δx над пленкой будет совершена работа $\Delta A = 2f\Delta x$. При этом площадь поверхности пленки увеличится на $2\Delta F = 2a\Delta x$, где a — длина перемычки CD, а ΔF — увеличение поверхности каждой простой пленки, из которых состоит двойная пленка. По определению поверхностного натяжения работа ΔA может быть представлена в виде $\Delta A = 2\sigma\Delta F = 2\sigma a\Delta x$. Приравнивая оба выражения, получим

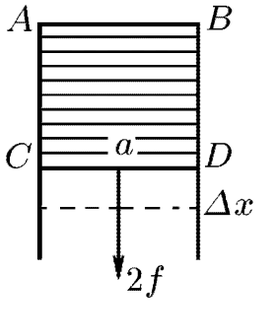


Рис. 111

$$\sigma = f/a. \tag{106.2}$$

Отсюда видно, что пленка находится в *состоянии натяжения*. В таком же состоянии натяжения находится вообще поверхность всякой жидкости. Это означает следующее. Разрежем мысленно пленку вдоль

отрезка прямой линии. Тогда каждая из половин разрезанной пленки будет действовать на линию разреза с определенной силой, имеющей характер натяжения, касательной к пленке и перпендикулярной к линии разреза. Сила, отнесенная к единице длины линии разреза, и есть поверхностное натяжение σ.

Таким образом, жидкость ведет себя так, как если бы она была помещена в эластичную (например, резиновую) пленку. Однако такая аналогия является внешней. При растяжении резиновой пленки увеличиваются расстояния между частицами в направлении растяжения, для чего требуется затрата работы против сил сцепления, действующих между этими частицами. Напротив, при увеличении поверхности жидкости работа идет на перемещение молекул из внутренних частей жидкости в поверхностный слой. Кроме того, для растяжения резиновой пленки требуется тем большая сила, чем больше пленка растянута, тогда как сила, требующая увеличение поверхности жидкости, от степени растяжения поверхностной пленки не зависит. При растяжении поверхностной пленки увеличивается только ее поверхность, в остальном пленка остается такой же, поскольку она пополняется молекулами, поступающими из внутренних слове жидкости.

3. Определение поверхностного натяжения надо еще дополнить указанием среды, с которой граничит жидкость. Дело в том, что на молекулы поверхностного слоя действуют молекулярные силы не только со стороны молекул рассматриваемой жидкости, но и со стороны молекул окружающей среды. Работа, которую надо затратить для извлечения молекулы из внутренних слоев жидкости на ее поверхность, а следовательно, и поверхностное натяжение зависят не только от самой жидкости, но и *от той среды, с которой она граничит*. В таблицах обычно приводят значение поверхностного натяжения жидкости на границе с воздухом или с ее насыщенным паром. Так, поверхностное натяжение воды при 0 °C на границе с воздухом равно 75,7 дин/см, а на границе с насыщенным паром 73,2 дин/см. Можно говорить также о поверхностном натяжении на границе раздела двух не смешивающихся жидкостей или на границе раздела жидкости и твердого тела.

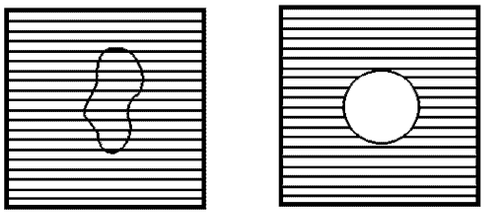


Рис. 112

4. Опишем некоторые опыты и явления, объясняющиеся поверхностным натяжением.

Возьмем проволочный каркас с натянутой на него мыльной пленкой, на которую положим петлю из нитки (рис. 112). Пока мыльная пленка внутри петли цела, петля может принимать любую форму. Но если пленку внутри петли проколоть, так что жидкая пленка останется только между нитью и наружной проволокой, то нить под действием сил поверхностного натяжения принимает форму окружности, как это показано на рис. 112 справа.

Соединим вершины A и C четырехугольного проволочного каркаса $ABCD$ гибкой ненатянутой нитью (рис. 113). Площадь каркаса затянем мыльной пленкой. Находящаяся на пленке нить AC примет форму неправильной кривой линии, не оказывающей никакого сопротивления при ее

деформации. Если, однако, уничтожить проколом часть пленки по одну сторону нити, то под действием поверхностного натяжения оставшейся части пленки нить примет форму дуги окружности.

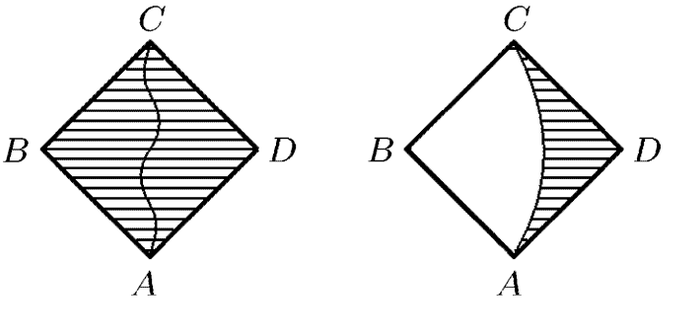


Рис. 113

Подвесим палочку AB (рис. 114) на двух параллельных гибких нитях AC и BD. Нити натянуты и прямолинейны. Если площадь $ABDC$ затянуть мыльной пленкой, то палочка поднимется, а нити искривятся, как показано на рис. 114 справа.

Когда волосяная щетка погружена в воду, то волосы щетки оказываются разделенными. Если же щетку вынуть из воды, то волосы слипаются вместе. В промежутках между волосами остается немного воды, свободная поверхность которой сокращается под действием поверхностного натяжения и стягивает волосы.

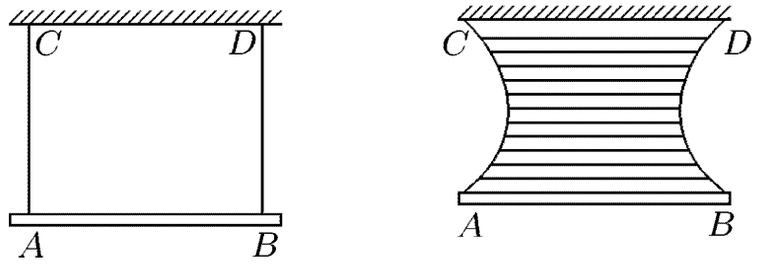


Рис. 114

Насыплем на свободную поверхность ртути, налитой в широкий сосуд, какой-либо порошок. Затем погрузим в нее вертикальную толстую стеклянную палочку. Порошок увлекается в углубление вокруг палочки, как будто бы ртуть была покрыта пленкой, не разрывающейся при погружении в нее палочки. Опыт удается и с водой, если палочку предварительно слегка смазать жиром.

Многие явления объясняются тем, что поверхностное натяжение у разных жидкостей не одинаково. Возьмем, например, маленькую модель кораблика из жести, в корме которой сделано небольшое отверстие. Нальем в кораблик немного эфира, чтобы уровень свободной поверхности эфира был несколько выше отверстия. Затем пустим кораблик плавать на поверхности воды. Тогда он начнет двигаться вперед, т. е. в направлении от кормы к носу. Явление объясняется тем, что поверхностное натяжение эфира меньше поверхностного натяжения воды. Эфир, просачивающийся через отверстие в корме, покрывает поверхность воды сзади кораблика тонким слоем. Поверхностное натяжение эфира тянет кораблик назад. Однако эта сила меньше силы поверхностного натяжение воды, которая тянет его вперед. Результирующая этих двух сил направлена вперед и приводит кораблик в движение.

То же явление наблюдается, когда меленькие кусочки камфоры бросают на поверхность чистой воды. Камфора медленно растворяется в воде и уменьшает ее поверхностное натяжение. Скорость растворения в различных местах кусочка камфоры не одинакова и зависит от его формы. Она больше вблизи выступающих острых концов кусочка. В результате образуется разность поверхностных натяжений, приводящая кусочек камфоры в беспорядочное и интенсивное вращательное и поступательное движение. Явление наблюдается только тогда, когда поверхность воды достаточно чистая. Тонкий слой жира уменьшает поверхностное натяжение воды настолько, что растворение камфоры не может уже произвести его дальнейшего уменьшения. Согласно Рэлею, движение камфоры прекращается, когда

толщина слоя масла на поверхности воды достигает $2 \cdot 10^{-7}$ см. В этом случае поверхностное натяжение воды на 28 % меньше, чем для чистой поверхности воды. Слой масла толщиной $1 \cdot 10^{-7}$ см и ниже, согласно Рэлею, практически не оказывает никакого влияния на поверхностное натяжение воды.

Плато (1801–1883) обнаружил, что колебания магнитной стрелки затухают значительно сильнее, когда она подвешена так, что лежит на поверхности воды, чем в том случае, когда она целиком погружена в воду. Это можно объяснить тем, что поверхность воды всегда немного загрязнена жиром. Когда стрелка движется по поверхности воды, жир сосредоточивается впереди нее, а сзади поверхность воды становится более чистой. В результате возникает разность сил поверхностного натяжения, тормозящая движение стрелки.

Волнение на море можно в значительной степени ослабить, выливая на поверхность воды масло. Ветер, действуя на поверхность воды, покрытую маслом, гонит масло вперед и оставляет за собой более чистую поверхность воды. Появляется обратно направленная сила поверхностного натяжения, стремящаяся затормозить поверхность воды, приведенную в движение. Даже когда волны уже образовались, слой масла создает дополнительную силу, стремящуюся затормозить волнение. С волнением связано увеличение толщины масляной пленки в одних местах и уменьшение в соседних. Это создает разность поверхностных натяжений, стремящуюся уменьшить волнение.

Насыплем на поверхность воды в сосуде споры ликоподия. Возьмем стакан, нальем в него несколько капель эфира, которыми смочим дно и стенки стакана, а остаток эфира выльем. В стакане останутся тяжелые пары эфира. Выльем их на поверхность воды, покрытой порошком ликоподия. Порошок быстро расходится во все стороны от того места, на которое попали пары эфира. Явление весьма эффектно, так как пары эфира не видны, и кажется, что опыт производится с пустым стаканом. Явление объясняется тем, что пары эфира уменьшают поверхностное натяжение воды. Возникшая разность поверхностных натяжений и заставляет порошок разбегаться к краям сосуда. Капля воды, висящая у нижнего конца вертикальной трубки, отрывается и падает, если вблизи нее поместить эфир.

5. *В состоянии равновесия свободная энергия системы должна быть минимальной.* Допустим, что жидкость совершенно свободна, т. е. не ограничена стенками сосуда и не подвержена действию внешних силовых полей. Тогда под действием сил поверхностного натяжения она должна принять *форму шара*, так как из всех тел заданного объема шар имеет наименьшую поверхность, а потому и наименьшую поверхностную свободную энергию. Получению шаровой формы капли жидкости мешает сила тяжести. Потенциальная энергия силы тяжести пропорциональна объему жидкости, тогда как энергия поверхностного натяжения пропорциональна ее поверхности. Поэтому относительное влияние силы тяжести по сравнению с влиянием поверхностного натяжения тем больше, чем больше объем капли. Малые капли и в особенности капли ртути практически шарообразны. Но в условиях невесомости капля любой жидкости, независимо от ее размеров, в состоянии равновесия имеет сферическую форму. Это действительно наблюдалось во время полета космических кораблей. В земных условиях можно полностью исключить влияние силы тяжести, погрузив определенную массу жидкости в сосуд с другой жидкостью, не смешивающейся с первой. Плотности обеих жидкостей должны быть одинаковы. Такой опыт был поставлен Плато. Он взял смесь спирта и

воды и в нее погрузил большую каплю оливкового масла. Плотности обеих жидкостей были подобраны одинаковыми, и оливковое масло приняло форму шара. Опыт Плато можно повторить и со многими другими жидкостями. Удобно, например, взять раствор поваренной соли в воде и подобрать концентрацию так, чтобы плотность раствора сделалась равной плотности анилина. Анилиновая капля в растворе принимает форму шара. Так как трудно достигнуть точного равенства плотностей обеих жидкостей, то для демонстрации явления удобно в растворе создать концентрацию соли, медленно убывающую с высотой. Центр анилиновой капли будет находиться в равновесии на такой высоте, где плотность раствора равна плотности анилина.

6. Висящая капля напоминает жидкость, подвешенную в резиновом мешочке. Однако натяжение резиновой пленки меняется с изменением количества заключенной в ней жидкости, тогда как поверхностное натяжение от размеров капли не зависит. С этим связано явление *отрыва капель*. Последовательные стадии роста капли схематически

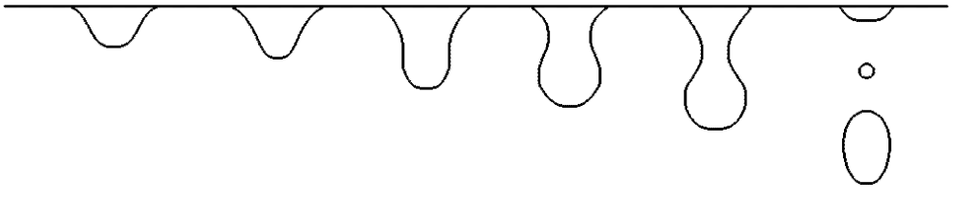


Рис. 115

показаны на рис. 115. Когда капля достигает определенного размера, на ней образуется шейка (сужение). По мере увеличения размеров капли она становится все тоньше и тоньше. В определенный момент на шейке появляется второе сужение, и наконец, капля разрывается сразу в двух местах. Вследствие этого, когда отрывается большая капля, за ней всегда следует маленькая капелька, называемая *шариком Плато*, который первым подметил это явление. Большая капля во время своего падения обнаруживает периодические изменения формы, вызываемые поверхностным натяжением. Она вначале имеет форму вытянутого эллипсоида вращения, затем эллипсоид переходит в шар, в сплюснутый эллипсоид и т. д.

Демонстрация явления образования и отрыва капель затруднительна тем, что явление происходит очень быстро. Дарлинг преодолел это затруднение искусственным уменьшением силы тяжести совершенно так же, как поступил Плато в своем известном опыте. При температурах выше 80 °C плотность жидкого анилина меньше, чем у воды, а при низших температурах — больше. Кроме того, анилин не смешивается с водой. Для опыта берется большой стеклянный стакан высотой 20–25 см и диаметром 10–12 см, наполненный дистиллированной водой. В воду добавляется 80–100 $см^3$ анилина. Нагретый анилин всплывает вверх и собирается в виде капель (диаметром 2-3 см) у свободной поверхности воды. Так как температура воды на поверхности ниже 80 °C, то при остывании капли анилина отрываются и медленно падают на дно. Здесь они снова нагреваются и поднимаются вверх. Процесс повторяется неограниченно долго, пока поддерживается необходимая разность температур между дном и поверхностью воды в сосуде.

7. Принцип минимума свободной поверхности энергии жидкости в применении к растворам приводит к новому интересному заключению. Состав поверхностного слоя в растворе может существенно отличаться от состава основной массы жидкости. Дело в том, что различные чистые вещества имеют различные поверхностные натяжения σ. Величина σ, а потому и свободная поверхностная энергия жидкости зависят от межмолекулярных сил, которые в свою очередь определяются строением молекул. Минимум поверхностной энергии жидкости достигается не только путем сокращения ее поверхности, но и путем укомплектования поверхностного слоя такими молекулами, чтобы σ была минимальной. Вещества, адсорбирующиеся на поверхности раздела с соответствующим понижением свободной поверхностной энергии, называются *поверхностно-активными*. Их добавление даже в небольшой концентрации существенно уменьшает поверхностное натяжение. Молекулы таких веществ прежде всего поступают в поверхностный слой и лишь после того, как он окажется заполненным ими достаточно плотно, проникают в основной объем жидкости.

К поверхностно-активным веществам обычно принадлежат молекулы органических веществ, имеющие вытянутую форму. Они состоят из полярной группы (например, $-\mathrm{OH}$, $-\mathrm{COOH}$, $-\mathrm{NH_2}$), к которой присоединяется неполярная часть молекулы (например, углеводородная цепь или кольцо). Полярная группа обладает большим дипольным моментом, который является центром силового поля молекулы. Дипольный момент неполярной части молекулы практически равен нулю. В поверхностном слое жидкости молекулы поверхностно-активных веществ выстраиваются параллельно друг другу полярными группами, обращенными внутрь жидкости, а неполярными частями — наружу.

Наиболее известным примером поверхностно-активного вещества является мыло. С резким уменьшением поверхностного натяжения мыльного раствора (а также с увеличением его вязкости) связана возможность выдувать большие мыльные пузыри, чего никогда не удается сделать с чистой водой. В последнем случае молекулы на границах пленки с силой втягиваются внутрь пленки, благодаря чему она становится слишком тонкой и лопается.

Существуют и такие вещества (соль, сахар), добавление которых к воде повышает ее поверхностное натяжение. Благодаря этому подсаливание мыльного раствора выталкивает в поверхностный слой жидкости еще больше молекул мыла, чем в пресной воде. Это используется в технике мыловарения для выделения мыла из раствора путем его «высаливания».

§ 107. Термодинамика поверхностного натяжения

Для увеличения поверхности пленки на dF над ней надо произвести работу σdF. Сама пленка при этом совершает работу $\delta A = -\sigma dF$. По первому началу термодинамики $\delta Q = dU + \delta A$ или

$$\delta Q = dU - \sigma dF. \tag{107.1}$$

Введя энтропию S, перепишем это равенство в виде

$$dU = T\,dS + \sigma dF. \tag{107.2}$$

Свободная энергия

$$\Psi = U - TS, \tag{107.3}$$

а потому

$$d\Psi = -S\,dT + \sigma\,dF. \tag{107.4}$$

Отсюда $S = -(\partial\Psi/\partial T)_F$, и следовательно,

$$\Psi = U + T\left(\frac{\partial\Psi}{\partial T}\right)_F. \tag{107.5}$$

Подставим сюда $\Psi = \sigma F$. Поверхностное натяжение зависит от температуры пленки, но не зависит от ее площади. Поэтому после подстановки получим

$$U = \left(\sigma - T\,\frac{d\sigma}{dT}\right)F. \tag{107.6}$$

Если пленка расширяется изотермически, то ей надо сообщить теплоту

$$Q = \Delta U - \sigma F = -T\,\frac{d\sigma}{dT}\,dF.$$

Теплота, сообщенная единице поверхности пленки при изотермическом расширении,

$$q = -T\,\frac{d\sigma}{dT}. \tag{107.7}$$

Она положительна, потому что, как показывает опыт, σ уменьшается с повышением температуры. Величина q называется *теплотой образования единицы поверхности пленки*. (Другой вывод формулы (107.7) см. в задаче 1 к этому параграфу.)

ЗАДАЧИ

1. Получить соотношение (107.7) методом циклов.

Р е ш е н и е. Рассмотрим пленку жидкости и проведем с ней бесконечно малый цикл Карно. Будем откладывать по горизонтальной оси площадь пленки F, а по вертикальной оси — поверхностное натяжение σ (рис. 116). При постоянной температуре поверхностное натяжение также постоянно. Поэтому на нашей диаграмме изотермы изобразятся горизонтальными прямыми. Начальное состояние пленки характеризуется точкой *1*. Приведем пленку в тепловой контакт с нагревателем, температура которого равна температуре пленки в состоянии *1*. Затем квазистатически растянем пленку до состояния *2*. На это надо затратить работу. Работа самой пленки отрицательна и равна $A_1 = -\sigma(T_1)\Delta F$, где ΔF — приращение площади пленки при растяжении по изотерме *12*. При изотермическом растяжении к пленке надо подводить теплоту. Величина подведенной теплоты $Q_1 = q\Delta F$. В состоянии *2* изолируем пленку от нагревателя и

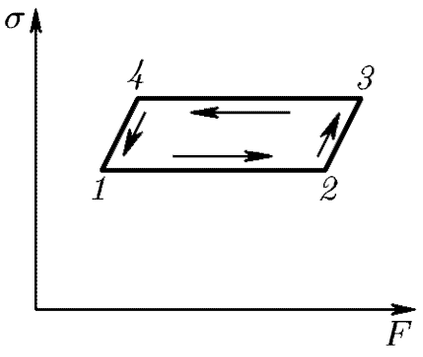


Рис. 116

адиабатически бесконечно мало растянем ее до состояния *3*, в котором пленка примет температуру холодильника T_2. Предполагается, что температуры T_1 и T_2 отличаются друг от друга бесконечно мало. В состоянии *3* приведем пленку в тепловой контакт с холодильником и изотермически переведем ее в состояние *4*. Поверхность пленки уменьшится на ΔF, и она совершит положительную работу $A_2 = \sigma(T_2)\Delta F$. Из состояния *4* вернем пленку в исходное состояние *1*. Работой пленки на адиабатах *23* и *41* можно пренебречь, как величиной более высокого порядка малости. Полная работа, совершенная пленкой во время кругового процесса,

$$A = A_1 + A_2 = [\sigma(T_2) - \sigma(T_1)]\,\Delta F = \frac{d\sigma}{dT}\,(T_2 - T_1)\Delta F.$$

по теореме Карно

$$\frac{A}{Q_1} = \frac{T_1 - T_2}{T_1}.$$

Подставляя сюда найденные выше выражения для A и Q_1, после сокращения получим формулу (107.7).

2. Определить изменение температуры пленки при адиабатическом расширении.

Ответ.

$$dT = -\frac{q}{c_F}\,dF, \tag{107.8}$$

где c_F — теплоемкость единицы поверхности пленки при постоянном значении F, а q определяется формулой (107.7). При адиабатическом расширении пленка охлаждается.

3. Мыльная пленка имеет толщину $h = 10^{-3}$ мм и температуру $T = 300$ К. Вычислить уменьшение температуры этой пленки, если ее растянуть адиабатически настолько, чтобы площадь пленки удвоилась. Поверхностное натяжение мыльного раствора убывает на 0,15 дин/см при повышении температуры на 1 К.

Ответ. Считая $c_F = c_v h$ (c_v — удельная теплоемкость воды), получим

$$\Delta T = \frac{2T}{c_v h}\,\frac{d\sigma}{dT} \approx -0{,}02 \text{ К}.$$

(Коэффициент 2 учитывает то обстоятельство, что пленка — двухсторонняя.)

4. Показать, что вблизи абсолютного нуля поверхностное натяжение жидкости перестает зависеть от температуры, т. е.

$$\lim_{T\to 0}\frac{d\sigma}{dT} = 0. \tag{107.9}$$

(Конкретно речь может идти только о гелии — единственном веществе, остающемся жидким при абсолютном нуле температуры.)

Решение. Подставим в формулу (107.7) $q = T\Delta S$, где ΔS — приращение энтропии пленки при увеличении ее поверхности на единицу. Получим

$$\frac{d\sigma}{dT} = -\Delta S.$$

Согласно теореме Нернста при абсолютном нуле температуры все процессы идут без изменения энтропии, т. е. $\Delta S = 0$. Отсюда и следует (107.9).

§ 108. Краевые углы. Смачивание и несмачивание

1. Допустим, что три жидких среды *1*,*2*,*3* (одна из них может быть газообразной) попарно граничат между собой вдоль тех поверхностей, пересекающихся вдоль некоторой линии O (на рис. 117 изображено

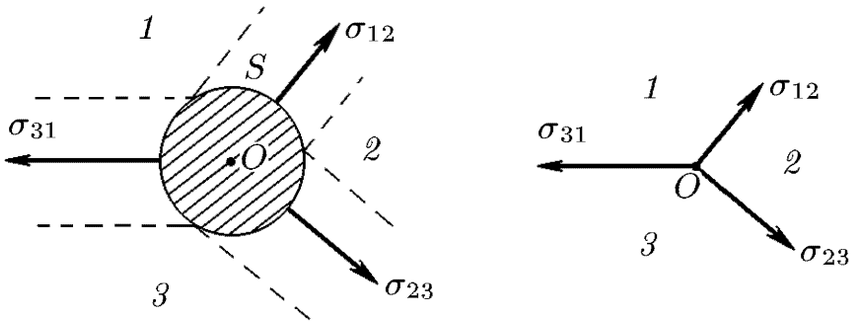


Рис. 117

сечение рассматриваемой системы плоскостью рисунка, перпендикулярной к линии O; линия O не изображена, а указана только точка пересечения ее с плоскостью рисунка). Возможно ли и при каких условиях механическое равновесие между этими средами? При ответе на этот вопрос надо иметь в виду, что границы раздела между жидкостями являются не геометрическими поверхностями, а представляют собой переходные слои, в которых и действуют силы поверхностного натяжения. Толщина этих слоев порядка радиуса действия молекулярных сил. Возьмем на линии O отрезок единичной длины и окружим его цилиндрической поверхностью S, как указано на рис. 117 *а*. Границы поверхностных слоев между жидкостями обозначены штриховыми линиями. Для равновесия необходимо, чтобы силы, действующие на жидкость внутри цилиндра S, уравновешивались. Эти силы состоят из сил поверхностного натяжения σ_{12}, σ_{23}, σ_{31}, действующих вдоль границ раздела между жидкостями, сил гидростатического давления на поверхность S и силы веса жидкости, заключенной внутри объема, ограниченного этой поверхностью. По закону Архимеда результирующая сил гидростатического давления того же порядка, что и вес жидкости в цилиндре S. Обе эти силы пропорциональны объему цилиндра. По порядку величины они равны $f \sim \rho\pi r^2 g$, где ρ — плотность жидкости, а r — радиус действия молекулярных сил. Полагая $\rho \sim 1$ г/см^3, $r \sim 10^{-6}$ см, получим $f \sim 10^{-8}$ дин/см, тогда как поверхностное натяжение σ составляет десятки дин на сантиметр. Ясно поэтому, что силой веса и гидростатического давления можно полностью пренебречь и записать условие равновесия в виде

$$\sigma_{12} + \sigma_{23} + \sigma_{31} = 0. \tag{108.1}$$

Таким образом, все происходит так, если бы речь шла о равновесии трех сил σ_{12}, σ_{23}, σ_{31}, приложенных в одной точке O (рис. 117 *б*).

Геометрически условие (108.1) означает, что из отрезков с длинами $\sigma_{12}, \sigma_{23}, \sigma_{31}$ можно составить замкнутый треугольник. Длины сторон треугольника однозначно определяют и сам треугольник. Поэтому углы, под которыми сходятся поверхности раздела на линии O при равновесии, однозначно определяются поверхностными натяжениями $\sigma_{12}, \sigma_{23}, \sigma_{31}$. Если одно из этих поверхностных натяжений больше суммы двух остальных, то треугольник построить нельзя, и равновесие невозможно.

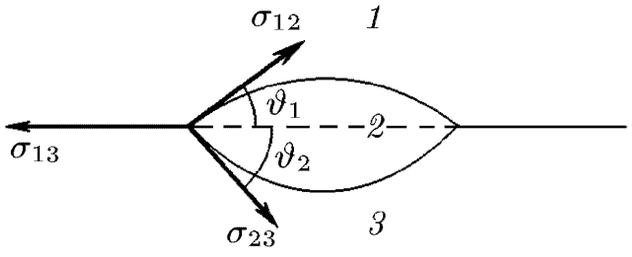


Рис. 118

2. Примером, когда три среды граничат между собой, может служить капля жидкости на поверхности другой жидкости. Капля имеет форму чечевицы (рис. 118). В этом случае векторное условие равновесия (108.1) распадается на два скалярных уравнения:

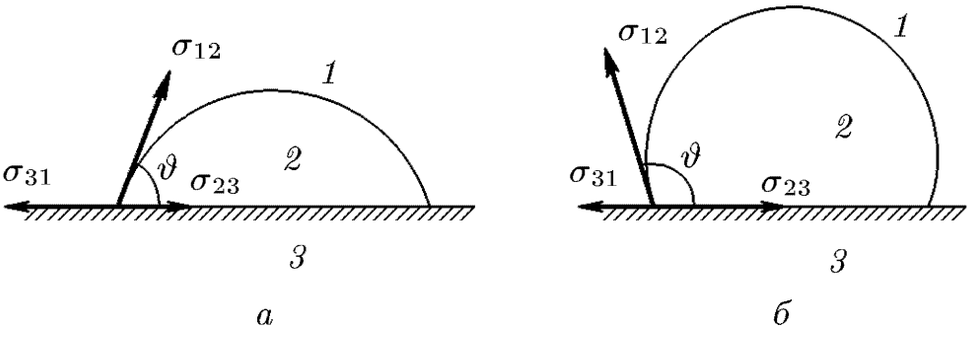
$$\sigma_{13} = \sigma_{12}\cos\vartheta_1 + \sigma_{23}\cos\vartheta_2, \quad \sigma_{12}\sin\vartheta_1 = \sigma_{23}\sin\vartheta_2. \tag{108.2}$$

Из них получаем

$$\cos\vartheta_1 = \frac{\sigma_{13}^2 + \sigma_{12}^2 - \sigma_{23}^2}{2\sigma_{13}\sigma_{12}}, \quad \cos\vartheta_2 = \frac{\sigma_{13}^2 + \sigma_{23}^2 - \sigma_{12}^2}{2\sigma_{13}\sigma_{23}}. \tag{108.3}$$

Этими формулами однозначно определяются углы ϑ_1 и ϑ_3. Равновесие возможно только в том случае, когда $\sigma_{13} < \sigma_{12} + \sigma_{23}$, как это видно из первого уравнения (108.2). В этом случае капля действительно имеет форму чечевицы. Так ведет себя, например, капля жира на поверхности воды. Если $\sigma_{13} > \sigma_{12} + \sigma_{23}$, то не существует углов ϑ_1 и ϑ_2, удовлетворяющих условиям (108.2). Равновесие капли невозможно, и она растекается по поверхности жидкости *3*, покрывая ее тонкой пленкой. Примером может служить пленка бензина или керосина на поверхности воды. Такие пленки обычно имеют радужную окраску, что объясняется интерференцией света. В рассматриваемом случае говорят, что *жидкость 3 полностью смачивается жидкостью 2* (или наоборот).

3. Аналогично ведет себя капля жидкости на поверхности твердого тела (рис. 119). Разница только в том, что поверхность твердого тела не



а *б*

Рис. 119

может деформироваться. Равнодействующая сил поверхностного натяжения $\sigma_{12}+\sigma_{23}+\sigma_{31}$ уравновешивается силой нормального давления или натяжения на границе жидкости с твердым телом. Поэтому вместо условия (108.1) надо требовать лишь обращения в нуль касательной составляющей результирующей силы $\sigma_{12}+\sigma_{23}+\sigma_{31}$. В рассматриваемом случае угол ϑ_2 равен нулю, и для определения одного угла $\vartheta_1\equiv\vartheta$ достаточно первого уравнения (108.2). Оно дает

$$\cos\vartheta=\frac{\sigma_{13}-\sigma_{23}}{\sigma_{12}}. \tag{108.4}$$

Угол ϑ называется *краевым углом*. Его обычно выбирают так, чтобы он включал в себя область, занятую жидкостью *2*. Когда $\dfrac{\sigma_{13}-\sigma_{23}}{\sigma_{12}}>1$, т. е. $\sigma_{13}>\sigma_{12}+\sigma_{23}$, условие (108.4) не может быть удовлетворено. Капля жидкости *2* не находится в равновесии, а растекается по поверхности твердого тела, покрывая его тонкой пленкой (например, керосин на поверхности жести или стекла). В этом случае говорят, что *жидкость полностью смачивает поверхность твердого тела*. В другом случае, когда $\dfrac{\sigma_{13}-\sigma_{23}}{\sigma_{12}}<-1$, т. е. $\sigma_{23}>\sigma_{13}+\sigma_{12}$, также не существует никакого угла ϑ, который бы удовлетворял условию (108.4). Жидкость стягивается в шаровую каплю, несколько сплюснутую силой тяжести. (Например, капля ртути на поверхности стекла или капля воды на поверхности парафина.) В этом случае говорят, что *жидкость полностью не смачивает поверхность твердого тела*. В большинстве случаев имеет место *частичное смачивание* (когда $0<\vartheta<\pi/2$, рис. 119 *а*) или частичное несмачивание (когда $\pi/2<\vartheta<\pi$, рис. 119 *б*).

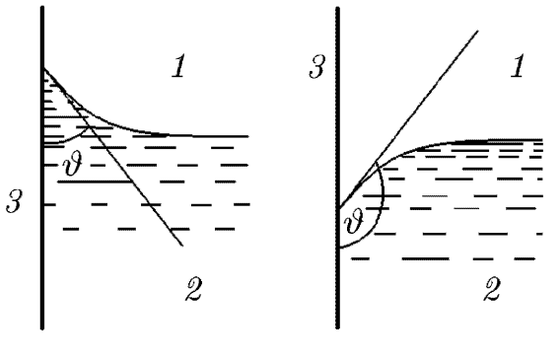


Рис. 120

Явление краевого угла наблюдается у стенок сосудов, когда в них налита жидкость (рис. 120). Значение краевого угла здесь также определяется формулой (108.4).

4. Несмачиванием твердых тел жидкостями объясняются многие явления. Приведем некоторые из них. На поверхность воды положить лист алюминия. Он не утонет, даже если на него положить небольшой груз. Стальная иголка (в особенности если она покрыта тонким слоем парафина) не тонет, если ее осторожно положить на поверхность воды. Возьмем сито с металлической сеткой. Погрузим сетку в расплавленный парафин, а затем встряхнем, чтобы парафин не заполнял отверстия в сетке. После этого положим на дно сита лист бумаги и нальем в него воды. Осторожно вытянем лист бумаги. Вода не будет выливаться через отверстия сетки.

Возьмем ареометр или похожий на него прибор, плавающий в вертикальном положении на поверхности воды, выступая немного наружу. На верхний конец трубки ареометра наденем и закрепим плоский кружок из проволочной сетки и погрузим весь прибор в воду. Ареометр начнет всплывать. Однако когда сетка дойдет до поверхности воды, то прибор остановится,

встречая у самой поверхности сопротивление пропускающей его пленки. Если слегка наклонить прибор, чтобы край сетки отделился от воды, то он всплывет, и сетка окажется значительно выше поверхности воды. Можно заставить ареометр всплывать и не наклоняя его. Для этого достаточно капнуть на воду несколько капель эфира.

§ 109. Разность давлений по разные стороны изогнутой поверхности жидкости. Формула Лапласа

1. Если поверхность жидкости — кривая, то при равновесии давления по разные стороны ее должны быть разными. Явление обусловлено силами поверхностного натяжения. Рассмотрим сначала простейший случай, когда жидкость ограничена боковой поверхностью прямого круглого цилиндра. Поперечное сечение цилиндра представлено на рис. 121. Выберем на его поверхности бесконечно малый участок AB, стягиваемый центральным углом φ. На его боковые стороны действуют касательные силы $b\sigma$, где b — длина цилиндра. Равнодействующая этих сил направлена параллельно радиусу CO цилиндра и равна $F = 2b\sigma \sin(\varphi/2)$, или $F = d\sigma\varphi$, так как угол φ выбран бесконечно малым. Подставляя сюда $\varphi = a/R$, где a — длина дуги AB, а R — радиус цилиндра, получим

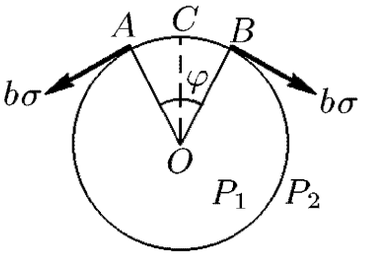


Рис. 121

$$F = \frac{\sigma}{R} S. \tag{109.1}$$

Здесь $S = ab$ — площадь бесконечно малого прямоугольного участка на боковой поверхности цилиндра. Разделив силу F на площадь S, найдем разность давлений внутри и снаружи жидкости

$$P_2 - P_1 = \sigma/R.$$

2. Обобщим теперь эту формулу на случай, когда жидкость ограничена поверхностью двойной кривизны. С этой целью возьмем на поверхности жидкости четыре точки A, B, C, D, находящиеся в вершинах бесконечно малого прямоугольника (рис. 122). Проведем через A и B, B и C и т. д. плоскости, перпендикулярные к поверхности жидкости. (На рис. 122 изображены только две плоскости, проведенные через AD и BC. Они пересекаются вдоль бесконечно малого отрезка OO'.) В результате получится бесконечно малый криволинейный прямоугольник $ABCD$. Его стороны могут рассматриваться как бесконечно малые дуги окружностей. Пусть R_1 — радиус кривизны

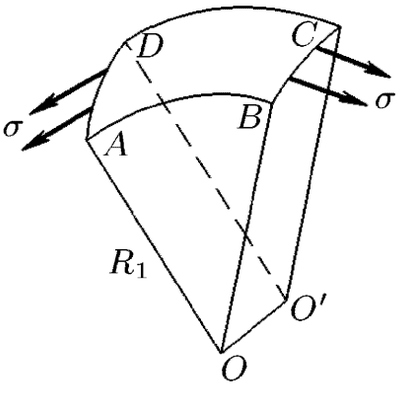


Рис. 122

дуги AB. Радиус кривизны дуги DC отличается от R_1 бесконечно мало. На противоположные стороны AD и BC действуют касательные силы поверхностного натяжения. Результирующая этих сил нормальна к поверхности жидкости. Согласно формуле (109.1) ее модуль $F_1 = (\sigma/R_1)S$, где S — площадь прямоугольника $ABCD$. Рассуждая так же, найдем, что результирующая касательных сил поверхностного натяжения, действующих на противоположные стороны AB и DC, тоже нормальна к поверхности жидкости и ее модуль $F_2 = (\sigma/R_2)S$, где R_2 — радиус кривизны дуги AD. (Радиус кривизны дуги BC отличается от него бесконечно мало.) Таким образом, модуль результирующей всех сил поверхностного натяжения, действующих на границах прямоугольника $ABCD$,

$$F = F_1 + F_2 = \sigma S\Big(\frac{1}{R_1} + \frac{1}{R_2}\Big).$$

Разделив ее на S, получим искомую разность давлений:

$$P_2 - P_1 = \sigma\Big(\frac{1}{R_1} + \frac{1}{R_2}\Big). \tag{109.2}$$

Формула (109.2) называется *формулой Лапласа*. Величины R_1 и R_2 суть радиусы кривизны двух взаимно перпендикулярных нормальных сечений поверхности жидкости. Радиус кривизны считается положительным, если соответствующее нормальное сечение вогнуто в сторону жидкости. В противном случае он считается отрицательным. Величина $1/R_1 + 1/R_2$ называется *средней кривизной поверхности*. Средняя кривизна не зависит от выбора взаимно перпендикулярных нормальных сечений поверхности жидкости. В противном случае от такого выбора зависела бы и разность давлений $P_2 - P_1$, что физически бессмысленно. В дифференциальной геометрии независимость средней кривизны любой поверхности от выбора нормальных сечений доказывается чисто математическими методами без использования физических или каких-либо других соображений. Эта независимость составляет содержание так называемой *теоремы Эйлера*.

3. Если поверхность жидкости — сферическая, то $R_1 = R_2 = R$, и формула (109.2) переходит в

$$P_2 - P_1 = \frac{2\sigma}{R}. \tag{109.3}$$

Для мыльного пузыря разность давлений воздуха внутри и вне пузыря вдвое больше по сравнению с тем, что дает формула (109.3), т. е.

$$P_2 - P_1 = \frac{4\sigma}{R}.$$

Это связано с тем, что оболочка пузыря имеет две поверхности: наружную и внутреннюю. Она действует как пленка с удвоенным поверхностным натяжением.

Таким образом, чем больше кривизна поверхности пузыря, тем больше давление газа в нем. Положение здесь противоположно тому,

с которым мы сталкиваемся, надувая футбольный мяч: с увеличением размеров мяча увеличивается и давление газа внутри него. Указанное различие связано с тем, что поверхностное натяжение пленки пузыря не зависит от его размеров, тогда как натяжение оболочки камеры футбольного мяча возрастает по мере ее надувания.

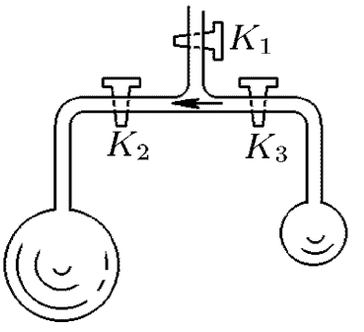


Рис. 123

Рассмотрим два мыльных пузыря, сообщающихся друг с другом. Их можно получить при помощи тройника, снабженного необходимыми кранами (рис. 123). Сначала каждый пузырь выдувается в отдельности. Затем кран K_1 закрывается, а краны K_2 и K_3 открываются. Давление в меньшем пузыре больше, и воздух из этого пузыря будет перетекать в больший пузырь. Больший пузырь будет расти за счет меньшего.

4. Применим формулу Лапласа для расчета высоты поднятия жидкости в цилиндрическом капилляре радиуса a (рис. 124). Пренебрежем изменением давления жидкости при изменении высоты на величину порядка a. В этом приближении разность давлений $P_2 - P_1$ будет одной и той же во всех точках мениска. То же относится к средней кривизне $1/R_1 + 1/R_2$, как это следует из формулы Лапласа (109.2). Кроме того, ввиду симметрии $R_1 = R_2$. Поэтому в рассматриваемом приближении мениск можно считать сферическим. Его радиус кривизны

$$R = \frac{a}{\cos\vartheta},$$

где ϑ — краевой угол. В рассматриваемом случае P_1 есть атмосферное давление, а P_2 — давление жидкости на уровне мениска. Эти давления связаны соотношением

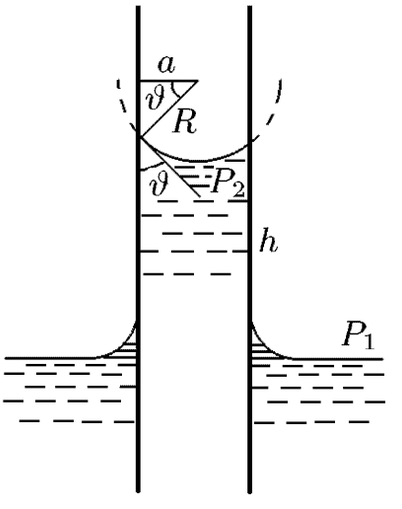


Рис. 124

$$P_1 - P_2 = \rho g h,$$

где h — высота поднятия, а ρ — плотность жидкости. Сравнивая эту формулу с формулой (109.3), получим

$$h = \frac{2\sigma}{\rho g R} = \frac{2\sigma}{\rho g a}\cos\vartheta. \qquad (109.4)$$

Высота поднятия обратно пропорциональна радиусу капилляра. Когда угол ϑ тупой, т. е. мениск выпуклый, величина h отрицательна, т. е. имеет место не поднятие, а опускание жидкости в капилляре.

ЗАДАЧИ

1. Капля воды массы $m = 0{,}1$ г введена между двумя плоскими и параллельными между собой стеклянными пластинками, смачиваемыми водой,

причем краевой угол $\vartheta = 0$. Как велика сила притяжения между пластинками F, если они находятся друг от друга на расстоянии $d = 10^{-4}$ см? Поверхностное натяжение воды (при 18°) $\sigma = 73$ дин/см.

Решение. Капля примет форму диска с вогнутой периферийной поверхностью. Кривизной сечения этой поверхности плоскостью, параллельной пластинкам, можно пренебречь. Радиус кривизны нормального к нему сечения $r = d/2$. Средняя кривизна боковой поверхности диска $1/R_1 + 1/R_2 = 2/d$. Давление жидкости между дисками меньше атмосферного на $\Delta P = 2\sigma/d$. Площадь диска $S = m/\rho d$, где ρ — плотность жидкости. Пластинки будут прижиматься друг к другу с силой

$$F = S\,\Delta P = 2m\sigma/\rho d^2 = 1{,}46 \cdot 10^9 \text{ дин} = 1{,}46 \cdot 10^4 \text{ Н}.$$

2. Грамм ртути помещен между двумя плоскими стеклянными пластинками. Какую силу F надо приложить к верхней пластинке, чтобы ртуть приняла форму круглой лепешки однородной толщины и радиуса $R = 5$ см? Поверхностное натяжение ртути (при 15 °C) $\sigma = 487$ дин/см, краевой угол между ртутью и стеклом $\Theta = 40°$.

Ответ. $F = \dfrac{2\sigma\rho\cos\Theta}{m}\pi^2 R^4 = 630$ Н (m — масса ртути).

3. С какой силой F притягиваются две вертикальные и параллельные стеклянные пластинки, частично погруженные в воду так, что расстояние между ними равно $d = 0{,}1$ мм? Ширина пластинок $l = 10$ см, $\sigma = 73$ дин/см, $\vartheta = 0$. Высота пластинок такова, что поднявшаяся вода не доходит до их верхних краев.

Ответ. $F = \dfrac{2\sigma^2 l}{\rho g d^2}\cos^2\vartheta \approx 10$ Н.

4. Бесконечно длинная прямоугольная пластинка кладется на поверхность смачивающей ее жидкости, а затем слегка приподнимается, увлекая за собой некоторое количество жидкости (рис. 125). Найти уравнение боковой поверхности жидкости, устанавливающейся под влиянием капиллярных сил и силы тяжести.

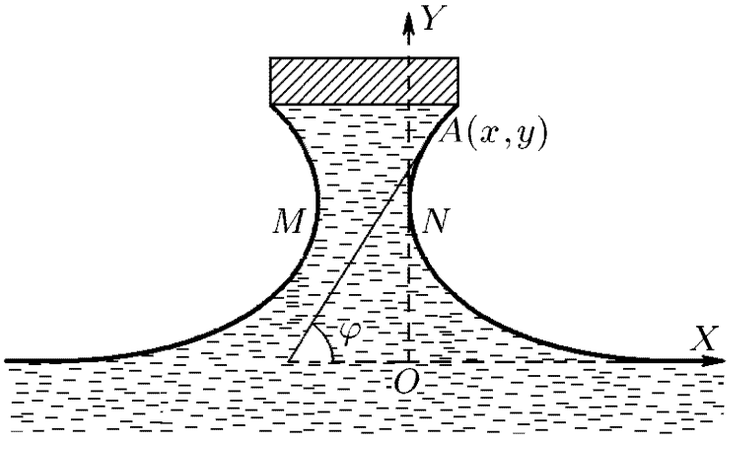


Рис. 125

Решение. Примем за ось X прямую, перпендикулярную к длинной стороне пластинки и лежащую на горизонтальной поверхности жидкости, а за ось Y — вертикальную прямую, касающуюся правой цилиндрической поверхности жидкости. Пусть x и y означают текущие координаты точки, лежащей на искомой поверхности. Давление внутри жидкости на уровне точки A (рис. 125) равно $P = P_0 - \rho g y$, где P_0 — атмосферное давление. То же давление можно выразить по формуле Лапласа $P = P_0 - \sigma K$, где K — абсолютное значение кривизны поверхности жидкости в точке A. Следовательно,

$$\rho g y = \sigma K. \tag{109.5}$$

По определению кривизны $K = -d\varphi/ds$, где ds — элемент длины дуги, считаемый положительным, когда он проходит в направлении снизу вверх. Он связан с dx и dy соотношениями: $dx = ds\cos\varphi$; $dy = ds\sin\varphi$. Таким образом,

$$K = -\frac{d\varphi}{dx}\cos\varphi = -\frac{d\varphi}{dy}\sin\varphi.$$

Подставляя эти выражения в (109.5), получим два уравнения:

$$\rho g y\,dy + \sigma\sin\varphi\,d\varphi = 0, \tag{109.6}$$

$$\rho g y\,dx + \sigma\cos\varphi\,d\varphi = 0. \tag{109.7}$$

Интегрируя (109.6) при начальном условии $\varphi = \pi$ при $y = 0$, получим

$$y = 2\sqrt{\frac{\sigma}{\rho g}}\cos\frac{\varphi}{2}. \tag{109.8}$$

Подстановка этого выражения в (109.7) приводит к уравнению

$$dx = -\sqrt{\frac{\sigma}{\rho g}}\cos\frac{\varphi}{2}\,d\varphi + \frac{1}{2}\sqrt{\frac{\sigma}{\rho g}}\,\frac{d\varphi}{\cos(\varphi/2)},$$

интегрирование которого при начальном условии $x = 0$ при $\varphi = \pi/2$ дает

$$x = 2\sqrt{\frac{\sigma}{\rho g}}\left(\frac{1}{\sqrt{2}} - \sin\frac{\varphi}{2}\right) + \frac{1}{2}\sqrt{\frac{\sigma}{\rho g}}\ln\frac{(1+\sin(\varphi/2))(\sqrt{2}-1)}{(1-\sin(\varphi/2))(\sqrt{2}+1)}. \tag{109.9}$$

Формулы (109.8) и (109.9) выражают уравнение искомой поверхности в параметрической форме.

5. Определить в предыдущей задаче максимально возможную высоту поднятия пластинки над уровнем жидкости h и толщину приподнятого столба жидкости D в наиболее узком месте MN (рис. 125) при той же высоте поднятия. Найти также силу F, которую необходимо приложить к единице длины пластинки, чтобы оторвать последнюю от жидкости. Вес единицы длины пластинки равен q, ее ширина a.

Р е ш е н и е. Минимальная толщина столба жидкости $D = MN$ при максимально возможной высоте поднятия h (рис. 125) определится из требования $\varphi = 0$ при $y = h$. Подставляя в формулу (109.9) $x = (a - D)/2$, $\varphi = 0$, получим

$$D = a - 2\sqrt{\frac{\sigma}{\rho g}}\,[\sqrt{2} - \ln(\sqrt{2}+1)] = a - 1{,}066\sqrt{\frac{\sigma}{\rho g}}. \tag{109.10}$$

Если $a < 1{,}066\sqrt{\sigma/\rho g}$, то минимальное значение D равно нулю. В этом случае предельное значение угла $\varphi = 0$ не достигается.

Пусть $a > 1{,}066\sqrt{\sigma/\rho g}$. Тогда максимально возможная высота поднятия определится из формулы (109.8), если положить $\varphi = 0$:

$$h = 2\sqrt{\frac{\sigma}{\rho g}}. \tag{109.11}$$

Разность атмосферного и гидростатического давлений на пластинку направлена вниз и равна $\rho g\,h$. Поэтому

$$F = q + \rho g h a = q + 2a\sqrt{\rho g\sigma}. \tag{109.12}$$

Рассмотрим теперь второй случай: $a < 1{,}066\sqrt{\sigma/\rho g}$. В этом случае

$$h = 2\sqrt{\frac{\sigma}{\rho g}}\cos\frac{\varphi}{2}, \tag{109.13}$$

где φ определяется из трансцендентного уравнения

$$\frac{a}{2} = 2\sqrt{\frac{\sigma}{\rho g}}\left(\frac{1}{\sqrt{2}} - \sin\frac{\varphi}{2}\right) + \frac{1}{2}\sqrt{\frac{\sigma}{\rho g}}\ln\frac{(1+\sin(\varphi/2))(\sqrt{2}-1)}{(1-\sin(\varphi/2))(\sqrt{2}+1)}. \tag{109.14}$$

При нахождении F необходимо учесть, что в рассматриваемом случае пластинку тянет вниз дополнительная сила поверхностного натяжения $2\sigma\sin\varphi$. С учетом этой силы

$$F = q + 2a\sqrt{\rho g\sigma}\cos\frac{\varphi}{2} + 2\sigma\sin\varphi. \tag{109.15}$$

Если $a \ll \sqrt{\sigma/\rho g}$, то вторым членом в этой формуле можно пренебречь. Пренебрегая также в (109.14) членом $a/2$, находим $\varphi = \pi/2$. Таким образом, при $a \ll \sqrt{\sigma/\rho g}$

$$F = q + 2\sigma, \quad h = \sqrt{\frac{2\sigma}{\rho g}}.$$

6. Бесконечно длинная прямоугольная пластинка ширины a положена на поверхность несмачивающей ее жидкости с поверхностным натяжением σ. Плотность вещества пластинки ρ_0 больше плотности жидкости ρ. Найти максимальную толщину пластинки h, при которой она еще не утонет.

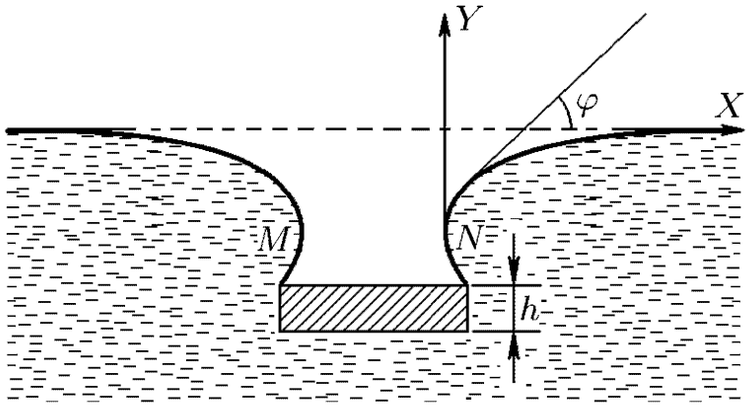


Рис. 126

Р е ш е н и е. Примем за ось Y вертикальную прямую, касающуюся боковой поверхности жидкости, а за ось X — горизонтальную прямую, перпендикулярную к длине пластинки и касающуюся поверхности жидкости в бесконечности (рис. 126). Уравнение боковой поверхности жидкости будет

$$x = 2\sqrt{\frac{\sigma}{\rho g}}\left(\frac{1}{\sqrt{2}} - \cos\frac{\varphi}{2}\right) + \frac{1}{2}\sqrt{\frac{\sigma}{\rho g}}\ln\frac{(1+\cos(\varphi/2))(\sqrt{2}-1)}{(1-\cos(\varphi/2))(\sqrt{2}+1)}, \tag{109.16}$$

$$y = -2\sqrt{\frac{\sigma}{\rho g}}\sin\frac{\varphi}{2}. \tag{109.17}$$

Минимальное расстояние $D = MN$ при максимально возможной глубине погружения пластинки $|y|_{\text{макс}}$ определится из требования $\varphi = \pi$, которое дает

$$D = a - 2\sqrt{\frac{\sigma}{\rho g}}(\sqrt{2} - \ln(\sqrt{2}+1)) = a - 1{,}066\sqrt{\frac{\sigma}{\rho g}}. \tag{109.18}$$

Если $a < 1{,}066\sqrt{\sigma/\rho g}$, то $D = 0$ и предельное значение угла $\varphi = \pi$ не достигается.

Рассмотрим сначала случай $a > 1{,}066\sqrt{\sigma/\rho g}$. В этом случае максимальная глубина погружения верхнего основания пластинки определится из (109.17), если положить $\varphi = \pi$. Она равна $|y|_{\text{макс}} = 2\sqrt{\sigma/\rho g}$. При этом на основание пластинки будет действовать направленная вверх разность давлений $\rho g(h + |y|_{\max})$, которая должна быть уравновешена весом пластинки. Максимальная толщина пластинки, при которой она еще не утонет, определится из условия $\rho g(h + |y|_{\text{макс}}) = \rho_0 g$, которое дает

$$h = \frac{2}{\rho_0 - \rho}\sqrt{\frac{\sigma\rho}{g}}. \tag{109.19}$$

Теперь рассмотрим случай $a < 1{,}066\sqrt{\sigma/\rho g}$. В этом случае

$$|y|_{\text{макс}} = 2\sqrt{\frac{\sigma}{\rho g}}\,\sin\frac{\varphi}{2}, \tag{109.20}$$

где φ определяется из уравнения

$$\frac{a}{2} = 2\sqrt{\frac{\sigma}{\rho g}}\Big(\frac{1}{\sqrt{2}} - \cos\frac{\varphi}{2}\Big) + \frac{1}{2}\sqrt{\frac{\sigma}{\rho g}}\,\ln\frac{(1+\cos(\varphi/2))(\sqrt{2}-1)}{(1-\cos(\varphi/2))(\sqrt{2}+1)}. \tag{109.21}$$

Для максимальной толщины пластинки получаем

$$h = \frac{2}{\rho_0 - \rho}\sqrt{\frac{\sigma\rho}{g}}\,\sin\frac{\varphi}{2} + \frac{2\sigma\sin\varphi}{ga(\rho_0 - \rho)}. \tag{109.22}$$

Если $a \ll \sqrt{\sigma/\rho g}$, то первым членом справа можно пренебречь. При этом, как видно из (109.21), $\varphi = \pi/2$, и мы находим

$$h = \frac{2\sigma}{ga(\rho_0 - \rho)}, \quad \text{или} \quad 2\sigma = gah(\rho_0 - \rho), \tag{109.23}$$

т. е. вес пластинки уравновешивается поверхностным натяжением и архимедовой подъемной силой.

7. Определить силу F, необходимую для отрыва круглой невесомой пластинки радиуса $r = 8$ см, положенной на поверхность воды. Поверхностное натяжение воды $\sigma = 73$ дин/см. Пластинка смачивается водой.

Ответ. Пренебрегая кривизной окружности, ограничивающей пластинку, получим

$$F \approx 2\pi r^2\sqrt{\rho g\sigma} \approx 1{,}1\ \text{Н}.$$

8. Найти высоту поднятия h жидкости у вертикальной бесконечной пластинки, смачиваемой жидкостью. Краевой угол равен ϑ.

Указание. См. решение задачи **4**.

Ответ. $h = \sqrt{\dfrac{2\sigma}{\rho g}\,(1 - \sin\vartheta)}$.

9. Определить глубину h ртутной лужицы на плоском горизонтальном стекле. Поперечные размеры лужицы велики по сравнению с ее глубиной. Поверхностное натяжение ртути на границе с воздухом $\sigma = 490$ дин/см, краевой угол на стекле $\vartheta = 140^\circ$. Плотность ртути $\rho = 13{,}6$ г/см^3.

Ответ. $h = 2\sqrt{\dfrac{\sigma}{\rho g}}\,\sin\dfrac{\vartheta}{2} = 3{,}6$ мм.

10. Стальная иголка, покрытая тонким слоем парафина или жира, может плавать на поверхности воды (рис. 127). Найти радиус иголки r, ширину зазора $D = MN$ между боковыми поверхностями жидкости в наиболее

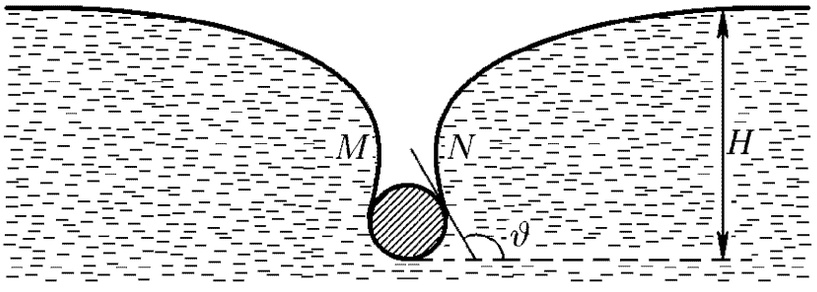


Рис. 127

узком месте, а также глубину погружения H для различных значений угла ν, образуемого общей касательной к поверхности иголки и жидкости с горизонтальной плоскостью. Плотность стали $\rho_0 = 7{,}8$ г/см^3, поверхностное натяжение воды $\sigma = 73$ дин/см. Определить максимальный радиус иголки, при котором она еще не утонет. Найти максимально возможную глубину погружения и соответствующий ей радиус иголки. При расчете иголку заменить бесконечно длинным цилиндром.

Решение. В точке A (рис. 128) поверхности жидкости и иголки тангенциально расходятся. На единицу длины иголки вверх действует сила поверхностного натяжения $F_1 = 2\sigma \sin\vartheta$. Кроме того, на нее действует

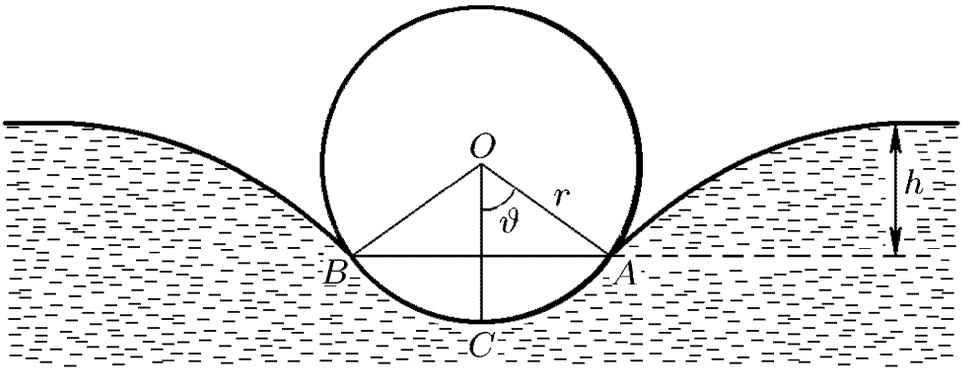


Рис. 128

сила гидростатического давления, также направленная вверх. Если бы часть ACB иголки была заменена жидкостью, то сила гидростатического давления была бы равна $F_2 = \rho g h \cdot AB = 2\rho g h r \sin\vartheta$, где r — радиус иголки, а ρ — плотность жидкости. Благодаря тому что часть ABC погружена в жидкость, на иголку дополнительно действует сила гидростатического давления F_3, равная весу воды, вытесненной частью ACD, т. е. $F_3 = \rho g r^2(\vartheta - \sin\vartheta\cos\vartheta)$. Сумма трех сил F_1, F_2 и F_3 должна равняться весу единицы длины иголки. Это дает

$$2\sigma \sin\vartheta + 2\rho g h r \sin\vartheta + \rho g r^2(\vartheta - \sin\vartheta\,\cos\vartheta) = \rho_0 g \pi r^2.$$

Между углом ϑ и высотой h существует соотношение $h = 2\sqrt{\dfrac{\sigma}{\rho g}}\,\sin\dfrac{\vartheta}{2}$ (см. решение задачи 6), и предыдущее уравнение принимает вид

$$\left[\pi\rho_0 - \rho\left(\vartheta - \frac{1}{2}\sin 2\vartheta\right)\right] r^2 - 4r\sqrt{\frac{\rho\sigma}{g}}\,\sin\vartheta\sin\frac{\vartheta}{2} - \frac{2\sigma\sin\vartheta}{g} = 0. \qquad (109.24)$$

Для D и H получаем

$$D = 2r\sin\vartheta + 2\sqrt{\frac{\sigma}{\rho g}}\left(2\cos\frac{\vartheta}{2} + \ln\operatorname{tg}\frac{\vartheta}{4}\right) - 2\sqrt{\frac{\sigma}{\rho g}}\,(\sqrt{2} - \ln(\sqrt{2}+1)), \quad (109.25)$$

$$H = 2r\sin^2\frac{\vartheta}{2} + 2\sqrt{\frac{\sigma}{\rho g}}\,\sin\frac{\vartheta}{2}. \quad (109.26)$$

После подстановки числовых значений:

$$\left[24{,}5 - \left(\vartheta - \frac{1}{2}\sin 2\vartheta\right)\right]r^2 - 1{,}091\sin\vartheta\sin\frac{\vartheta}{2}\cdot r - 0{,}1488\sin\vartheta = 0, \quad (109.27)$$

$$D = 2r\sin\vartheta + 1{,}091\cos\frac{\vartheta}{2} + 1{,}256\lg\operatorname{tg}\frac{\vartheta}{4} - 0{,}291. \quad (109.28)$$

(Предполагается, что здесь длины выражаются в сантиметрах.) Придавая ϑ различные значения, получим следующую таблицу.

ϑ, град	r, мм	H, мм	D, мм	ϑ, град	r, мм	H, мм	D, мм
0	0	0	—	80	0,955	4,29	—
10	0,328	0,481	—	90	0,990	4,85	1,98
20	0,471	0,975	—	100	1,005	5,35	1,91
30	0,583	1,49	—	110	1,001	5,82	1,68
40	0,680	2,03	—	120	0,977	6,20	1,24
50	0,763	2,58	—	130	0,922	6,45	0,65
60	0,840	3,15	—	139	0,846	6,59	0,04
70	0,903	3,72	—	139°30′	0,842	6,60	0,00

Наибольший радиус r получается при $\vartheta \approx 100^\circ$ и составляет приблизительно 1 мм. Если $r > 0{,}842$ мм, то существуют два положения равновесия иголки: одно при $\vartheta \lesssim 100^\circ$, другое при $\vartheta \gtrsim 100^\circ$. Если же $r < 0{,}842$ мм, то существует только одно положение равновесия с $\vartheta \lesssim 60^\circ$, так как в этом случае при $\vartheta \gtrsim 60^\circ$ формула (109.28) дает для D отрицательное значение. Наибольшая глубина погружения H получается при $r \approx 0{,}842$ мм и равна приблизительно 6,60 мм.

11. Две вертикальные параллельные пластинки частично погружены в жидкость. Показать, что между ними будет наблюдаться притяжение, когда обе пластинки либо смачиваются, либо не смачиваются жидкостью, и отталкивание, когда одна пластинка смачивается жидкостью, а другая нет.

Р е ш е н и е. В случае смачивания жидкость между пластинками поднимается (рис. 129 *а*). Давление в поднявшейся части жидкости становится меньше давления окружающей атмосферы. Атмосферное давление стремится прижать пластинки друг к другу. В случае несмачивания (рис. 129 *б*) давление жидкости снаружи пластинок больше давления воздуха между ними. Появляется разность давлений, стремящаяся сблизить пластинки. Рассмотрим теперь случай, когда левая пластинка смачивается жидкостью, а правая не смачивается (рис. 129 *в*). Если пластинки расположены достаточно близко друг к другу, то поверхность жидкости между ними ни в одной точке не становится горизонтальной. Она имеет точку перегиба где-то между пластинками. Вследствие этого жидкость между пластинками поднимается

ниже у левой пластинки и опустится меньше у правой пластинки, чем наружная жидкость. С этим обстоятельством и связано в рассматриваемом случае появление отталкивания между пластинками. Давление жидкости

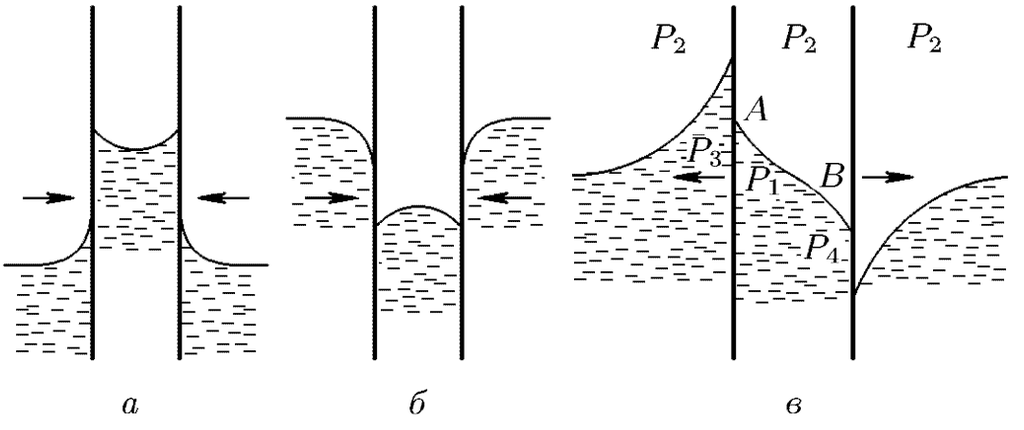


Рис. 129

между пластинками в точке A равно делению наружной жидкости P_3 на той же высоте. Давление воздуха P_2 больше P_1, так как поверхность жидкости у левой пластинки обращена к воздуху вогнутой стороной. Давление P_3 убывает с высотой, тогда как P_2 остается практически постоянным. Поэтому разность давлений стремится переместить левую пластинку влево. В точке B давление жидкости P_4 больше P_2, так как поверхность жидкости в этой точке обращена к воздуху выпуклой стороной. Тем более это справедливо для давления ниже этой точки. В результате разность давлений $P_4 - P_2$ будет перемещать правую пластинку вправо. Действием рассмотренных сил объясняется концентрация в кучи пузырьков воздуха, листьев, мелких щепок и прочих смачиваемых тел, плавающих на поверхности воды в стоячих водоемах.

12. Столбик жидкости, помещенный в коническую трубку, сам движется к более узкой части, когда он смачивает стенки трубки, и к более широкой части, когда не смачивает. Объяснить явление.

13. Если в трубке находится ряд капель (столбиков) какой-либо жидкости, то требуется значительное давление, чтобы продвинуть их вдоль трубки, независимо от того, смачивают они стенки трубки или не смачивают. Сопротивление смачивающих капель еще более увеличивается, когда канал трубки попеременно суживается и расширяется. При этом капли собираются в суженных частях канала. Объяснить явление.

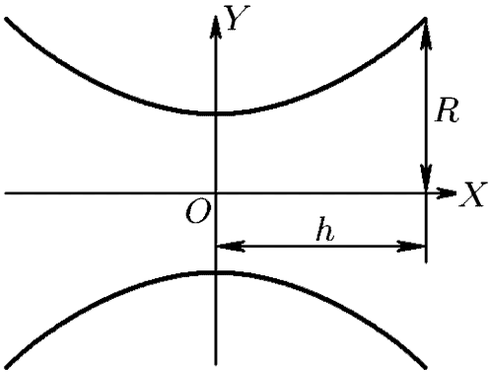


Рис. 130

14. Определить форму мыльной пленки, края которой закреплены на двух одинаковых кольцах радиуса R, удаленных друг от друга на расстояние $2h$. Центры колец лежат на общей прямой, перпендикулярной к их плоскостям. Плоскости колец не затянуты пленками.

Р е ш е н и е. Ввиду симметрии пленка будет поверхностью вращения вокруг прямой, на которой лежат центры колец. Пересечем поверхность пленки произвольной плоскостью, проходящей через эту ось, и примем ее за координатную плоскость XY (рис. 130). Так как давления по обе стороны пленки

одинаковы, то ее полная кривизна $1/R_1 + 1/R_2$ должна равняться нулю. Радиус кривизны R_1 нормального сечения пленки, лежащего в плоскости XY, определяется формулой $\dfrac{1}{R_1} = -\dfrac{y''}{(1+y'^2)^{3/2}}$ (величина отрицательная). Радиус кривизны перпендикулярного к нему нормального сечения легко определить с помощью известной из дифференциальной геометрии теоремы Менье, согласно которой $y = R_2 \cos\alpha$, где α — угол между плоскостью нормального сечения и координатной плоскостью YZ. Подставляя значение $\cos\alpha$, получим $R_2 = y\sqrt{1+y'^2}$ (величина положительная). Таким образом, дифференциальное уравнение, определяющее форму осевого сечения пленки, принимает вид

$$\frac{y''}{1+y'^2} - \frac{1}{y} = 0. \tag{109.29}$$

Введем подстановку $y' = \operatorname{sh}\Theta$. Тогда $1 + y'^2 = \operatorname{ch}^2\Theta$, $y = \dfrac{\operatorname{ch}\Theta}{d\Theta/dx}$. Дифференцируя последнее соотношение и принимая во внимание, что $y' = \operatorname{sh}\Theta$, находим $d^2\Theta/dx^2 = 0$, откуда $\Theta = ax + b$, где a и b — постоянные. Они определятся из граничных условий: $y = R$ при $x = \pm h$. Очевидно, $b = 0$, так как ввиду симметрии y должна быть четной функцией от x. Окончательно:

$$y = \frac{1}{a}\operatorname{ch} ax = \frac{1}{2a}\left(e^{ax} + e^{-ax}\right), \tag{109.30}$$

где постоянная a определяется уравнением

$$aR = \operatorname{ch} ah. \tag{109.31}$$

Поверхность пленки получается вращением кривой (109.30) вокруг оси X. Она называется *катеноидом*. Уравнение (109.31) легче всего исследовать и решать графически. Применяя этот метод, нетрудно доказать, что оно имеет решение только при условии $R/h > 1{,}51$. Значит, чтобы между кольцами могла образоваться пленка, необходимо, чтобы расстояние между ними $2h$ не превышало $(2/1{,}51)\,R = 1{,}32\,R$.

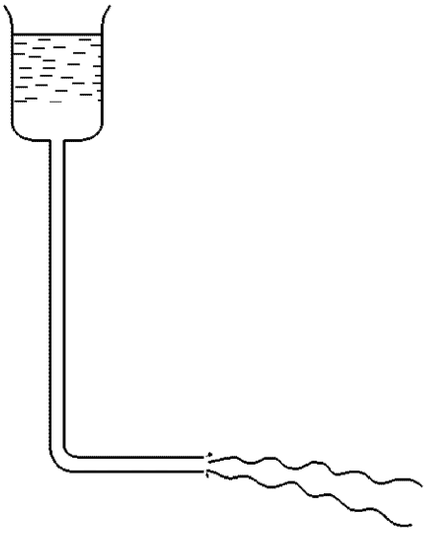


Рис. 131

15. Струя жидкости вытекает через трубку в дне сосуда (рис. 131). Поперечное сечение трубки имеет форму эллипса, вытянутого в горизонтальном направлении. Струя принимает форму цепи, звенья которой попеременно то вытянуты, то сплюснуты в горизонтальном направлении. Объяснить явление. Пользуясь соображениями размерности, найти зависимость длины звена l в начальной части струи от плотности жидкости ρ, поверхностного натяжения σ, расстояния h между основанием трубки и уровнем жидкости, а также от ускорения свободного падения g, если поперечное сечение трубки остается неизменным. На наблюдении этого явления основан метод Рэлея измерения поверхностного натяжения жидкостей.

Ответ. $l \sim \sqrt{\rho g h/\sigma}$

16. Мыльный пузырь выдут через цилиндрическую трубку с внутренним радиусом $r = 1$ мм и длиной $l = 10$ см. В тот момент, когда радиус

пузыря достигает значения $R_0 = 10$ см, перестают дуть, и воздух из пузыря начинает выходить через трубку. Через какое время, начиная с этого момента, пузырь исчезнет? Поверхностное натяжение мыльного раствора $\sigma = 50$ дин/см, вязкость воздуха $\eta = 1{,}8 \cdot 10^{-4}$ г/(с · см). Изменением плотности воздуха за время процесса пренебречь.

Ответ. Время t связано с радиусом пузыря соотношением

$$t = \frac{2\eta l}{\sigma r^4}(R_0^4 - R^4).$$

Пузырь исчезнет через время

$$t = \frac{2\eta l}{\sigma r^4} R_0^4 = 7{,}2 \cdot 10^3 \text{ с} = 2 \text{ ч}.$$

17. На какую величину ΔT температура воздуха внутри мыльного пузыря должна превышать температуру окружающего воздуха T, чтобы пузырь стал подниматься? Радиус пузыря равен r, поверхностное натяжение мыльной пленки σ. Массой пленки можно пренебречь. Учесть, что давление воздуха внутри пузыря мало отличается от атмосферного давления P.

Ответ. $\Delta T > 4\sigma T/Pr$.

§ 110. Капиллярно-гравитационные волны малой амплитуды

1. *Капиллярно-гравитационными волнами называются волны, распространяющиеся по поверхности жидкости под действием сил поверхностного натяжения и силы тяжести.*

Для понимания настоящего параграфа требуется знакомство с некоторыми понятиями, относящимися к учению о волнах, которые будут подробно изложены в третьем томе нашего курса. Читатель, не знакомый с этими понятиями, может пропустить этот параграф без ущерба для понимания дальнейшего. Ограничимся рассмотрением капиллярно-гравитационных *волн малой амплитуды*. Так называются волны, амплитуда которых мала по сравнению с длиной волны. Мы будем также считать жидкость *глубокой*, т. е. рассмотрим случай, когда глубина жидкости значительно больше длины волны.

Найдем выражение для скорости распространения капиллярно-гравитационных волн. Это можно сделать очень просто, если воспользоваться следующим результатом, вытекающим из уравнений гидродинамики несжимаемой жидкости. *В плоской бегущей синусоидальной волне малой амплитуды каждая частица жидкости движется по окружности, расположенной в вертикальной плоскости, проходящей через направление распространения волны.* Радиус окружности r мал по сравнению с длиной волны λ. Он убывает экспоненциально при удалении от поверхности жидкости. Однако знание конкретного закона, по которому происходит такое убывание, для последующих рассуждений не требуется. Существенно только то, что на поверхности жидкости амплитуда колебаний максимальна, а далеко от нее (на расстояниях во много раз больших λ) обращается в нуль.

Если точки поверхности жидкости, расположенные на некоторой прямой, заставить совершать гармоническое колебательное движение, то вдоль поверхности жидкости перпендикулярно к этой прямой побежит капиллярно-гравитационная волна, скорость распространения которой обозначим через c. В неподвижной системе отсчета, как уже сказано, каждая частица жидкости движется по окружности. Рассмотрим явление в системе отсчета,

равномерно движущейся со скоростью c. В этой системе волны будут стоять на месте. Движение частицы будет слагаться из равномерно-поступательного со скоростью c и равномерного вращения по окружности радиуса r. Так как радиус r предполагается малым по сравнению с длиной волны λ, то можно пренебречь горизонтальными колебаниями частицы. Если ось X направить по невозмущенной поверхности жидкости в сторону распространения волны, а ось Z — вертикально вниз, то в указанном приближении движение частицы на поверхности жидкости изобразится уравнениями

$$x = ct, \quad z = r \sin \frac{2\pi ct}{\lambda}. \tag{110.1}$$

Форма траектории найдется отсюда исключением времени t, что дает

$$z = r \sin \frac{2\pi x}{\lambda}. \tag{110.2}$$

Это — синусоида. Частицы, расположенные не на поверхности, а в глубине жидкости, также движутся по синусоидам. Но для них радиус r меньше — он убывает с глубиной.

2. На рис. 132 верхняя синусоида ABC представляет траекторию частицы на поверхности жидкости, а $A'B'C'$ — бесконечно близкой к ней частицы в глубине жидкости. В рассматриваемой системе отсчета течение

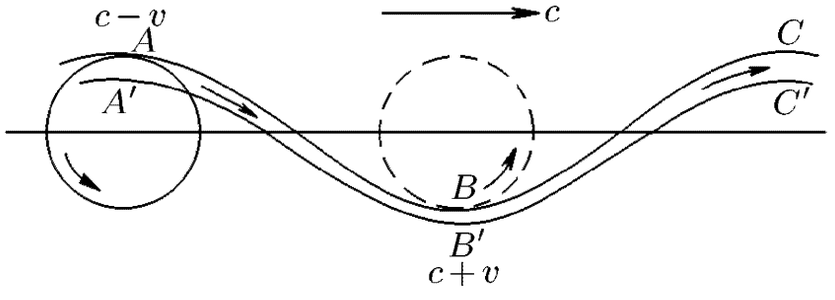


Рис. 132

жидкости стационарно. Пространство между поверхностями ABC и $A'B'C'$ представляет собой трубку тока. Применим к ней уравнение Бернулли. Если v — скорость движения жидкости по окружности, то в точке A, где поступательное и вращательное движения вычитаются, полная скорость жидкости будет $c - v$, а в точке B, где они складываются, $c + v$. Разность высот точек A и B равна $h = 2r$. Поэтому по уравнению Бернулли

$$P_A + \frac{\rho}{2}(c - v)^2 + 2\rho g r = P_B + \frac{\rho}{2}(c + v)^2,$$

или

$$2\rho c v = 2\rho g r + (P_A - P_B). \tag{110.3}$$

Очевидно,

$$v = \frac{2\pi r}{T} = \frac{2\pi r c}{\lambda}. \tag{110.4}$$

Давления жидкости в точках A и B по формуле Лапласа равны соответственно

$$P_A = P_0 + \sigma K, \quad P_B = P_0 - \sigma K, \tag{110.5}$$

где K — абсолютное значение кривизны синусоиды в точке A или B. Поскольку в этих точках первая производная dz/dx равна нулю, для кривизны

K получаем из (110.2)

$$K = \left|\frac{d^2z}{dx^2}\right| = \frac{4\pi^2 r}{\lambda^2}. \tag{110.6}$$

Из (110.3) с учетом (110.4), (110.5) и (110.6) получаем формулу для скорости распространения капиллярно-гравитационных волн:

$$c = \sqrt{\frac{g\lambda}{2\pi} + \frac{2\pi\sigma}{\rho\lambda}}. \tag{110.7}$$

Заметим, что в теории волн величина c называется *фазовой скоростью*, т. е. скоростью, с которой распространяется фаза волны. Эта скорость зависит от длины волны, т. е. *капиллярно-гравитационные волны обладают дисперсией.*

3. Для длинных волн, когда $g\lambda/2\pi \gg 2\pi\sigma/\rho\lambda$, т. е. $\lambda \gg 2\pi\sqrt{\sigma/\rho g}$, поверхностное натяжение не играет роли, и формула (110.7) переходит в

$$c = \sqrt{g\lambda/2\pi}. \tag{110.8}$$

В этом случае волны называются *гравитационными.*

В другом предельном случае, когда $\lambda \ll 2\pi\sqrt{\sigma/\rho g}$, наоборот, несущественно действие силы тяжести. В этом случае волны называются *капиллярными.* Для их скорости распространения получаем

$$c = \sqrt{2\pi\sigma/\rho\lambda}. \tag{110.9}$$

Наблюдение капиллярных волн дает удобный метод измерения поверхностного натяжения жидкостей. На поверхности жидкости возбуждаются круговые капиллярные волны колебаниями погруженного в нее штифта. Измеряется частота колебаний $\nu = c/\lambda$ и длина волны λ. Поверхностное натяжение рассчитывается по формуле

$$\sigma = \rho\lambda^3\nu^2/2\pi. \tag{110.10}$$

4. В качестве дополнения к настоящему параграфу докажем те следствия уравнений гидродинамики, на которых основывались наши рассуждения. Во-первых, мы исходим из условия сохранения массы жидкости. Если жидкость несжимаемая, то это условие в нашем случае записывается в виде

$$\frac{\partial v_x}{\partial x} + \frac{\partial v_z}{\partial z} = 0 \tag{110.11}$$

(уравнение непрерывности). Во-вторых, мы пользуемся уравнением Эйлера для малых колебаний жидкости

$$\rho\frac{\partial v}{\partial t} = -\operatorname{grad} P. \tag{110.12}$$

В этих уравнениях скорость рассматривается как функция времени и координат точки пространства, к которой эта скорость относится. В точном уравнении Эйлера в левой части (110.12) должно было бы стоять ускорение частицы dv/dt, а не частная производная $\partial v/\partial t$. Эта частная производная показывает лишь, как меняется во времени скорость различных частиц, проходящих через одну и ту же точку пространства. Для вычисления же ускорения надо было бы сравнивать скорости одной и той же частицы в различные моменты времени (в которые частица занимает различные положения в пространстве). Однако для малых колебаний это различие можно

не принимать во внимание и писать уравнение Эйлера в упрощенной форме (110.12). Переходя к координатной форме записи, получим из (110.12)

$$\rho\frac{\partial v_x}{\partial t}=-\frac{\partial P}{\partial x},\quad \rho\frac{\partial v_z}{\partial t}=-\frac{\partial P}{\partial z}.$$

Дифференцируя первое уравнение по z, а второе по x, исключим P:

$$\frac{\partial}{\partial t}\Big(\frac{\partial v_x}{\partial z}-\frac{\partial v_z}{\partial x}\Big)=0. \tag{110.13}$$

Допустим теперь, что в жидкости распространяется синусоидальная волна:

$$v_x=v_{0x}(z)\cos(\omega t-kx),\quad v_z=v_{0z}(z)\sin(\omega t-kx+\delta),$$

где ω, k, δ — постоянные. Дифференцируя и подставляя полученные выражения в уравнение (110.11), получим

$$kv_{0x}(z)\sin(\omega t-kx)+\frac{dv_{0z}}{dz}\sin(\omega t-kx+\delta)=0.$$

Это соотношение должно соблюдаться в любой момент времени, что возможно лишь при выполнении условий $\delta=0$ и

$$kv_{0x}(z)+\frac{dv_{0z}}{dz}=0.$$

(Другую возможность: $\delta=\pi$, можно не рассматривать, так как она сводится к предыдущей изменением знака у v_{0z}.) Аналогичным образом из уравнения (110.13) получаем

$$kv_{0z}(z)+\frac{dv_{0x}}{dz}=0.$$

Сравнивая эти два уравнения, находим

$$v_{0x}\,dv_{0x}=v_{0z}\,dv_{0z}.$$

Отсюда $v_{0x}^2=v_{0z}^2+\text{const}$. Постоянная интегрирования здесь равна нулю, так как на дне сосуда, в который налита жидкости, скорость обращается в нуль. Итак, $v_{0x}=v_{0z}\equiv C(z)$, а потому

$$v_x=C(z)\cos(\omega t-kx),\quad v_z=C(z)\sin(\omega t-kx).$$

Интегрируя по времени, находим

$$x-x_0=\frac{C(z)}{\omega}\sin(\omega t-kx),\quad z-z_0=-\frac{C(z)}{\omega}\cos(\omega t-kx).$$

Отсюда видно, что траекторией частицы является окружность радиуса $r=C(z)/\omega$.

Глава X

ФАЗОВЫЕ РАВНОВЕСИЯ И ФАЗОВЫЕ ПРЕВРАЩЕНИЯ

§ 111. Фазы и фазовые превращения

1. *Фазой называется макроскопическая физическая однородная часть вещества, отделенная от остальных частей системы границами раздела, так что она может быть извлечена из системы механическим путем.*

Допустим, например, что в закрытом сосуде заключена некоторая масса воды, над которой находится смесь воздуха с водяными парами. Эта система является *двухфазной*. Она состоит из двух фаз: *жидкой* (вода) и *газообразной* (смесь воздуха с водяными парами). Если бы воздуха не было, то в системе также было бы две фазы: жидкая (вода) и газообразная (водяные пары). Бросим в воду кусочки льда. Система превратится в *трехфазную* и будет состоять из *твердой фазы* (лед), *жидкой* (вода) и *газообразной* (смесь воздуха с водяными парами). Добавим к воде некоторое количество спирта. Число фаз не изменится, так как вода смешивается со спиртом, образуя физически однородную жидкость. Однако если к воде добавить ртуть, то последняя не смешивается с водой, и получится система с *двумя жидкими фазами*: ртутью и водой. Газообразная фаза по-прежнему будет одна; она состоит из смеси воздуха, паров воды и паров ртути. Бросив в воду кусочки поваренной соли, получим систему с *двумя твердыми фазами*: льдом и твердой поваренной солью. При подсчете числа фаз не имеет значения, является та или иная фаза единым телом или состоит из нескольких частей, отделенных одна от другой. Так, капельки тумана в воздухе образуют вместе с ним двухфазную систему, состоящую из жидкой фазы (вода) и газообразной фазы (смесь воздуха с водяными парами). *В системе может быть несколько твердых или жидких фаз. Но она не может содержать более одной газообразной фазы, так как все газы смешиваются между собой.*

2. Важнейшим вопросом в учении о фазах является выяснение условий, при которых система, состоящая из двух или нескольких фаз, находится в *равновесии*. Последнее включает в себя *механическое* и *тепловое равновесия*. Для теплового равновесия необходимо, чтобы все фазы системы имели одну и ту же температуру. Необходимым условием механического равновесия является равенство давлений по разные стороны границы раздела соприкасающихся фаз. Впрочем, последнее условие строго справедливо только в случае плоских границ

раздела. В случае границ оно нарушается действием сил поверхностного натяжения. Так, на поверхности раздела жидкости и ее пара при равновесии существует разность давлений $P_2 - P_1 = \sigma K$, где $K = 1/R_1 + 1/R_2$ — средняя кривизна этой поверхности (см. § 109). Сначала мы будем пренебрегать кривизной поверхностей раздела фаз, предполагая, что они либо плоские, либо их кривизна мала. Влияние кривизны будет учтено особо.

Равенство давлений и температур еще не означает, что система находится в равновесии, так как соприкасающиеся фазы могут превращаться друг в друга. Такие превращения называются *фазовыми превращениями*. При фазовых превращениях одни фазы растут другие уменьшаются и даже могут совсем исчезнуть. Состояние равновесия характеризуется тем, что массы всех фаз системы остаются неизменными. Следовательно, должно быть выполнено еще одно необходимое условие равновесия — *равновесие по отношению к взаимным превращениям фаз*. Это основное условие в учении о равновесии фаз и фазовых превращениях.

3. Примерами фазовых превращений могут служить изменения *агрегатного состояния вещества*. Под агрегатными состояниями понимают твердое, жидкое и газообразное состояния вещества. Твердое и жидкое состояния называются *конденсированными*. *Испарением*, или *парообразованием*, в широком смысле слова называют переход вещества из конденсированного состояния в газообразное. Обратный переход называется *конденсацией*. В узком смысле испарение есть переход вещества из жидкого состояния в газообразное. Переход из твердого состояния непосредственно в газообразное называется *сублимацией* или *возгонкой*. Переход из твердого состояния в жидкое называется *плавлением*, а обратный переход из жидкого состояния в твердое — *затвердеванием* или *кристаллизацией*.

Хорошо известным примером сублимации является превращение льда в пар — мокрое белье высыхает на морозе. Приведем другой пример. Поместим кристаллики йода в стеклянную колбу и нагреем их на спиртовой или газовой горелке. Плавления не наблюдается. Однако в колбе появятся фиолетовые пары йода, получившиеся в результате его сублимации. При охлаждении пары йода, соприкасаясь с холодными стенками колбы кристаллизуются и оседают на них в виде маленьких кристалликов.

Твердое состояние вещества может реализоваться в различных *кристаллических модификациях*. Это явление называется *полиморфизмом*. Например, твердый углерод может существовать в виде графита или алмаза, которые отличаются друг от друга кристаллической структурой. Существует несколько разновидностей льда, т. е. твердой воды. Твердое железо может существовать в четырех различных модификациях (α-, β-, γ- и δ-железо). Некоторые жидкости могут также существовать в виде различных модификаций (жидкие кристаллы, жидкие гелий I и гелий II). При изменении температуры и давления

модификации могут превращаться в другие. Такие превращения относятся также к числу фазовых превращений. Превращение вещества из одной кристаллической модификации в другую называется *полиморфным превращением.*

Приведем примеры полиморфных превращений. Возьмем пробирку с двуйодистой ртутью (HgI_2). При комнатной температуре это порошок красного цвета. При нагревании до 126 °С он превращается в кристаллики желтого цвета. При охлаждении происходит обратное превращение. При температуре 253 °С желтая модификация плавится, превращаясь в красновато-бурую жидкость.

Другим примером может служить превращение α-железа в γ-железо. Берется длинная горизонтально натянутая железная проволока, один конец которой закрепляется неподвижно, а другой конец перекидывается через неподвижный блок. К свободному концу подвешивается груз, чтобы проволока была в натянутом состоянии. К блоку прикреплена стрелка, по вращению которой можно судить об удлинении или укорочении проволоки. Проволока нагревается электрическим током до красного каления. При некоторой температуре происходит превращение α-железа в γ-железо. Однако это превращение заметить трудно. Легче наблюдать обратное превращение. Раскаленная проволока остывает и укорачивается. Но в некоторый момент внезапно снова происходит ее удлинение. В этот момент γ-железо опять превращается в α-железо. При этом на мгновение проволока начинает светиться ярче, что объясняется выделением теплоты перехода.

4. Рассмотрим еще простейший пример фазовых превращений — испарение и конденсацию. На этом примере лучше всего уяснить смысл равновесия между фазами. Допустим, что в закрытом сосуде заключена некоторая масса жидкости, над которой находится ее пар. Объем сосуда остается неизменным, а температура поддерживается постоянной. Молекулы вещества совершают движение и все время пересекают границу раздела между жидкостью и паром. Происходит непрерывный обмен молекулами между этими двумя фазами. Если из жидкости в пар переходит больше молекул, чем из пара в жидкость, то количество жидкости уменьшается, т. е. идет процесс испарения. Тогда говорят, что пар над жидкостью *ненасыщенный*, или *перегретый*. Если, наоборот, число молекул, переходящих из пара в жидкость, превышает число молекул, переходящих в обратном направлении, то пар конденсируется в жидкость. В этом случае пар называется *пересыщенным*. Наконец, когда число молекул, переходящих из жидкости в пар, равно числу молекул, переходящих за то же время из пара в жидкость, наступает состояние *динамического*, или *статистического*, *равновесия*, в котором количество вещества в каждой фазе в среднем остается неизменным. Это и есть состояние фазового равновесия. Вообще, *фазовое равновесие между любыми фазами 1 и 2 не есть статическое состояние, в котором полностью прекратились фазовые превращения, а характеризуется равенством средних скоростей двух взаимно противоположных процессов: превращения фазы 1 в фазу 2 и обратного превращения фазы 2 в фазу 1*. При равновесии эти противоположные процессы взаимно компенсируют друг друга, как того требует принцип детального равновесия. Благодаря этому количество вещества в каждой фазе в среднем остается неизменным.

§ 112. Условие равновесия фаз химически однородного вещества

1. В этой главе мы ограничимся рассмотрением фазовых превращений только *химически однородных веществ*. Фазовые превращения в растворах и смесях будут рассмотрены в следующей главе. Условие равновесия фаз можно получить из теорем термодинамики. Как уже говорилось выше, при равновесии системы температуры и давления всех ее фаз одинаковы. Если их поддерживать постоянными, то термодинамический потенциал системы может только убывать (см. § 50). При равновесии он принимает минимальное значение. Этим положением мы и воспользуемся для вывода условия равновесия фаз.

Рассмотрим систему, состоящую из двух фаз 1 и 2, которые могут превращаться друг в друга. Пусть m_1 — масса первой, а m_2 — масса второй фазы. Обозначим через φ_1 и φ_2 *удельные термодинамические потенциалы* вещества в этих фазах. Термодинамический потенциал всей системы представится в виде $\Phi = m_1\varphi_1 + m_2\varphi_2$. Пусть давление и температура системы поддерживаются постоянными. Тогда при фазовых превращениях величины φ_1 и φ_2 не будут изменяться, так как они являются однозначными функциями только температуры и давления. Не будет, разумеется, изменяться и полная масса вещества $m = m_1 + m_2$. Могут изменяться только массы m_1 и m_2. И эти изменения должны происходить в таком направлении, чтобы термодинамический потенциал Φ принял наименьшее значение, возможное в рассматриваемых условиях. Если $\varphi_1 > \varphi_2$, то всякое превращение фазы 1 в фазу 2 сопровождается уменьшением Φ. Это превращение и будет происходить, пока вся фаза 1 не перейдет в более устойчивую фазу 2. Тогда система сделается однофазной, а ее термодинамический потенциал достигнет минимального значение $m\varphi_2$. Наоборот, если $\varphi_1 < \varphi_2$, то фаза 2 в конце концов превратится в фазу 1. Только при условии

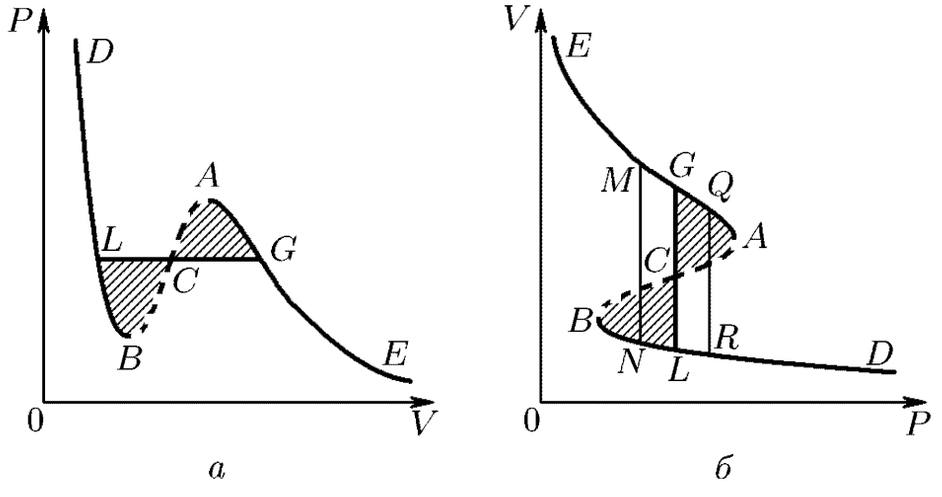
$$\varphi_1(P, T) = \varphi_2(P, T) \tag{112.1}$$

фазы будут находиться в равновесии друг с другом. Таким образом, *условием равновесия фаз является равенство их удельных термодинамических потенциалов.*

Внутренняя энергия U и энтропия S тела определены с точностью до произвольных аддитивных постоянных. Поэтому термодинамический потенциал $\Phi = U - TS + PV$ и его удельное значение $\varphi(P, T)$ определены с точностью до произвольной линейной функции температуры. Возникающая благодаря этому неоднозначность должна быть исключена из условия (112.1). Для этого достаточно условиться определять удельные термодинамические потенциалы $\varphi_1(P, T)$ и $\varphi_2(P, T)$ путем интегрирования выражения: $d\varphi = -s\,dT + v\,dP$, исходя из *одного и того же состояния*. Смысл условия (112.1) состоит в том, что *при любых фазовых превращениях величина удельного термодинамического потенциала всегда изменяется непрерывно*. В этом отношении он отличается от других физических величин — удельного объема,

удельной энтропии и теплоемкости, диэлектрической и магнитной проницаемостей, электрической проводимости и пр., которые при фазовых превращениях, как правило, меняются скачкообразно.

2. Применим условие (112.1) к процессам превращения жидкости в пар и обратно. Возьмем какую-либо изотерму вещества, подчиняющегося уравнению Ван-дер-Ваальса, при температуре ниже критической (рис. 133 *а*). Жирные ветви изотермы EG и LD изображают



а *б*

Рис. 133

устойчивые состояния вещества: газообразное и жидкое. Участок GA соответствует перенасыщенному пару, участок LB — перегретой жидкости. Они изображены тонкими сплошными линиями. Штриховой линией представлена ветвь ACB, точки которой изображают *абсолютно неустойчивые* состояния вещества. Жирный горизонтальный участок LCG представляет изотерму вещества в двухфазном состоянии. Положение всех этих участков было определено в § 101. То же самое, и притом более полно, можно сделать с помощью условия (112.1). Когда точка, изображающая состояние вещества, перемещается вдоль изотермы, то ввиду постоянства температуры ($dT = 0$) для изменения удельного термодинамического потенциала можно написать

$$d\varphi = v\,dP, \quad \varphi = \int v\,dP. \tag{112.2}$$

Поскольку независимой переменной теперь является P, удобнее повернуть координатные оси так, чтобы ось давлений стала горизонтальной, а ось объемов — вертикальной (рис. 133 *б*). Исследуем изменение функции $\varphi(P, T)$ при перемещении изображающей точки вдоль теоретической изотермы $EACBD$ (рис. 133 *а*). На участке EA дифференциал dP положителен, а потому потенциал φ возрастает. На участке ACB dP меняет знак, а φ начинает убывать. При дальнейшем изменении состояния вещества вдоль изотермы BD потенциал φ снова начинает монотонно возрастать. Величина φ будет проходить через те же значения, которые она принимала раньше на ветви AE. Отсюда следует, что существует такая изобара GL, в точках которой L и G значения

удельного термодинамического потенциала φ одинаковы: $\varphi_G = \varphi_L$. Следовательно, $\int_{GACBL} v\,dP = 0$, или

$$\int\limits_{GAC} v\,dP = \int\limits_{LBC} v\,dP.$$

Это значит, что площади $GACG$ и $CBLC$, заштрихованные на рис. 133, одинаковы. Проведем далее изобары MN и QR левее и правее изобары GL. Тогда

$$\varphi_M = \varphi_G - \int\limits_{P_M}^{P_G} v_{\text{п}}\,dP, \quad \varphi_N = \varphi_L - \int\limits_{P_N}^{P_L} v_{\text{ж}}\,dP,$$

где $v_{\text{п}}$ — удельный объем пара, а $v_{\text{ж}}$ — жидкости. Так как $\varphi_G = \varphi_L$, $v_{\text{п}} > v_{\text{ж}}$, а пределы интегрирования одни и те же, то $\varphi_M < \varphi_N$. Так же доказывается, что $\varphi_R < \varphi_G$.

Итак, удельный термодинамический потенциал газа на ветви изотермы EG меньше соответствующего удельного термодинамического потенциала жидкости на ветви BL. Поэтому из двух возможных состояний M и N, в которых вещество может существовать при заданных температуре и давлении, газообразное состояние M является более устойчивым. Оно и реализуется в действительности. Наоборот, на ветви LD жидкость имеет меньший удельный термодинамический потенциал, чем газ на участке изотермы GA. Из двух возможных состояний вещества R и Q, в которых давления и температуры одинаковы, жидкое состояние R здесь более устойчиво, чем газообразное Q. Если жидкость и газ в таких состояниях граничат между собой, то газ будет конденсироваться в жидкость, пока все вещество не окажется в жидком состоянии. В точках G и L удельные термодинамические потенциалы газа и жидкости одинаковы. Поэтому состояния G и L одинаково устойчивы. Газ в состоянии G и жидкость в состоянии L находятся в фазовом равновесии друг с другом. Такой газ является насыщенным паром этой жидкости.

Таким образом, мы снова пришли к правилу Максвелла, полученному в § 101 непосредственно из равенства Клаузиуса. Однако новый подход позволяет не только указать положение равновесной изобары LG, но и понять, почему именно в точках L и G должно происходить фазовое превращение. При таком подходе можно было бы найти положение равновесной изобары LG и в том случае, когда изотерма состоит из двух изолированных ветвей DB и AE, не связанных между собой промежуточной кривой ACB (см. рис. 99). Разумеется, в этом случае правило Максвелла теряет смысл, а равновесная изобара LG должна быть определена из условия $\varphi_L = \varphi_G$. Однако для объяснения возможности существования метастабильных состояний — перенасыщенного пара и перегретой жидкости (участки GA и LB изотермы) —

изложенные соображения недостаточны. Этот вопрос подробно разобран в § 119.

ЗАДАЧИ

1. В толстостенном закрытом сосуде помещен кусок льда, над которым находится насыщенный водяной пар. В сосуд можно нагнетать воздух до высокого давления. На сколько надо повысить давление воздуха в сосуде, чтобы давление насыщенного пара над льдом повысилось на один процент, если температура ($T = 250$ К) поддерживается постоянной? Удельный объем льда $v_л = 1{,}1$ см^3/г.

Р е ш е н и е. Изотермическое увеличение внешнего давления на ΔP увеличивает удельный термодинамический потенциал льда на $\Delta\varphi_л = v_л \Delta P$, причем сжимаемостью льда можно пренебречь. Чтобы равновесие не нарушалось, на столько же должен возрасти удельный термодинамический потенциал пара. Но для пара $\Delta\varphi_п = v_п \Delta P_п = \dfrac{PT}{\mu}\dfrac{\Delta P_п}{P_п}$. Приравнивая оба выражения, получим

$$\Delta P = \frac{PT}{\mu v_л}\frac{\Delta P_п}{P_п} = 10{,}5 \text{ атм}.$$

2. В цилиндре под поршнем помещена вода, над которой находится смесь воздуха и насыщенных водяных паров. Начальное давление на поршень равно атмосферному (1 атм). Затем давление на поршень увеличивают в два раза. На сколько процентов изменится давление насыщенного водяного пара в цилиндре, если температура ($T = 300$ К) сохраняется неизменной?

О т в е т. $\dfrac{\Delta P_п}{P_п} = \dfrac{\mu v_ж}{RT}\Delta P = 0{,}08\,\%$.

§ 113. Уравнение Клапейрона–Клаузиуса. Испарение и конденсация. Плавление и кристаллизация

1. Перейдем к рассмотрению дальнейших следствий уравнения (112.1), выражающего условие равновесия фаз. Ради определенности будем иметь в виду процессы испарения и конденсации. Однако основные результаты, которые мы получим, без всяких изменений применимы и к другим фазовым превращениям. Состояние вещества будем изображать точкой на плоскости TP (рис. 134). Каждая точка этой плоскости соответствует однородному (однофазному) состоянию вещества — либо жидкости, либо ее пару. Исключение составляют точки линии DK. Это — линия, представляемая уравнением (112.1). На линии DK удельные термодинамические потенциалы жидкости и пара одинаковы, здесь эти фазы находятся в равновесии друг с другом. Каждая точка линии

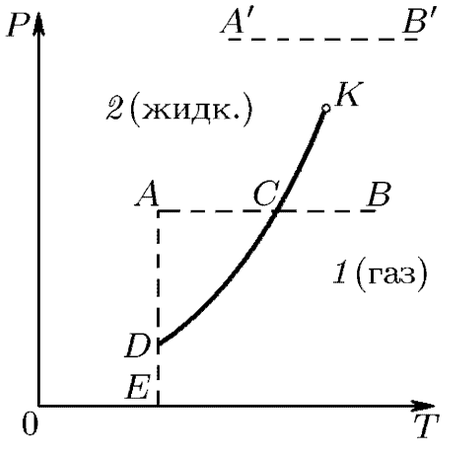


Рис. 134

DK изображает либо жидкость, либо ее пар, либо смесь этих фаз в любых пропорциях. Если решить уравнение (112.1) относительно P, то уравнение кривой DK представится в виде

$$P = P(T).$$

Это уравнение дает зависимость давления насыщенного пара от температуры. Кривая DK называется *кривой равновесия жидкости и ее насыщенного пара* или *кривой испарения*.

2. Пересечем кривую испарения горизонтальной прямой, т. е. изобарой AB. Пусть давление на изобаре AB меньше критического. В точке A вещество находится в жидком состоянии. Действительно, здесь давление $P = EA$ выше давления насыщенного пара $P = ED$ при той же температуре. Под таким давлением пар существовать не может. Он сконденсируется в жидкость. При нагревании жидкости под постоянным давлением изображающая точка перемещается вправо. В точке C пересечения изобары AB с кривой испарения DK начнется испарение жидкости. Во все время испарения температура жидкости и ее насыщенного пара будет оставаться неизменной. Когда вся жидкость испарится, изображающая точка при дальнейшем нагревании будет перемещаться по изобаре вдоль отрезка CB — этому соответствует нагревание пара при постоянном давлении. Следовательно, точки, лежащие левее кривой испарения DK, изображают жидкое состояние вещества, а точки, лежащие правее этой кривой, — газообразное состояние. Допустим теперь, что давление на изобаре выше критического, т. е. изобара $A'B'$ проходит выше критической точки K. Тогда при изобарическом нагревании или охлаждении никаких превращений жидкости в пар или обратно не произойдет. Поэтому кривая испарения DK должна оканчиваться сверху в критической точке K. В этом можно убедиться также, проведя вертикальные прямые, т. е. изотермы. Если изотерма пересекает кривую dK, то соответствующий изотермический процесс сопровождается превращениями жидкости в пар или обратно. В этом случае температура ниже критической. Если же температура выше критической, то изотерма не может пересечь кривую испарения. Значит, последняя должна оканчиваться в какой-то точке, именно в критической точке K, причем рассматриваемая изотерма проходит правее этой точки.

Следствием обрыва кривой испарения в критической точке является непрерывность жидкого и газообразного состояний вещества. Действительно, из произвольной начальной точки A можно перейти в произвольную конечную точку B так, чтобы при переходе пересечь кривую испарения. Тогда произойдет фазовое превращение. Но можно перейти в то же конечное состояние, обойдя критическую точку K без пересечения кривой испарения DK. Тогда не наступит никаких фазовых превращений. Вещество все время останется физически однородным, а его свойства будут меняться непрерывно.

3. Найдем наклон кривой испарения. Для этого вычислим производную давления насыщенного пара по температуре dP/dT. При

смещении вдоль кривой испарения (112.1) $d\varphi_1 = d\varphi_2$. Так как $d\varphi = -s\,dT + v\,dP$, то это соотношение можно записать в виде

$$v_1\,dP - s_1\,dT = v_2\,dP - s_2\,dT,$$

или

$$\frac{dP}{dT} = \frac{s_1 - s_2}{v_1 - v_2}, \tag{113.1}$$

где s_1, v_1; s_2, v_2 — удельные энтропии и удельные объемы пара и жидкости. Фазовые превращения, вообще говоря, сопровождаются скачкообразными изменениями энтропии. Это означает, что при таких превращениях поглощается или выделяется теплота. Например, при переходе единицы массы вещества из газообразного состояния *1* в жидкое *2* выделяется теплота

$$q = T(s_1 - s_2).$$

При обратном переходе из жидкого состояния *2* в газообразное *1* такая же теплота поглощается. Предполагается, что переход совершается квазистатически при постоянной температуре, а следовательно, и при постоянном давлении. Теплота q называется *удельной теплотой испарения*. В общем случае она называется *удельной теплотой фазового превращения*. Например, говорят о *теплоте плавления*, *теплоте возгонки* и пр. Смысл этих терминов не требует пояснений. Если теплоту испарения q ввести в уравнение (113.1), то получится

$$\frac{dP}{dT} = \frac{q}{T(v_1 - v_2)}. \tag{113.2}$$

Это важное соотношение называется *уравнением Клапейрона–Клаузиуса*.

Почему для испарения жидкости требуется затрата теплоты — это легко понять с молекулярной точки зрения. Скорости молекул жидкости распределены по закону Максвелла. Вылететь из жидкости в окружающее пространство могут только наиболее быстрые молекулы, так как лишь они в состоянии преодолеть силы притяжения, действующие в поверхностном слое жидкости. Проходя через поверхностный слой, молекулы замедляются, так что температура пара оказывается равной температуре жидкости (см. § 77, п. 7). В результате ухода быстрых молекул жидкость охлаждается. Для поддержания ее температуры постоянной требуется подвод теплоты. Естественно ожидать, что теплота перехода должна наблюдаться и в других фазовых превращениях, хотя механизм явления здесь не столь прост, как в случае испарения.

4. Уравнение Клапейрона–Клаузиуса легко также получить методом циклов. Проведем цикл Карно с двухфазной системой, состоящей из жидкости и ее насыщенного пара. Так как давление насыщенного пара однозначно определяется его температурой, то для такой системы изотерма $T = \text{const}$ является в то же время изобарой $P = \text{const}$. На диаграмме VP изотермы изображаются горизонтальными прямыми.

Пусть начальное состояние двухфазной системы изображается точкой *1* (рис. 135). Приведя систему в тепловой контакт с нагревателем, будем квазистатически подводить к ней теплоту. Жидкость начнет испаряться, а система совершать работу, например поднимая нагруженный поршень. Когда испарится единица массы жидкости, устраним тепловой контакт и адиабатически изолируем систему. После этого заставим ее расширяться по бесконечно короткой адиабате *23*, пока температура системы не сравняется с температурой холодильника T_2. В качестве холодильника возьмем тепловой резервуар, температура которого T_2 бесконечно мало отличается от температуры нагревателя T_1. Из состояния *3* вернем систему по изотерме *34* и адиабате *41* в исходное состояние *1*. В результате система совершит бесконечно малый цикл Карно. Количество теплоты, полученное системой от нагревателя $Q_1 = q$. На изотерме *12* система совершила положительную работу $A_1 = P(T_1)(v_1 - v_2)$, так как ее объем увеличился на $v_1 - v_2$, а на изотерме *34* — отрицательную, $A_2 = -P(T_2)\,(v_1 - v_2)$. Работой на адиабатах *23* и *41* можно пренебречь, как величиной более высокого порядка малости. Полная работа системы

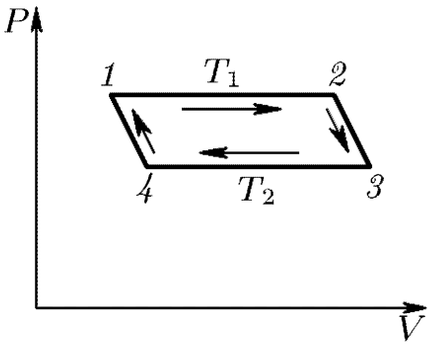


Рис. 135

$$A = A_1 + A_2 = (v_1 - v_2)[P(T_1) - P(T_2)] = \frac{dP}{dT}\,(v_1 - v_2)(T_1 - T_2).$$

По теореме Карно

$$\frac{A}{Q_1} = \frac{T_1 - T_2}{T_1}.$$

Подставляя сюда значения A и Q_1 и заменяя T_1 на T, получим уравнение (113.2).

5. Уравнение Клапейрона–Клаузиуса, как ясно из его вывода, справедливо не только для испарения, но и для других фазовых превращений, сопровождающихся выделением или поглощением теплоты, например для плавления, сублимации и пр. В случае плавления, например, можно написать

$$\frac{dP}{dT} = \frac{q_{23}}{T(v_2 - v_3)}, \tag{113.3}$$

где q_{23} — удельная теплота плавления, v_2 и v_3 — удельные объемы жидкой и твердой фаз, T — температура плавления при давлении P, величина q_{23} существенно положительна. Поэтому если $v_2 > v_3$, то $dP/dT > 0$. Это значит, что с повышением давления точка плавления повышается. Если же $v_2 < v_3$, то $dP/dT < 0$, т. е. при увеличении давления температура плавления понижается. Последний случай имеет место для воды. Разность удельных объемов льда и воды при 0 °C равна приблизительно

$$v_3 - v_2 = 9{,}1 \cdot 10^{-2}\ \text{см}^3/\text{г}.$$

Теплота плавления

$$q = 80 \text{ кал/г} = 3{,}35 \cdot 10^9 \text{ эрг/г}.$$

Используя эти данные, получаем

$$\frac{dP}{dT} = -\frac{3{,}35 \cdot 10^9}{273 \cdot 9{,}1 \cdot 10^{-2}} = -1{,}35 \cdot 10^8 \text{ дин/(см}^2 \cdot \text{К)} = 134 \text{ атм/К}.$$

Отсюда видно, что с увеличением давления на одну атмосферу точка плавления льда понижается приблизительно на 0,0075 градуса. Дьюар опытным путем нашел 0,0072 К/атм, что хорошо согласуется с вычисленным значением.

Если на брусок льда, лежащий своими концами на неподвижных опорах, накинуть проволочную петлю и к ней тяжелый груз, то лед под проволокой начинает плавиться. Вода выдавливается из-под проволоки и замерзает на ней. Проволока постепенно проходит через брусок, однако брусок остается неразрезанным.

ЗАДАЧИ

1. В закрытом сосуде при 0 °С находится моль (18 г) воды. Какое требуется количество теплоты Q, чтобы повысить температуру системы до 100 °С, если объем сосуда таков, что при этом вся вода превращается в насыщенный пар? Теплота испарения воды при 100 °С и постоянном давлении $q = 539$ кал/г. Упругостью насыщенного пара при 0 °С и теплоемкостью стенок сосуда пренебречь. Пренебречь также объемом воды по сравнению с объемом насыщенного пара.

Решение. Объем системы не меняется, а, следовательно, работа не производится. Поэтому Q равно изменению внутренней энергии системы и не зависит от способа перехода ее из начального в конечное состояние. Произведем этот переход в два этапа. 1) Введем твердую подвижную перегородку (с бесконечно малой теплоемкостью) между жидкостью и паром и нагреем систему до 100 °С. Работой расширения жидкости можно пренебречь. Вся система в целом, как сказано выше, работы не производит. Поэтому на этом этапе надо подвести теплоту $Q_1 = 18 \cdot 100 = 1800$ кал. 2) Перемещая перегородку при постоянной температуре, превратим всю воду в насыщенный пар. Работа по перемещению перегородки будет равна нулю, так как давления по обе стороны ее одинаковы. На этом этапе надо подвести теплоту $Q_2 = U_\text{п} - U_\text{ж}$, где $U_\text{п}$ и $U_\text{ж}$ — внутренние энергии пара и жидкости при 100 °С. По первому началу термодинамики $q_\text{мол} = U_\text{п} - U_\text{ж} + A$, где $q_\text{мол} = 18q = 9700$ кал/моль — молярная теплота испарения жидкости, а A — работа против постоянного внешнего давления ($A = PV_\text{п} \approx RT = 1{,}98 \cdot 373 = 739$ кал/моль). Таким образом, $Q_2 = U_\text{п} - U_\text{ж} = q_\text{мол} - A = 8960$ кал/моль, $Q = Q_1 + Q_2 = 1800 + 8960 = 10760$ кал/моль.

2. При 0 °С упругость водяного пара над льдом равна 4,58 мм рт. ст. Теплота плавления льда при 0 °С 80 кал/г, теплота испарения воды при той же температуре 596 кал/г. Найти упругость пара над льдом при температуре $t = -1$ °С.

Ответ. 4,20 мм рт. ст.

3. В следующей таблице приведены давления насыщенных паров азота при трех температурах. Пользуясь ими, вычислить удельную теплоту испарения q жидкого азота при температуре $t = -196$ °С. Считать, что газообразный азот вплоть до температуры конденсации подчиняется уравнению

Клапейрона. Удельным объемом жидкого азота по сравнению с газообразным пренебречь.

t, °C	P, мм рт. ст
−195	833
−196	741
−197	657

Ответ. $q = \dfrac{RT^2}{\mu}\left(\dfrac{1}{P}\dfrac{dP}{dT}\right) = 50$ кал/г.

4. В закрытом сосуде объемом 5 л находится 1 кг воды при температуре $T = 373$ К. Пространство над водой занято насыщенным паром (воздух выкачан). Найти увеличение массы насыщенного пара Δm при повышении температуры на $\Delta T = 1$ К.

Ответ. Если пренебречь изменением объема пара V при нагревании на ΔT, то $\Delta m = \dfrac{\mu V P}{RT^2}\left(\dfrac{q\mu}{RT} - 1\right)\Delta T = 0{,}075$ г.

5. Кусок льда помещен в адиабатическую оболочку при температуре 0 °C и атмосферном давлении. Как изменится температура льда, если его адиабатически сжать до давления $P = 100$ атм? Какая доля льда $\Delta m/m$ при этом расплавится? Удельный объем воды $v_в = 1$ см^3/г, льда $v_л = 1{,}09$ см^3/г. Теплоемкости воды и льда связаны соотношением $c_л \approx 0{,}6c_в$.

Ответ. $\Delta T \approx \dfrac{PT}{q}(v_в - v_л) = -0{,}72\,^\circ$C, $\dfrac{\Delta m}{m} = -c_л\dfrac{\Delta T}{q} = 0{,}0054$.

6. Найти выражение для молярной теплоты испарения жидкости при постоянной температуре T под давлением ее насыщенного пара в предположении, что жидкость и пар подчиняются уравнению состояния Ван-дер-Ваальса.

Ответ. $q_{мол} = V_п\left(\dfrac{RT}{V_п - b} - \dfrac{2a}{V_п^2}\right) - V_ж\left(\dfrac{RT}{V_ж - b} - \dfrac{2a}{V_ж^2}\right)$, где $V_п$ и $V_ж$ — молярные объемы пара и жидкости.

7. Показать, что вблизи абсолютного нуля касательная к кривой плавления $P = P(T)$ на диаграмме (P, T) становится горизонтальной. Точнее, $\lim\limits_{T\to 0}\dfrac{dP}{dT} = 0$. Это утверждение справедливо, когда удельные объемы твердой и жидкой фаз различны. Конкретно речь может идти о гелии II — единственном веществе, которое может оставаться жидким вплоть до абсолютного нуля температур (см. рис. 139).

Решение. Из теоремы Нернста следует, что $q = 0$. Дальнейшее вытекает из уравнения (113.2).

§ 114. Зависимость давления насыщенного пара от температуры

1. Если известны как функции температуры удельная теплота испарения q и удельные объемы v_1 и v_2, то уравнение (113.2) можно проинтегрировать и найти в явном виде зависимость давления насыщенного пара от температуры. В самом грубом приближении можно считать, что величина q не зависит от температуры, а удельным

объемом жидкости по сравнению с удельным объемом пара можно пренебречь. Кроме того, можно считать, что к пару применимо уравнение состояния Клапейрона–Менделеева $Pv = RT/\mu$. (Мы опустили индекс 1 у v). Эти упрощения допустимы, если интервал изменения температуры не слишком широк [1]. В принятом приближении уравнение (113.2) перейдет в

$$\frac{dP}{dT} = \frac{q}{Tv} = \frac{\mu q}{RT^2}\,P, \tag{114.1}$$

или

$$\frac{dP}{P} = \frac{\mu q}{R}\,\frac{dT}{T^2}.$$

Интегрируя его, получим

$$\ln P = -\frac{\mu q}{RT} + C.$$

Постоянную интегрирования C можно найти, если известно давление насыщенного пара P_0 при какой-либо одной температуре T_0. При этой температуре

$$\ln P_0 = -\frac{\mu q}{RT_0} + C.$$

Исключая постоянную C, получим

$$P = P_0 \exp \frac{\mu q}{R}\left(\frac{1}{T_0} - \frac{1}{T}\right). \tag{114.3}$$

2. Для воды при нормальном атмосферном давлении температура кипения $T = 373$ К, теплота парообразования $q = 539$ кал/(г · К). Подставляя в формулу (114.1) $R = 1{,}9858$ кал/(моль · К), $\mu = 18$ г/моль, получим

$$\frac{1}{P}\,\frac{dP}{dT} = \frac{18 \cdot 539}{1{,}986 \cdot 373^2} = 0{,}03521\ \text{К}^{-1}.$$

Отсюда следует, что при нагревании на один градус давление насыщенного водяного пара возрастает на 0,0351 атм, или 27 мм рт. ст.

Часть теплоты испарения q идет на приращение внутренней энергии системы, другая часть — на производство внешней работы A. Последняя, очевидно, $A = P(v_1 - v_2)$ или $A = Pv \approx RT/\mu$, если пренебречь удельным объемом жидкости, а водяной пар считать идеальным газом. Для отношения работы A ко всей теплоте испарения q получаем

$$\frac{A}{q} = \frac{RT}{\mu q} = 0{,}076.$$

Таким образом, на внешнюю работу тратится лишь очень небольшая часть теплоты испарения.

3. Более точную формулу давления насыщенного пара можно получить, если учесть зависимость удельной теплоты парообразования от температуры. Как уже сказано, величина q слагается из двух частей. Первая часть есть

[1] Однако вблизи критической температуры оба допущения даже приблизительно не соответствуют действительности.

разность удельных внутренних энергий пара и жидкости $u^{\text{п}} - u^{\text{ж}}$. Вторая часть есть работа против внешнего давления и равна $P(v^{\text{п}} - v^{\text{ж}})$. Таким образом,

$$q = (u^{\text{п}} - u^{\text{ж}}) + P(v^{\text{п}} - v^{\text{ж}}). \tag{114.4}$$

Пренебрежем удельным объемом жидкости и будем считать, что пар подчиняется уравнению Клапейрона. В этом приближении

$$q = (u^{\text{п}} - u^{\text{ж}}) + \frac{RT}{\mu}.$$

В частности, при какой-то фиксированной температуре $T = T_1$

$$q_1 = q(T_1) = (u_1^{\text{п}} - u_1^{\text{ж}}) + \frac{RT_1}{\mu}.$$

Внутренняя энергия пара, поскольку он считается идеальным газом, зависит только от температуры, а потому

$$u^{\text{п}} = u_1^{\text{п}} + \int_{T_1}^{T} c_v^{\text{п}}(T)\,dT.$$

Если пренебречь работой расширения жидкости при нагревании, то

$$u^{\text{ж}} = u_1^{\text{ж}} + \int_{T_1}^{T} c^{\text{ж}}(T)\,dT,$$

где $c^{\text{ж}}$ — удельная теплоемкость жидкости под давлением своих насыщенных паров (практически она равна теплоемкости жидкости при постоянном давлении). Считая в температурном интервале (T_1, T) величины $c_v^{\text{п}}$ и $c^{\text{ж}}$ постоянными и воспользовавшись для пара соотношением Майера $c_v^{\text{п}} + R/\mu = c_P^{\text{п}}$, получим

$$q = q_1 + (c^{\text{ж}} - c_P^{\text{п}})T_1 - (c^{\text{ж}} - c_P^{\text{п}})T.$$

Подставляя это значение в формулу (113.4) и пренебрегая удельным объемом v_2, получим

$$\frac{dP}{dT} = \frac{q_1 + (c^{\text{ж}} - c_P^{\text{п}})T_1}{RT^2}\,\mu - \frac{c^{\text{ж}} - c_P^{\text{п}}}{RT}\,\mu,$$

а после интегрирования

$$\ln P = -\mu\,\frac{q_1 + (c^{\text{ж}} - c_P^{\text{п}})T_1}{RT} - \mu\,\frac{c^{\text{ж}} - c_P^{\text{п}}}{R}\,\ln T + A.$$

Таким образом,

$$\ln P = A - \frac{B}{T} - C\ln T, \tag{114.5}$$

где A, B, C — постоянные. Это уравнение было получено Кирхгофом и широко используется для обработки экспериментальных данных.

4. К формуле, аналогичной (114.5), приводят и простые молекулярно-кинетические соображения. Молекула пара обладает большей потенциальной энергией, чем молекула жидкости. Пусть b означает работу, которую надо затратить против молекулярных сил, чтобы перевести молекулу из области, занятой жидкостью, в область занятую паром. По формуле Больцмана

$$n = n_0 \exp\Big(-\frac{b}{kT}\Big),$$

где n — концентрация молекул пара, а n_0 — концентрация молекул жидкости. Примем, что величина b не зависит от температуры. (Только в этом случае формула Больцмана строго справедлива.) Кроме того, пренебрежем работой расширения RT/μ по сравнению с приращением внутренней энергии $u^{\text{п}} - u^{\text{ж}}$. Тогда

$$b = \frac{u^{\text{п}} - u^{\text{ж}}}{N}\,\mu = \frac{q\mu}{N}, \quad n = n_0 \exp\Big(-\frac{\mu q}{NkT}\Big) = n_0 \exp\Big(-\frac{\mu q}{RT}\Big).$$

Для давления насыщенного пара получаем

$$P = nkT = n_0 kT \exp\Big(-\frac{\mu q}{RT}\Big),$$

откуда

$$\ln P = -\frac{\mu q}{RT} + \ln T + \text{const}. \tag{114.6}$$

Логарифм T есть медленно меняющаяся функция температуры. Если ее заменить постоянной, то (114.6) перейдет в формулу (114.3).

ЗАДАЧИ

1. В закрытом сосуде при температуре $t = 20\,^\circ\text{C}$ находится влажный воздух с относительной влажностью $f = 80\,\%$. На сколько градусов надо понизить температуру стенок сосуда, чтобы на них начала выпадать роса? Удельная теплота парообразования воды при $20\,^\circ\text{C}$ $q = 600$ кал/г. Водяной пар рассматривать как идеальный газ.

Решение. Для приближенной оценки в уравнении (114.1) заменим производную dP/dT отношением конечных приращений. Получим

$$\frac{T_2 - T_1}{P_2 - P_1} = \frac{RT_1^2}{\mu q P_1},$$

где P_1 и P_2 — давления насыщенного пара при температурах T_1 и T_2. Давление пара в воздухе при температуре T_1 и относительной влажности f будет fP_1, а потому $P_2 = (T_2/T_1)fP_1$. Подставляя эти значения в предыдущее соотношение, найдем

$$T_2 - T_1 = \frac{f-1}{\mu q - fRT_1}\,RT_1^2 = -3{,}3\ \text{K}.$$

Для нахождения более точного решения из формулы (114.2) получаем

$$\ln\frac{fT_2}{T_1} = \frac{\mu q}{RT_1T_2}\,(T_2 - T_1).$$

Подставляя числовые значения и переходя к десятичным логарифмам, преобразуем это уравнение к виду

$$T_2 - T_1 = 0{,}124\,T_2 \ln\frac{fT_2}{T_1}. \tag{A}$$

Для решения уравнения применяем метод последовательных приближений. В нулевом приближении полагаем $T_2 - T_1$. Пользуясь этим, находим первое приближение:

$$T_2 - T_1 = 0{,}124\,T_1 \lg f = -3{,}52\ \text{K}.$$

Вычислив отсюда T_2 и подставив в правую часть уравнения (А), найдем второе приближение: $T_2 - T_1 = -3{,}66705$ K. Поступая так дальше, получим

третье приближение: $T_2 - T_1 = 3{,}67313$ K, четвертое приближение: $T_2 - T_1 = -3{,}67360$ K. С точностью до трех значащих цифр $T_2 - T_1 = -3{,}67$ K. Таким образом, замена производной dP/dT отношением конечных приращений приводит к погрешности около 10 %.

2. В следующей таблице приведены значения (в мм рт. ст.) давления насыщенных паров над водой и льдом при трех температурах. Используя эти данные, вычислить удельную теплоту замерзания воды q_{23} при 0 °C.

Решение. Пренебрегая разностью значений q_{23}, q_{13}, и q_{12} в тройной точке и в точке $t = 0\,^\circ$C, $P = 1$ атм, можем написать $q_{23} = q_{13} - q_{12}$ и далее

$$q_{23} = \frac{RT_0}{\mu \Delta T}\left(T_1 \ln \frac{P_0}{P_{13}} - T_2 \ln \frac{P_{12}}{P_0}\right) \approx 81 \text{ кал/г}$$

($T_0 = 273$ K, $T_1 = 263$ K, $T_2 = 283$ K, P_0, P_{13}, P_{12} — давления насыщенных паров при этих температурах, $\Delta T = T_0 - T_1 = T_2 - T_1$).

3. В тонкостенный металлический шар радиуса $r = 10$ см, из которого выкачан воздух, налита вода. Давление воздуха вне шара равно атмосферному. До какой максимальной температуры можно нагреть воду, чтобы стенки шара не разорвались, если предельное натяжение на разрыв, которое они могут выдержать, $\sigma = 84$ Н/см? Количество воды в шаре таково, что при этой температуре еще не вся вода испаряется, однако объем воды мал по сравнению с объемом пара.

	t, °C	P
Лед	−10	1,950
	0	4,579
Вода	0	4,579
	+10	9,209

Ответ. $\dfrac{1}{T} = \dfrac{1}{T_0} - \dfrac{R}{\mu q}\ln\left(1 + \dfrac{2\sigma}{P_0 r}\right)$, $T = 404$ K, $t = 131\,^\circ$C (T_0 — температура кипения при нормальном атмосферном давлении P_0).

4. По одной из теорий гейзеры представляют собой большие подземные резервуары, наполненные грунтовой водой и прогреваемые подземным теплом. Выход из них на поверхность Земли осуществляется через узкий канал, который в «спокойный» период практически полностью заполнен водой. Считая, что «активный» период наступает, когда закипает вода в подземном резервуаре, и что во время извержения гейзера канал практически заполнен только паром, который и выбрасывается наружу, оценить, какую часть воды теряет резервуар гейзера во время одного извержения. Глубина канала, т. е. расстояние от подземного резервуара до поверхности Земли, $h = 90$ м.

Ответ. $\dfrac{\Delta m}{m} \approx \dfrac{c(T_{10} - T_1)}{q} \approx 14\,\%$ (c — удельная теплоемкость воды, T_1 и T_2 — температуры кипения воды при давлениях 1 и 10 атм соответственно).

§ 115. Теплоемкость насыщенного пара

1. Допустим, что насыщенный пар нагревается и одновременно меняется его объем таким образом, что пар все время остается насыщенным. Пусть при повышении температуры на dT пару надо сообщить

количество теплоты δQ. Отношение $\delta Q/dT$ называется *теплоемкостью насыщенного пара*. Если масса пара равна единице, то это есть *удельная теплоемкость* c; если же пар взят в количестве одного моля, то получается *молярная теплоемкость* C.

Чтобы при нагревании пар оставался насыщенным, его одновременно необходимо подвергать сжатию, так как плотность насыщенного пара с возрастанием температуры возрастает. При сжатии же происходит нагревание газа. Могут представиться три случая. 1) Теплота, выделяющаяся при сжатии, настолько значительна, что пар становится ненасыщенным (перегретым), и для сохранения его в состоянии насыщения от него надо отводить теплоту. В этом случае теплоемкость c отрицательна. 2) Теплота, выделяющаяся при сжатии, слишком мала, чтобы сжатый пар при отсутствии притока теплоты извне не сделался бы пересыщенным. Для сохранения состояния насыщения к пару требуется подводить теплоту. В этом случае теплоемкость пара c положительна. 3) Теплоты сжатия как раз достаточно, чтобы сохранить пар в состоянии насыщения без дополнительного притока или отвода теплоты. В этом случае $c = 0$.

2. Вычислим теперь удельную теплоемкость насыщенного пара. Первое начало термодинамики для единицы массы пара можно записать в виде $\delta Q = di^{\text{п}} - v\,dP$, где $i^{\text{п}}$ — удельная энтальпия, а v — удельный объем пара. Мы применяем это уравнение к процессу, в котором P не остается постоянным. Однако если пар считать идеальным газом, то его энтальпия будет зависеть только от температуры. Тогда для любого квазистатического процесса $di^{\text{п}}/dT = c_P^{\text{п}}$. Поэтому для искомой теплоемкости насыщенного пара получаем $c = c_P^{\text{п}} - -v\,dP/dT$. Поскольку нагревание производится так, что пар все время остается насыщенным, производная dP/dT определяется формулой (114.1), пользуясь которой, получаем

$$c = c_P^{\text{п}} - q/T. \tag{115.1}$$

Согласно классической теории молярная теплоемкость водяного пара при постоянном давлении равна 8 кал/(моль $\cdot$ К), а удельная теплоемкость $c_P^{\text{п}} = 8/18 = 0{,}444$ кал/(г $\cdot$ К). Теплота парообразования при $T = 373$ К для воды $q = 539$ кал/г. Пользуясь этими данными получаем по формуле (115.1) $c = -1$ кал/(г $\cdot$ К). Теплоемкость c отрицательна. Значит, при *адиабатическом расширении насыщенного водяного пара он охлаждается и становится пересыщенным.*

3. Формула (115.1) выведена в предположении, что пар ведет себя как идеальный газ, т. е. подчиняется уравнению Клапейрона. Выведем теперь более точную формулу, не вводя этого предположения. При выводе, однако, будем по-прежнему пренебрегать удельным объемом жидкости по сравнению с удельным объемом пара. Как и раньше, для единицы массы пара можно написать $\delta Q = di^{\text{п}} - v^{\text{п}}\,dP$. Отсюда

$$c = \frac{\delta Q}{dT} = \frac{di^{\text{п}}}{dT} - v^{\text{п}}\,\frac{dP}{dT}.$$

Производные $di^{\text{п}}/dT$ и dP/dT должны вычисляться при условии, что пар при нагревании все время остается насыщенным. Подставляя для последней производной ее значение из формулы (114.1), получим

$$c = \frac{di^{\text{п}}}{dT} - \frac{q}{T}.$$

Для вычисления производной $di^{\text{п}}/dT$ пользуемся формулой

$$q = u^{\text{п}} - u^{\text{ж}} + P(v^{\text{п}} - v^{\text{ж}}) = i^{\text{п}} - i^{\text{ж}}.$$

Поскольку это соотношение написано для процесса, в котором пар все время остается насыщенным, величины, входящие в него, могут зависеть только от температуры. Дифференцируя его по температуре, находим

$$\frac{di^{\text{п}}}{dT} = \frac{dq}{dT} + \frac{di^{\text{ж}}}{dT}.$$

Для жидкости $di^{\text{ж}} = c_P^{\text{ж}}\,dT + v^{\text{ж}}\,dP$, причем последним слагаемым можно пренебречь. В этом приближении $di^{\text{ж}}/dT = c_P^{\text{ж}}$, а потому

$$\frac{di^{\text{п}}}{dT} = c_P^{\text{ж}} + \frac{dq}{dT}.$$

Подставляя эту величину в выражение для c, получим

$$c = c_P^{\text{ж}} - \frac{q}{T} + \frac{dq}{dT}. \tag{115.2}$$

Для воды при $T = 373$ К

$$q = 539 \text{ кал/г}, \quad \frac{dq}{dT} = -0{,}64 \text{ кал/(г} \cdot \text{К)}, \quad c_P^{\text{ж}} = 1{,}01 \text{ кал/(г} \cdot \text{К)}.$$

Используя эти значения, найдем $c = -1{,}07$ кал/(г · К), что отличается от ранее полученного значения на 7 %.

ЗАДАЧИ

1. Определить изменение энтропии системы, состоящей из воды и насыщенного пара, при переходе ее в насыщенный пар. Начальная температура системы T_1, конечная T_2. Начальная масса пара m_1, конечная m_2. Зависимостью удельной теплоты парообразования воды q от температуры пренебречь. Пар рассматривать как идеальный газ.

Решение. Квазистатически и изотермически испарим жидкость при температуре T_1. Изменение энтропии в этом процессе

$$\Delta_1 S = \frac{q}{T_1}(m_2 - m_1).$$

Затем будем квазистатически менять температуру пара и притом так, чтобы он все время оставался насыщенным. Элементарное количество теплоты, которое требуется подводить к пару в этом процессе,

$$\delta Q = m_2 c\,dT = m_2\left(c_P^{\text{п}} - \frac{q}{T}\right)dT.$$

Так как $dS = \delta Q/T$, то, интегрируя и пренебрегая при этом зависимостью q от T, найдем для соответствующего изменения энтропии

$$\Delta_2 S = m_2\Big[c_P^{\text{п}} \ln \frac{T_2}{T_1} + q\Big(\frac{1}{T_2} - \frac{1}{T_1}\Big)\Big].$$

2. Найти выражение для удельной энтропии насыщенного пара, рассматривая его как идеальный газ и пренебрегая зависимостью q от температуры.

Ответ. $s = c_P^{\text{п}} \ln T + q/T + \text{const}$.

§ 116. Тройные точки. Диаграммы состояния

1. Допустим теперь, что число фаз химически однородного вещества, находящихся в равновесии друг с другом, равно трем. Примером может служить система, состоящая из твердой фазы, жидкости и ее пара. Для равновесия необходимо выполнение трех условий:

$$\begin{aligned} \varphi_1(P,T) &= \varphi_2(P,T), \\ \varphi_2(P,T) &= \varphi_3(P,T), \\ \varphi_3(P,T) &= \varphi_1(P,T). \end{aligned} \tag{116.1}$$

Первое — условие равновесия между жидкостью и ее паром; второе — между жидкостью и твердой фазой; третье — между твердой фазой и паром. Эти три условия не независимы. Каждое из них является следствием двух остальных. Первое уравнение системы (116.1) изображает на плоскости кривую равновесия между газом и жидкостью, т. е. *кривую испарения 12* (рис. 136). Второе изображает кривую равновесия твердой и жидкой фаз *23*. Она называется *кривой плавления*. Кривая плавления пересекается с кривой испарения в точке A, называемой *тройной точкой*. Через тройную точку должна проходить и *кривая возгонки 31*, т. е. кривая равновесия между твердой и газообразной фазами. Это непосредственно следует из третьего уравнения (116.1). Таким образом, *три фазы могут находиться в равновесии друг с другом, вообще говоря, лишь в одной, а именно тройной точке, т. е. при вполне определенных значениях температуры и давления.*

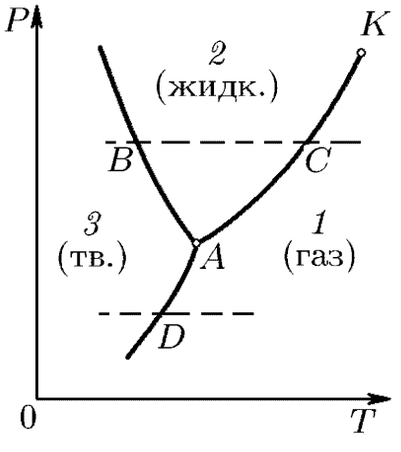


Рис. 136

В тройной точке кривая возгонки *31* поднимается круче кривой испарения *12*. В самом деле, наклоны этих кривых определяются уравнениями Клапейрона–Клаузиуса:

$$\frac{dP_{12}}{dT} = \frac{q_{12}}{T(v_1 - v_2)}, \qquad \frac{dP_{13}}{dT} = \frac{q_{13}}{T(v_1 - v_3)}.$$

Знаменатели этих выражений практически совпадают, так как удельными объемами жидкой и твердой фаз можно пренебречь. Кроме того,

в тройной точке, как это следует из первого начала термодинамики, $q_{13} = q_{12} + q_{23}$. Значит, $q_{13} > q_{12}$, а потому $dP_{13}/dT > dP_{12}/dT$.

Для воды тройная точка лежит примерно на 0,008 °С выше точки плавления при нормальном атмосферном давлении. Давление в тройной точке составляет около 4,58 мм рт. ст. Тройная точка воды является основной реперной точкой при построении абсолютной термодинамической шкалы температур Кельвина, а также практической международной шкалы температур Цельсия (см. § 4, 6, 31).

2. Кривые испарения, плавления и возгонки делят плоскость TP на три области *1*,*2*,*3* (см. рис. 136). Точкам области *1* соответствует газообразное, области *2* — жидкое, области *3* — твердое состояние вещества. Плоскость TP с указанными тремя кривыми равновесия называется *диаграммой состояния*. Диаграмма состояния позволяет судить, какие будут происходить фазовые превращения при том или ином процессе. Допустим, например, что производится нагревание при постоянном давлении. На диаграмме состояния такой процесс представляется горизонтальной прямой. Если эта прямая проходит выше тройной, но ниже критической точки, то в точке B она пересечет кривую плавления, а в точке C — кривую испарения. Значит, при нагревании твердое тело сначала расплавится (точка B), а затем жидкость испарится (точка C). Если же указанная прямая проходит ниже тройной точки, то она пересечет только кривую возгонки в некоторой точке D, в которой и произойдет непосредственное превращение твердого тела в газообразное состояние. Промежуточного состояния не будет. В § 104, п. 4 был описан опыт, в котором газообразная углекислота охлаждалась при дросселировании и непосредственно переходила в твердое состояние. Получившаяся твердая углекислота не плавится, а непосредственно превращается в газ (сублимируется). Явление объясняется тем, что давление в тройной точке углекислоты выше атмосферного давления, при котором производится опыт. Для углекислоты температура в тройной точке равна −56,6 °С, а давление 4,8 атм. Поэтому при атмосферном давлении углекислота не может находиться в жидком состоянии. Для этого требуется повысить давление.

3. Поставим теперь вопрос о возможности сосуществования четырех или большего числа фаз химически однородного вещества. В случае четырех фаз для равновесия необходимо выполнение шести уравнений типа (116.1), из которых, однако, независимы только три. Геометрически задача сводится к нахождению общей точки пересечения трех кривых равновесия фаз, выражаемых, например, уравнениями

$$\varphi_1(P,T) = \varphi_2(P,T), \quad \varphi_2(P,T) = \varphi_3(P,T), \quad \varphi_3(P,T) = \varphi_4(P,T).$$

Но три кривые пересекаются, вообще говоря, в трех, а не в одной точке. Пересечение в одной точке является исключительным случаем, с которым практически можно не считаться. Физически это означает, что *четыре или большее число фаз химически однородного вещества не*

могут находиться в равновесии между собой ни при каких давлениях и температурах. Максимальное число фаз, находящихся в равновесии друг с другом, не может превышать трех.

4. Если число фаз, в которых может находиться химически однородное вещество при всевозможных значениях температуры и давления, превышает три, то все равновесные состояния системы можно также изобразить диаграммой на плоскости TP, называемой попрежнему *диаграммой состояния*. Плоскость TP разбивается на ряд областей. Каждая точка плоскости TP, если она не лежит на границе области, изображает однофазное состояние вещества. Области граничат между собой вдоль кривых, каждая из которых является *кривой равновесия* соответствующих двух фаз. Всякая точка, лежащая на кривой равновесия, изображает двухфазное состояние вещества, причем в этом состоянии фазы могут быть представлены в любых пропорциях. Кривые равновесия фаз могут пересекаться по три в отдельных точках. Это тройные точки, в которых находятся в равновесии три граничащие друг с другом фазы.

На диаграмме состояния сразу видно, в каких равновесных состояниях может находиться вещество при тех или иных значениях температуры и давления, а также когда и какие оно будет испытывать фазовые превращения при том или ином процессе. В качестве примера на рис. 137 представлена в упрощенном виде диаграмма состояния серы. Сера может существовать в двух кристаллических модификациях — моноклинной и ромбической. В соответствии с этим на диаграмме состояния имеются три тройных точки, а именно S, T, L. Область моноклинной модификации ограничена треугольником STL. Область ромбической модификации лежит выше кривой $GSLF$. Возьмем ромбические кристаллы серы при комнатной температуре и нормальном давлении и будем нагревать их, сохраняя давление постоянным. Этот процесс изобразится горизонтальной прямой MN_1N_2Q. В точке N_1, где эта прямая пересекает кривую равновесия между двумя кристаллическими модификациями, ромбические кристаллы превращаются в моноклинные (при атмосферном давлении точке N_1 соответствует температура 95,5 °C). В точке N_2 (при температуре 119,2 °C) моноклинные кристаллы плавятся. В точке P (при температуре 444,60 °C) жидкая сера закипает. При охлаждении вещества те же превращения будут происходить в обратном порядке. Если ромбические кристаллы взять в состоянии M' под давлением выше давления в тройной точке L (1280 атм), то изобара $M'N'$ пройдет выше этой точки. Поэтому

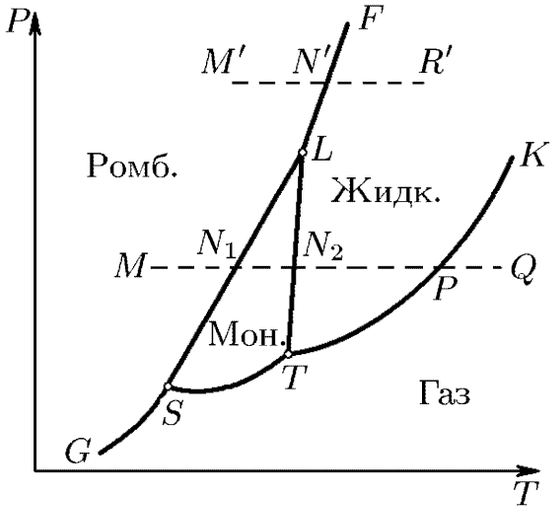


Рис. 137

превращения ромбических кристаллов в моноклинные происходить не будет. Ромбические кристаллы в точке N' будут сразу плавиться.

5. Закончим этот параграф следующим замечанием. Кривая испарения, как мы видели, оканчивается в критической точке. Только благодаря этому возможен непрерывный переход вещества из жидкого состояния в газообразное или обратно, т. е. такой переход, который не сопровождается фазовыми превращениями. Это связано с тем, что различие между газом и жидкостью является чисто *количественным*. Газ и жидкость отличаются друг от друга только большей или меньшей ролью взаимодействия между молекулами. Но оба эти состояния *изотропны* и характеризуются *одинаковой симметрией* внутреннего строения. Совсем другой характер имеет различие между кристаллической и жидкой (газообразной) фазами или между двумя различными кристаллическими фазами. Эти фазы отличаются друг от друга не только количественно, но и *качественно*, а именно *симметрией внутреннего строения*. О всяком же свойстве симметрии можно сказать, что оно либо есть, либо его нет. То или иное свойство симметрии может появиться или исчезнуть только сразу, скачком, а не непрерывно. В каждом состоянии тело будет обладать либо одной, либо другой симметрией, а потому можно указать, к которой из двух фаз оно относится. Кривая равновесия таких фаз поэтому не может обрываться в изолированной (критической) точке. Она может либо заканчиваться в точке пересечения ее с другой кривой равновесия, либо уходить в бесконечность.

ЗАДАЧА

Показать, что в тройной точке (рис. 137) справедливо соотношение

$$(v_1 - v_2)\frac{dP_{12}}{dT} + (v_2 - v_3)\frac{dP_{23}}{dT} + (v_3 - v_1)\frac{P_{31}}{dT} = 0.$$

§ 117. Кипение и перегревание жидкости

1. Если жидкость в сосуде нагревать при постоянном внешнем давлении, то сначала образование пара носит спокойный характер. Оно идет лишь со свободной поверхности жидкости. Такой процесс парообразования называется *испарением*. По достижении определенной температуры, называемой *температурой кипения*, образование пара начинает происходить не только со свободной поверхности, но и изнутри жидкости. Внутри жидкости возникают, растут и поднимаются на поверхность пузыри пара, увлекая за собой и саму жидкость. Процесс парообразования приобретает бурный, неспокойный характер. Это явление называется *кипением*.

По существу кипение есть особый вид испарения. Дело в том, что жидкость никогда не бывает физически однородной. В ней всегда имеются пузырьки воздуха или других газов, но часто настолько малые, что они невидимы невооруженным глазом. На поверхности

каждого пузырька непрерывно идет испарение жидкости и конденсация пара, пока не наступит состояние динамического равновесия, в котором эти два противоположно направленные процесса компенсируют друг друга. В состоянии механического равновесия сумма давлений воздуха и пара внутри пузырька должна равняться внешнему давлению вне пузырька. Последнее слагается из давления атмосферы и гидростатического давления окружающей жидкости. Если нагреть жидкость до такой температуры, чтобы давление насыщенного пара превзошло давление вне пузырька, то пузырек начнет расти за счет испарения жидкости с его внутренней поверхности и подниматься вверх под действием архимедовой подъемной силы. Двухфазная система — жидкость с воздушными пузырьками становится механически неустойчивой, и начинается процесс кипения. Граница устойчивости определяется такой температурой, при которой давление насыщенного пара равно сумме атмосферного и гидростатического давления на рассматриваемой высоте. Это и есть температура кипения.

2. В отличие от температуры тройной точки, которая для всякого вещества является вполне определенной величиной, температура кипения жидкости зависит от внешнего давления. Она повышается при увеличении внешнего давления и понижается при уменьшении. Так, воду можно заставить кипеть при комнатной температуре. Для демонстрации стеклянную колбу с водопроводной водой помещают под колпак воздушного насоса. При откачке воздуха давление на поверхность воды понижается, и при достижении определенной степени разрежения вода закипает. Теплота, необходимая для превращения жидкости в пар, заимствуется у самой жидкости, поэтому она охлаждается. При продолжительной откачке вода может замерзнуть. Для ускорения процесса замерзания воду наливают в мелкое блюдце, чтобы увеличить свободную поверхность, с которой происходит испарение. Для той же цели под колпак воздушного насоса помещается крепкий раствор серной кислоты, поглощающий водяные пары. После одной-двух минут откачки вода в блюдце замерзает.

Понижение температуры кипения жидкости при уменьшении внешнего давления можно демонстрировать и без воздушного насоса. Берется круглодонная колба среднего размера, наполненная наполовину водопроводной водой. Вода в колбе кипятится в течение примерно 15 минут, чтобы образовавшиеся водяные пары вытеснили из колбы воздух. Затем колба снимается, быстро закупоривается каучуковой пробкой, переворачивается вверх дном и помещается на кольцеобразную подставку. Если колбу сверху поливать холодной водой, то часть водяных паров конденсируется в жидкость, давление на поверхность воды уменьшается, и она закипает.

3. Из изложенного следует, что кипение возможно только тогда, когда внутри жидкости имеются пузырьки газа. Если же таковых нет, т. е. жидкость вполне физически однородна, то парообразование внутри жидкости, т. е. кипение, становится невозможным. Такую жидкость можно нагреть выше температуры кипения. Физически однородная

жидкость, температура которой при заданном внешнем давлении превосходит температуру кипения, называется *перегретой*. Можно сказать иначе. Перегретой называется жидкость, находящаяся под давлением ниже давления ее насыщенных паров при заданной температуре. На изотерме Ван-дер-Ваальса (см. рис. 133) перегретая жидкость изображается точками участка LB, так как давление жидкости на этом участке ниже давления на изотерме — изобаре LCG, где оно равно давлению насыщенного пара. Перегретая жидкость *метастабильна*, или малоустойчива. Пока нет зародышей более устойчивой парообразной фазы, перегретая жидкость может существовать как физически однородное тело. Однако при наличии таких зародышей, например пузырьков воздуха, она становится неустойчивой и переходит в более устойчивое при данной температуре состояние — пар.

Перегретую воду можно получить, например, в кварцевой колбе с гладкими стенками. Колба тщательно промывается сначала серной, азотной или какой-либо другой кислотой, а затем дистиллированной водой. В промытую колбу наливается дистиллированная вода, из которой продолжительным кипячением удаляется растворенный в ней воздух. После этого воду в колбе можно нагреть на газовой горелке до температуры, значительно превышающей температуру кипения, и тем не менее она не будет кипеть, а только интенсивно испаряться со свободной поверхности. Лишь изредка на дне колбы образуется пузырек пара, который быстро растет, отделяется от дна и поднимается на поверхность жидкости, причем размеры его при поднятии сильно возрастают. Затем вода длительное время остается спокойной. Если в такую воду ввести зародыши газообразной формы, например бросить щепотку чая, то она бурно закипает, а ее температура быстро понижается до температуры кипения. Это эффективный опыт носит характер взрыва. Для успеха опыта важно, чтобы стенки колбы были гладкими. Всякие шероховатости и острые края способствуют образованию зародышей газообразной фазы. От них непрерывно отделяются и поднимаются на поверхность воды пузырьки пара — вода кипит со дна или стенки колбы, перегревание ее получить трудно и даже совсем невозможно.

Возникает, однако следующий вопрос. Сколько бы ни очищали воду от растворенного в ней воздуха, последний всегда остается в каком-то, хотя и ничтожном, количестве в виде мельчайших пузырьков. Если даже воду полностью очистить от растворенных в ней газов, то в ней все же могут возникать пузырьки пара флуктуационного происхождения. Почему же в таком случае воду еще можно перегреть? Ответ на этот вопрос будет дан в двух следующих параграфах.

ЗАДАЧА

Стакан наполнен водой до высоты 10 см. На дне его лежат капиллярные трубки, запаянные с одного конца и заполненные воздухом. Когда вода кипит, на открытых концах капилляров образуются пузырьки пара, диаметр которых в момент отрыва равен 0,2 мм. Чему равна температура воды на дне сосуда во время кипения, если атмосферное давление равно 760 мм рт. ст.?

Поверхностное натяжение кипящей воды 57 дин/см, а упругость водяного пара вблизи 100 °C возрастает на 2,7 мм рт. ст. при повышении температуры на 1 К.

Ответ. 100,59 °C.

§ 118. Зависимость давления насыщенного пара от кривизны поверхности жидкости

1. Допустим, что в сообщающихся сосудах находится жидкость. Один из сосудов возьмем узким, а другой широким (рис. 138 *а*). Поверхность жидкости AB в широком сосуде можно считать плоской. В узком сосуде жидкость поднимется, если она смачивает стенки, и

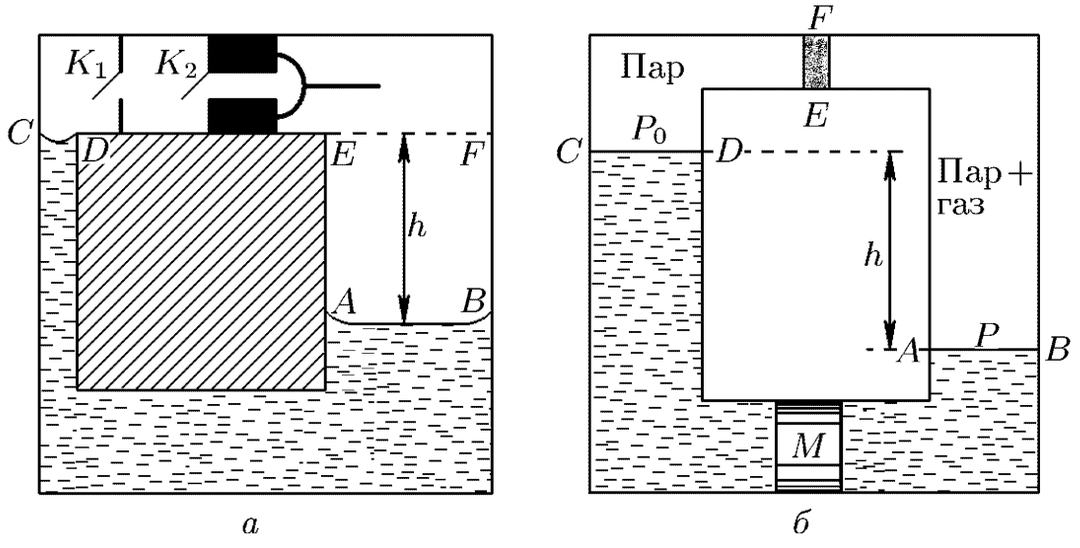


Рис. 138

опустится в противоположном случае. Проведем рассуждения в предположении, что жидкость смачивает стенки. Случай несмачивания может быть рассмотрен совершенно так же. Соединим оба эти сосуда сверху цилиндрической трубкой. Затем поместим всю систему в термостат, поддерживающий ее температуру постоянной. В состоянии равновесия давление насыщенного пара на одной и той же высоте должно быть одинаковым. Если бы это было не так, то под действием разности давлений пар пришел бы в движение, и этим можно было бы воспользоваться для построения перпетуум мобиле второго рода. Действительно, допустим, например, что давление насыщенного пара над вогнутой поверхностью жидкости CD больше, чем над горизонтальной плоскостью EF, находящейся на том же уровне. Воспользуемся приспособлением, устройство которого ясно из рис. 138 *а*. Пусть в начальный момент клапан K_1 открыт, а клапан K_2 закрыт. По предположению давление на поршень слева больше давления справа. Под действием этой разности давлений поршень будет перемещаться вправо, и его можно заставить совершать работу, например поднимать груз. Когда поршень достигает крайнего правого положения, закроем клапан K_1 и откроем клапан K_2. Произойдет неравновесный процесс

выравнивания давлений пара по обе стороны поршня. После того как он закончится, вернем поршень с открытым клапаном K_2 в исходное положение. Поскольку давления по разные стороны поршня одинаковы, на это не потребуется дополнительной работы. Затем закроем клапан K_2 и откроем клапан K_1. После наступления равновесия вся система вернется в начальное состояние. Результатом описанного кругового процесса является положительная работа, совершенная системой за счет теплоты, заимствованной от термостата. Но это есть процесс Томсона–Планка, противоречащий второму началу термодинамики. Получившееся противоречие и доказывает, что давление насыщенного пара над вогнутой поверхностью CD не может быть больше давления на уровне плоскости EF. Точно так же мы пришли бы к противоречию со вторым началом термодинамики, если бы допустили, что давление пара над поверхностью CD меньше давления над плоскостью EF. Следовательно, оба давления должны быть одинаковы.

Но благодаря действию силы тяжести давление пара на уровне плоскости EF, а следовательно, и равное ему давление P над поверхностью жидкости CD меньше давления пара P_0 над плоской поверхностью жидкости AB. Стенки сосуда сами по себе, конечно, не могут влиять на величину давления насыщенного пара. Непосредственная причина изменения давления насыщенного пара есть искривление поверхности жидкости. Следовательно, мы приходим к заключению, что *давление насыщенного пара над вогнутой поверхностью жидкости должно быть меньше, чем над плоской поверхностью.* Рассуждая аналогично, найдем, что *давление насыщенного пара над выпуклой поверхностью жидкости больше, чем над плоской.*

2. Разность давлений определяется выражением $P_0 - P = \rho_{\text{п}} g h$, где $\rho_{\text{п}}$ — плотность пара, а h — разность уровней жидкости в сообщающихся сосудах. При этом мы пренебрегли изменением плотности пара с высотой. Введя вместо плотности удельный объем пара $v_{\text{п}} = 1/\rho_{\text{п}}$, получим

$$P = P_0 - \frac{gh}{v_{\text{п}}}. \tag{118.1}$$

Остается найти величину h. Обозначим через P' давление внутри жидкости под поверхностью CD. Тогда $P_0 - P' = \rho_{\text{ж}} g h = gh/v_{\text{ж}}$, где $\rho_{\text{ж}}$ и $v_{\text{ж}}$ — плотность и удельный объем жидкости. По формуле Лапласа $P' - P = \sigma K$, где K — кривизна поверхности CD. Она считается положительной для выпуклой и отрицательной для вогнутой поверхности. Исключая P', находим

$$gh = v_{\text{ж}}(P_0 - P - \sigma K).$$

Подставляя это значение в (118.1), получим

$$P = P_0 + \frac{v_{\text{ж}}}{v_{\text{п}} - v_{\text{ж}}}\,\sigma K. \tag{118.3}$$

Эта формула называется *формулой Вильяма Томсона.* Она справедлива не только в том случае, когда радиусы кривизны R_1 и R_2 имеют одинаковые знаки, т. е. для выпуклых и вогнутых поверхностей,

но также и тогда, когда эти знаки разные. Для доказательства достаточно в нашем рассуждении заменить узкий сосуд с однородными стенками сосудом, стенки которого сделаны из разных материалов — смачиваемых и несмачиваемых жидкостью.

3. Когда кривизна очень велика, уже нельзя пренебрегать изменением плотности пара с высотой. В этом случае вместо формулы (118.1) надо пользоваться барометрической формулой

$$P = P_0 \exp\Big(-\frac{\mu g h}{RT}\Big). \tag{118.4}$$

Однако изменениями плотности жидкости с высотой по-прежнему можно пренебречь. Поэтому, исключая gh из (118.2) и (118.4), получим

$$\ln \frac{P_0}{P} = \frac{\mu v_{\text{ж}}}{RT}\,(P_0 - P - \sigma K). \tag{118.5}$$

Если $|P - P_0| \ll P_0$, то формула (118.5) переходит в более простую формулу (118.3). В этом легко убедиться с помощью формулы $\ln(1+x) = x$ (при $|x| \ll 1$), если еще воспользоваться соотношением $v_{\text{п}} = RT/\mu P$.

В другом предельном случае, когда соблюдаются условия

$$|\sigma K| \gg |P - P_0|, \tag{118.6}$$

$$\frac{\mu v_{\text{ж}}}{RT}\,|P - P_0| \ll 1, \tag{118.7}$$

формула (118.5) переходит в

$$P = P_0 \exp\Big(\frac{\mu v_{\text{ж}}}{RT}\,\sigma K\Big). \tag{118.8}$$

Для примера вычислим давление насыщенного водяного пара над поверхностью сферической капли воды с радиусом $R = 10^{-5}$ см (капелька тумана) при температуре 20 °C. При такой температуре для воды $\sigma = 72{,}7$ дин/см, $v_{\text{ж}} = 1{,}002$ см^3/г, $P_0 = 17{,}5$ мм рт. ст. $= 2{,}34 \cdot 10^4$ дин/см. Для удельного объема водяного пара получаем $v_{\text{п}} = RT/\mu P_0 = 5{,}77 \times 10^4$ см^3/г. Кривизна $K = 2/R = 2 \cdot 10^5$ см$^{-1}$. Подставляя эти данные в формулу (118.3), найдем

$$P - P_0 = 252 \text{ дин/см}^2 \approx 0{,}19 \text{ мм рт. ст.}$$

Условие $|P - P_0| \ll P_0$ в рассматриваемом случае хорошо выполняется, чем и оправдывается применение формулы (118.3). Таким образом, давление насыщенного пара над поверхностью капельки тумана превышает давление над плоской поверхностью примерно на 1 %.

Рассмотрим теперь капельку воды с радиусом $R = 10^{-7}$ см. Для такой капли формула (118.3) неприменима и надо пользоваться формулой (118.8). Она дает $P/P_0 \approx 2{,}9$. Давление над выпуклой поверхностью превышает соответствующее давление над плоской поверхностью примерно в три раза. Впрочем, на приведенный расчет следует смотреть как на оценочный, так как радиус капли того же порядка, что и радиус действия молекулярных сил. При таких условиях следовало бы учитывать зависимость поверхностного натяжения σ от кривизны поверхности, чего мы не делали.

4. Изложенный метод можно применять и для решения задач, аналогичных разобранной в этом параграфе. Допустим, например, что поверх-

ность жидкости находится под давлением $П$ какого-либо нейтрального газа. Исследуем зависимость давления насыщенного пара жидкости P от величины $П$. С этой целью возьмем сообщающиеся сосуды, изображенные на рис. 138 *б*. Пусть полупроницаемая перегородка EF свободно пропускает молекулы пара, но не пропускает молекулы нейтрального газа. Поршень M может свободно перемещаться вправо и влево. Он не позволяет газу, растворяющемуся в левом сосуде, переходить в правый. Вся система помещена в термостат, температура которого поддерживается постоянной. Если бы давления насыщенного пара по разные стороны перегородки EF были разными, то можно было бы осуществить перпетуум мобиле второго рода. Рассуждая, как раньше, можем написать

$$П + P - P_0 = \frac{gh}{v_ж}, \quad P - P_0 = \frac{gh}{v_п},$$

где P_0 — давление насыщенного пара над поверхностью жидкости CD, P — над поверхностью AB, подвергающейся давлению $П$. Исключая gh, получим

$$P = P_0 + \frac{v_п}{v_п - v_ж} П.$$

Эта формула вполне аналогична формуле (118.3). Если при выводе пользоваться барометрической формулой, то получится выражение типа (118.5). Повышение давления насыщенного пара с увеличением внешнего давления $П$ можно объяснить следующим образом. При возрастании $П$ возрастает противодавление жидкости, а с ним и число молекул жидкости, ударяющихся о ее поверхность. Значит, должно возрасти и число молекул, переходящих из жидкости в пар.

ЗАДАЧИ

1. Туман состоит из капелек воды с радиусом $R = 0{,}0005$ мм. На сколько должен быть пересыщен водяной пар в окружающем пространстве, температура которого 10 °С, чтобы капельки находились в равновесии с паром? Упругость пара, насыщающего пространство при 10 °С, равна 9,2 мм рт. ст. Поверхностное натяжение $\sigma = 70$ дин/см.

Ответ. $P - P_0 = \dfrac{\rho_п}{\rho_ж}\dfrac{2\sigma}{R} = 0{,}02$ мм рт. ст.

2. Капля воды с радиусом $r = 2$ мм находится в атмосфере насыщенного водяного пара при температуре $t = 20$ °С. Сколько молекул в начальный момент времени испаряется с поверхности капли в одну секунду? Плотность насыщенного водяного пара при 20 °С $\rho = 1{,}7 \cdot 10^{-5}$ г/см^3, поверхностное натяжение $\sigma = 72{,}5$ дин/см.

Ответ. $n = 4\pi\sigma r N_A \dfrac{\rho_п}{\rho_ж}\sqrt{\dfrac{2}{\pi\mu RT}} = 2{,}2 \cdot 10^{15}$ с$^{-1}$. (Здесь N_A — постоянная Авогадро, μ — молекулярная масса воды. Мы пренебрегли плотностью пара по сравнению с плотностью жидкости.)

3. Найти время испарения сферической водяной капли с начальным радиусом $a_0 = 1$ мм в воздухе с относительной влажностью $f = 40\,\%$ при температуре $t = 20$ °С. Плотность насыщенного водяного пара над плоской поверхностью при этой температуре $\rho_{нас} = 1{,}7 \cdot 10^{-5}$ г/см, коэффициент диффузии пара $D = 0{,}22$ см^2/с.

Решение. Масса пара, ежесекундно диффундирующая через сферическую поверхность радиуса r, концентрическую с поверхностью капли,

равна $m = -D \cdot 4\pi r^2 \, d\rho/dr$, где ρ — плотность пара. Если считать процесс стационарным, то эта величина не будет зависеть от радиуса r. Это приводит к уравнению $r^2 \, d\rho/dr = -A$, где $A = m/(4\pi D)$. После интегрирования получим $\rho = A/r + \rho_\infty$, где ρ_∞ — плотность пара на бесконечном расстоянии от капли. Величина A найдется из требования, что при $r = a$ (a — радиус капли, меняющийся во времени) пар должен быть насыщенным. Это дает

$$\rho = (\rho_{\text{нас}} - \rho_\infty)\frac{a}{r} + \rho_\infty, \tag{118.9}$$

$$m = 4\pi D a(\rho_{\text{нас}} - \rho_\infty). \tag{118.10}$$

Подставляя последнее выражение в уравнение $m = -\dfrac{4\pi}{3}\,\rho_{\text{ж}}\,\dfrac{da^3}{dt}$, получим

$$a\,\frac{da}{dt} = -\frac{D}{\rho_{\text{ж}}}\,(\rho_{\text{нас}} - \rho_\infty).$$

Пренебрегая зависимостью $\rho_{\text{нас}}$ от кривизны поверхности капли, найдем после интегрирования

$$a^2 = -\frac{2D(\rho_{\text{нас}} - \rho_\infty)}{\rho_{\text{ж}}}\,t + a_0^2. \tag{118.11}$$

Капля исчезает за время

$$t = \frac{\rho_{\text{ж}} a_0^2}{2D(\rho_{\text{нас}} - \rho_\infty)} = \frac{\rho_{\text{ж}} a_0^2}{2D(1-f)\rho_{\text{нас}}} \approx 37 \text{ мин}.$$

4. Решить предыдущую задачу для капли с начальным радиусом $a_0 = 0{,}1$ мм в предположении, что воздух насыщен водяными парами.

Р е ш е н и е. Введя в формулу (118.10) вместо плотности пара его давление, получим

$$m = \frac{4\pi D a \mu}{RT}\,[P_{\text{нас}}(a - P_\infty)].$$

По формуле Томсона (118.3), если в ней в знаменателе пренебречь $v_{\text{ж}}$, получим

$$P_{\text{нас}} - P_\infty = \frac{2\sigma}{a}\,\frac{\rho_{\text{п}}}{\rho_{\text{ж}}} = \frac{2\sigma}{a}\,\frac{\rho_\infty}{\rho_{\text{ж}}}.$$

Это дает

$$m = \frac{8\pi D \mu \sigma \rho_\infty}{RT\rho_{\text{ж}}}.$$

Величина m не зависит от размеров капли, а потому

$$t = \frac{(4\pi/3)\rho_{\text{ж}} a_0^3}{m} = \frac{RT\rho_{\text{ж}}^2 a_0^3}{6D\mu\sigma\rho_\infty} \approx 225 \text{ часов}.$$

§ 119. Метастабильные состояния

1. Теперь легко понять, почему можно перегреть жидкость даже при наличии в ней пузырьков какого-либо газа или пара самой жидкости. Для перегревания необходимо только, чтобы пузырьки были достаточно малы. Допустим, что пузырек настолько мал, что давление насыщенного пара внутри него значительно ниже соответствующего

давления пара над плоской поверхностью при той же температуре. Если пузырек состоит из пара, то он будет сжат гидростатическим давлением окружающей жидкости и в конце концов исчезнет даже в том случае, когда температура жидкости заметно превышает температуру кипения. Если же пузырек газовый, то по той же причине он не может увеличиваться в объеме за счет испарения жидкости. Отсюда следует, что слишком малые пузырьки, как зародыши парообразной фазы, не эффективны. Для того чтобы жидкость при заданной температуре закипела, необходимо, чтобы размеры пузырьков были не меньше определенного предела.

2. Совершенно также обстоит дело с пересыщенным паром. Это тоже метастабильное состояние вещества. На изотерме Ван-дер-Ваальса ему соответствует участок AG (см. рис. 133). Давление пересыщенного пара P больше давления насыщенного пара при той же температуре. Допустим, что в перенасыщенном паре образовались капельки жидкости, например из-за тепловых флуктуаций. Если их размеры меньше определенного предела, то они испарятся. Действительно, давление пара, находящегося в равновесии с жидкой каплей, тем больше, чем меньше ее радиус. Если это давление превосходит P, то капля будет испаряться и в конце концов исчезнет. Такие малые капли, как центры конденсации, неэффективны. Капля будет расти, а, следовательно, пар конденсироваться в жидкость, если равновесное давление пара над ее поверхностью меньше давления окружающего перенасыщенного пара. Это имеет место для достаточно больших капель. Наличие пыли и других мелких частиц в перенасыщенном паре способствует конденсации. Дело в том, что капелька жидкости, образовавшаяся на пылинке, не будет иметь сферическую форму. Ее форма определяется формой и размерами самой пылинки. Ввиду этого кривизна поверхности капли, даже при очень малых размерах последней, может быть невелика. Такие капли являются эффективными центрами конденсации.

Еще более эффективными центрами конденсации являются электрически заряженные частицы, или ионы. Когда проводящий шар заряжен электричеством, то электрические заряды, отталкиваясь друг от друга, располагаются на его поверхности. Но и находясь на поверхности, они продолжают отталкиваться. Таким образом, на поверхностные заряды действуют силы, направленные наружу шара. Такие силы, как это будет показано в учении об электрических явлениях (см. т. III, § 33), действуют и на границе заряженного диэлектрического шара, если даже заряды располагаются не на поверхности, а по объему шара. Эти выталкивающие силы существуют и в случае заряженной капли. Они направлены противоположно силам лапласова капиллярного давления, обусловленного кривизной поверхности капли. Таким образом, влияние заряда капли эквивалентно уменьшению поверхностного натяжения. Вследствие этого давление насыщенного пара над заряженной каплей меньше, чем над незаряженной тех же размеров. Этим и объясняется, почему заряд капли способствует конденсации пара.

3. Перенасыщенный водяной пар, как это подробно было выяснено в § 115, можно получить быстрым адиабатическим расширением влажного воздуха. Этот принцип используется в одном из основных приборов ядерной физики и физики элементарных частиц — *камере Вильсона*. Камера Вильсона (1869–1959) представляет собой геометрически замкнутый объем, заполненный неконденсирующимся газом (гелий, аргон, азот и пр.) и насыщенными парами некоторых жидкостей (вода, этиловый спирт и пр.). Одна из стенок камеры делается подвижной (в виде поршня или упругой диаграммы), что позволяет менять величину рабочего объема камеры. При адиабатическом расширении в рабочем объеме создается перенасыщенный пар жидкости. Он не конденсируется, пока нет центра конденсации. Если, однако, через пар пролетает заряженная частица, то на своем пути она создает много ионов, на которых перенасыщенный пар конденсируется в виде маленьких капелек, достигающих видимых размеров. Получаются цепочки капелек, расположенные вдоль траектории ионизующей частицы. Они называются *треками*. Их можно осветить и сфотографировать. Изучение параметров трека (его длины, кривизны в магнитном поле, отклонения от прямолинейности за счет многократного рассеяния ионизующей частицы, количества капель на единице длины трека и т. д.) дает возможность делать заключение о природе и свойствах пролетающих частиц.

4. Явление перегревания жидкости также используется для наблюдения следов ионизующих частиц. На нем основано устройство *пузырьковой камеры*, сконструированной Д. А. Глезером (р. 1926) в 1952 г. Жидкость в пузырьковой камере сначала находится при температуре, превышающей температуру кипения. От закипания она удерживается высоким давлением, передаваемым на поршень или упругую мембрану, соприкасающуюся с жидкостью. Для приведения камеры в рабочее состояние внезапно понижают давление. Тогда жидкость оказывается перегретой и в течение короткого времени может находиться в этом метастабильном состоянии. Если в это время через камеру пролетит ионизующая частица, то она вызовет резкое вскипание жидкости в узкой области вдоль траектории частицы. В результате путь частицы будет отмечен цепочкой пузырьков пара. Явление можно объяснить тем, что ионизующая частица теряет на своем пути энергию, переходящую главным образом в тепло. А так как жидкость перегрета, то этого добавочного тепла достаточно для интенсивного образования пузырьков пара на пути частицы. Одним из важнейших преимуществ пузырьковой камеры перед камерой Вильсона является высокая плотность рабочего вещества. Это дает возможность получать большое количество взаимодействий частицы с ядрами рабочего вещества. В качестве жидкостей в пузырьковой камере применяются жидкий водород, жидкий пропан, фреоны, жидкий гелий, CO_2 и т. д.

5. Перегретой жидкостью и перенасыщенным паром не исчерпывается многообразие метастабильных состояний вещества. Другими примерами являются *переохлажденные жидкости* и различные крис-

таллические модификации твердых тел. Возьмем, например, воду, очищенную от посторонних вкраплений, и будем охлаждать ее при постоянном давлении и без сотрясений. Вода останется жидкой, если даже ее температура достигнет $-10\,^\circ$C и ниже. Если, однако, в такую воду бросить кристаллики льда (играющие роль зародышей кристаллической фазы) или встряхнуть сосуд, в котором она находится, то происходит быстрое затвердевание, причем температура резко поднимается до $0\,^\circ$C. Жидкость, охлажденная до температуры ниже температуры затвердевания, называется *переохлажденной жидкостью*. Она является менее устойчивой фазой, чем кристаллическая фаза при той же температуре и давлении. Поэтому если жидкость не очищена от посторонних вкраплений, способных выполнять функции зародышей кристаллической фазы, то переохлаждения не наблюдается. При охлаждении жидкости кристаллизация начинается при температуре плавления, соответствующей тому давлению, под которым находится жидкость. (Это не совсем точно, так как температура кристаллизации зависит от размеров и вида зародышей.)

Скорость кристаллизации переохлажденной жидкости при наличии в ней зародышей сильно зависит от температуры. При понижении температуры за точку кристаллизации скорость кристаллизации сначала растет, достигает максимума, а затем падает. При очень больших переохлаждениях скорость кристаллизации делается практически равной нулю. При достаточно больших переохлаждениях начинается самопроизвольное образование зародышей. Скорость этого процесса также сначала растет с понижением температуры, достигает максимума и стремится к нулю при дальнейшем понижении температуры. Обычно максимум скорости самопроизвольного образования зародышей находится при более низкой температуре, чем максимум скорости кристаллизации.

Переохлажденными жидкостями являются расплавленный сахар и мед. Здесь скорость кристаллизации очень мала. Однако процесс кристаллизации происходит, хотя и медленно. Мед и варенье с течением времени «засахариваются», т. е. переходят в кристаллическую модификацию. Многие тела, которые в обиходе называются твердыми, не обладают внутренней кристаллической структурой. Их лучше рассматривать как сильно переохлажденные жидкости. Таковы асфальт, сапожный вар, стекло, различные пластмассы и пр. Истинно твердыми телами являются только кристаллы. В стекле переохлаждение настолько сильно, что практически нет ни образования зародышей, ни кристаллизации на существующих зародышах. Однако даже и здесь процесс кристаллизации идет, хотя и очень медленно. Он приводит к тому, что по истечении десятков лет стекло может стать мутным (*«расстекловывание»* стекла).

Ярким примером, иллюстрирующим роль зародышей в полиморфных превращениях, может служить явление, известное под названием *«оловянной чумы»*. Существуют две модификации твердого олова — обычное, или белое, олово и порошкообразное, или серое, олово. При атмосферном давлении эти модификации находятся в равновесии при

температуре 18 °C. Выше 18 °C более устойчиво белое олово, ниже 18 °C — серое. После сильного мороза при потеплении оловянные предметы, если в них есть подходящий зародыш, могут рассыпаться в порошок. Это явление и называется «оловянной чумой». Оно очень редкое, так как обычно таких зародышей нет, самопроизвольно они образуются только при очень низких температурах. Зародыши недейственны, пока стоит сильный мороз, ввиду ничтожной скорости превращения. Но при потеплении скорость превращения сильно возрастает, и оловянный предмет рассыпается в порошок. Скорость превращения белого олова в серое (при наличии зародыша последнего) максимальна около 0 °C и быстро спадает при более низких температурах.

Иногда после суровой зимы при потеплении наблюдались «эпидемии» оловянной чумы. Такое явление произошло, например, в конце прошлого века в Петербурге. На одном из складов военного обмундирования находился большой запас солдатских пуговиц, которые в то время изготовлялись из белого олова. Склад не отапливался, пуговицы «простудились» и «заболели оловянной чумой». Сначала слегка потемнели несколько пуговиц. Пуговицы быстро теряли блеск и через несколько дней рассыпались в порошок. «Заболевшие» пуговицы «заражали» своих соседей из белого олова. Болезнь распространялась быстро, как чума. В несколько дней горы блестящих оловянных пуговиц превратились в бесформенную кучу серого порошка.

Оловянная чума явилась одной из основных причин гибели антарктической экспедиции Роберта Скотта (1868–1912): разрушились банки с горючим, полярники остались без горячей пищи. Достигнув южного полюса, они уже не смогли вернуться на свою базу.

ЗАДАЧА

При прохождении через перегретую жидкость ионизующей частицы вдоль ее траектории образуются мельчайшие пузырьки пара. Те из пузырьков, радиус которых больше «критического радиуса» $R_{\text{кр}}$, быстро вырастают до видимых размеров, а пузырьки меньших размеров захлопываются силами поверхностного натяжения. Определить $R_{\text{кр}}$ для жидкого пропана (C_3H_8), если в камере он находится под давление $P_{\text{ж}} = 5$ атм и температуре $T = 328$ К. Давление насыщенного пара пропана при этой температуре $P_{\text{п}} = 15$ атм, поверхностное натяжение пропана $\sigma = 4{,}46$ дин/см.

Ответ. $R_{\text{кр}} = \dfrac{2\sigma}{P_{\text{п}} - P_{\text{ж}}} = 9 \cdot 10^{-7}$ см. Легко показать, что в рассматриваемой задаче влияние кривизны поверхности пузырька на давление насыщенного пара не существенно.

§ 120. Фазовые превращения второго рода

1. Каждое фазовое превращение сопровождается скачкообразными изменениями каких-то величин, характеризующих свойства вещества. Удельный термодинамический потенциал $\varphi(T, P)$ остается непрерывным при любых превращениях (см. § 112). Однако его производные

могут испытывать разрыв непрерывности. *Фазовые превращения, при которых первые производные функции $\varphi(T, P)$ меняются скачкообразно, называются фазовыми превращениями первого рода. Фазовые превращения, при которых первые производные той же функции остаются непрерывными, а вторые производные меняются скачкообразно, называются фазовыми превращения второго рода.*

Рассмотрим сначала фазовые превращение первого рода. Так как

$$s = -\left(\frac{\partial\varphi}{\partial T}\right)_P, \quad v = \left(\frac{\partial\varphi}{\partial P}\right)_T, \tag{120.1}$$

то такие превращения характеризуются скачкообразными изменениями либо удельной энтропии s, либо удельного объема v, либо обеих этих величин вместе. Скачкообразное изменение удельной энтропии означает, что фазовое превращение сопровождается выделением или поглощением тепла (например, теплоты плавления, парообразования или возгонки). Количество теплоты q, которое надо сообщить единице массы вещества, чтобы квазистатически перевести ее из состояния 1 в состояние 2, определяется выражением

$$q = T(s_2 - s_1). \tag{120.2}$$

Все фазовые превращения, которые мы рассматривали до сих пор (плавление, испарение, возгонка, кристаллизация), сопровождаются выделением или поглощением теплоты, а потому они относятся к фазовым превращениям первого рода.

Рассмотрим теперь фазовые превращения второго рода. Из формул (120.1) следует, что при таких превращениях величины s и v остаются непрерывными. Это означает, что *фазовые превращения второго рода не сопровождаются выделением или поглощением теплоты, а также изменением удельного объем вещества. При фазовых превращениях второго рода претерпевают разрыв все или некоторые вторые производные удельного термодинамического потенциала.* Для каждой фазы эти производные непрерывны и могут быть представлены в виде

$$\frac{\partial^2\varphi}{\partial T^2} = -\left(\frac{\partial s}{\partial T}\right)_P = -\frac{c_P}{T}, \quad \frac{\partial^2\varphi}{\partial T\,\partial P} = \frac{\partial^2\varphi}{\partial P\,\partial T} = \left(\frac{\partial v}{\partial T}\right)_P,$$
$$\frac{\partial^2\varphi}{\partial P^2} = \left(\frac{\partial v}{\partial P}\right)_T.$$

Они претерпевают разрыв лишь при фазовых превращениях. Из этих формул видно, что *фазовые превращения второго рода сопровождаются скачкообразными изменениями одной или нескольких из следующих величин*: 1) *удельной теплоемкости* c_P; 2) *температурного коэффициента объемного расширения* $\alpha = \dfrac{1}{v_0}\left(\dfrac{\partial v}{\partial T}\right)_P$; 3) *изотермического коэффициента сжатия вещества* $\gamma = -\dfrac{1}{v}\left(\dfrac{\partial v}{\partial P}\right)_T$.

2. К фазовым превращениям второго рода относится, например, превращение железа, никеля, кобальта или какого-либо магнитного

сплава из *ферромагнитного состояния* в *парамагнитное*. Оно происходит при нагревании материала при определенной температуре, называемой *точкой Кюри*. Аналогичные фазовые превращения, при которых меняются диэлектрические свойства вещества, испытывает при нагревании и охлаждении *сегнетова соль* и многие другие *сегнетоэлектрики*. Температура превращения здесь также называется точкой Кюри. Фазовые превращения второго рода испытывают многие металлы и сплавы при переходе в *сверхпроводящее состояние*. Этот процесс происходит при низких температурах и характеризуется скачкообразным уменьшением электрического сопротивления до нуля. Явление называется *сверхпроводимостью*. Для металлов и сплавов критическая температура T_K, ниже которой происходит переход в сверхпроводящее состояние, не превышает 23 К. В 1986–1987 гг. были получены керамики (например, иттрийбариевая керамика) с температурой перехода T_K около 100 К. Указанные явления подробнее будут рассмотрены в учении об электричестве и атомной физике.

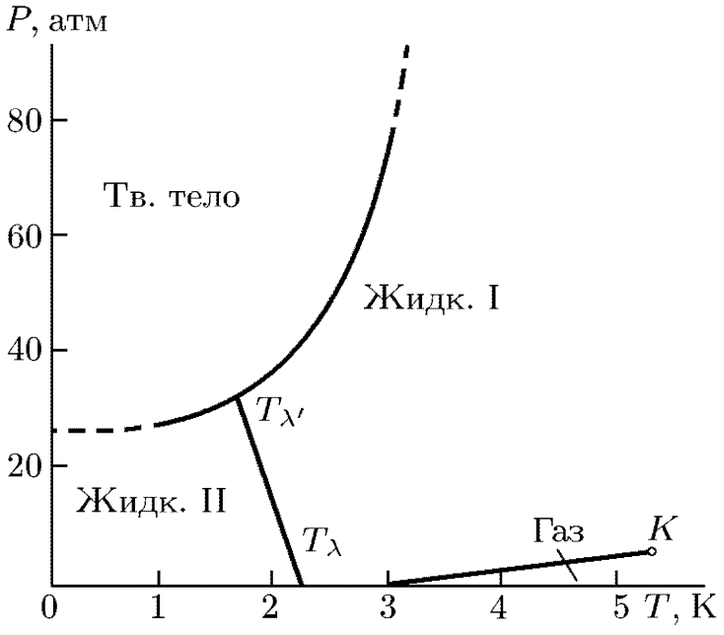


Рис. 139

Замечательным примером фазового превращения второго рода является превращение обыкновенного жидкого гелия (так называемого гелия I) в другую жидкую модификацию, называемую гелием II. Диаграмма состояний гелия представлена на рис. 139. Гелий может существовать в газообразной, твердой и двух жидких модификациях: I и II. Характерной особенностью диаграммы состояний гелия является отсутствие на ней кривой возгонки. Если охлаждать жидкий гелий I, то при вполне определенной температуре (зависящей от внешнего давления), называемой λ-точкой, он испытывает фазовое превращение второго рода и переходит в жидкий гелий II. Эта жидкая модификация гелия продолжает оставаться жидкой вплоть до температуры абсолютного нуля. Таким свойством обладает только гелий. Все прочие вещества при абсолютном нуле температуры могут находиться только в твердом состоянии.

Приведем значения параметров основных точек на диаграмме состояний гелия (рис. 139):

$$P_K = 1718 \text{ мм рт. ст.}, \quad T_K = 5{,}20 \text{ К}, \quad T_{\lambda'} = 1{,}778 \text{ К},$$

$$P_{\lambda'} = 29{,}96 \text{ атм}, \quad T_\lambda = 2{,}186 \text{ К}, \quad P_\lambda = 38{,}3 \text{ мм рт. ст.}$$

Жидкий гелий II обладает удивительным свойством *сверхтекучести*. Это явление, открытое П. Л. Капицей, состоит в том, что жидкий гелий II не обладает вязкостью. Для него вязкость равна нулю. Явление сверхтекучести тесно связано с явлением сверхпроводимости. Последнее явление может быть охарактеризовано как сверхтекучесть электронного газа в металлах. Оба явления могут быть понятны только на основе квантовых представлений.

3. Уравнение Клапейрона–Клаузиуса (113.2) в случае фазовых превращений второго рода теряет смысл. Для таких превращений числитель и знаменатель в правой части уравнения (113.2) обращаются в нуль, и оно принимает неопределенный вид $0/0$. При фазовых превращениях второго рода уравнение Клапейрона–Клаузиуса надо заменить *соотношениями Эренфеста* (1880–1933), к выводу которых мы и перейдем.

Соотношения Эренфеста являются следствиями непрерывности удельной энтропии s и удельного объема v при фазовых превращениях второго рода. Если рассматривать удельную энтропию s какой-либо фазы как функцию температуры и давления, то для ее дифференциала можно написать

$$ds = \left(\frac{\partial s}{\partial T}\right)_P dT + \left(\frac{\partial s}{\partial P}\right)_T dP,$$

или, используя соотношения

$$\left(\frac{\partial s}{\partial T}\right)_P = \frac{c_P}{T}, \quad \left(\frac{\partial s}{\partial P}\right)_T = -\left(\frac{\partial v}{\partial T}\right)_P, \tag{120.3}$$

$$ds = \frac{c_P}{T}\, dT - \left(\frac{\partial v}{\partial T}\right)_P dP. \tag{120.4}$$

Напишем это соотношение для каждой фазы:

$$ds_1 = \frac{c_{1P}}{T}\, dT - \left(\frac{\partial v_1}{\partial T}\right)_T dP, \quad ds_2 = \frac{c_{2P}}{T}\, dT - \left(\frac{\partial v_2}{\partial T}\right)_P dP.$$

Возьмем точки (T, P) и $(T + dT, P + dP)$ на кривой равновесия. Тогда величина dP/dT определит наклон этой кривой. Кроме того, ввиду непрерывности удельной энтропии при фазовых превращениях второго рода $ds_1 = ds_2$. Это дает

$$(c_{2P} - c_{1P})\frac{dT}{T} = \left[\left(\frac{\partial v_2}{\partial T}\right)_P - \left(\frac{\partial v_1}{\partial T}\right)_P\right] dP,$$

или сокращенно

$$\Delta c_P = T\Delta\left(\frac{\partial v}{\partial T}\right)_P \frac{dP}{dT}, \tag{120.5}$$

где Δc_P и $\Delta(\partial v/\partial T)_P$ означают скачки, испытываемые величинами c_P и $(\partial v/\partial T)_P$ при фазовых превращениях. Соотношение (120.5) и есть первое соотношение Эренфеста.

Второе соотношение Эренфеста получается таким же путем. Надо только удельную энтропию s рассматривать как функцию температуры и удельного объема. Оно имеет вид

$$\Delta c_v = -T\Delta\left(\frac{\partial P}{\partial T}\right)_v \frac{dv}{dT}. \tag{120.6}$$

Третье соотношение Эренфеста получается также из условия непрерывности удельной энтропии, но рассматриваемой как функцию v и P. Таким путем находим

$$\Delta\left(\frac{\partial v}{\partial T}\right)_P = \Delta\left(\frac{\partial P}{\partial T}\right)_v \frac{dv}{dP}. \tag{120.7}$$

Наконец, четвертое соотношение Эренфеста получается из условия непрерывности удельного объема v, если его рассматривать как функцию T и P. Оно имеет вид

$$\Delta\left(\frac{\partial v}{\partial T}\right)_P = -\Delta\left(\frac{\partial v}{\partial P}\right)_T \frac{dP}{dT}. \tag{120.8}$$

Разумеется, производные dv/dT, dv/dP, dP/dT в формулах (120.6), (120.7) и (120.8) берутся вдоль соответствующих кривых равновесия.

Следует, однако, заметить, что приведенная выше классификация Эренфеста фазовых переходов и основанная на ней термодинамическая теория имеют ограниченную область применимости. Классификация предполагает, что вторые производные термодинамического потенциала в точках фазовых превращений остаются *конечными*. А это, как показали экспериментальные и теоретические исследования, по-видимому, не всегда имеет место. Так, в случае перехода вещества из ферромагнитного в парамагнитное состояние или обратно, а также при переходах гелия I в гелий II и обратно теплоемкость c_P, по-видимому, логарифмически стремится к бесконечности, когда температура стремится к соответствующей температуре перехода. А это, как видно из формул (120.3), означает стремление к бесконечности также производной $(\partial s/\partial T)_P$, а с ней и производной $(\partial^2\varphi/\partial^2 T^2)_P$. Однако к явлениям сверхпроводимости теория Эренфеста, по-видимому, применима.

§ 121. Конвективная устойчивость жидкостей и газов

1. Если жидкость (или газ), помещенная в поле тяжести, нагрета неравномерно, то не при всяком распределении температур она может находиться в механическом равновесии. Вообще говоря, в такой жидкости будет происходить перемешивание (*конвекция*) различно нагретых частей. Для простоты предположим, что температура жидкости меняется только с высотой. Поле тяжести будем считать однородным. Выясним, при каких условиях конвекции не будет. Будем пренебрегать процессами теплопроводности в жидкости. Тогда всякое перемешивание элемента жидкости из одного положения в другое может рассматриваться как адиабатический процесс, в котором энтропия не меняется.

2. В состоянии механического равновесия температура T, удельный объем v и давление P жидкости являются функциями только высоты z над земной поверхностью. Действительно, ускорение свободного падения g направлено вниз, т. е. в отрицательную сторону оси Z, и зависит только от z. Оно не имеет слагающих вдоль координатных осей X и Y. Поэтому при механическом равновесии $\partial P/\partial x = \partial P/\partial y = 0$, т. е. P зависит также только от z. Далее, $dP/dz = -\rho g(z)$, откуда видно, что плотность $\rho \equiv 1/v$ является функцией одной только координаты z. То же относится и к температуре T, как это видно из уравнения состояния $T = T(R, \rho)$.

Пусть dv, dT, dP означают бесконечно малые приращения v, T, P в покоящейся жидкости при изменении высоты на dz. В силу уравнения состояния эти величины связаны соотношением

$$dv = \left(\frac{\partial v}{\partial T}\right)_P dT + \left(\frac{\partial v}{\partial P}\right)_T dP. \tag{121.1}$$

Допустим теперь, что под действием какого-то бесконечно малого возмущения элемент жидкости переместился вверх на dz. Так как указанное перемещение происходит адиабатически, то для изменения удельного объема жидкости при таком перемещении можно написать

$$dv_{\text{ад}} = \left(\frac{\partial v}{\partial T}\right)_P dT_{\text{ад}} + \left(\frac{\partial v}{\partial P}\right)_T dP. \tag{121.2}$$

Здесь $dT_{\text{ад}}$ и dP означают приращения температуры и давления внутри рассматриваемого элемента жидкости при адиабатическом поднятии его на высоту dz. (Индекс «ад» у dP мы опустили, так как приращение давления в элементе жидкости — такое же, что и приращение давления в окружающей жидкости.) Если $dz > 0$, т. е. элемент жидкости сместился действительно вверх, и $dv_{\text{ад}} > dv$, то сместившийся элемент окажется относительно более легким, чем окружающая жидкость. Он будет подниматься еще выше, и равновесие жидкости окажется неустойчивым. В противоположном случае, когда $dv_{\text{ад}} < dv$, давление окружающей жидкости вернет элемент в исходное положение, т. е. равновесие будет устойчивым. Воспользовавшись выражениями (121.1), (121.2) и поделив неравенство на положительную величину dz, условие устойчивости равновесия можно записать в виде

$$\left(\frac{\partial v}{\partial T}\right)_P \left(\frac{dT}{dz}\right)_{\text{ад}} < \left(\frac{\partial v}{\partial T}\right)_P \frac{dT}{dz}. \tag{121.3}$$

Требование $dz > 0$, использованное при выводе, теперь можно снять, так как в неравенство (121.3) входят только производные $(dT/dz)_{\text{ад}}$ и dT/dz, значения которых от знака dz не зависят. Для большинства тел температурный коэффициент объемного расширения положителен, и вместо условия (121.3) можно написать более простое условие

$$\frac{dT}{dz} > \left(\frac{dT}{dz}\right)_{\text{ад}}. \tag{121.3а}$$

Для тел с отрицательным температурным коэффициентом объемного расширения знак неравенства надо заменить на противоположный. Ниже предполагается, что имеет место первый случай.

3. Таким образом, чем больше температурный градиент dT/dz, тем более затруднена конвекция, тем устойчивее механическое равновесие жидкости. Нижней границей dT/dz, при которой конвекция еще может отсутствовать, является «*адиабатический температурный градиент*» $(dT/dz)_{\text{ад}}$. Для его вычисления замечаем, что при адиабатическом процессе удельная энтропия s не меняется. Рассматривая ее как функцию T и P, можно

написать

$$\left(\frac{ds}{dz}\right)_{\text{ад}} = \left(\frac{\partial s}{\partial T}\right)_P \left(\frac{dT}{dz}\right)_{\text{ад}} + \left(\frac{\partial s}{\partial P}\right)_T \frac{dP}{dz} = 0.$$

Воспользовавшись термодинамическими соотношениями (120.3) и уравнением гидростатики $dP/dz = -\rho g = -g/v$, получим

$$\left(\frac{dT}{dz}\right)_{\text{ад}} = -\frac{gT}{vc_P}\left(\frac{\partial v}{\partial T}\right)_P. \tag{121.4}$$

4. Для воздуха, если его рассматривать как идеальный газ, объем v пропорционален температуре T (при $P = \text{const}$), а потому $(\partial v/\partial T)_P = v/T$. Это дает

$$\left(\frac{dT}{dz}\right)_{\text{ад}} = -\frac{g}{c_P}. \tag{121.5}$$

Считая воздух двухатомным газом, имеем по классической теории теплоемкостей $c_P = 7R/2\mu$, где μ — средняя молярная масса воздуха ($\mu \approx 28{,}8$ г/моль). Подстановка числовых значений дает

$$\left(\frac{dT}{dz}\right)_{\text{ад}} = -\frac{2\mu g}{7R} = -9{,}7 \cdot 10^{-5}\ \text{К/см} \approx -10^{-2}\ \text{К/м}.$$

Если температура воздуха повышается с высотой, то атмосфера в механическом отношении устойчива. Но устойчивое равновесие возможно и тогда, когда с высотой температура воздуха понижается. Однако это понижение не может превосходить примерно одного градуса на каждые сто метров высоты.

5. Мы не учитывали влияние водяных паров, всегда имеющихся в атмосфере. Основной интерес представляют случаи, когда температура воздуха значительно ниже температуры кипения воды. При таких условиях количество водных паров относительно мало. Влияние их на величину адиабатического температурного градиента было бы ничтожно, если бы при адиабатических процессах не происходила конденсация водяного пара. В действительности при адиабатическом поднятии воздух охлаждается, становится насыщенным, а затем пересыщенным. В результате водяные пары конденсируются на ионах, пыли и других центрах конденсации. При этом выделяется теплота парообразования. Это обстоятельство существенно меняет дело.

Рассмотрим какую-либо порцию воздуха, насыщенного водяными парами. Массу воздуха в ней обозначим через $m_{\text{в}}$, массу водяного пара — $m_{\text{п}}$, массу жидкой воды — $m_{\text{ж}}$. При адиабатическом поднятии энтропия рассматриваемой системы меняться не будет:

$$m_{\text{в}} s_{\text{в}} + m_{\text{п}} s_{\text{п}} + m_{\text{ж}} s_{\text{ж}} = \text{const}, \tag{121.6}$$

где $s_{\text{в}}$, $s_{\text{п}}$, $s_{\text{ж}}$ — удельные энтропии воздуха, водяного пара и жидкой воды соответственно. При этом полное количество воды остается постоянным: $m_{\text{п}} + m_{\text{ж}} = \text{const}$, так что $dm_{\text{ж}} = -dm_{\text{п}}$. Массу жидкой воды мы должны положить равной нулю, если в рассматриваемом состоянии вся вода существует в виде насыщенного водяного пара. Но, конечно, величина $dm_{\text{ж}}$ должна считаться отличной от нуля, так как при поднятии вверх водяные пары конденсируются в жидкие капли. Имея это в виду, из условия (121.6) получим

$$m_{\text{в}}\, ds_{\text{в}} + m_{\text{п}}\, ds_{\text{п}} + (s_{\text{п}} - s_{\text{ж}})\, dm_{\text{п}} = 0.$$

Разность удельных энтропий выразим через удельную теплоту испарения $q = T(s_{\text{п}} - s_{\text{ж}})$. Масса пара $m_{\text{п}}$ в рассматриваемой системе зависит только

от температуры T, так что $dm_{\text{п}} = \dfrac{dm_{\text{п}}}{dT}\,dT$. Для дифференциала удельной энтропии воздуха $ds_{\text{в}}$ с учетом, что воздух может считаться идеальным газом, получим такое же выражение, как и в случае сухого воздуха:

$$ds_{\text{в}} = \frac{c_{P\text{в}}}{T}\,dT - \frac{v_{\text{в}}}{T}\,dP_{\text{в}}.$$

То же можно написать и для водяного пара. Однако надо учесть, что давление насыщенного пара зависит только от температуры, а потому

$$ds_{\text{п}} = \left[\frac{c_{P\text{п}}}{T} - \frac{v_{\text{п}}}{T}\,\frac{dP_{\text{п}}}{dT}\right] dT.$$

Наконец, применим к воздуху уравнение гидростатики:

$$dP_{\text{в}} = -\rho_{\text{в}} g\,dz = -\frac{g}{v_{\text{в}}}\,dz.$$

С учетом всего этого получим

$$\left[c_{P\text{в}} + \frac{m_{\text{п}}}{m_{\text{в}}}\left(c_{P\text{п}} - v_{\text{п}}\,\frac{dP_{\text{п}}}{dT} + \frac{q}{m_{\text{п}}}\,\frac{dm_{\text{п}}}{dT}\right)\right]\frac{dT}{dz} = -g.$$

В этом соотношении $m_{\text{в}}$ и $m_{\text{п}}$, очевидно, можно заменить на плотности воздуха и водяного пара $\rho_{\text{в}}$ и $\rho_{\text{п}}$. Из уравнения Клапейрона–Менделеева $\rho = \mu P/RT$, а потому

$$\frac{m_{\text{п}}}{m_{\text{в}}} = \frac{\mu_{\text{п}} P_{\text{п}}}{\mu_{\text{в}} P_{\text{в}}},$$

$$\frac{1}{m_{\text{п}}}\,\frac{dm_{\text{п}}}{dT} = \frac{1}{\rho_{\text{п}}}\,\frac{d\rho_{\text{п}}}{dT} = \frac{T}{P_{\text{п}}}\,\frac{d}{dT}\left(\frac{P_{\text{п}}}{T}\right) = \frac{1}{P_{\text{п}}}\,\frac{dP_{\text{п}}}{dT} - \frac{1}{T}.$$

Выполнив соответствующую подстановку, получим

$$\left[c_{P\text{в}} + \frac{\mu_{\text{п}} P_{\text{п}}}{\mu_{\text{в}} P_{\text{в}}}\left(c_{P\text{п}} - v_{\text{п}}\,\frac{dP_{\text{п}}}{dT} - \frac{q}{T} + \frac{q}{P_{\text{п}}}\,\frac{dP_{\text{п}}}{dT}\right)\right]\frac{dT}{dz} = -g.$$

Наконец, воспользуемся уравнением Клапейрона–Клаузиуса в упрощенном виде:

$$\frac{dP_{\text{п}}}{dT} = \frac{q}{Tv_{\text{п}}}.$$

В результате найдем

$$\left(\frac{dT}{dz}\right)_{\text{ад}} = -\frac{g}{c_{P\text{в}}}\,\frac{1}{1 + \dfrac{\mu_{\text{п}} P_{\text{п}}}{\mu_{\text{в}} P_{\text{в}} c_{P\text{в}}\left[c_{P\text{в}} - \dfrac{2q}{T} + \dfrac{\mu_{\text{п}}}{R}\left(\dfrac{q}{T}\right)^2\right]}}. \tag{121.7}$$

В окончательной формуле (121.7) мы ввели у температурного градиента индекс «ад», опущенный в промежуточных расчетах. Удельные теплоемкости воздуха и водяного пара вычислим по классической теории, считая воздух двух-, а водяной пар — трехатомным газами. Тогда

$$c_{P\text{в}} = \frac{7R}{2\mu_{\text{в}}}, \quad c_{P\text{п}} = \frac{4R}{\mu_{\text{п}}}.$$

Учтем, далее, что множитель $-g/c_{P\text{в}}$ дает адиабатический градиент температуры для сухого воздуха, который мы обозначим через $(dT/dz)_{\text{ад, сух}}$. Тогда формула (121.7) представится в виде

$$\left(\frac{dT}{dz}\right)_{\text{ад}} = f \cdot \left(\frac{dT}{dz}\right)_{\text{ад, сух}}, \tag{121.8}$$

где коэффициент f определяется выражением

$$\frac{1}{f} = 1 + \frac{8}{7}\frac{P_{\text{п}}}{P_{\text{в}}}\left[1 - \frac{g\mu_{\text{п}}}{2RT} + \left(\frac{q\mu_{\text{п}}}{2RT}\right)^2\right]. \tag{121.9}$$

Теми же формулами можно пользоваться и в тех случаях, когда при охлаждении водяные пары не конденсируются, а превращаются в лед. Только в этих случаях под q следует понимать теплоту возгонки, равную сумме теплот парообразования и плавления.

6. В табл. 12 приведены вычисленные значения коэффициента f при различных температурах для двух значений полного давления: $P_{\text{в}} + P_{\text{п}} = 760$ мм рт. ст. и $P_{\text{в}} + P_{\text{в}} = 380$ мм рт. ст. (высота над уровнем моря

Таблица 12

$P_{\text{в}} + P_{\text{п}} = 760$ мм рт. ст.		$P_{\text{в}} + P_{\text{п}} = 760$ мм рт. ст.		$P_{\text{в}} + P_{\text{п}} = 380$ мм рт. ст.	
T, °C	f	T, °C	f	T, °C	f
−30	0,94	10	0,47	−30	0,88
−25	0,91	15	0,40	−25	0,83
−20	0,86	20	0,33	−20	0,76
−15	0,81	25	0,28	−15	0,68
−10	0,74	30	0,23	−10	0,57
−5	0,65	40	0,16	−5	0,47
0	0,62	50	0,10	0	0,44
5	0,54				

около 5,5 км, если атмосферу Земли считать изотермической при $t = 0\,^{\circ}$C). Из таблицы видно, насколько существенно влияние влажности, если адиабатические процессы в атмосфере сопровождаются конденсацией или замерзанием водяных паров.

Таблица показывает, что в случае «влажной» адиабаты адиабатическое охлаждение воздуха с высотой происходит в 2-3 раза медленнее, чем в случае «сухой» адиабаты. С этим связано возникновение *фёна*, т. е. сухого и теплого ветра, дующего с гор. В СССР фёны наиболее распространены на Кавказе и в Средней Азии. Допустим, что насыщенная водяными парами масса воздуха переваливает через горный хребет. При поднятии воздух охлаждается по «влажной» адиабате, т. е. сравнительно медленно, ибо по мере поднятия конденсация паров все время увеличивается, а выделяющееся при этом скрытое тепло замедляет охлаждение. Отдельные капли воды настолько увеличиваются, что начинают падать на поверхность Земли в виде дождя. В результате масса воздуха, перевалившая через хребет, оказывается обедненной водой. При опускании в долину она адиабатически нагревается, причем это нагревание сначала идет опять по «влажной» адиабате, т. е. медленно, так как значительная часть тепла затрачивается на испарение еще существующих облаков. Но как только облака испарятся, дальнейшее нагревание

воздуха начнет происходить по «сухой» адиабате, т. е. быстро. Обедненный влагой воздух спускается в долину значительно нагретым. Таким образом, большие горные цепи могут становиться как бы разделами погоды. Области восходящих влажных потоков воздуха являются дождливыми. Места же позади горных хребтов, куда воздух поступает значительно обедненным влагой и нагретым из-за адиабатического сжатия, являются сухими и бедными дождями. Примером может служить западный берег Южной Америки, где преобладают преимущественно западные ветры, обогащенные влагой, поскольку они дуют с Тихого океана. Узкая полоса, лежащая к западу от Кордильер, исключительно богата дождями, в то время как местность по другую сторону этого горного хребта напоминает пустыню.

Глава XI
РАСТВОРЫ

§ 122. Общие сведения

1. *Растворами называют физически однородные (гомогенные) смеси двух или нескольких веществ.* Физическая однородность достигается равномерным перемешиванием молекул. В этом отношении растворы отличаются от механических смесей, в которых перемешаны не молекулы, а макроскопические частицы вещества.

Если одного вещества в растворе больше, чем других, то оно называется *растворителем*, а прочие вещества — *растворенными веществами*.

Растворы называются *слабыми* или *разбавленными*, если число молекул растворенных веществ очень мало по сравнению с числом молекул растворителя. Растворы же, содержащие много растворенных веществ, называются *крепкими*. Очень крепкие растворы называются *концентрированными*.

2. От химических соединений растворы отличаются тем, что в химические соединения вещества вступают в строго определенных пропорциях, тогда как относительные количества веществ в растворах могут меняться в более или менее широких пределах.

Но растворы нельзя считать и механическими смесями молекул. Из-за взаимодействия последних растворы по некоторым признакам приближаются к химическим соединениям. Так, при смешении спирта с водой наблюдается некоторое уменьшение объема. Растворение обычно сопровождается выделением или поглощением теплоты. Тепловой эффект считается положительным, если при растворении теплота выделяется, и отрицательным, если она поглощается. Тепловой эффект зависит не только от количества растворенного вещества, но и от количества растворителя.

Под теплотой растворения обычно понимают количество теплоты, которое выделяется или поглощается при растворении моля вещества в настолько большом количестве растворителя, что дальнейшее добавление растворителя уже не приводит к выделению или поглощению теплоты. Следующие данные дают представление о теплоте растворения некоторых веществ в воде при 18 °С и атмосферном давлении:

нашатырь (NH_4Cl, тв.)	−16,5 кДж/моль
азотнокислый аммоний (NH_4NO_3, тв.)	−26,5 кДж/моль
гидроокись калия (KOH, тв.)	+54,2 кДж/моль
серная кислота (H_2SO_4, жидк.)	+74,5 кДж/моль

Эти величины примерно лишь на порядок меньше соответствующих величин при химических реакциях (см. § 17).

При смещении соли со снегом или мелко раздробленным льдом происходит образование раствора, сопровождающееся сильным охлаждением. С помощью NaCl можно добиться понижения температуры до $-21\,^\circ$C, а с помощью $CaCl_2 \cdot 6H_2O$ — до $-55\,^\circ$C. Это явление используется в холодильной технике (см. § 105).

В слабых растворах взаимодействие между молекулами растворенных веществ не играет заметной роли. В теории им обычно пренебрегают подобно тому, как это делается в учении об идеальных газах. Имеет значение только взаимодействие молекул растворенных веществ с молекулами растворителя.

В электролитах молекулы растворенных веществ полностью или частично диссоциируют (распадаются) на ионы. Этим объясняется электрическая проводимость электролитов и связанное с ней явление электролиза.

Во многих жидких растворах установлено существование *сольватов*. Так называют более или менее непрочные соединения непостоянного состава молекул или ионов растворенных веществ с молекулами растворителя. В водных растворах сольваты называются *гидратами*. Сольваты и гидраты непрерывно разрушаются (диссоциируют) и образуются вновь.

3. Вещества, входящие в раствор, называются его *компонентами*. Относительное содержание компонентов в растворе характеризуется их *концентрациями*. Различают *массовые*, *молярные* и *объемные* концентрации. *Массовая концентрация есть отношение массы рассматриваемого компонента к общей массе раствора*. Она обычно выражается в процентах. В теоретических исследованиях особенно удобна молярная концентрация. Это есть *отношение числа молей рассматриваемого компонента к общему числу молей раствора. Объемной концентрацией компонента называется количество его (в граммах или молях) в единице объема раствора.*

§ 123. Растворимость тел

1. Всякое вещество растворяется в другом веществе далеко не всегда в неограниченном количестве. *Раствор, содержащий наибольшее количество вещества, которое может в нем растворится, называется насыщенным*. Если к насыщенному раствору добавить порцию растворяемого вещества, то концентрация раствора меняться не будет. Возникает термодинамическое или статистическое равновесие между растворяемым телом и раствором: число молекул, переходящих от тела в раствор, в среднем будет равно числу молекул, возвращающихся обратно из раствора к телу. Доказательством статистического характера равновесия может служить, например, такой опыт. Если в плотно закрывающийся сосуд с насыщенным раствором NaCl погрузить крупный, неправильной формы кусок кристалла поваренной соли

и длительно выдерживать всю систему при постоянной температуре, то форма куска заметно изменится (приближаясь к характерной для NaCl кубической), хотя масса его останется без изменения.

Раствор называется *ненасыщенным*, если концентрация растворенного вещества в нем меньше, чем в насыщенном растворе. Если же она больше, то раствор называется *пересыщенным*. Пересыщенные растворы *метастабильны*. В присутствии зародышей растворяемого вещества избыток его выпадает из раствора, и раствор становится насыщенным.

Концентрация насыщенного раствора может служить мерой способности рассматриваемого вещества растворяться в растворителе. Ее называют *растворимостью*.

2. Растворимость зависит от температуры. Характер этой зависимости определяется знаком теплового эффекта при растворении. Рассмотрим равновесную систему, состоящую из насыщенного раствора и граничащего с ним растворяемого вещества. Если ее нагреть, то равновесие нарушится. Согласно принципу Ле-Шателье–Брауна для восстановления равновесия в системе должны начаться процессы, стремящиеся уменьшить это нагревание. Отсюда следует, что если тепловой эффект при растворении положителен, то часть растворенного вещества должна выпасть из раствора. Если же он отрицателен, то некоторое количество вещества должно перейти в раствор. Таким образом, *для веществ с положительным тепловым эффектом растворимость убывает с температурой, а для веществ с отрицательным тепловым эффектом — возрастает.*

При растворении газов в жидкостях и твердых телах молекулы переходят из газа, где они слабо взаимодействуют друг с другом, в раствор, где они подвергаются сильному притяжению со стороны молекул растворителя. С этим связано то обстоятельство, что в подавляющем большинстве случаев тепловой эффект при растворении газов положителен. Значит, *растворимость газов* (при постоянном давлении) должна убывать с повышением температуры.

3. Рассмотрим газ, находящийся в равновесии со своим раствором. Если давление газа над раствором повысить, то равновесие нарушится. Принцип Ле-Шателье–Брауна требует, чтобы для восстановления равновесия начались процессы, которые бы уменьшали это давление. Часть газа должна перейти в раствор. Значит, *с повышением давления (при неизменной температуре) растворимость газов должна возрастать, а с понижением давления — убывать.*

В состоянии статистического равновесия скорость перехода молекул из газа в раствор должна равняться скорости обратного перехода молекул из раствора в газ. Скорость первого перехода пропорциональна числу ударов молекул газа о поверхность раствора, т. е. (при неизменной температуре) давлению газа над раствором. Для слабых растворов скорость обратного перехода пропорциональна концентрации газа в растворе, т. е. его растворимости. (Для крепких растворов

это может оказаться неверным из-за взаимодействия между молекулами самого газа в растворе.) Если над раствором находится смесь газов, то это рассуждение применимо для каждого газа в отдельности. Таким образом, *при неизменной температуре растворимость газа пропорциональна парциальному давлению его над раствором*. Этот закон был установлен экспериментально в 1803 г. английским химиком Уильямом Генри (1774–1836). *Закон Генри* применим только к слабым растворам при отсутствии химического взаимодействия молекул газа между собой и с молекулами растворителя. Он применим, например, к плохо растворимым в воде кислороду и азоту, но не применим к углекислому газу и аммиаку, хорошо растворяющимся в воде.

4. Аналогичные соображения применимы к распределению растворенного вещества между двумя несмешивающимися растворителями, например водой и керосином. Пусть водный раствор некоторого вещества смешан с керосином. Тогда часть вещества перейдет из воды в керосин. В состоянии равновесия число молекул растворенного вещества, переходящих из воды в керосин, должно равняться числу молекул, переходящих за то же время в обратном направлении из керосина в воду. Отсюда следует, что *в случае слабых растворов отношение равновесных концентраций растворенного вещества в двух несмешивающихся растворителях не зависит от концентрации, а является функцией только температуры и давления*. Это положение называется *законом распределения*.

5. Газы в обычных условиях способны смешиваться между собой в любых пропорциях. Иными словами, каждый газ обладает неограниченной растворимостью в другом газе. Не существует насыщенных растворов одного газа в другом. Только при очень высоких давлениях, когда плотности газов приближаются к плотностям соответствующих жидкостей, наблюдаются в некоторых случаях отступления от этого правила: смеси газов распределяются на две фазы разного состава. Такая ограниченная взаимная растворимость газов впервые наблюдалась на смеси N_2 и NH_3 при 140 °C и давлении, равном 5 тыс. атм.

6. Растворимость твердых тел в жидкостях всегда ограничена. Что касается растворимости жидкостей в других жидкостях, то здесь возможны оба случая: встречаются жидкости, смешивающиеся друг с другом в любых пропорциях (например, вода и спирт), а также жидкости, смешивающиеся в ограниченной степени и притом в самых разнообразных соотношениях. Например, вода и бензол или вода и сероуглерод практически совсем не растворяют друг друга. Эфир же в воде и вода в эфире растворяются заметно (при 20 °C раствор эфира в воде может содержать до 6,5 % эфира по весу, а раствор воды в эфире — до 1,2 % воды).

Состояние смеси двух или нескольких веществ удобно изображать на *диаграммах состояния*, откладывая по осям координат значения параметров, характеризующих состояние системы. Так как обычно давление (атмосферное) бывает фиксировано, то в качестве таких параметров можно взять температуру и концентрации компонентов

смеси. Для бинарных смесей число независимых параметров равно двум (так как концентрации компонентов не независимы, а связаны между собой условием нормировки). Диаграмма состояния получается плоской (двумерной). На рис. 140 схематически изображена такая диаграмма для смеси фенола с водой. Линия MAK соответствует насыщенному раствору воды в феноле, линия NBK — насыщенному раствору фенола в воде. Точки левее линии MAK изображают насыщенный раствор воды в феноле, правее линии NBK — фенола в воде. Области ниже кривой MKN соответствует система из двух фаз — насыщенного раствора воды в феноле и насыщенного раствора фенола в воде. Первая фаза, как более тяжелая, опускается вниз, вторая, как более легкая, поднимается вверх. При 66 °C обе кривые MAK и NBK смыкаются в общей точке K. Выше этой точки фенол и вода смешиваются друг с другом в неограниченных количествах. Температура, соответствующая точке K, называется *критической температурой смешения*.

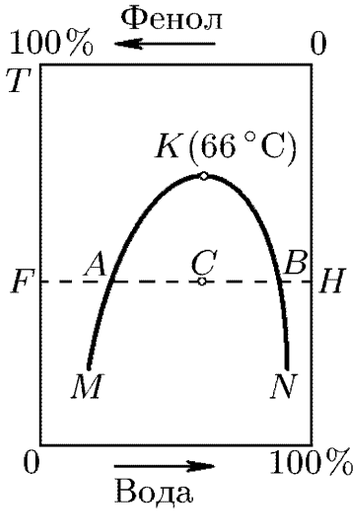


Рис. 140

Соотношение между массами насыщенных растворов фенола в воде $m_{\text{фв}}$ и воды в феноле $m_{\text{вф}}$ в какой-либо точке C двухфазной области (рис. 140) определяется *«правилом рычага»*. Согласно этому правилу точка C делит горизонтальный отрезок AB на части, обратно пропорциональные массам соответствующих насыщенных растворов, т. е.

$$\frac{m_{\text{фв}}}{m_{\text{вф}}} = \frac{AC}{CB}.$$

Действительно, пусть $m_{\text{фв}}$ и $m_{\text{вф}}$ означают относительные массы соответствующих растворов в точке C. Тогда $m_{\text{фв}} + m_{\text{вф}} = 1$. Относительное содержание фенола в насыщенном растворе фенола в воде соответствует точке B и численно равно длине отрезка HB. Содержание же фенола в насыщенном растворе воды в феноле представляется длиной отрезка HA. Таким образом, относительное количество фенола в точке C будет

$$m_{\text{фв}} \cdot HB + (1 - m_{\text{фв}}) \cdot HA.$$

С другой стороны, та же величина представляется длиной отрезка HC. Приравнивая оба выражения, получим

$$m_{\text{фв}} = \frac{HA - HC}{HA - HB} = \frac{AC}{AB}.$$

Отсюда

$$m_{\text{вф}} = 1 - m_{\text{фв}} = \frac{AB - AC}{AB} = \frac{BC}{AB}.$$

Почленным делением получаем требуемый результат.

Существуют жидкости, для которых область, где они неограниченно смешиваются друг с другом, лежит ниже некоторой определенной температуры (*нижняя критическая температура смешения*). Примером может служить триэтиламин [$N(C_2H_5)_3$] и вода (рис. 141).

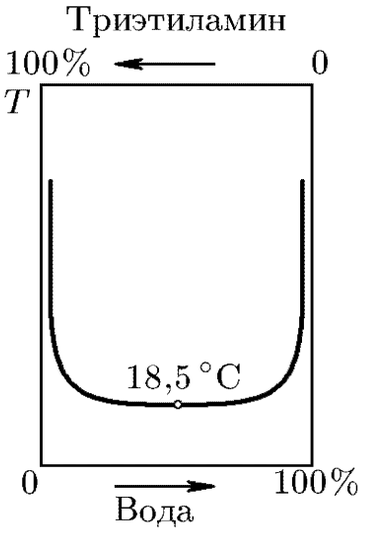


Рис. 141

Наиболее общий случай взаимной растворимости двух жидкостей был изучен на системе никотин-вода. Здесь существуют *две* критические температуры смешения: *нижняя* K_2 и *верхняя* K_1 (рис. 142). Ниже 61 °C и выше 208 °C обе жидкости смешиваются в любых соотношениях, а при промежуточных температурах происходит расслоение смеси на две фазы: водную с большим содержанием воды и никотиновую с большим содержанием никотина.

Впрочем, для большинства систем критические температуры смешения практически не достигаются, и растворимость при всех возможных условиях опыта остается либо ограниченной, либо неограниченной. Достижению нижней критической температуры смешения препятствует затвердевание одного из компонентов смеси, достижению верхней — закипание.

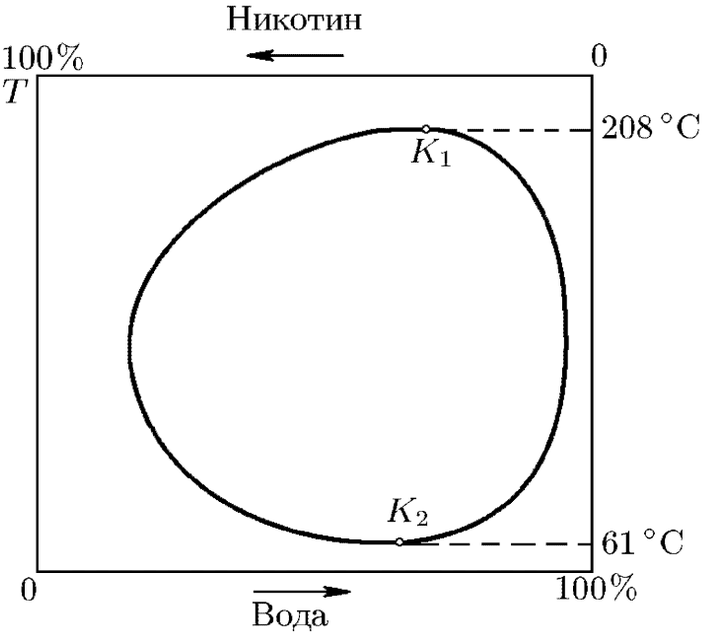


Рис. 142

7. Твердые тела растворяются в других твердых телах очень редко. Подавляющее большинство твердых тел совсем не растворимы друг в друге. Однако встречаются исключения из этого правила. Существуют твердые тела, образующие растворы в других твердых телах. Такие растворы называются *твердыми растворами*. Чаще всего приходится встречаться с твердыми растворами химических элементов. Некоторые химические элементы растворяются друг в друге в неограниченных количествах. Таковы, например, золото и серебро или медь и серебро. Твердые растворы бывают двух типов: «*типа внедрения*» и «*типа замещения*».

В растворах типа внедрения атомы растворенного вещества внедряются между узлами кристаллической решетки растворителя, несколько раздвигая при этом атомы последнего. Естественно, что это происходит в тех случаях, когда атомы растворенного вещества

значительно меньше атомов растворителя. К рассматриваемому типу относится большинство растворов углерода. Таков, например, раствор углерода в железе — аустенит.

Более распространенными являются растворы типа замещения. В этих твердых растворах атомы растворенного вещества вытесняют из кристаллической решетки некоторые атомы растворителя, а сами становятся на их место. В результате часть узлов кристаллической решетки оказывается занятой атомами растворителя, а другая часть — атомами растворенного вещества. К растворам типа замещения относится огромное большинство твердых сплавов металлов друг с другом, например сплав золота и серебра. По типу замещения всегда образуются растворы химических соединений друг в друге, так как молекулы химических соединений слишком велики и по этой причине не способны занимать места между узлами кристаллической решетки.

ЗАДАЧА

Показать, что при одной и той же температуре насыщенный пар имеет один и тот же состав и одинаковые давления над насыщенным раствором жидкости 1 в жидкости 2 и над насыщенном раствором жидкости 2 в жидкости 1.

Указание. Рассмотреть сообщающиеся сосуды, один из которых наполнен первым из указанных растворов, а другой — вторым. Пренебречь разницей давлений насыщенных паров над уровнями жидкостей в сосудах, обусловленной силой тяжести.

§ 124. Осмос и осмотическое давление

1. Пусть два раствора отделены один от другого пористой перегородкой, через которую могут проходить как молекулы растворителя, так и молекулы растворенного вещества. Если концентрации растворенного вещества по разные стороны перегородки различны, то начнется переход молекул из одного раствора в другой. Макроскопически он будет продолжаться до тех пор, пока концентрации обоих растворов не выравняются.

Существуют перегородки, проницаемые для молекул растворителя, но непроницаемые для молекул растворенного вещества. Они называются *полупроницаемыми*. К ним относятся различные оболочки и ткани растительного и животного происхождения. Для физических опытов более удобны искусственные полупроницаемые перегородки. Таковы, например, пленки из железисто-синеродистой меди $[Cu_2Fe(CN)_6]$, получаемые при взаимодействии растворов сернокислой меди и железистосинеродистого калия:

$$2CuSO_4 + K_4Fe(CN)_6 = Cu_2Fe(CN)_6 + 2K_2SO_4.$$

Они проницаемы для воды, но непроницаемы для многих растворенных веществ, например сахара. Однако эти пленки непрочны. Поэтому железистосинеродистую медь обычно осаждают в стенках мелкопористых глиняных или фарфоровых сосудов. Для этого сосуд наполняют

раствором $K_4Fe(CN)_6$, а затем опускают на некоторое время в раствор $CuSO_4$. Таким путем приготовлял полупроницаемые перегородки Пфеффер (1845–1920), впервые систематически изучивший законы осмотического давления.

Когда раствор отделен от чистого растворителя полупроницаемой перегородкой, то через нее молекулы чистого растворителя переходят в область, занятую раствором. Это явление называется *осмосом*. В результате осмоса возникает разность давлений между раствором и чистым растворителем. Когда она достигает определенного значения, то осмос прекращается. Разность давлений, при которой осмос прекращается, называется *осмотическим давлением*. Осмос происходит также в тех случаях, когда полупроницаемая перегородка разделяет два раствора различной концентрации.

Если, например, животный пузырь, наполненный спиртом, крепко завязать и опустить в воду, то он начнет раздуваться и даже может лопнуть под действием сил осмотического давления. Если же пузырь наполнить водой и опустить в спирт, то, наоборот, он будет сжиматься. Дело в том, что стенки пузыря проницаемы для воды, но непроницаемы для спирта. Подобные же явления набухания и разрыва оболочек наблюдаются при погружении в воду семян растений, фруктов или овощей. Ими объясняется «оживление» увядших цветов, опущенных в воду.

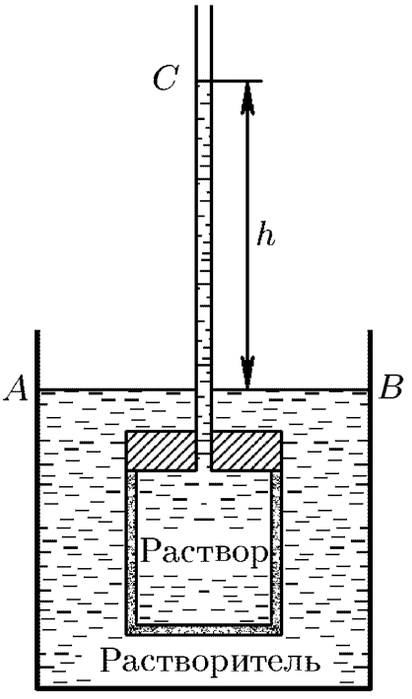


Рис. 143

2. Механизм осмоса и действия полупроницаемой перегородки нельзя считать полностью выясненным. Согласно одной из точек зрения прохождение молекул растворителя через полупроницаемую перегородку осуществляется путем предварительного растворения их в материале перегородки. Этот процесс имеет тенденцию к насыщению по отношению к чистому растворителю. Однако еще до достижения этого насыщения раствор жидкости в перегородке оказывается пересыщенным по отношению к раствору, находящемуся по другую сторону перегородки (так как последний обеднен растворителем благодаря присутствию растворенного вещества). Тогда перегородка начнет отдавать в жидкий раствор часть растворенной в ней жидкости, растворяя при этом новые порции чистой жидкости. Растворенное вещество не проникает из жидкого раствора через перегородку, так как она не растворяется в ней. Таким образом, свойство полупроницаемости обусловлено не размерами пор перегородки, а различной растворимостью в ее материале соприкасающихся с ней веществ.

3. Приборы, служащие для измерения осмотического давления, называются *осмометрами*. На рис. 143 изображен осмометр, которым

пользовался Пфеффер в своих исследованиях по осмотическому давлению (1877 г.). В порах глиняного сосуда осаждена железистосинеродистая медь по способу, описанному выше. Сосуд наполняется водным раствором исследуемого вещества, например сахара.

Горло сосуда заткнуто плотно пригнанной пробкой с отверстием, в которое вставлена длинная вертикальная трубка. Если сосуд погрузить в чистую воду, то она начнет проникать внутрь сосуда через полупроницаемые стенки. Когда уровень жидкости в трубке достигнет определенной высоты, дальнейшее проникновение воды в сосуд прекратится, так как этому будет препятствовать гидростатическое давление столба жидкости в трубке. Вес этого столба между уровнями AB и C и будет служить мерой осмотического давления: $P_{\text{осм}} = \rho g h$, где ρ — плотность раствора, а h — высота указанного столба жидкости.

4. Нетрудно понять происхождение осмотического давления. Так как растворитель может свободно проходить через полупроницаемую перегородку, то в состоянии равновесия числа молекул растворителя, ударяющихся о перегородку с противоположных сторон, должны быть одинаковы. Эти удары, уравновешивая друг друга, не будут оказывать никакого давления на перегородку. Но удары о перегородку молекул растворенного вещества никакими противоположно направленными молекулярными ударами не уравновешиваются. Они и приводят к возникновению осмотического давления.

Для разбавленных растворов нетрудно вычислить и величину осмотического давления. Согласно теореме о равномерном распределении кинетической энергии по степеням свободы, скорость теплового движения молекул зависит только от температуры. При одной и той же температуре средняя кинетическая энергия теплового движения молекул одинакова для жидкостей и газов. Кроме того, если раствор слабый, то число ударов молекул растворенного вещества о стенку сосуда будет таким же, как и для идеального газа с той же концентрацией молекул. Поэтому будет таким же и давление молекул растворенного вещества, т. е. осмотическое давление. Таким образом,

$$P_{\text{осм}} = \frac{NkT}{V} = \nu \frac{RT}{V}, \tag{124.1}$$

где N и ν — числа молекул и молей растворенного вещества соответственно в объеме раствора V. Формула (124.1) показывает, что *осмотическое давление в слабых растворах подчиняется тем же законам, что и давление идеальных газов*. Этот закон был установлен Вант Гоффом (1852–1911) на основе опытов Пфеффера.

Величина осмотического давления слабых растворов не зависит от природы растворителя и растворенного вещества, а только от молярной объемной концентрации последнего. Например, если в литре раствора содержится 1 моль растворенного вещества, то по формуле (124.1) осмотическое давление при 15 °C (288 K) будет равно

$$P_{\text{осм}} = \frac{288 \cdot 0{,}0821}{1} = 23{,}6 \text{ атм}.$$

Мы видим, что осмотическое давление довольно значительно. Однако стенки сосуда, содержащего раствор, его не воспринимают. Дело в том, что осмотическое давление действует и на свободную поверхность жидкости. Это приводит к растяжению жидкости. Возникающие силы натяжения компенсируют осмотическое давление. На стенки сосуда действует только гидростатическое давление.

В растворах электролитов молекулы растворенного вещества диссоциируют на ионы. Из-за этого осмотическое давление оказывается больше вычисленного по формуле (124.1) без учета диссоциации. Однако если под N понимать общее число растворенных частиц (ионов обоих знаков и нейтральных молекул), то формула (124.1) опять становится верной (разумеется, для слабых растворов). Ионы и нейтральные молекулы могут сольватироваться или гидратироваться. Но это обстоятельство не меняет общего числа частиц, производящих осмотическое давление, а потому не сказывается на справедливости формулы (124.1).

Формула (124.1) служит в химии для определения молекулярных масс таких химических соединений, которые не могут быть получены в газообразном состоянии, например белков и полимеров.

5. Осмос играет важную роль в жизнедеятельности животных и растений. Обмен веществ в организме осуществляется посредством соков и крови — растворов, омывающих полупроницаемые перегородки растительных и животных клеток. Осмотическое давление клеток многих растений составляет 5–20 атм. Только благодаря этому вода из почвы может поступать на большую высоту по стволам деревьев (например, эвкалиптов).

Величина осмотического давления крови человека составляет 7,6–7,9 атм. Однако разность осмотических давлений крови и лимфы, имеющая значение для перехода воды между ними, составляет всего 0,03–0,04 атм. Вообще, благодаря наличию сложных механизмов регулирования, клетки обладают лишь незначительно повышенным или равным осмотическим давлением по отношению к омывающим их внутренним жидкостям организма. Падение осмотического давления в клетках, например при обезвоживании организма, приводит к их коллапсу. Обессоливание организма может привести к набуханию и разрыву клеток (осмотический шок).

§ 125. Закон Рауля

Если в жидкости растворено нелетучее вещество, то свободная поверхность жидкости будет вести себя как полупроницаемая перегородка. Через нее могут свободно проходить молекулы растворителя, но не могут проходить молекулы растворенного вещества. Легко понять, что при одной и той же температуре давление насыщенного пара над раствором будет меньше, чем над растворителем. Убедиться в этом и вычислить величину понижения давления насыщенного пара можно с помощью тех же рассуждений, какие мы применяли в § 118 при

исследовании зависимости давления насыщенного пара от кривизны поверхности жидкости.

Пусть раствор помещен в сосуде осмометра и вся система покрыта сверху колпаком (рис. 144). Обозначим через P_0 давление насыщенного пара над поверхностью растворителя, а через P — давление над поверхностью раствора в трубке осмометра. В состоянии равновесия давление насыщенного пара в трубке осмометра должно равняться давлению насыщенного пара снаружи на той же высоте. Если бы это было не так, то с помощью осмометра, как это следует из рассуждений § 118, можно было бы осуществить перпетуум мобиле второго рода. Таким образом, если пренебречь изменением плотности пара в пределах высоты h, то должно быть $P_0 - P = \rho_{\text{п}} g h$. Кроме того, $P_{\text{осм}} = \rho_{\text{ж}} g h$. Из этих соотношений получаем

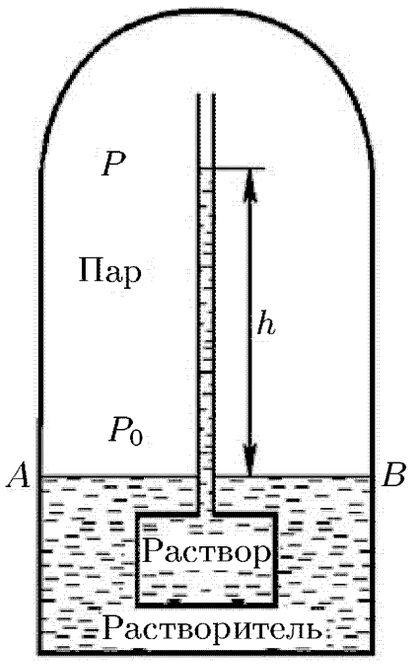


Рис. 144

$$\frac{P_0 - P}{P_0} = \frac{\rho_{\text{п}}}{\rho_{\text{ж}}} \frac{P_{\text{осм}}}{P_0}.$$

Если раствор слабый, то различием между плотностью раствора и растворителя можно пренебречь. В этом приближении $\rho_{\text{п}}/\rho_{\text{ж}} = n_{\text{п}}/n$, где $n_{\text{п}}$ — число молекул пара, а n — растворителя в единице объема. Кроме того, по закону Вант Гоффа $P_{\text{осм}} = n'kT$, а по уравнению Клапейрона $P_0 = nkT$ (n' — число молекул растворенного вещества в единице объема раствора). Используя эти соотношения, нетрудно получить

$$\frac{P_0 - P}{P_0} = \frac{n'}{n} = \frac{\nu'}{\nu}, \tag{125.1}$$

где ν' и ν — числа молекул растворенного вещества и растворителя в единице объема раствора. Таким образом, *относительное понижение давления насыщенного пара растворителя над поверхностью слабого раствора нелетучего вещества равно отношению числа молей растворенного вещества к числу молей растворителя*. Этот закон был установлен французским химиком Раулем (1830–1901) и носит его имя.

§ 126. Повышение точки кипения и понижение точки замерзания раствора

1. Пусть AB — кривая испарения чистого растворителя (рис. 145). Кривая испарения $A'B'$ для насыщенного раствора нелетучего вещества должна идти ниже, так как давление насыщенного пара над раствором ниже, чем над чистым растворителем. Пусть P означает внешнее давление. Поддерживая его постоянным, будем нагревать растворитель и раствор, т. е. изменять их состояние вдоль изобары AC.

В точках A и A', где изобара пересекает кривые испарения растворителя и раствора, эти жидкости закипят. Обозначим температуры этих точек через T и T' соответственно. Мы видим, что *при одном и том же давлении температура кипения раствора T' выше температуры кипения чистого растворителя T.*

Для вычисления T' через точку A' проведем изотерму $A'D$. Она пересечет кривую испарения чистого растворителя в точке B, давление в которой обозначим через P'. Очевидно, T' будет температурой кипения чистого растворителя под давлением P'. Если раствор слабый, то стороны треугольника $AA'B$ можно считать прямолинейными. По закону Рауля

$$\frac{P'-P}{P'}=\frac{\nu'}{\nu}.$$

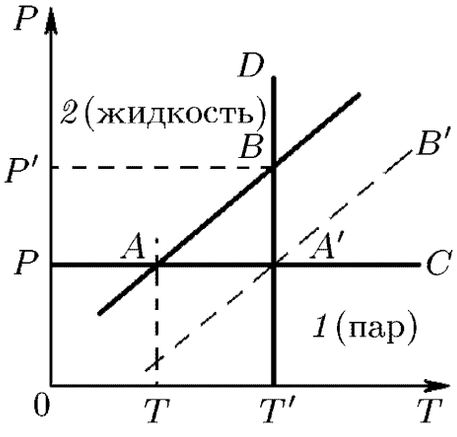


Рис. 145

По уравнению Клапейрона–Клаузиуса, если пренебречь удельным объемом жидкости по сравнению с удельным объемом пара,

$$\frac{P'-P}{T'-T}=\frac{q_{12}}{Tv_{\text{п}}}=\frac{q\mu P}{RT^2},$$

где q_{12} — удельная теплота испарения, а μ — молярная масса пара растворителя. Почленным делением находим

$$\frac{T'-T}{P'}=\frac{\nu'}{\nu}\,\frac{RT^2}{\mu q_{1\,2}P},$$

или, пренебрегая различием P и P',

$$T'-T=\frac{\nu'}{\nu}\,\frac{RT^2}{\mu q_{1\,2}}. \qquad (126.1)$$

Этот результат можно также представить в виде

$$T'-T=\frac{P_{\text{осм}}T}{\nu\mu q_{1\,2}}. \qquad (126.2)$$

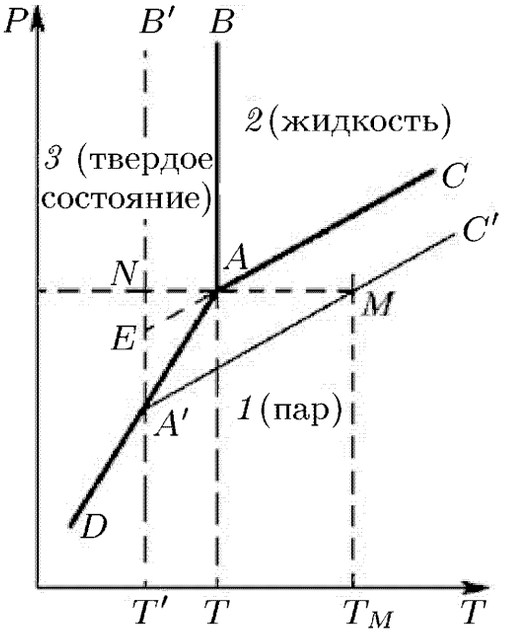


Рис. 146

2. Аналогично покажем, что *температура замерзания раствора нелетучего вещества T' ниже температуры замерзания чистого растворителя T.* Пусть A — тройная точка чистого растворителя (рис. 146), в которой сходятся кривая плавления AB, кривая испарения AC и кривая возгонки AD. Поскольку в дальнейшем мы будем иметь дело с малыми изменениями температуры и давления, нам понадобятся только малые отрезки этих кривых. Их, как и в предыдущем рассмотрении, можно считать прямолинейными.

Существенно, что кривая возгонки DA поднимается круче кривой испарения AC (см. §116). Что касается кривой плавления AB, то она поднимается настолько круто ($|v_{ж} - v_{тв}| \ll v_{п}$), что ее можно считать вертикальной. Иными словами, можно считать, что температура плавления не зависит от давления и равна температуре в тройной точке A. Кривая испарения раствора $A'C'$, как мы видели, идет ниже кривой испарения чистого растворителя AC. Поэтому тройная точка чистого растворителя A' лежит левее тройной точки раствора A. Проведя через A' вертикальную прямую, найдем кривую плавления раствора $A'B'$. Ее смещение AN относительно AB дает понижение точки плавления раствора. С другой стороны, длина отрезка AM дает повышение точки кипения раствора. Как видно из рисунка,

$$\frac{NA}{AM} = \frac{EN}{A'E}.$$

Из уравнения Клапейрона–Клаузиуса для возгонки и испарения следует

$$\frac{A'N}{EN} = \frac{q_{13}}{q_{12}}.$$

Отсюда

$$\frac{A'N - EN}{EN} = \frac{A'E}{EN} = \frac{q_{13} - q_{12}}{q_{12}} = \frac{q_{23}}{q_{12}}.$$

Замечая еще, что $AM = \dfrac{\nu'}{\nu}\dfrac{RT^2}{\mu q_{12}}$, $NA = T - T'$, получаем

$$T' - T = -\frac{\nu'}{\nu}\frac{RT^2}{\mu q_{23}}, \tag{126.3}$$

или

$$T' - T = -\frac{P_{осм}T}{\nu\mu q_{23}}. \tag{126.4}$$

Согласно (126.2) и (126.4) *повышение точки кипения и понижение точки затвердевания раствора зависит только от числа молей растворенного вещества, приходящихся на единицу объема растворителя, но не зависит от химической природы этого вещества.* В частности, если растворено два или несколько веществ, то изменения температур кипения и затвердевания растворов равны сумме изменений, вызванных каждым из этих веществ в отдельности. Однако эти изменения зависят от химической природы растворителя.

ЗАДАЧИ

1. Один грамм обыкновенного сахара $C_{12}H_{22}O_{11}$ растворен в 100 см^3 воды. Определить повышение точки кипения этого раствора при нормальном атмосферном давлении. Плотность воды при 100 °С равна 0,96 г/см^3.

Решение. Молярная масса сахара 342 г/моль; $\nu' = \dfrac{1}{342 \cdot 10^2}$; $\mu\nu = 0{,}96$; $q = 539$ кал/г; $R = 1{,}98$ кал/моль; $T = 373$ К.

Подставляя эти данные в формулу (126.1), получим $T' - T = 0{,}0156\,^\circ\mathrm{C}$.

2. В предыдущей задаче определить понижение точки замерзания раствора сахара.

Ответ. $T' - T = -0{,}0543\,^\circ\mathrm{C}$.

§ 127. Правило фаз

1. Рассмотрим термодинамическую систему, состоящую из нескольких фаз. Для характеристики состава фазы надо указать количества химически однородных веществ, из которых она состоит. Если фазы не находятся в равновесии, то эти количества могут меняться в более или менее широких пределах и притом независимо друг от друга. Однако это будет не так, если система находится в термодинамическом равновесии. В этом случае между веществами, входящими в состав фазы, будут существовать определенные количественные соотношения. Поэтому достаточно указать содержание не всех веществ, входящих в фазу, а только некоторых. Количества остальных определятся условиями термодинамического равновесия. *Минимальное число химически однородных веществ, заданием которых однозначно определяется состав каждой фазы системы в состоянии термодинамического равновесия, называется числом компонентов системы.* Сами эти вещества называются *независимыми компонентами*, или просто *компонентами*, системы. Их количества в системе могут задаваться произвольно и независимо друг от друга. Выбор компонентов не однозначен. Какие вещества принять за компоненты — это не имеет значения. Важно их число. Если число независимых компонентов одно, то система называется *однокомпонентной*, если два, то *двухкомпонентной*, или *бинарной*, и т. д.

Возьмем, например, равновесную систему, состоящую из жидкой воды, льда и водяного пара при определенных давлениях и температуре. Это однокомпонентная трехфазная система. В качестве компонента можно взять всю массу воды в системе, ибо условия термодинамического равновесия однозначно определяют, как эта масса распределяется между жидкой, твердой и газообразной фазами. В частности, в зависимости от значений P и T может оказаться, что одна или две из этих фаз будут отсутствовать. Но в качестве компонента можно взять, например, и весь водород, содержащийся в системе, так как его заданием также однозначно определится состав всех фаз системы. Действительно, число атомов водорода должно вдвое превышать число атомов кислорода, поскольку эти атомы не свободны, а входят в систему в связанном состоянии в виде молекул воды H_2O.

В анализе этого примера можно пойти дальше. При повышении температуры молекулы H_2O диссоциируют с образованием молекул O_2 и H_2. Водород и кислород растворяются в жидкой воде, причем растворимость кислорода больше. Кроме того, в воде присутствуют комплексы молекул H_2O типа $(H_2O)_n$. Если бы все эти частицы образовывались в системе в результате диссоциации и ассоциации молекул

воды, то их число в каждой фазе однозначно определялось бы условиями равновесия. Система по-прежнему была бы однокомпонентной. Именно так обстоит дело в отношении комплексов молекул $(H_2O)_n$. С молекулами O_2 и H_2 положение другое. К системе можно добавлять произвольное количество водорода и кислорода. При этом условии система становится двухкомпонентной. За независимые компоненты можно взять общую массу водорода и общую массу кислорода. Условия термодинамического равновесия позволяют однозначно определить, сколько молекул O_2 и H_2 останутся в свободном состоянии, а сколько соединятся в молекулы воды H_2O; сколько молекул O_2 и H_2 останется в газообразной фазе, а сколько перейдет в раствор, и т. д. Короче говоря, эти условия однозначно определяют состав каждой фазы системы.

В качестве другого примера рассмотрим насыщенный водный раствор какой-либо соли с нерастворенным твердым осадком в нем и насыщенным паром над этим раствором. В отношении твердой фазы этой системы предположим, что нерастворенная соль не содержит воды в произвольных количествах. Последняя может входить в молекулы соли в совершенно определенной пропорции в виде кристаллизационной воды. Рассматриваемая система является двухкомпонентной. Состав каждой фазы ее в состоянии равновесия однозначно определится, если задать, например, количество воды и безводной соли. В самом растворе могут присутствовать комплексы молекул воды $(H_2O)_n$, ионы, на которые распадаются молекулы соли, гидраты и пр. Однако, как уже указывалось выше, наличие таких частиц не влияет на число компонентов системы, так как мы не можем произвольно менять их количества.

Трудно указать общее правило для определения числа компонентов, хотя в частных случаях это обычно не встречает особых затруднений.

2. Каждый компонент в системе распределяется по фазам, из которых она состоит. Если фазы граничат друг с другом, то компоненты могут переходить из одной фазы в другую. В состоянии термодинамического равновесия устанавливается вполне определенное распределение компонентов по фазам. Выведем необходимые условия термодинамического равновесия. Допустим, что равновесие установилось. Необходимо, чтобы оно сохранялось по отношению к переходу каждого компонента из одной фазы в другую. Поэтому при рассмотрении нарушений равновесия можно предположить, что количества компонентов во всех фазах остаются неизменными, за исключением одного, который может переходить из одной фазы в другую. Тогда система будет вести себя как однокомпонентная, и к ней применимо условие фазового равновесия (112.1), выведенное для химически однородной (однокомпонентной) системы. Таким образом, *если фазы граничат друг с другом, то в состоянии равновесия удельные термодинамические потенциалы каждого компонента во всех фазах должны быть одинаковы*. Кроме того, во всех фазах должны быть одинаковы температуры и давление.

3. Пусть число компонентов в системе равно k, а число фаз, находящихся в равновесии, n. Тогда условия равновесия можно записать в виде

$$\begin{aligned} \varphi_1^{(1)} &= \varphi_1^{(2)} = \ldots = \varphi_1^{(n)}, \\ \varphi_2^{(1)} &= \varphi_2^{(2)} = \ldots = \varphi_2^{(n)}, \\ \varphi_k^{(1)} &= \varphi_k^{(2)} = \ldots = \varphi_k^{(n)}. \end{aligned} \tag{127.1}$$

Нижний индекс означает номер компонента, а верхний, заключенный в скобки, — номер фазы. Так, $\varphi_2^{(3)}$ означает удельный термодинамический потенциал второго компонента в третьей фазе. Каждый из этих потенциалов определяется составом соответствующей фазы и зависит также от температуры T и давления P, которые одинаковы для всей системы. Состав фаз определяется концентрациями компонентов в ней, т. е. отношениями масс компонентов в фазе к массе всей фазы. Эти концентрации, однако, не независимы, так как их сумма равна единице. Поэтому состав фазы определяется не k, а $k-1$ независимыми аргументами. Число таких аргументов во всех n фазах будет $n(k-1)$. Присоединив к ним температуру и давление, получим всего $n(k-1)+2$ неизвестных, входящих в уравнения (127.1). Число уравнений равно $k(n-1)$. Чтобы эти уравнения не были противоречивыми, необходимо, вообще говоря, чтобы это число не превышало числа неизвестных, т. е. $k(n-1) \leqslant n(k-1)+2$. Отсюда получаем

$$n \leqslant k+2. \tag{127.2}$$

Таким образом, *число фаз, которые могут находиться в равновесии между собой, может превышать число компонентов не более чем на два*. Это положение было установлено Гиббсом и называется *правилом фаз Гиббса*. В частном случае однокомпонентных систем оно уже было доказано нами в § 116.

Приведенное доказательство предполагает, что $n(k-1)$ уравнений, образующих систему (127.1), *независимы*. Обоснованием этого предположения может служить следующее замечание. Так как число различных атомов, а с ними и химических соединений конечно, то крайне невероятно, что среди функций $\varphi_1^{(1)}, \varphi_2^{(1)}, \ldots$ найдутся одинаковые. Столь же невероятно, что между этими функциями существует какая-то функциональная связь. Можно не считаться с возможностью таких связей по той же причине, по которой мы не считаемся с возможностью столкновений между молекулами газа, когда рассматриваем их как геометрические точки.

Кроме того, доказательство предполагает, что физическое состояние фазы, помимо ее состава, определяется еще *двумя* параметрами: давлением P и температурой T. Если систему поместить в магнитное поле, то к этим параметрам надо добавить также напряженность магнитного поля. Тогда вместо (127.2) мы получили бы $n \leqslant k+3$. Вообще

$$n \leqslant k+r,$$

где r — число независимых параметров, определяющих физическое состояние фазы при заданном ее составе. Но мы ограничимся в дальнейшем случаем $r = 2$.

4. Физическое состояние каждой фазы определяется $k+1$ координатами: температурой, давлением и $k-1$ концентрациями компонентов. На геометрическом языке это означает что состояние определяется точкой в пространстве $k+1$ измерений, по координатным осям которого отложены значения указанных координат. Такое пространство будем называть *изображающим пространством*, а его точки, изображающие состояние системы — *изображающими точками*. Для однокомпонентных систем изображающее пространство будет двухмерным (плоскость), для двухкомпонентных — трехмерными, для систем с большим числом компонентов — многомерным. С другой стороны, общее число параметров, определяющих состояние неравновесной системы, как мы видели, равно $k(n-1)+2$. Если n фаз, из которых состоит система, находится в равновесии друг с другом, то эти параметры не независимы, а связаны $k(n-1)$ соотношениями. Число независимых параметров, называемое *числом степеней свободы*, или *вариантностью*, термодинамической системы, будет $f = n(k-1) + 2 - k(n-1)$, т. е.

$$f = k + 2 - n. \tag{127.3}$$

Отсюда следует, что совокупность точек изображающего пространства, в которых находятся в равновесии n фаз, образует в этом пространстве подпространство f измерений. Для однофазной системы ($n = 1$) получаем $f = k + 1$, т. е. возможные изображающие точки заполняют область $k+1$ измерений. При $n = 2$ ($f = k$) изображающие точки располагаются в подпространстве k измерений, при $n = 3$ ($f = k - 1$) — в подпространстве $k-1$ измерений и т. д. Наконец, при $n = k+2$ ($f = 0$) подпространство изображающих точек вырождается в точку. По аналогии с тройной точкой ее можно называть $(k+2)$-*кратной точкой*. При $f = 0$ система называется *инвариантной*, или *нонвариантной*. Для нее давление, температура и все концентрации определены однозначно. При $f = 1$ ($n = k + 1$) система называется *моновариантной*, или *универиантной*, при $f = 2$ ($n = k$) — *дивариантной*, или *бивариантной*, при $f \geqslant 3$ ($n \leqslant k-1$) — *поливариантной*.

§ 128. Диаграммы состояния бинарных смесей

1. Как мы видели, состояние многокомпонентной системы можно представлять точкой в пространстве $k + 1$ измерений (k — число компонентов системы). По осям координат откладываются температура T, давление P и концентрации каких-либо $k-1$ компонентов. В изображающем пространстве можно выделить подпространства меньшего числа измерений, точки которых соответствуют равновесию двух или нескольких фаз. Если это сделать, то получится так называемая *диаграмма состояния системы*. При $k = 1$ и $k = 2$ изображающими пространствами будут плоскость и обыкновенное трехмерное

пространство. В этих случаях указанный графический метод обладает геометрической наглядностью и часто применяется. Случай $k = 1$ был рассмотрен в предыдущей главе. Обратимся теперь к случаю $k = 2$ (двухкомпонентные, или бинарные, системы). Диаграмма состояния трехмерна. Обычно рассматривают сечение ее либо плоскостью $T =$ $=$ const, либо плоскостью $P =$ const. Для конкретности будем пользоваться сечениями $P =$ const. Так мы поступали уже в § 123 при рассмотрении взаимной растворимости тел. Для иллюстрации метода диаграмм состояния рассмотрим кипение жидких смесей и плавление твердых сплавов.

2. Закономерности кипения жидких смесей эмпирически были исследованы русским химиком Д. П. Коноваловым (1856–1929). Их термодинамическое обоснование было дано Гиббсом. Допустим, что две жидкости смешиваются друг с другом в произвольных количествах. Примером может служить смесь жидких азота и кислорода (рис. 147). Как показали экспериментальные исследования, диаграмма состояния такой смеси (точнее, ее сечение плоскостью $P =$ const) состоит из тех областей. Заштрихованная область изображает двухфазное состояние системы, состоящее из жидкой смеси и ее пара. Эта область ограничена снизу *кривой жидкости* ALB, сверху — *кривой пара* AGB. Ниже кривой ALB система однофазна и состоит из однородной смеси жидкостей, выше кривой AGB — также однофазна, но состоит из смеси паров тех же жидкостей. Диаграмма состояния такого типа называется «сигарой». Левая вертикальная прямая соответствует стопроцентному содержанию жидкости *1* (азот), правая — стопроцентному содержанию жидкости *2* (кислород). Пересечем «сигару» горизонтальной прямой M_1GCLM_2. В точке L система находится в жидком состоянии, причем количества жидкостей *1* и *2* в ней относятся как длины отрезков M_2L и M_1L. В точке G система газообразна с относительными содержаниями жидкостей *1* и *2*, пропорциональными соответственно отрезками M_2G и M_1G. Наконец, в точке C имеется смесь пара и жидкости. Массы азота и кислорода в ней относятся как длины отрезков M_2C и M_1C, а массы жидкости и пара — как длины отрезков GC и LC соответственно (правило рычага).

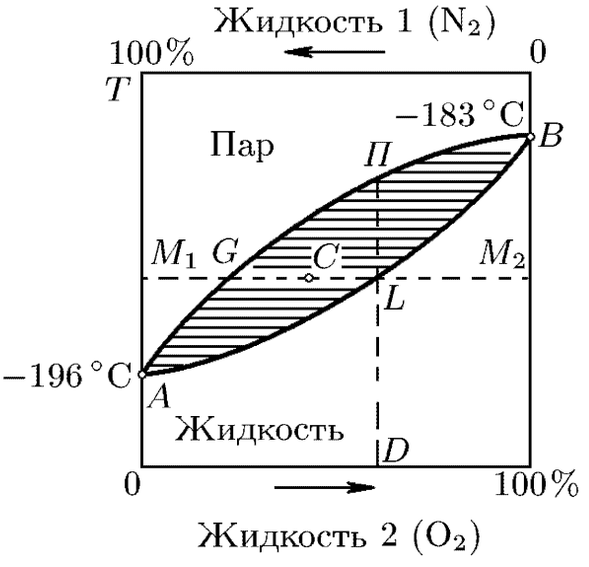


Рис. 147

Исследуем теперь характер кипения жидкости. Допустим сначала, что жидкость нагревается в закрытом сосуде, например в цилиндре с поршнем, причем внешнее давление на поршень поддерживается постоянным. Образующиеся пары остаются в системе, так что ее хими-

ческий состав не меняется. Поэтому процесс нагревания изображается вертикальной прямой $DL\Pi$. В точке L жидкость начинает кипеть. По мере нагревания температура кипения повышается и одновременно увеличивается содержание пара в системе. В точке Π, где система состоит только из одного пара, кипение прекращается. Таким образом, у смеси жидкостей нет определенной точки кипения. Все точки кипения (при данном составе и давлении) располагаются в пределах вертикального отрезка $L\Pi$, причем каждой точке кипения соответствует вполне определенное соотношение между количествами пара и жидкости.

Допустим теперь, что нагревание производится в открытом сосуде. Образующиеся пары удаляются из системы. Достигнув точки L, смесь жидкостей снова закипит. Однако состояние насыщенного пара смеси при температуре точки L изображается не этой точкой, а точкой G, расположенной левее. Насыщенный пар смеси обладает бóльшим содержанием азота, чем смесь в точке L. Образующиеся пары являются насыщенными, а потому уносят из системы относительно больше азота, чем кислорода. Система при нагревании будет обогащаться кислородом. Нагревание изображается линией LB. В конце концов, когда будет достигнута точка B, в системе останется один только кислород. На этом явлении основана *ректификация*, т. е. разделение жидкостей, состоящих из двух иди большего числа компонентов. В этом методе смесь, подлежащая разделению, подвергается многократному испарению и конденсации образующихся паров.

Если проанализировать экспериментальные диаграммы состояния бинарных смесей органических и неорганических жидкостей, то под тип «сигары» подходит столь большое количество их, что этот тип с полным основанием можно назвать «*нормальным*». Диаграммы такого типа дают в особенности вещества, близкие по химическому составу (бензол-толуол, гексан-октан, метиловый и этиловый спирты и т. д.).

3. Существуют, однако, диаграммы состояния другого типа. Они представлены на рисунках 148 и 149. Например, для смеси ацетона

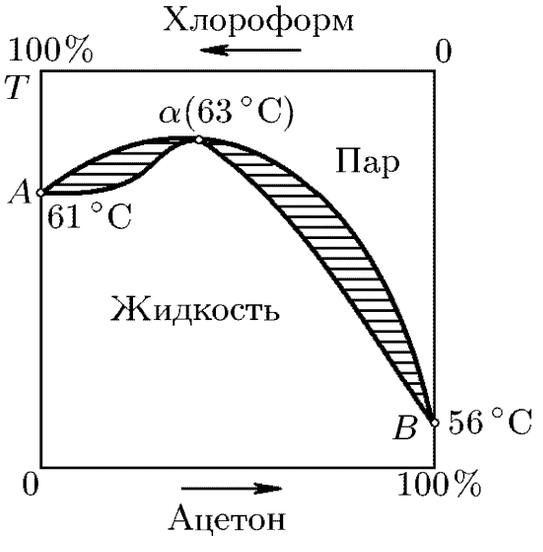


Рис. 148

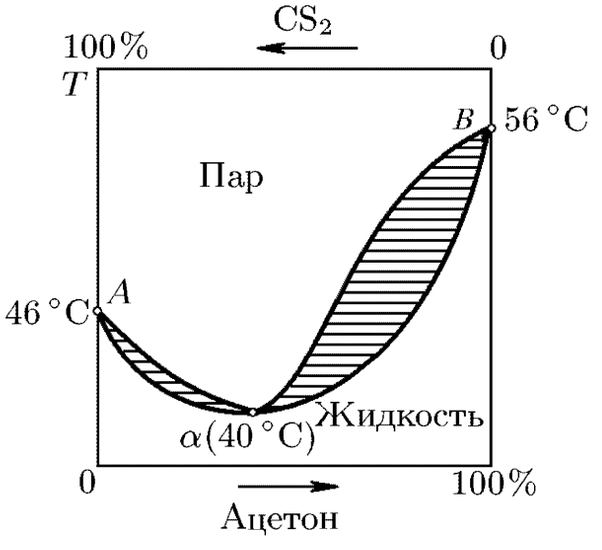


Рис. 149

и хлороформа или смеси H_2O с HNO_3 кривые жидкости и пара имеют максимум в точке α, где они касаются друг друга. Для смеси ацетона и сероуглерода или смеси воды и этилового спирта эти кривые имеют минимум, где они также касаются между собой. Диаграмма состояния с максимумом встречаются реже, чем с минимумом. На диаграммах состояния как того, так и другого типа состав жидкости и пара в точке касания α один и тот же. Смесь, состав которой соответствует точке α, называется *азеотропной*.

Допустим, что нагревание производится в открытом сосуде. Азеотропная смесь будет иметь вполне определенную температуру кипения, выкипая целиком, подобно химически чистому веществу. Если же смесь не азеотропна, то в случае рис. 148 кипение приводит к перемещению изображающей точки в положение α, независимо от того, каково было начальное состояние. Таким образом, в результате кипения получается азеотропная смесь жидкостей, которая дальше кипит как химически однородная жидкость. В случае рис. 149 нагревание жидкой смеси приводит к перемещению изображающей точки либо в положение B, либо в положение A, в зависимости от того, находилась ли начальная точка правее или левее точки α.

4. Приведенная выше диаграмма состояния смеси азота и кислорода (рис. 147) имеет форму сигары при атмосферном давлении. При повышении давления характер диаграммы не изменяется. Так будет, однако, происходить до тех пор, пока мы не достигнем критического давления азота, равного 33,5 атм (критическое давление кислорода выше и равно 49,7 атм). При таком и более высоких давлениях разделение чистого азота на жидкую и газообразную фазы, становится уже невозможным. Поэтому ясно, что левый конец «сигары» должен оторваться от соответствующей вертикальной прямой, и диаграмма состояния приобретает вид, схематически представленный на рис. 150. Кривые пара и жидкости смыкаются в некоторой точке K, называемой *критической*. Наличие такой точки означает, что разделение смеси на жидкую и газообразную фазы возможно только для состояний, ограниченных замкнутой кривой $BGKB$. Вне этой кривой двухфазное состояние невозможно, а подразделение вещества на жидкость и газ приобретает условный характер.

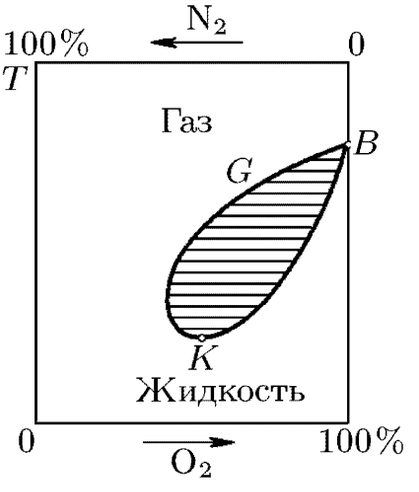


Рис. 150

Отметим в связи с этим своеобразное явление, относящееся к конденсации газообразных смесей. При рассмотрении этого явления более удобно пользоваться диаграммой не T, c, а P, c (c — концентрация). Возьмем участок диаграммы состояния в окрестности критической точки (рис. 151). В переменных P, c область двухфазного состояния расположена ниже, а не выше критической точки. При изотермическом сжатии состояние системы

будет изменяться вдоль вертикальной прямой AB. Конденсация газовой смеси начинается в точке A. Здесь количество жидкости равно нулю. По мере сжатия жидкая фаза будет расти. В точке C отношение масс жидкости и пара равно отношению отрезков MC и CN. Однако при дальнейшем сжатии количество жидкости опять начинает уменьшаться, пока в точке B жидкость не исчезнет совсем. Это явление называется *обратной конденсацией*.

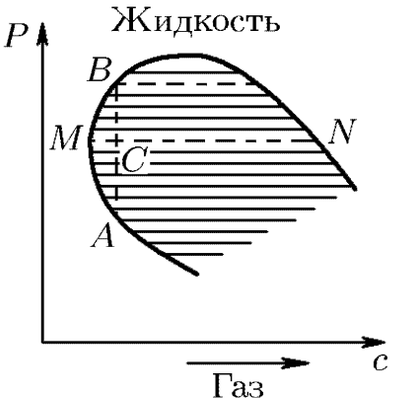


Рис. 151

5. Рассмотрим теперь диаграммы состояния твердых и жидких смесей. Возможны диаграммы типа сигары (например, для Ag—AU, Cu—Ni, AgCl—NaCl), диаграммы с минимумом (например, Cr—Fe, Ti—Zr, LiCl—NaCl) и максимумом (например, Na_2SO_4—$CaSO_4$, стеарин–пальмитин). Ко всем этим диаграммам применимы рассуждения, приведенные выше для жидких и газообразных смесей. Например, на рис. 152 приведена диаграмма состояния для сплава золота и серебра. Область над верхней кривой изображает жидкое, под нижней кривой — твердое состояние тела. Заштрихованная область между обеими кривыми — двухфазную систему, состоящую из жидкости и твердого тела. При нагревании твердого сплава на

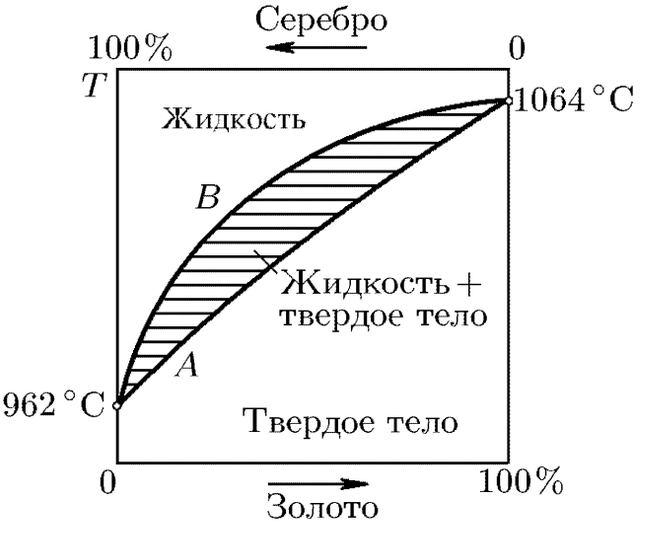


Рис. 152

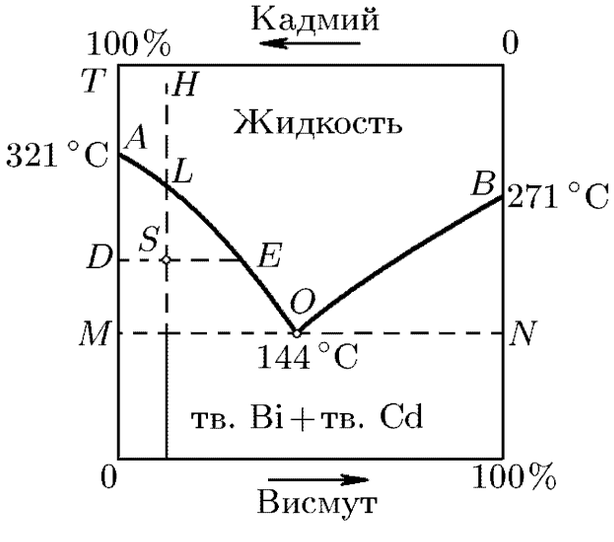


Рис. 153

кривой A начинается плавление. По мере повышения температуры увеличивается содержание жидкой смеси, пока на кривой B вся смесь не перейдет в жидкое состояние. Точки плавления смеси при заданном составе заполняют конечный температурный интервал, причем точке плавления соответствует вполне определенное соотношение между твердой и жидкой фазами.

6. Совсем другой характер имеют диаграммы состояния для систем висмут–кадмий (рис. 153) или свинец–сурьма. Они характеризуются тем, что оба компонента, входящие в состав системы, не образуют

смешанных кристаллов. Точки A и B являются точками плавления чистых кадмия и висмута. Добавление к кадмию висмута понижает точку затвердевания жидкого раствора висмута и кадмия. С увеличением концентрации висмута кривая затвердевания раствора идет вниз. Аналогично ведет себя кривая затвердевания раствора кадмия в висмуте. Обе кривые пересекаются в точке O, называемой *эвтектической точкой*. Область выше кривой AOB изображает однородную жидкую смесь висмута и кадмия. Область ниже горизонтальной прямой MON изображает двухфазную систему, состоящую из кристаллов висмута и кристаллов кадмия. Область, ограниченная замкнутой кривой $AOMA$, представляет двухфазную систему из кристаллов кадмия и жидкого расплава кадмия с висмутом. Состав жидкого расплава один и тот же во всех точках всякой горизонтальной прямой DE, пересекающей рассматриваемую область. Он определяется точкой пересечения E указанной прямой с участком кривой плавления AEO, где все вещество существует в виде жидкого расплава. Содержание жидкой и твердой частей в какой-либо точке S рассматриваемой области определяется обычным правилом рычага: количества твердого кадмия и жидкого расплава обратно пропорциональны длине отрезков DS и SE. Аналогично, область, ограниченная замкнутой кривой $OBNO$, представляет двухфазную смесь состоящую из кристаллов висмута и жидкого расплава этого металла с кадмием.

Рассмотрим процесс затвердевания жидкой смеси при ее охлаждении вдоль вертикальной прямой HL. В точке L из расплава начнут выделяться кристаллы вдоль кадмия. Система обогащается висмутом, и температура затвердевания ее понижается. Дальше процесс затвердевания идет вдоль кривой LEO до эвтектической точки O. В этой точке понижение температуры прекращается, пока не затвердеет вся жидкость. В точке O будут продолжать выпадать кристаллы оставшегося в жидкости кадмия и начнет кристаллизоваться висмут. При этом относительное содержание кадмия и висмута остается неизменным в течение всего процесса кристаллизации. Эвтектическая точка O является *равновесной точкой сосуществования трех фаз*: жидкой смеси кадмия с висмутом, кристаллов кадмия и кристаллов висмута. Поскольку в эвтектической точке кристаллизация обоих компонентов происходит одновременно, кристаллическая смесь в этой точке состоит из очень мелких кристалликов кадмия и висмута. Она называется *эвтектической смесью*, или *эвтектикой*. Слева от эвтектической точки в области, расположенной ниже прямой MN, в эвтектическую смесь вкраплены ранее выделившиеся более крупные кристаллы кадмия, а справа — висмута.

Глава XII

СИММЕТРИЯ И СТРОЕНИЕ КРИСТАЛЛОВ

§ 129. Симметрия тел

1. Симметрия тела выражает свойства его совмещаться с самим собой при определенных перемещениях, называемых *преобразованиями* или *операциями симметрии*. Эти перемещения не должны сопровождаться растяжениями, сжатиями, сдвигами и другими деформациями, при которых изменяются расстояния между точками тела. К преобразованиям симметрии относятся: 1) *параллельный перенос* всех точек тела на определенное расстояние (*трансляция*); 2) *поворот тела* вокруг некоторой оси на определенный угол; 3) *отражение в плоскости*; 4) *инверсия* или *отражение в точке*, а также все комбинации таких преобразований.

Под отражением тела в плоскости $П$ понимают операцию, в результате которой каждая точка тела переходит в точку, симметрично расположенную с ней относительно этой плоскости (рис. 154). Если плоскость $П$ принять за координатную плоскость XY прямоугольной системы координат, то при отражении в этой плоскости точка (x, y, z) переходит в точку $(x, y, -z)$. В случае двукратного и вообще четного числа отражений в одной и той же плоскости получается тождественное преобразование, при котором тело возвращается в исходное положение. Примером отражения в плоскости может служить переход от тела к его мнимому оптическому изображению в плоском зеркале.

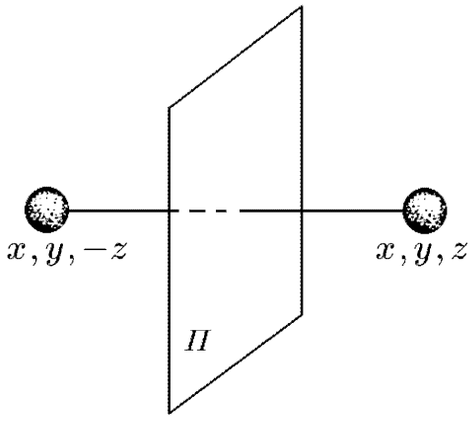


Рис. 154

Все операции симметрии могут быть сведены к последовательно выполняемым операциям отражения в плоскости. Так, поворот на угол α можно получить путем двух последовательных отражений в двух плоскостях OA и OB, пересекающихся на оси поворота O под углом $(1/2)\,\alpha$ (рис. 155). Если ось поворота O удалить в бесконечность, т. е. выполнить два отражения в двух параллельных плоскостях, то поворот перейдет в параллельный перенос (трансляцию). Если выполнить последовательно три отражения в координатных плоскостях $x = \text{const}$, $y = \text{const}$, $z = \text{const}$, то точка (x, y, z) перейдет в точку $(-x, -y, -z)$. В результате получается инверсия, или отражение в начале координат. Таким образом, *симметрию любого тела можно описать с помощью одних только операций отражения*. Од-

нако для большей наглядности удобнее при таком описании пользоваться и сложными операциями симметрии, сводящимися к последовательно выполняемым отражениям в плоскостях.

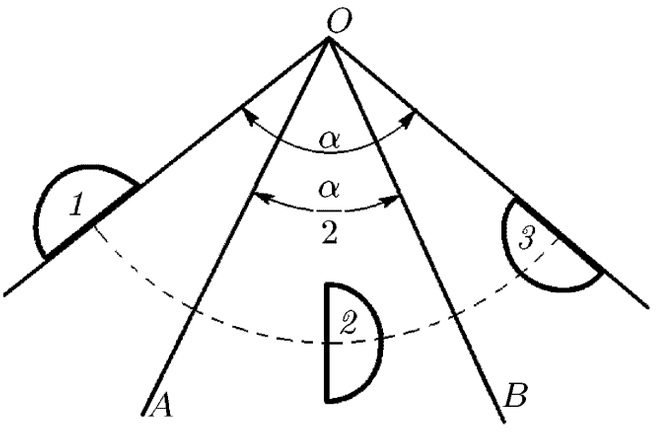


Рис. 155

Определенные геометрические точки, прямые и плоскости, симметрично расположенные относительно тела, называются его *элементами симметрии*. К ним относятся *ось симметрии*, *плоскость симметрии*, *зеркально-поворотная ось*, *центр симметрии* и пр. Совокупность всех элементов симметрии тела называется его *группой симметрии*. Группы симметрии, содержащие только операции отражения, поворота и инверсии, но не содержащие трансляций, называются *точечными группами*. Такие группы оставляют на месте по крайней мере одну точку тела и описывают симметрию *конечных фигур*: атомов, молекул, многогранников и пр. Группы симметрии, содержащие, наряду с перечисленными операциями, также трансляции, описывают симметрию *бесконечных систем* с периодической структурой. Они называются *пространственными группами*.

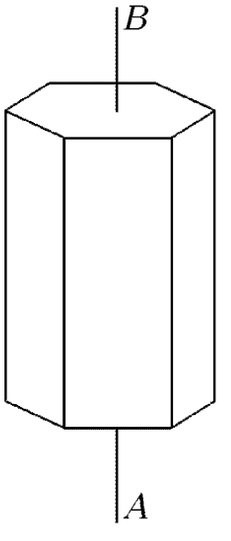


Рис. 156

2. Если тело переходит само в себя при повороте на угол $\varphi_n = 2\pi/n$ $(n = 1, 2, 3, 4, \ldots)$ вокруг некоторой оси, то эта ось называется *поворотной осью* или *осью симметрии n-го порядка*. В дальнейшем ради краткости условимся обозначать одним и тем же символом элемент симметрии и соответствующее ему преобразование. Так, поворотную ось n-го порядка и поворот вокруг нее на угол $2\pi/n$ будем обозначать одним и тем же символом C_n. Если $n = 1$, тело поворачивается на угол $\varphi_1 = 2\pi$, т. е. возвращается в исходное положение. Такой поворот, следовательно, есть тождественное преобразование. Ему самому по себе не соответствует никакая симметрия. При повороте на угол $\varphi = p(2\pi/n)$, где n — целое число, тело, очевидно, также переходит само в себя. Угол φ можно представить в виде $\varphi = 2\pi : (n/p)$. Отсюда видно, что если n кратно p, то рассматриваемая поворотная ось C_n будет одновременно поворотной осью более низкого порядка n/p, т. е. осью $C_{n/p}$. Так, геометрическая ось AB правильной шестигранной призмы (рис. 156) является поворотной осью шестого, третьего и второго порядков.

3. Если тело переходит само в себя в результате зеркального отражения в некоторой плоскости, то эту плоскость называют *плоскостью симметрии*. Ее, а также соответствующую операцию отражения обо-

значают буквой σ. Так, человеческое тело, если отвлечься от расположения внутренних органов (сердце находится слева), имеет плоскость симметрии, которая делит его на две похожие половины: правую и левую.

Наличие в теле поворотной оси любого порядка еще не означает, что в нем есть плоскость симметрии, проходящая через эту ось. Так, правильная шестигранная призма (рис. 156) имеет шесть плоскостей симметрии, проходящих через ось AB. Если взять совокупность таких призм с общей осью, произвольно повернутых относительно друг друга, то поворотная ось сохранится, однако плоскостей симметрии, проходящих через эту ось, вообще говоря, не будет.

4. Операция поворота тела вокруг неподвижной оси на угол $2\pi/n$ с одновременным отражением его в плоскости, перпендикулярной к той же оси, называется *зеркально-поворотным преобразованием*. Если в результате такого преобразования тело переходит само в себя, то соответствующую ось называют *зеркально-поворотной осью n-го порядка*. Так, система из четырех точек $ABCD$ на рис. 157 обладает зеркально-поворотной осью четвертого порядка. Очевидно, эта ось является также обычной поворотной осью второго порядка. Зеркально-поворотные преобразования и ось будем обозначать через S_n.

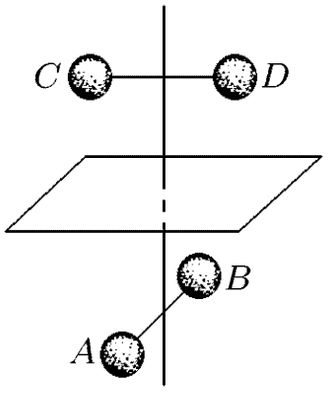


Рис. 157

Легко видеть, что при нечетном n зеркально-поворотная ось n-го порядка не является новым элементом симметрии, а сводится к комбинации поворотной оси n-го порядка C_n и перпендикулярной к ней плоскости симметрии σ. (Поэтому при рассмотрении зеркально-поворотных осей достаточно ограничиться осями четных порядков.) Действительно, повторим операцию S_n n раз. Тело повернется на угол 2π, претерпев нечетное число отражений. Все эти операции эквивалентны однократному отражению в плоскости σ. Так как каждая из них совмещает тело само с собой, то тем же свойством обладает и операция σ. Это показывает, что плоскость σ является плоскостью симметрии тела. Учтя это, произведем над телом преобразования σ и S_n — тело перейдет само в себя. Но совокупность операций σ и S_n эквивалентна одному повороту C_n. Значит, ось S_n является поворотной осью тела n-го порядка.

5. Если при инверсии относительно некоторой точки O тело переходит само в себя, то точка O называется *центром симметрии* тела. Допустим, что тело имеет зеркально-поворотную ось второго порядка S_2. Докажем, что в этом случае точка O, в которой ось S_2 пересекает соответствующую ей плоскость Π, будет центром симметрии тела. Действительно, поворот вокруг оси S_2 на 180° (рис. 158) переводит точку A в положение A'. Отражение в плоскости Π переводит точку A' в положение B. Так как S_2 есть зеркально-поворотная ось, то в положении B должна находиться точка тела, идентичная с точкой A.

Но точки A и B симметрично расположены относительно точки O. Это значит, что точка O является центром симметрии тела. Таким образом, *если в теле есть зеркально-поворотная ось второго порядка, то в нем есть и центр симметрии.* Обратно, *если в теле есть центр симметрии, то в нем имеются и зеркально-поворотные оси второго порядка и притом, как нетрудно заметить, таких осей бесконечно много.*

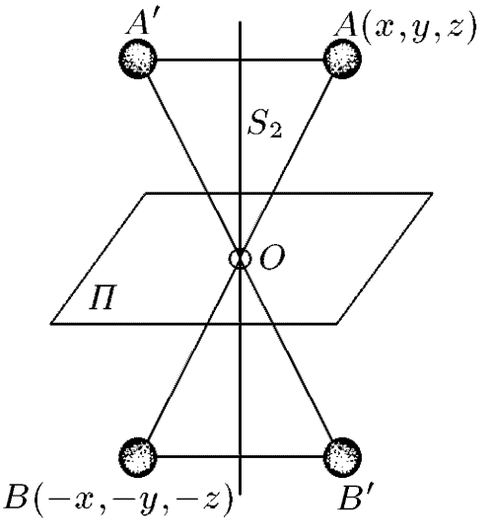


Рис. 158

Из рис. 158 видно также, что три элемента: центр симметрии, поворотная ось второго порядка и перпендикулярная к ней плоскость симметрии не независимы. Существование любой пары таких элементов симметрии влечет и существование третьего. Действительно, допустим, например, что в теле есть центр симметрии и поворотная ось второго порядка. Тогда, если тело содержит точку A, то из-за наличия поворотной оси оно должно содержать также точку A', а из-за наличия центра симметрии — точки B и B'. Но точки A' и B, A и B' симметрично расположены относительно плоскости $П$. Следовательно, $П$ есть плоскость симметрии.

6. При отражении тела в плоскости может получиться тело, вполне идентичное с исходным — оба тела можно совместить друг с другом одними только поворотами. Но может получиться тело, хотя и похожее на исходное, однако отличающееся от него примерно так, как правая рука отличается от левой. Никакими поворотами такие два тела нельзя совместить друг с другом. Первый случай будет иметь место всегда, когда тело достаточно симметрично, а именно имеет плоскость симметрии или зеркально-поворотную ось, в частности центр симметрии; второй — когда тело этими элементами симметрии не обладает. Органическая химия дает многочисленные примеры, когда молекулы вещества состоят из одних и тех же атомов, но встречаются в двух неидентичных модификациях: одна может быть получена из другой путем зеркального отражения. Такие молекулы называются *зеркальными изомерами*, или *стереоизомерами*. Само явление получило название *зеркальной изомерии*. Оно было открыто Пастером (1822–1895). Приведем простейший пример. Возьмем молекулу метана CH_4. Она имеет форму тетраэдра с атомом углерода в центре и четырьмя атомами водорода в его вершинах. Такая молекула имеет плоскости симметрии, а потому зеркальной изомерией не обладает. Не обладает зеркальной изомерией и молекула, в которой один из атомов водорода заменен хлором, а также молекула, в которой один из атомов водорода заменен хлором, а другой — бромом. (В таких молекулах есть, по крайней мере, одна плоскость симметрии.) Однако если третий атом водорода заменить йодом, то получится молекула CHClBrI, которая

может существовать в виде двух различных стереоизомеров, как это видно из рис. 159 (атом C не изображен).

Физические свойства одного изомера можно представить себе как свойства другого изомера, полученные отражением в плоскости. Эти свойства могут отличаться друг от друга только в том же отношении, в каком правое отличается от левого. В частности, в растворах стереоизомеры вращают плоскость поляризации линейно поляризованного света в противоположных направлениях: один вправо, другой — влево. (Предполагается, что свет распространяется к глазу наблюдателя.) По этому признаку один из изомеров называют *правым*, другой — *левым*.

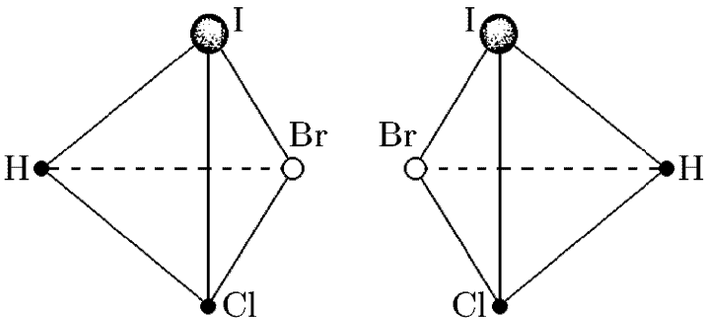


Рис. 159

Химические реакции между правыми изомерами различных веществ A и B протекают так же, как и между левыми, — один из этих процессов является зеркальным изображением другого. По той же причине одинаково реагируют правый изомер A с левым изомером B и правый изомер B с левым изомером A. Однако реакции в двух первых случаях протекают существенно иначе, чем в последних двух. Все это имеет важное значение в биологии, так как молекулы, входящие в состав живых организмов, асимметричны и способны к образованию стереоизомеров.

Пусть молекулы A и B зеркальной изомерией не обладают, а при химическом соединении друг с другом образуют молекулы, обладающие этим свойством. Тогда возможны две реакции:

$$1)\quad A+B=(AB)_{\text{прав}};\qquad 2)\quad A+B=(AB)_{\text{лев}}.$$

Если эти реакции отразить в зеркале, то молекулы A и B перейдут сами в себя, правый изомер AB заменится левым, а левый — правым. В результате рассматриваемые реакции перейдут в

$$1)\quad A+B=(AB)_{\text{лев}};\qquad 2)\quad A+B=(AB)_{\text{прав}},$$

т. е. первая реакция заменится второй, а вторая — первой. Отсюда следует, что при соединении A с B образуется столько же правых стереоизомеров, сколько и левых.

§ 130. Кристаллические решетки

1. Основной особенностью кристаллов, отличающих их от жидкостей и аморфных твердых тел, является *периодичность пространственного расположения атомов, молекул и ионов*, из которых состоит кристалл. Такая периодичность получила название *дальнего*

порядка [1]. В дальнейшем ради краткости мы будем говорить, что кристаллы построены из атомов, хотя роль атомов могут выполнять также молекулы или ионы. Совокупность таких периодически расположенных атомов образует периодическую структуру, называемую *кристаллической решеткой*. Точки, в которых расположены сами атомы (точнее, точки, относительно которых они совершают тепловые и нулевые колебания), называются *узлами кристаллической решетки*. Если нас интересует только пространственная периодичность в расположении атомов, то можно отвлечься от их внутренней структуры и рассматривать атомы как геометрические точки. В этом смысле говорят о *пространственной решетке*. Представление о пространственной решетке в кристаллографии было введено французским кристаллографом и математиком Огюстом Браве (1811–1863). Тем самым были заложены основы для систематического теоретического исследования симметрии кристаллов. Экспериментальное, хотя и несколько косвенное, доказательство указанного представления было впервые получено в 1912 г. в знаменитом опыте Лауэ (1879–1960) и его сотрудников Фридриха (1883–1968) и Книппинга (1883–1935) по дифракции рентгеновских лучей (см. т. IV, § 61). Позднее аналогичное доказательство принесло также исследование дифракции электронов и в особенности нейтронов.

2. Чтобы выявить внутреннюю симметрию кристалла, мы будем предполагать, что кристаллическая решетка неограниченная. Периодичность решетки проявляется в так называемой *трансляционной симметрии*. Трансляционная симметрия означает, что существуют три некомпланарных вектора $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$, характеризующиеся тем, что при смещении решетки на вектор

$$\mathbf{T} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3, \tag{130.1}$$

где n_1, n_2, n_3 — целые числа (в том числе и нули), она переходит сама в себя. Такие смещения называются *трансляциями*, а вектор $\mathbf{T}$ — *вектором трансляции*. Если при неизменных направлениях векторов $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ выбрать их длины минимальными, чтобы трансляциями вдоль этих направлений можно было получить всю кристаллическую решетку, то векторы $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ называются *основными*, или *базисными*, векторами, а их совокупность — *базисом решетки*. Параллелепипед с ребрами $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ называют *основным*, или *базисным*, *параллелепипедом*. Вместе с находящимися в нем атомами он образует так называемую *элементарную ячейку* кристаллической решетки. Длины ребер a, b, c называются *основными периодами решетки*.

Если элементарная ячейка содержит восемь атомов в вершинах основного параллелепипеда, но не содержит ни одного атома внутри объема или на гранях этого параллелепипеда, то она называется *примитивной* (рис. 160). Все прочие ячейки называются *сложными*.

[1] В аморфных и жидких телах упорядоченное расположение частиц может распространяться только на соседние атомы (ближний порядок).

Теми же терминами пользуются для названия соответствующих решеток и параллелепипедов. Поскольку к каждой вершине параллелепипеда примыкает восемь элементарных ячеек, на каждую примитивную ячейку приходится один атом. Примитивная пространственная решетка называется также *решеткой Браве*. Она может быть получена из одной точки, если подвергнуть последнюю всевозможным трансляциям параллельно ребрам основного параллелепипеда $\mathbf{a}, \mathbf{b}, \mathbf{c}$. Сложную кристаллическую решетку можно рассматривать как совокупность решеток Браве, вставленных друг в друга.

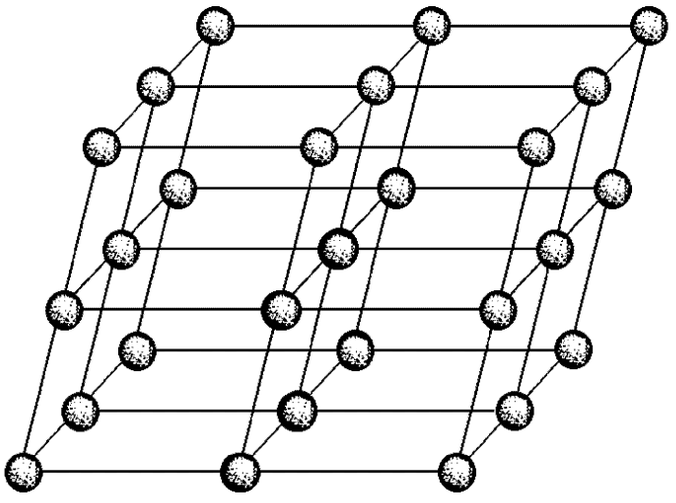


Рис. 160

3. Выбор базиса, а с ним и элементарной ячейки не однозначен. Поясним это на примере сетки, т. е. плоской решетки (рис. 161). В качестве элементарных ячеек можно выбрать, например, параллелограммы $ABCD$ и $ACED$. В обоих случаях в элементарную ячейку входит одно и то же число атомов каждого сорта (по одному атому сорта *1*, одному атому сорта *2* и одному атому сорта *3*). Вообще, в качестве базисных векторов $\mathbf{a}'_1$, $\mathbf{a}'_2$, $\mathbf{a}'_3$ можно взять любые линейные некомпланарные комбинации векторов $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ типа (130.1). Важно только, чтобы при фиксированных направлениях векторов $\mathbf{a}'_1$, $\mathbf{a}'_2$, $\mathbf{a}'_3$ длины их были минимальны. Тогда число атомов каждого сорта в обеих элементарных ячейках будет одно и то же. Будут одинаковы и объемы всех элементарных ячеек кристаллической решетки. Это видно из того, что объем элементарной ячейки представляется выражением $v = nV/N$, где n — полное число атомов в элементарной ячейке, N — число атомов всего кристалла, а V — объем последнего. Выражение же nV/N от выбора базиса не зависит.

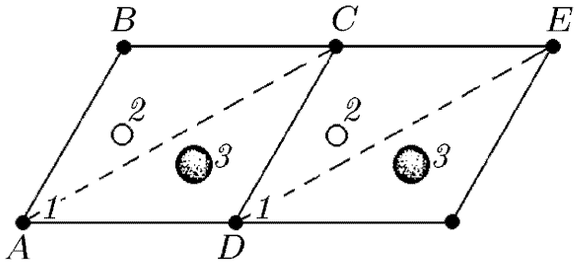


Рис. 161

4. По внешнему виду пространственной решетки не всегда просто определить, является ли она примитивной или сложной. Приведем пример, важный для последующего изложения. Рассмотрим примитивную пространственную решетку, основным параллелепипедом которой является прямоугольный параллелепипед с ребрами $\mathbf{a}, \mathbf{b}, \mathbf{c}$ (см. рис. 164, третий ряд). Такая решетка называется простой *ромбической решеткой*. Поместим в центре каждой элементарной ячейки

по одной точке — получится новая пространственная решетка, называемая *объемноцентрированной ромбической решеткой*. Если же поместить по одной точке в центрах граней элементарной ячейки, то получится решетка, называемая *ромбической гранецентрированной*. На первый взгляд кажется, что обе решетки — сложные. На самом деле это не так.

Действительно, узлы объемноцентрированной решетки могут быть представлены выражениями

$$\mathbf{r} = n_1\mathbf{a} + n_2\mathbf{b} + n_3\mathbf{c}, \tag{130.2}$$

$$\mathbf{r} = (n_1 + 1/2)\mathbf{a} + (n_2 + 1/2)\mathbf{b} + (n_3 + 1/2)\mathbf{c}. \tag{130.3}$$

Перейдем к новому базису:

$$\mathbf{a}' = \frac{-\mathbf{a} + \mathbf{b} + \mathbf{c}}{2}, \quad \mathbf{b}' = \frac{\mathbf{a} - \mathbf{b} + \mathbf{c}}{2}, \quad \mathbf{c}' = \frac{\mathbf{a} + \mathbf{b} - \mathbf{c}}{2}, \tag{130.4}$$

т. е. примем за базисные векторы, соединяющие вершину одного из параллелепипедов с центрами примыкающих к ней параллелепипедов. Тогда

$$\mathbf{a} = \mathbf{b}' + \mathbf{c}', \quad \mathbf{b} = \mathbf{c}' + \mathbf{a}', \quad \mathbf{c} = \mathbf{a}' + \mathbf{b}',$$

и выражения (130.2) и (130.3) примут вид

$$\mathbf{r} = (n_2 + n_3)\mathbf{a}' + (n_3 + n_1)\mathbf{b}' + (n_1 + n_2)\mathbf{c}', \tag{130.2а}$$

$$\mathbf{r} = (n_2 + n_3 + 1)\mathbf{a}' + (n_3 + n_2 + 1)\mathbf{b}' + (n_1 + n_2 + 1)\mathbf{c}'. \tag{130.2б}$$

Но оба они содержатся в выражении $\mathbf{r} = m_1\mathbf{a}' + m_2\mathbf{b}' + m_3\mathbf{c}'$, если только m_1, m_2, m_3 принимают всевозможные целочисленные значения. Отсюда следует, что объемноцентрированная решетка является примитивной.

Аналогичное рассуждение можно провести и для гранецентрированной решетки. Надо только принять за новые базисные векторы три вектора, соединяющие какую-либо вершину параллелепипеда с центрами примыкающих к ней трех граней, например

$$\mathbf{a}' = \frac{\mathbf{b} + \mathbf{c}}{2}, \quad \mathbf{b}' = \frac{\mathbf{c} + \mathbf{a}}{2}, \quad \mathbf{c}' = \frac{\mathbf{a} + \mathbf{b}}{2}. \tag{130.5}$$

Исходный прямоугольный параллелепипед, для простой ромбической решетки может быть принят за основной. Однако он не является основным для объемно- и гранецентрированных решеток. Действительно, в этих случаях путем трансляций базиса $\mathbf{a}, \mathbf{b}, \mathbf{c}$ нельзя получить все узлы решетки. Но трансляциями базисов (130.4) и (130.5) этого можно достигнуть. Поэтому для объемноцентрированной решетки за базисный параллелепипед можно принять параллелепипед (130.4), а для граненцентрированной — параллелепипед (130.5).

Для гранецентрированной и объемноцентрированной решеток базисные параллелепипеды, вообще говоря, будут косоугольными. Действительно, рассмотрим частный случай, когда исходный прямоугольный параллелепипед вырождается в куб (см. рис. 164, первый ряд). Тогда угол α между двумя соседними базисными векторами (130.4),

т. е. между пространственными диагоналями куба, будет определяться уравнением $\sin(\alpha/2) = 1/\sqrt{3}$, из которого получаем $\alpha = 70°32'$. Углы же между базисными векторами (130.5) (т. е. между диагоналями соседних граней куба) будут составлять 60°. Таким образом, решетки, получаемые из простой кубической решетки путем центрирования граней и объемов основных кубов, будут примитивными. Однако они не будут простыми кубическими решетками.

5. Из каждой примитивной решетки можно выделить параллелепипед, называемый *приведенным*. Он получается следующим образом. Рассмотрим совокупность всех векторов, соединяющих попарно узлы решетки. Из них выберем вектор минимальной длины и примем его за вектор $\mathbf{a}_1$. Из оставшихся векторов выберем вектор минимальной длины, не коллинеарный с $\mathbf{a}_1$. Его примем за вектор $\mathbf{a}_2$. Из всех остальных векторов выберем вектор минимальной длины $\mathbf{a}_3$, не компланарный с векторами $\mathbf{a}_1$ и $\mathbf{a}_2$. Базис $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ и построенный на нем параллелепипед и называются приведенными. *Основное свойство приведенного параллелепипеда состоит в том, что этот параллелепипед — примитивный.* Для доказательства заметим прежде всего, что вершины приведенного, как и всякого основного параллелепипеда, помещаются в узлах решетки. После этого предположим, что внутри приведенного параллелепипеда имеется узел решетки M (рис. 162). Опустим из M перпендикуляр MN на ближайшую грань параллелепипеда. Из основания перпендикуляра N опустим новый перпендикуляр NP на ближайшее ребро грани, проходящей через точку N. Наконец, соединим точку P с ближайшим концом O ребра, на котором она расположится. Из построения ясно, что длины взаимно перпендикулярных отрезков MN, NP и PO не могут превосходить половину длины максимального ребра приведенного параллелепипеда $a_3/2$. Отсюда и из теоремы Пифагора следует, что $OM \leqslant \sqrt{3}a_3/2 < a_3$. Это значит, что вектор $\overline{OM}$, соединяющий узлы O и M, короче вектора $\mathbf{a}_3$ и не компланарен с векторами $\mathbf{a}_1$ и $\mathbf{a}_2$. Но это противоречит предположению, что базис $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ — приведенный. Аналогичные рассуждения с линейной и плоской решетками показывают, что не может существовать внутренних узлов на ребрах и гранях приведенного параллелепипеда. Доказанная теорема позволяет в качестве основного брать приведенный параллелепипед.

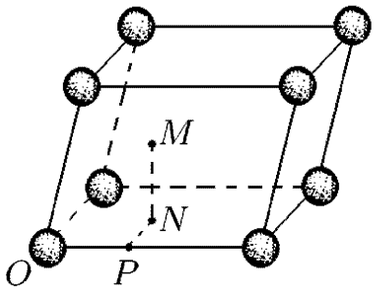


Рис. 162

6. В математических рассуждениях часто бывает удобно, наряду с самой пространственной решеткой, вводить вспомогательную систему точек, называемую *обратной решеткой*. Базисными векторами обратной решетки являются векторы, взаимные по отношению к векторам $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$, т. е.

$$\mathbf{a}_1^* = \frac{[\mathbf{a}_2\mathbf{a}_3]}{([\mathbf{a}_1\mathbf{a}_2]\mathbf{a}_3)}, \quad \mathbf{a}_2^* = \frac{[\mathbf{a}_3\mathbf{a}_1]}{([\mathbf{a}_1\mathbf{a}_2]\mathbf{a}_3)}, \quad \mathbf{a}_3^* = \frac{[\mathbf{a}_1\mathbf{a}_2]}{([\mathbf{a}_1\mathbf{a}_2]\mathbf{a}_3)} \tag{130.6}$$

(см. т. I, §7, задача 9). Объемы базисных параллелепипедов исходной и обратной решеток связаны соотношением

$$VV^* = 1. \tag{130.7}$$

ЗАДАЧА

Доказать, что обратная решетка не зависит от выбора базиса.

Решение. Перейдем к новому базису:

$$\mathbf{a}'_1 = \mathbf{a}_1 + n_2\mathbf{a}_2, \quad \mathbf{a}'_2 = \mathbf{a}_2, \quad \mathbf{a}'_3 = \mathbf{a}_3. \tag{130.8}$$

Так как объемы базисных параллелепипедов V в обоих базисах одни и те же, то новые базисные векторы обратной решетки будут $\mathbf{a}'^*_3 = (1/V)\,[\mathbf{a}'_2\mathbf{a}'_3] = \mathbf{a}^*_1$,

$$\mathbf{a}'^*_2 = \frac{1}{V}\,[\mathbf{a}'_3\mathbf{a}'_1] = \frac{1}{V}\,[\mathbf{a}_3\mathbf{a}_1] + \frac{n_2}{V}\,[\mathbf{a}_3\mathbf{a}_2] = \mathbf{a}^*_2 - n_2\mathbf{a}^*_1,$$

$$\mathbf{a}'^*_3 = \frac{1}{V}\,[\mathbf{a}'_1\mathbf{a}'_2] = \frac{1}{V}\,[\mathbf{a}_1\mathbf{a}_2] = \mathbf{a}^*_3.$$

В обратной решетке можно перейти к новому базису

$$\mathbf{b}^*_1 = \mathbf{a}^*_1, \quad \mathbf{b}^*_2 = \mathbf{a}'^*_2 + n_2\mathbf{a}^*_1 = \mathbf{a}^*_2, \quad \mathbf{b}^*_3 = \mathbf{a}^*_3.$$

Но этот базис совпадает с базисом (130.6). Поэтому будут совпадать и соответствующие им обратные решетки. В общем случае переход к новому базису может быть выполнен путем частных преобразований типа (130.8). Поэтому заключение остается в силе и в этом случае.

§ 131. Кристаллические системы

1. Помимо трансляционной симметрии кристаллическая решетка может обладать и другими элементами симметрии. Так, всякая примитивная пространственная решетка имеет центр симметрии. Центром симметрии, как легко видеть, является каждая вершина и центр примитивного параллелепипеда решетки, а также середины его ребер и центры граней. Базис однозначно определяет примитивную решетку. Обратное несправедливо — для одной и той же решетки базис может быть выбран бесконечным множеством способов. Поэтому симметрия базисного параллелепипеда, вообще говоря, не совпадает с симметрией построенной на нем решетки.

2. Тела конечных размеров, например молекулы, могут обладать поворотными и зеркально-поворотными осями симметрии любого порядка. Неограниченные кристаллические решетки, как примитивные, так и сложные, благодаря наличию у них трансляционной симметрии, ведут себя иначе. *Поворотные и зеркально-поворотные оси симметрии кристаллической решетки могут быть только осями 2-го, 3-го, 4-го и 6-го порядков. Другие оси в кристаллической решетке невозможны.* Для доказательства возьмем какие-либо два идентичные соседние узла A и B на узловой линии AB, т. е. прямой, содержащей бесконечное множество атомов решетки (рис. 163). Если через узел A проходит поворотная ось n-го порядка, то параллельная ей прямая, проходящая через узел B, будет также поворотной осью

того же порядка. Закрепим эти оси неподвижно в пространстве и условимся называть их осями A и B. Повернем всю решетку вокруг оси B, на угол $\varphi \equiv \varphi_n = 2\pi n$. Атом, находившийся в точке A, перейдет в A', причем вся решетка совместится сама с собой. Место атома в точке A займет другой в точности такой же атом. Произведем теперь поворот на тот же угол φ вокруг оси A, но в противоположном направлении. Решетка опять совместится сама с собой. Атом находившийся в точке B, перейдет в положение B'; атомы в точках B и A' заменятся другими совершенно такими же атомами. Если точки A' и B' совпадают между собой, то $\varphi = 60°$, и оси A и B будут поворотными осями шестого порядка. Если же точки A' и B' не совпадают между собой, то прямая $B'A'$ будет узловой прямой решетки, так как она параллельна прямой AB. По условию A и B — соседние идентичные узлы на прямой AB, а потому длина AB является основным периодом для обеих узловых линий AB и $B'A'$. Значит, длина $B'A'$ должна быть кратна длине AB. Но $B'A' = AB \times$ $\times (1 - 2\cos\varphi)$. Поэтому число $1 - 2\cos\varphi$ должно быть целым (случай, когда оно обращается в нуль, т. е. когда $\cos\varphi = 1/2$, рассмотрен выше). Это возможно тогда и только тогда, когда $\cos\varphi = 0, -1/2, -1$. Таким значениям соответствуют углы поворота $\varphi = 90°, 120°, 180°$, т. е. поворотные оси 4-го, 3-го и 2-го порядков. Сочетая соответствующий поворот с отражением в плоскости, легко распространить приведенное доказательство и на зеркально-поворотные оси симметрии.

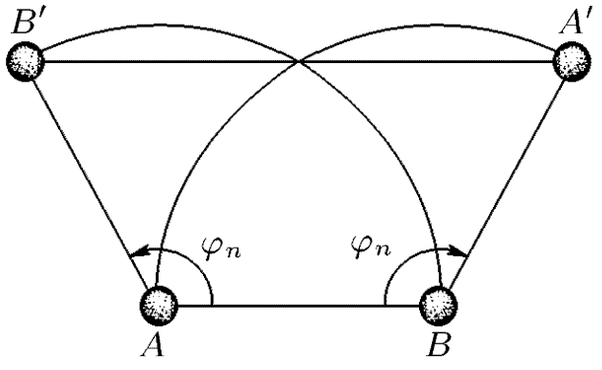
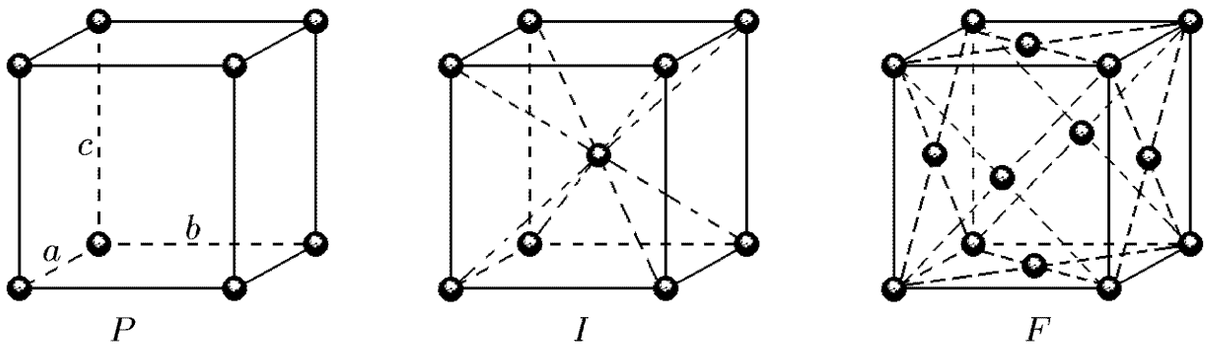


Рис. 163

3. Сложная пространственная решетка состоит из примитивных решеток (решеток Браве). По симметрии примитивных решеток все кристаллы разделяются на семь кристаллических систем. Под симметрией здесь понимается *точечная симметрия*, включающая в себя все элементы симметрии, за исключением трансляционных, т. е. центр, плоскости и поворотные оси симметрии различных порядков. В сущности, разделение кристаллов на кристаллические системы производится по числу поворотных осей симметрии различных порядков, которыми обладает решетка Браве.

Напомним, что симметрия пространственной решетки не всегда совпадает с симметрией основного параллелепипеда, на котором построена решетка. Однако Браве заметил, что из всякой примитивной решетки, за исключением гексагональной, можно выделить параллелепипед, содержащий все те элементы симметрии (за исключением, конечно, трансляционных), что и решетка в целом. Наименьший из таких параллелепипедов называется *параллелепипедом Браве*. Если он вырождается в куб, то мы будем называть его *кубом Браве*. Браве доказал, что могут существовать шесть типов решеток, для которых

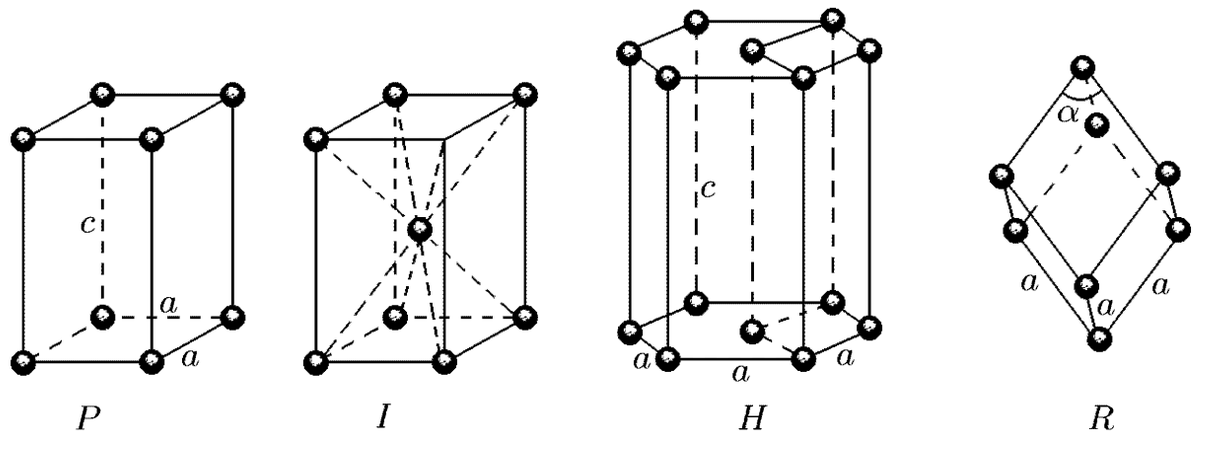
Кубическая система



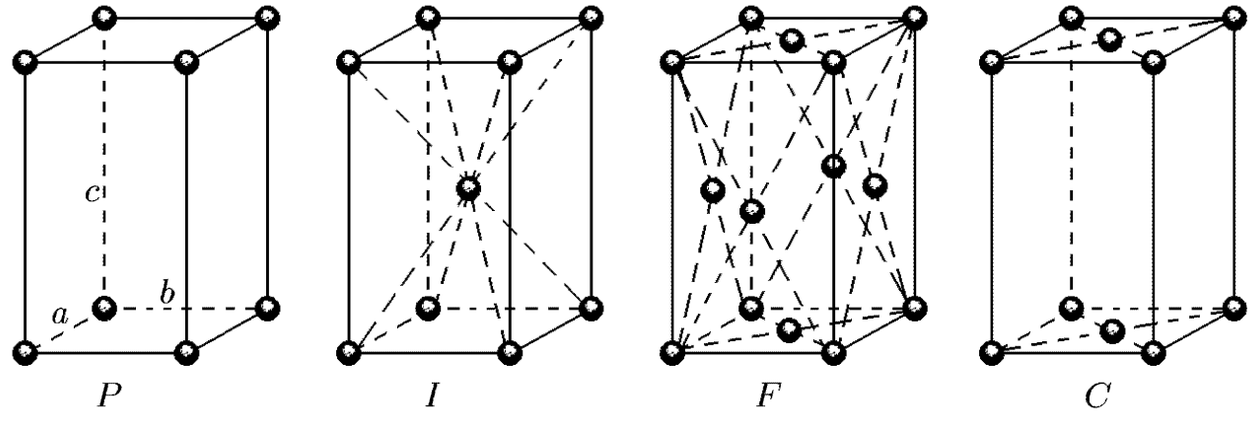
Тетрагональная система

Гексагональная система

Ромбоэдрическая система



Ромбическая система



Моноклинная система

Триклинная система

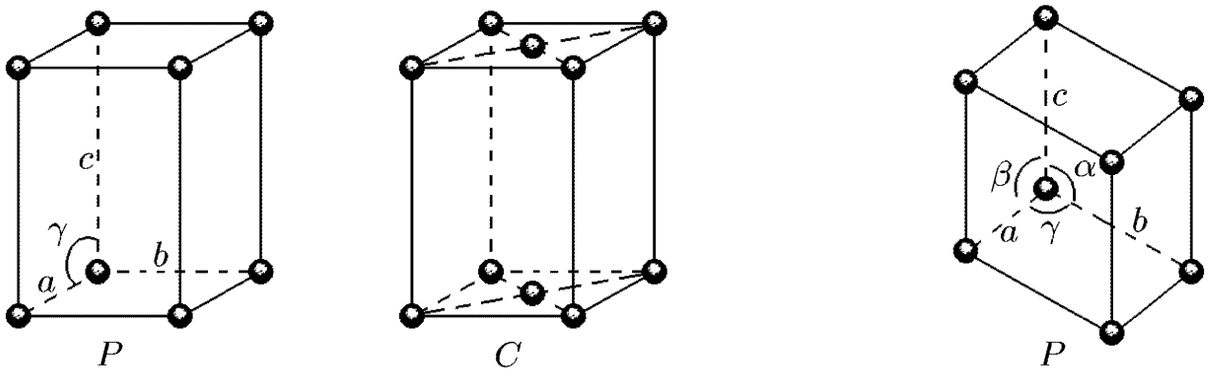


Рис. 164

параллелепипед Браве — примитивный. Если к ним присоединить гексагональную решетку, то получится всего 7 типов решеток, охватывающих всевозможные комбинации элементов симметрии решеток Браве. Центрирование граней и объемов параллелепипедов Браве не изменяет симметрию решетки. Однако оно приводит к появлению еще 7 новых типов решеток Браве. Таким образом, существует всего 14 типов решеток Браве, распределяющихся по 7 кристаллическим системам. Опишем эти системы и соответствующие им решетки Браве.

Кубическая система. Решетки этой системы наиболее симметричны. Параллелепипедом Браве является куб (рис. 164). Существуют три типа решеток Браве кубической системы: *простая* (обозначается через P), *объемноцентрированная* (обозначается через I) и *гранецентрированная* (обозначается через F). Как указывалось в § 130, п. 4, параллелепипед Браве простой кубической решетки является также основным параллелепипедом. Для остальных двух решеток основные параллелепипеды косоугольные, а потому эти параллелепипеды характеризуются более низкой симметрией, чем сами решетки. Длина ребра куба Браве для всех трех кубических решеток является единственным *пространственным параметром решетки. Эту длину называют постоянной решетки и обычно обозначают через* a. *Кристаллические решетки кубической системы имеют* 13 *поворотных осей симметрии*: 6 *осей второго порядка*, 4 *оси третьего порядка и* 3 *оси четвертого порядка. Оси симметрии второго порядка соединяют центры противоположных ребер куба Браве, третьего — противоположные вершины его, а четвертого — центры противоположных граней.*

Тетрагональная (или квадратная) система. Параллелепипед Браве имеет форму прямой квадратной призмы (рис. 164). Наряду с примитивной решеткой (P) существует еще объемноцентрированная решетка (I). Центрирование оснований не дает решеток нового типа. Оно приводит к разделению исходной решетки на две примитивные решетки того же типа, что и исходная решетка. Это видно из рис. 165. Не дает ничего нового и центрирование всех граней исходной решетки. Оно превращает последнюю в объемноцентрированные решетки той же системы. Таким образом, существуют только две решетки Браве тетрагональной системы: *простая* и *объемноцентрированная*. Они имеют 4 поворотных оси симметрии второго и одну ось симметрии четвертого порядков. Последняя соединяет центры квадратных оснований, а первые — центры боковых граней и середины ребер параллелепипеда Браве. Тетрагональная решетка определяется двумя параметрами: длиной стороны a квадратного основания параллелепипеда Браве и его высотой c.

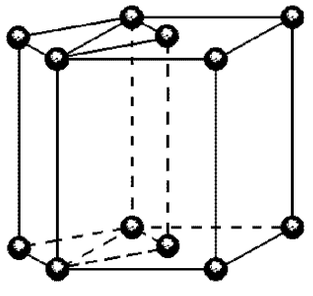


Рис. 165

Гексагональная система (ее решетка обозначается через H). Для кристаллов этой системы понятие параллелепипеда Браве теряет

смысл. Основной параллелепипед имеет форму прямой призмы, основанием которой служит ромб с острым углом 60° (рис. 166). Однако такой параллелепипед не передает симметрию пространственной решетки в целом. Для достижения этого три таких параллелепипеда соединяют вместе, чтобы они образовывали правильную шестигранную призму. Последняя полностью характеризует симметрию решетки. Узлы пространственной решетки располагаются в вершинах таких шестигранных призм и в центрах их оснований. Гексагональная решетка определяется двумя параметрами: длиной стороны основания a и высотой c призмы. Решетка имеет ось симметрии шестого порядка и 6 поворотных осей симметрии второго порядка, перпендикулярных к этой оси.

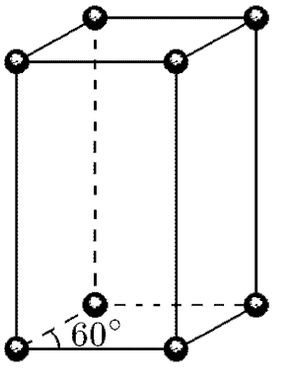


Рис. 166

Ромбоэдрическая система (решетка обозначается символом R). Параллелепипед Браве имеет форму ромбоэдра. Последний можно получить путем равномерного растяжения или сжатия куба в направлении его пространственной диагонали. Все грани ромбоэдра представляют собой одинаковые ромбы. Единственная решетка Браве такой системы является простой. Она характеризуется двумя параметрами: длиной a ребер параллелепипеда Браве и углом α между ними (при $\alpha = 90°$ ромбоэдр переходит в куб). Четыре пространственные диагонали куба являются поворотными осями симметрии третьего порядка. При растяжении или сжатии вдоль одной из этих диагоналей она продолжает оставаться осью симметрии третьего порядка. Другие три диагонали переходят в оси симметрии второго порядка. Оставшиеся семь осей симметрии куба утрачивают свойство симметрии. Таким образом, ромбоэдрическая решетка имеет четыре поворотных оси симметрии: одну третьего и три второго порядков.

Ромбическая или ортогональная система. Параллелепипед Браве — прямоугольный с тремя различными длинами ребер a, b, c, являющимися параметрами решетки. Существуют четыре типа решеток Браве рассматриваемой системы: *простая* (P), *объемноцентрированная* (I), *гранецентрированная* (F) и *базоцентрированная* (C), т. е. решетка с центрированными основаниями. Осей симметрии три. Они параллельны ребрам параллелепипеда Браве и являются осями второго порядка.

Моноклинная система. Параллелепипедом Браве является прямой параллелепипед. Основание его есть произвольный параллелограмм. Моноклинная решетка характеризуется четырьмя параметрами — длинами a, b, c ребер параллелепипеда Браве и углом γ между ребрами a и b (остальные углы — прямые). Она имеет единственную ось симметрии второго порядка, которая соединяет центры оснований параллелепипеда Браве.

Триклинная система. Решетки этой системы — только простые (P). Параллелепипед Браве может быть произвольной формы. Поэтому решетки триклинной системы характеризуются наименьшей

степенью симметрии. Они имеют только центр симметрии и не имеют осей симметрии. Параметрами решетки являются длины ребер параллелепипеда Браве a, b, c и углы между ними α, β, γ.

Принадлежность решетки Браве к какой-либо кристаллической системе однозначно определяется числом и характером осей симметрии. Это видно из табл. 13.

Таблица 13

Системы	Число осей симметрии				Общее число симметрии
	2-го порядка	3-го порядка	4-го порядка	6-го порядка	
Кубическая	6	4	3	—	13
Тетрагональная	4	—	1	—	5
Гексагональная	6	—	—	1	7
Ромбоэдрическая	3	1	—	—	4
Ромбическая	3	—	—	—	3
Моноклинная	1	—	—	—	1
Триклинная	—	—	—	—	—

Все 14 типов решеток Браве были найдены им из геометрических соображений без использования каких бы то ни было физических принципов. Поэтому они могут быть охарактеризованы как *геометрически возможные*. Все они действительно встречаются в природе. Но в этом мы убеждаемся в *результате наблюдений*, а не путем геометрических рассуждений. Это замечание полностью относится и к следующему изложению, где говорится о пространственных группах и кристаллических классах.

§ 132. Пространственные группы и кристаллические классы кристаллов

1. Примитивные пространственные решетки, из которых состоит сложная кристаллическая решетка, могут существенно отличаться от нее своей симметрий. Рассмотрим, например, тетрагональную примитивную решетку, основание которой изображено на рис. 167 *а*. Через каждый узел проходит поворотная ось симметрии четвертого порядка, перпендикулярная к плоскости рисунка. Вдвинем в эту решетку две такие же примитивные решетки, как показано на рис. 167 *б*. Если узлы исходной примитивной решетки окажутся посередине между узлами вновь вдвинутых примитивных решеток, то прежняя ось симметрии 4-го порядка в сложной решетке станет осью симметрии 2-го порядка. Если же этого не будет, то она вообще перестанет быть поворотной осью симметрии. В обоих случаях симметрия решетки понижается.

2. К этому надо добавить, что в сложной решетке возможны новые элементы симметрии: *винтовая ось* и *плоскость зеркального скольжения*. Винтовой осью n-го порядка называется прямая, при

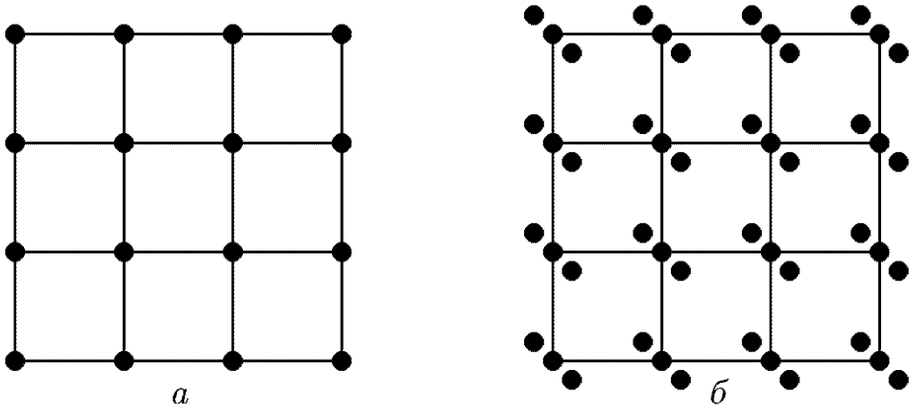


Рис. 167

повороте вокруг которой на угол $2\pi/n$ и одновременном параллельном смещении вдоль нее решетка совмещается сама с собой. В качестве примера на рис. 168 изображены три винтовые оси 4-го порядка. Из них первая является «*правой*», а вторая — «*левой*». Если смотреть вдоль винтовой оси в направлении смещения, то в первом случае для

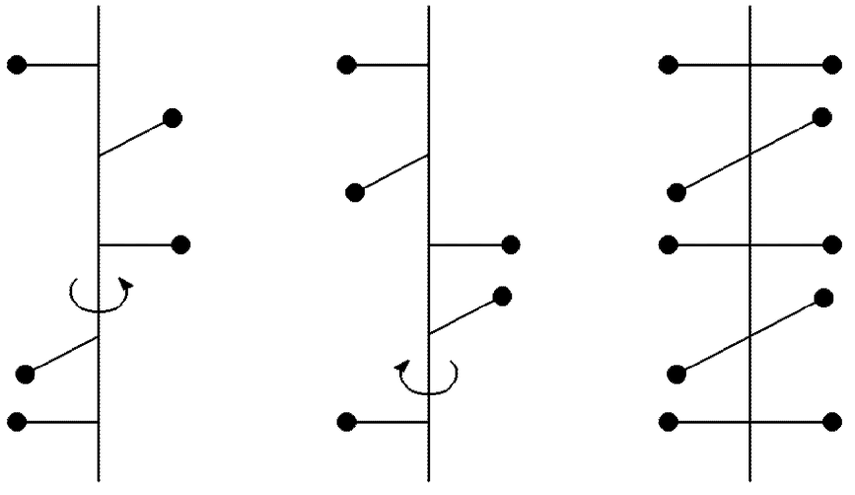


Рис. 168

совмещения самой с собой решетку надо поворачивать на 90° вправо, а во втором — влево. В третьем случае вращение может происходить и вправо, и влево. Так, кварц встречается в двух модификациях, одна из которых имеет правую, а другая — левую винтовую ось. Это явление есть частный случай так называемого *энантиоморфизма кристаллов*. Энантиоморфизм аналогичен явлению зеркальной изомерии молекул, описанному в § 129. Он состоит в том, что существуют кристаллические решетки, являющиеся зеркальными изображениями одна другой и притом такие, что они не могут быть совмещены друг с другом никакими поворотами в пространстве. Как и у молекул, энантиоморфизм возможен лишь для решеток, не содержащих плоскостей, центров и зеркально-поворотных осей симметрии.

Плоскостью зеркального скольжения называется такая плоскость, при отражении в которой и одновременном смещении на определенное расстояние в направлении, параллельном этой плоскости, решетка совмещается само с собой.

3. Таким образом, сложная пространственная решетка обладает трансляционной симметрией, а также может иметь и другие элементы симметрии: простые и винтовые оси симметрии, зеркально-поворотные оси, плоскости симметрии — простые и зеркального скольжения. *Совокупность всех элементов симметрии пространственной решетки называется ее пространственной группой.* Пространственная группа наиболее полно характеризует симметрию внутреннего строения кристаллов. Как показал в 1890 г. на основе геометрических соображений русский кристаллограф и минералог Е. С. Федоров (1853–1919), *может существовать всего* 230 *различных пространственных групп*, которые распределяются по кристаллическим системам следующим образом: кубическая — 36, тетрагональная — 68, гексагональная — 27, ромбоэдрическая — 25, ромбическая — 59, моноклинная — 13, триклинная — 2.

Среди 230 пространственных групп 11 пар отличаются только направлениями вращения винтовых осей. Это — *энантиоморфные группы*. По-видимому, не все пространственные группы Федорова реализуются в природе. Подтверждением этого может служить тот факт, что для 52 групп пока не найдено ни одного кристалла.

4. Выясним теперь понятие класса кристаллической решетки. Во многих физических явлениях атомистическая структура вещества непосредственным образом не проявляется. Такие явления, называемые *макроскопическими*, могут быть описаны в рамках представления о теле как о *сплошной среде* (*континууме*), характеризующейся определенными *макроскопическими параметрами*. Так поступают, например, при рассмотрении теплового расширения или деформаций тел.

В учении о кристаллах *континуум следует рассматривать как предельный случай кристаллической решетки*. Расстояния между соседними узлами решетки не могут входить в число параметров для характеристики свойств континуума. Такие расстояния должны считаться величинами бесконечно малыми, и ими всюду следует пренебрегать. Однако их отношения остаются величинами *конечными* и могут служить макроскопическими параметрами континуума. Возьмем, например, примитивную тетрагональную решетку. Ее макроскопическими параметрами являются длины ребер a и c соответствующего примитивного параллелепипеда (см. рис. 164). В качестве параметров решетки можно взять также, например, длину ребра a и отношение $\varepsilon = c/a$. При переходе к континууму параметр a обращается в нуль и может быть исключен из рассмотрения как величина фиксированная ($a = 0$). Остается только один (макроскопический) параметр ε, сохраняющий при предельном переходе конечное значение. Ясно, что

континуум, полученный в результате такого предельного перехода, будет *однородным*, но, вообще говоря, *анизотропным*.

Однородность означает, что все точки среды совершенно идентичны. При параллельном смещении среды на любое расстояние и в любом направлении она совмещается с собой. Поэтому при классификации кристаллов по симметрии их макроскопических свойств параллельные смещения можно совсем исключить из рассмотрения, поскольку они не могут выявить никаких специфических свойств, отличающих один кристалл от другого.

Анизотропия означает, что свойства кристаллов в различных направлениях разные. Но в некоторых направлениях они могут быть и одинаковыми. Тогда говорят о наличии *симметрии кристалла, рассматриваемого как континуум*. Например, если кристаллическая решетка имеет центр, плоскость и поворотную ось симметрии n-го порядка, то эти элементы симметрии сохраняются и для континуума, являющегося ее пределом. *Специфичность континуума состоит в том, что в нем* (ввиду обращения в нуль ребер основного параллелепипеда) *исчезает разница между простыми и винтовыми осями симметрии, а также между простыми плоскостями симметрии и плоскостями зеркального скольжения*. Для континуума остаются только следующие элементы симметрии: центр, плоскость, поворотные и зеркально-поворотные оси симметрии. Совокупность всех этих элементов симметрии кристаллической решетки, как континуума, называется ее *классом*. Ясно, что класс кристаллической решетки можно получить из ее пространственной группы, если игнорировать в ней все трансляции и не различать простые и винтовые оси, простые плоскости симметрии и плоскости зеркального скольжения.

Могут существовать 32 *кристаллических класса*, распределяющиеся по кристаллическим системам следующим образом: кубическая — 5, тетрагональная — 7, гексагональная — 7, ромбоэдрическая — 5, ромбическая — 3, моноклинная — 3, триклинная — 2. Среди классов, принадлежащих к данной системе, выделяется класс, обладающий *полной симметрий системы* (т. е. симметрией соответствующей ей примитивной решетки).

5. Приведем теперь пример физического явления, в котором проявляется анизотропия кристалла. Если решетка кубическая, то тепловое расширение ее по всем направлениям, параллельным ребрам куба, будет одно и то же. При нагревании кубическая решетка остается кубической. Если же решетка тетрагональная ($a \neq c$), то коэффициенты теплового расширения в направлениях ребер a и c будут разными. При нагревании отношение c/a будет изменяться.

Допустим, что при некоторой температуре длины ребер c и a отличаются незначительно ($a < c$). Допустим, далее, что при нагревании расширение кристалла в направлении c идет медленнее, чем в перпендикулярных направлениях. Тогда при некоторой температуре $T = T_c$ длины ребер a и c могут сравняться. Решетка изменяется непрерывно, никаких изменений плотности или выделения тепла при

температуре T_c не происходит. Однако при этой температуре *скачкообразно изменяется симметрия решетки*: из тетрагональной решетка становится кубической. Поэтому температура T_c в принципе может быть указана совершенно точно. Если при дальнейшем нагревании решетка продолжает оставаться также кубической, то можно сказать, что в точке $T = T_c$ произошел фазовый переход без изменения плотности и без выделения или поглощения теплоты перехода. Это — *фазовый переход второго рода*. Изменение симметрии решетки может привести к скачкообразному изменению коэффициента объемного расширения решетки, так как кубическая решетка расширяется иначе, чем тетрагональная, из которой она возникла. Точно так же скачкообразно может измениться и теплоемкость решетки. Приведенный воображаемый пример, принадлежащий Л. Д. Ландау (1908–1968), интересен в том отношении, что он может служить для разъяснения физической природы фазовых переходов второго рода. Заметим, что при фазовых переходах первого рода кристаллическая решетка либо разрушается (плавление), либо изменяется скачкообразно (полиморфные превращения). С этим и связано изменение объема тела и выделение теплоты при таких превращениях.

§ 133. Миллеровские индексы и индексы направлений

1. Для определения положения атомов в кристаллической решетке пользуются специальными прямолинейными системами координат, называемыми *кристаллографическими*. За начало координат принимается один из узлов решетки, а за координатные оси — ребра соответствующего параллелепипеда Браве. Для моноклинных и триклинных кристаллов выбор параллелепипеда Браве не однозначен. В гексагональных кристаллах за оси X и Y принимают стороны основания основного параллелепипеда, образующие угол $120°$, а за ось Z — ребро, перпендикулярное к этому основанию. В моноклинных кристаллах за ось Z принимают ребро, перпендикулярное к основанию параллелепипеда Браве. Мы видим, что в кубических, тетрагональных и ромбических кристаллах системы координат прямоугольные, в остальных кристаллах — косоугольные. Ребра параллелепипеда Браве принимаются за единицы длины в направлениях координатных осей. Такие единицы длины называются *осевыми*. Таким образом, в направлениях различных осей координат единицы длины разные. Так, атом в центре основного параллелепипеда Браве ромбического кристалла имеет координаты $(1/2, 1/2, 1/2)$, а атом в центре грани XY того же параллелепипеда — координаты $(1/2, 1/2, 0)$. Кристаллографические координаты применяются и для характеристики направлений кристаллических плоскостей и узловых линий решетки.

Кристаллической, или *узловой*, плоскостью называется всякая плоскость, в которой находится бесконечное множество атомов решет-

ки. Практическое значение имеют только плоскости, усеянные атомами достаточно густо. Именно они служат *естественными гранями кристалла* [1]. Кристаллические плоскости имеют большое значение для методов рентгеноструктурного и нейтроноструктурного анализа кристаллов. *Узловой линией* называется прямая, на которой расположено бесконечное множество атомов решетки. Основное значение имеют опять узловые линии, на которых атомы расположены достаточно густо.

2. Опишем теперь, как характеризуются направления кристаллических плоскостей в кристалле. Все параллельные плоскости имеют по определению одно и то же направление. Из них всегда можно выбрать плоскость, проходящую через любой узел решетки. Поэтому, не теряя общности, можно при рассмотрении вопроса о направлении кристаллических плоскостей ограничиться примитивными решетками. В таких решетках координаты всех узлов целочисленны. Каждую кристаллическую плоскость можно представить уравнением

$$\frac{x}{A}+\frac{y}{B}+\frac{z}{C}=1, \tag{133.1}$$

где A, B, C — длины отрезков (в осевых единицах), отсекаемые этой плоскостью на координатных осях. Эти длины всегда выражаются *рациональными числами* (положительными или отрицательными). Для доказательства выберем в плоскости (133.1) какие-либо три узла, не расположенные на одной прямой. Подставляя их в (133.1), получим три линейных уравнения с неизвестными $1/A, 1/B, 1/C$ и целочисленными коэффициентами. Эти уравнения однозначно определяют рассматриваемые неизвестные как рациональные числа. Следовательно, уравнение (133.1) всегда может быть приведено к виду

$$hx+ky+lz=D, \tag{133.2}$$

в котором коэффициенты h, k, l являются *целыми числами*. Можно считать, что они не имеют общего множителя, так как на таковой всегда можно сократить. Полученные таким образом целые числа h, k, l однозначно определяют направление кристаллической плоскости и называются *миллеровскими индексами*, или просто *индексами*, этой плоскости. Совокупность миллеровских индексов кристаллической плоскости принято заключать в круглые скобки, например (hkl). Если какой-либо индекс отрицателен, то знак минус пишут над ним, например $\overline{1}, \overline{3}$ и т. д.

[1] Внешняя поверхность кристалла приобретает правильную естественную форму лишь при условии *свободного роста его*. Неравномерность распределения температуры, неоднородность концентрации вещества в различных местах раствора, в котором растет кристалл, примеси посторонних веществ, механические препятствия и т. п. приводят к тому, что одни грани кристалла растут быстрее, другие — медленнее, чем требуется для того, чтобы кристалл принял естественную форму.

В качестве примера рассмотрим кубическую решетку (рис. 169). Заштрихованная грань куба представляется уравнением $y = 1$ или $0x + 1y + 0z = 1$. Значит, ее миллеровскими индексами будут (010). Индексы остальных граней: (100) и (001). Диагональная плоскость OBC представляется уравнением $x - y = 0$ и, следовательно, имеет индексы $(1\bar{1}0)$. Индексами плоскости ABC будут (111).

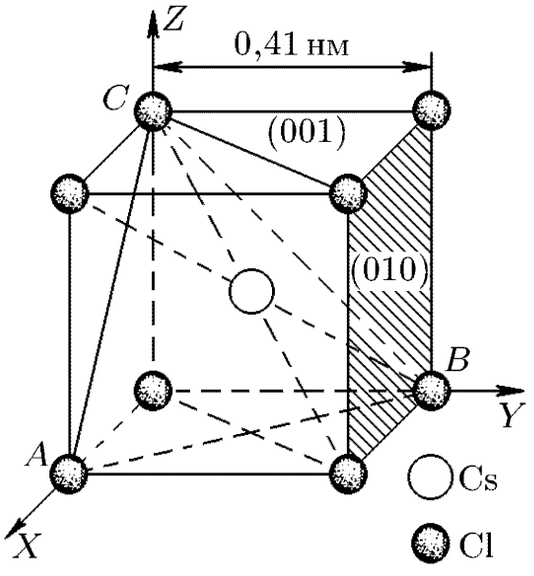


Рис. 169

3. Всякая естественная грань кристалла, как уже сказано выше, является кристаллической плоскостью. Рассмотрим какие-либо две естественные грани кристалла с миллеровскими индексами (hkl) и $(h'k'l')$. Пусть A, B, C и A', B', C' — длины отрезков (в осевых единицах), отсекаемые этими гранями на координатных осях решетки. Как видно из уравнения (133.2), эти длины обратно пропорциональны соответствующим миллеровским индексам, т. е.

$$A : B : C = \frac{1}{h} : \frac{1}{k} : \frac{1}{l}, \quad A' : B' : C' = \frac{1}{h'} : \frac{1}{k'} : \frac{1}{l'}.$$

Поделив одно соотношение на другое, получим

$$\frac{A'}{A} : \frac{B'}{B} : \frac{C'}{C} = \frac{h}{h'} : \frac{k}{k'} : \frac{l}{l'}. \tag{133.3}$$

Последнее соотношение должно выполняться независимо от того, в каких единицах измеряются отрезки A и A', B и B', C и C', так как отношения A'/A, B'/B, C'/C от выбора единиц не зависят. В частности, можно все отрезки измерять одними и теми же единицами. Умножением на общее кратное чисел h', k', l' правая часть соотношения (133.3) может быть приведена к отношению трех целых чисел. Таким образом, три отношения длин соответствующих отрезков, отсекаемых на осях кристаллической решетки какими-либо двумя гранями кристалла, относятся между собой как целые числа. Это правило называется *законом рациональности граней.*

4. Для указания направления какой-либо узловой линии кристаллической решетки достаточно указать разности координат двух соседних идентичных узлов, лежащих на этой линии. Первый узел обычно помещают в начале координат (для чего достаточно через начало координат провести прямую, параллельную рассматриваемому направлению). Полученные таким путем (целые) числа называются *индексами направлений* и заключают в квадратные скобки. Например, индексами направлений пространственной диагонали куба (рис. 169) будут [111].

§ 134. Решетки химических элементов и соединений

1. По роду частиц, из которых построена кристаллическая решетка, и по характеру сил взаимодействия между ними различают *ионные*, *атомные*, *металлические* и *молекулярные кристаллы*. Впрочем, между ними не всегда можно провести резкие границы. Некоторые кристаллы занимают как бы промежуточные положения между перечисленными видами кристаллов.

Кристаллическая решетка *ионных кристаллов* построена из противоположно заряженных ионов, кулоновское притяжение между которыми создает *«ионную» связь*. В такой кристаллической решетке невозможно выделить отдельные группы связанных атомов, т. е. молекулы. Весь кристалл представляет собой как бы одну гигантскую молекулу. Типичными представителями ионных кристаллов являются NaCl, CsCl, $CaCl_2$ (полевой шпат). Кристаллические решетки CsCl и NaCl изображены на рисунках 169 и 170. Обе они кубические: для CsCl — объемноцентрированная, для NaCl — гранецентрированная. Ионы натрия и цезия заряжены положительно, хлора — отрицательно. Решетка CaF_2 также кубическая — гранецентрированная; элементарная ячейка состоит из иона кальция и двух ионов фтора. К ионным кристаллам относятся некоторые интерметаллические соединения (т. е. соединения металла с металлом), например AuZn, MgAg, CdAg и др.

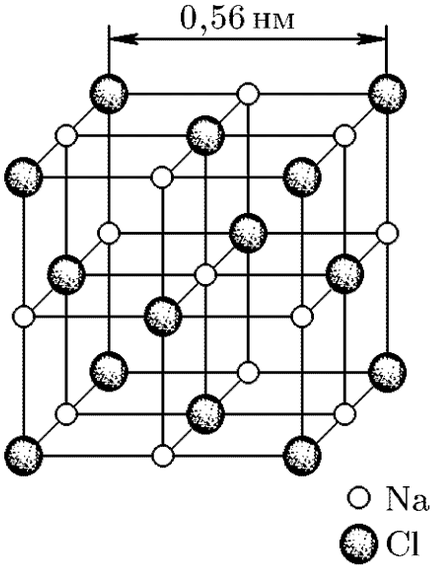


Рис. 170

Электростатическому взаимодействию между ионами кристалла соответствует определенная потенциальная энергия. В состоянии равновесия потенциальная энергия кристалла должна быть минимальна. Потенциальная энергия электростатического взаимодействия ионов одного знака всегда положительна. Если часть ионов заменить ионами противоположного знака, то потенциальная энергия уменьшится. Поэтому *гранями кристалла не могут быть кристаллические плоскости, состоящие из ионов одного знака*. Например, плоскости (100), (010) и (001) кристаллов CsCl состоят из ионов одного знака (рис. 169). Эти плоскости не могут служить гранями кристалла. Вот почему хлористый цезий не кристаллизуется в виде кубов. В случае NaCl кристаллические плоскости с теми же индексами состоят из ионов разных знаков. Благодаря этому NaCl и кристаллизуется в виде кубов.

Атомные кристаллы образуются атомами, которые связаны друг с другом так называемыми *гомеополярными* или *ковалентными связя-*

ми. Это те же связи, которые ведут к образованию молекул из одинаковых атомов, например H_2, O_2 и т. д. Гомеополярная связь, конечно, обусловлена электростатическим взаимодействием между электронами и атомными ядрами. Однако образование молекул невозможно объяснить «классически», т. е. на основе ньютоновской механики. Исчерпывающее понимание природы гомеополярной связи, как и всяких атомных явлений, стало возможным лишь с появлением квантовой теории (см. т. V, § 50, 53).

Решетки *металлических кристаллов* состоят из положительно заряженных ионов, между которыми находятся «свободные» электроны. Последние в металле «коллективизированы» и могут рассматриваться как своего рода «электронный газ». Металлы редко встречаются в виде одиночных кристаллов, называемых *монокристаллами.* Чаще всего они встречаются в виде *поликристаллов*, т. е. состоят из громадной совокупности мельчайших и беспорядочно ориентированных монокристаллов. Такая поликристаллическая структура металлов легко наблюдается с помощью микроскопа.

Молекулярные кристаллы состоят из молекул, связанных между собой *силами Ван-дер-Ваальса*, т. е. силами взаимодействия индуцированных молекулярных электрических диполей (см. т. V, § 52). Примером молекулярного кристалла может служить нафталин. Газы CO_2, O_2 и N_2 после затвердевания также образуют молекулярные кристаллы.

2. В некоторых твердых телах может осуществляться одновременно несколько видов связи. Примером может служить графит (рис. 171). Это единственный химический элемент, который кристаллизуется в гексагональной решетке. Решетка графита состоит из ряда плоских параллельных слоев, в которых атомы углерода располагаются в вершинах правильных шестиугольников. Расстояние между соседними слоями в 2,3 раза больше расстояния между соседними атомами отдельного слоя. Плоские слои связаны друг с другом силами Ван-дер-Ваальса. В пределах слоя три валентных электрона каждого атома углерода осуществляют гомеополярную связь с соседними атомами. Четвертый электрон остается свободным. Он «коллективизирован», однако не во всей решетке, а только в пределах одного слоя. Таким образом, в решетке графита сразу осуществляются три вида связи: гомеополярная и металлическая в пределах одного слоя и ван-дер-ваальсова между слоями. Этой особенностью связей объясняется своеобразная мягкость графита, на которой основано использование его для письма. Если

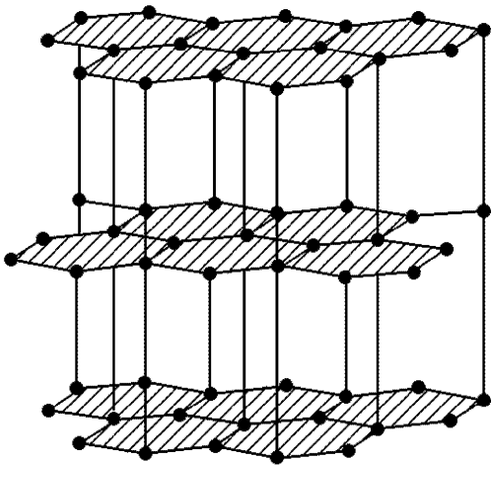


Рис. 171

давить на кристалл графита, слои решетки скользят и сдвигаются относительно друг друга.

Кристаллическая решетка алмаза, являющегося второй разновидностью углерода, построена совсем иначе. Она состоит из двух кубических гранецентрированных решеток, смещенных относительно друг друга на расстояние 1/4 пространственной диагонали куба (т. е. в направлении [111]), причем в узлах обеих решеток расположены атомы углерода (рис. 172). В результате каждый атом углерода оказывается окруженным четырьмя такими же атомами, которые располагаются на одинаковых расстояниях от него в вершинах тетраэдров. В противоположность графиту никаких плоских слоев решетка алмаза не содержит, и сдвинуть отдельный участок кристалла не удается. Поэтому алмаз много прочнее и тверже графита.

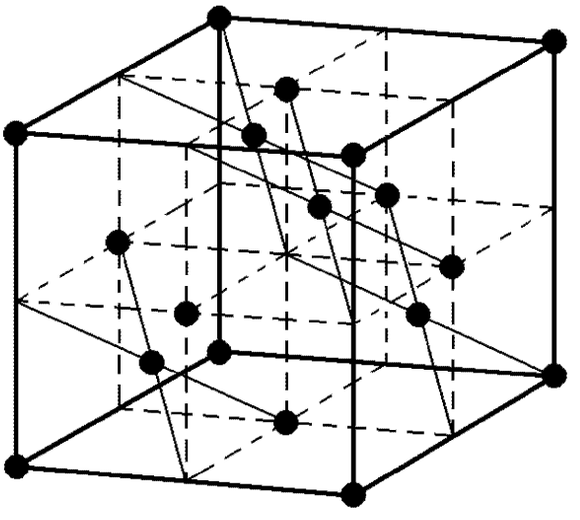


Рис. 172

3. Имеется около 20 химических элементов, которые кристаллизуются в виде кубических граненцентрированных решеток (Ag, Au, Cu, Al и др.). Около 15 элементов (Li, Na, K и т. д.) имеют кубические объемноцентрированные решетки. Однако ни один элемент не кристаллизуется в простой кубической решетке.

Чем же объясняется предпочтение, отдаваемое химическими элементами гранецентрированным и объемноцентрированным структурам по сравнению с простыми? Все дело в минимуме потенциальной энергии, при котором всякая система наиболее устойчива. Рассмотрим модель идеально твердых шаров, притягивающихся друг к другу. Минимум потенциальной энергии будет достигнут, когда шары «упакованы» наиболее плотно. Для этого необходимо (но не достаточно), чтобы шары соприкасались между собой. Посмотрим, в какой же из трех кубических решеток шары упакованы плотнее. Будем предполагать, что все шары одинаковы. В простой кубической решетке центры всех шаров располагаются в вершинах куба. На одну кубическую ячейку приходится один шар. Так как шары соприкасаются, то постоянная решетки a равна диаметру шара d. Объем шара $V_1 = \pi a^3/6 = 0{,}52a^3$. Упакуем теперь шары так, чтобы в центре каждого куба находилось по одному шару. Шары будут соприкасаться вдоль диагонали куба, так что длина ее станет равной $2d$. Но та же длина представляется выражением $a\sqrt{3}$, так что $2d = a\sqrt{3}$. Теперь на кубическую ячейку приходятся два шара с общим объемом $V_2 = \pi d^3/3 = \pi\sqrt{3}a^3/8 = 0{,}68a^3$. Наконец, упакуем шары так, чтобы в центре каждой грани находился центр шара. Тогда на кубическую ячейку будет приходиться

4 шара, и легко подсчитать, что их общий объем будет $V_4 = \sqrt{2}\pi a^3/6 = 0{,}74a^3$. Мы видим, что в первом случае 52% объема решетки заполнено шарами, во втором — 68%, в третьем 74%. Значит, из трех рассмотренных структур гранецентртированная упакована наиболее плотно.

4. Вообще из всех возможных структур гранецентрированная кубическая решетка упакована всего плотнее. Действительно, *наиболее плотную упаковку* можно получить следующим образом. Расположим сначала слой шаров в одной плоскости, как указано на рис. 173. Ясно, что в этой плоскости они будут упакованы наиболее плотно. Для того чтобы в дальнейшем можно было выражаться кратко, спроецируем центры шаров на плоскость, на которой они лежат. Эти проекции обозначим светлыми кружками (рис. 174). Спроецировав на ту же плоскость центры просветов между шарами, получим две системы точек, обозначенных на рис. 174 темными кружками и крестиками соответственно. Условимся далее всякий плотно упакованный слой называть слоем A, если центры его шаров расположены над светлыми кружками, слоем B, если они расположены над темными кружками, и слоем C, когда они расположены над крестиками. Теперь легко описать, как может быть получена наиболее плотная упаковка. Над первым слоем (A) помещаем второй такой же плотно упакованный слой, чтобы его шары расположились в просветах первого слоя. Это можно сделать двумя способами — взять в качестве второго слоя либо B, либо C. Поместим далее над вторым слоем третий плотно упакованный слой, что можно сделать также двумя способами, и т. д. Ясно, что всякая плотно упакованная структура может быть получена таким образом и записана в виде $ABCBAC\ldots$, причем в этой строке не могут стоять рядом слои, обозначенные одинаковыми буквами. Впрочем, из бесконечного множества мыслимых комбинаций реальное значение в учении о кристаллах имеют только два типа укладки, соответствующие схемам: 1) $ABCABC\ldots$ (*гранецентрированная кубическая структура*, рис. 175 *а*) и 2) $ABABAB\ldots$ (*гексагональная плотно упакованная структура*, рис. 175 *б*).

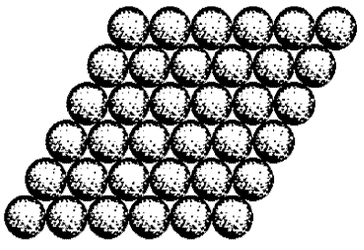


Рис. 173

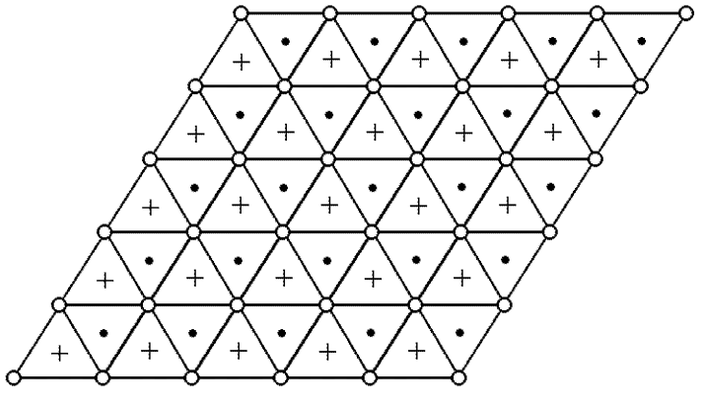
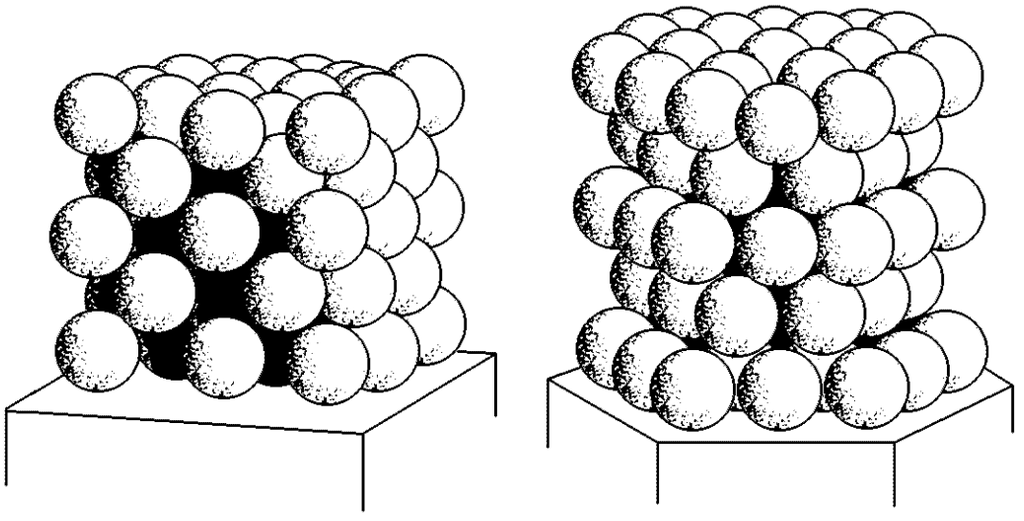


Рис. 174

Подсчитаем, какая часть пространства приходится на шары в первой из этих структур. Для этого, предполагая шары одинаковыми, будем строить все слои в виде ромбов, укладывая их друг к другу так,

чтобы получился ромбоэдр с острыми углами 60°. Если n — число шаров на ребре ромбоэдра, то полное число шаров в ромбоэдре будет $N = n^3$. Они занимают объем $v = \pi d^3 n^3/6 = \pi l^3/6$, где l — длина



а *б*

Рис. 175

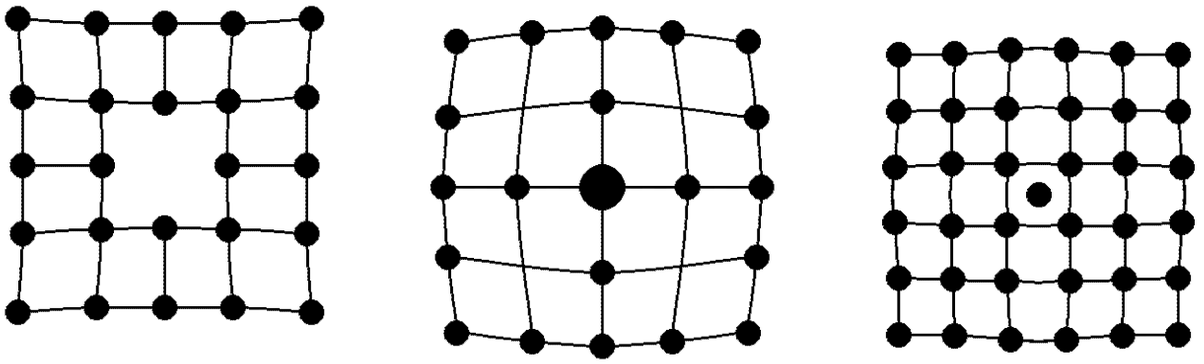
ребра ромбоэдра. Объем самого ромбоэдра $V = l^3/\sqrt{2}$. Таким образом, $v = \pi V/3\sqrt{2} = 0{,}74V$, т. е. на долю шаров приходится 74 % всего объема. Так и должно быть, ибо при рассмотренном способе укладки шаров они образуют гранецентрированную кубическую решетку.

§ 135. Дефекты в кристаллах

1. В реальных кристаллических решетках существуют отклонения от того идеального расположения атомов в решетке, которое мы до сих пор рассматривали. Все такие отклонения называются *дефектами кристаллической решетки*. Их можно подразделить на *макроскопические* и *микроскопические*. К макроскопическим дефектам относятся *поры*, *трещины*, *инородные макроскопические включения* и пр. Наиболее простыми микроскопическими дефектами являются *точечные дефекты*. К ним относятся: 1) отсутствие атома в каком-либо узле решетки (*вакансия*, рис. 176 *а*), 2) замена «своего атома» решетки каким-либо другим «чужим» атомом (рис. 176 *б*), 3) внедрение своего или чужого атома в межузельное пространство (*межузельный атом*, рис. 176 *в*). Точечные дефекты —вакансии и межузельные атомы — могут возникнуть в результате тепловых флуктуаций. Такие дефекты называются *термодинамически равновесными*. Очевидно, они столь же неизбежны, как и броуновское движение или любые флуктуации. При нагревании кристалла концентрация вакансий и межузельных атомов возрастает экспоненциально.

Энергия точечных дефектов много больше энергии тепловых колебаний решетки. Например, для меди энергия вакансии $E_в \approx 1$ эВ,

энергия межузельного атома $E_{\text{у}} \approx 3$ эВ, в то время как энергия тепловых колебаний kT даже вблизи температуры плавления (1084,5 °C)



а *б* *в*

Рис. 176

составляет всего 0,12 эВ. Поэтому равновесная концентрация точечных дефектов, как правило, невелика. Так, для меди вблизи температуры плавления равновесные концентрации вакансий и межузельных атомов, если их оценить по формуле Больцмана, будут соответственно

$$c_{\text{в}} \approx \exp\left(-\frac{E_{\text{в}}}{kT}\right) \sim 10^{-4}, \quad c_{\text{у}} \approx \exp\left(-\frac{E_{\text{у}}}{kT}\right) \sim 10^{-11}.$$

Однако в кристаллах в результате закалки, облучения нейтронами и пр. концентрация точечных дефектов часто бывает много выше равновесной. Такой «пересыщенный раствор» вакансий и межузельных атомов может распасться с образованием дислокаций.

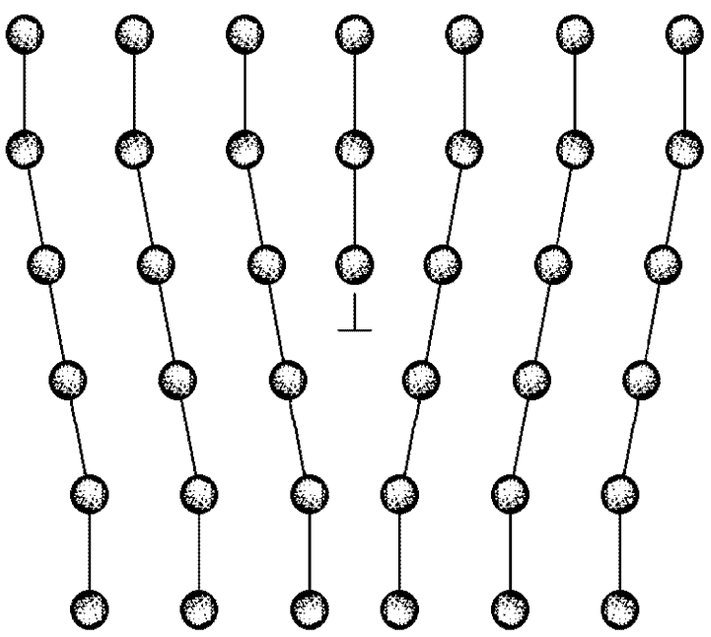


Рис. 177

2. *Дислокации* — это специфические *линейные дефекты* кристаллической решетки, нарушающие правильное чередование атомных плоскостей. В отличие от точечных дефектов, нарушающих *ближний порядок*, дислокации нарушают *дальний порядок* в кристалле, искажая всю его структуру. Поэтому именно дислокации играют наиболее важную роль в механических свойствах твердых тел.

Различают два главных типа дислокаций: *краевую* и *винтовую*. Схема краевой дислокации показана на рис. 177. Дислокация характеризуется лишней кристаллической плоскостью, вдвинутой между двумя соседними слоями атомов.

Линией дислокации в данном случае является прямая, перпендикулярная к плоскости рисунка и отмеченная на нем знаком $\perp$. «Лишний» слой атомов расположен над этим знаком. Краевая дислокация, образовавшаяся в результате неправильного наращивания кристаллической решетки, может существовать на протяжении десятков и сотен межатомных расстояний.

Винтовую дислокацию можно наглядно представить себе, произведя «разрез» решетки по полуплоскости и сдвинув части решетки по обе стороны разреза навстречу друг другу на один период параллельно краю разреза. Этот край называется *линией винтовой дислокации* и изображен на рис. 178 штриховой линией. Наличие винтовой дислокации превращает кристаллические плоскости решетки в *геликоидальную поверхность* (подобную винтовой линии без ступенек).

Для определения вида дислокации пользуются *методом Бюргерса*. Назовем *контуром Бюргерса* контур, составленный из основных векторов трансляции решетки так, чтобы он замыкался в идеальном кристалле. В дефектном кристалле при обходе вокруг линии дислокации контур Бюргерса окажется разомкнутым. Вектор, соединяющий его конечную точку с начальной, называется *вектором Бюргерса*. На рис. 178 в качестве примера показано, как строится вектор Бюргерса $\overrightarrow{EF}$ для винтовой дислокации. *В случае краевой дислокации вектор Бюргерса перпендикулярен, а в случае винтовой — параллелен линии дислокации.*

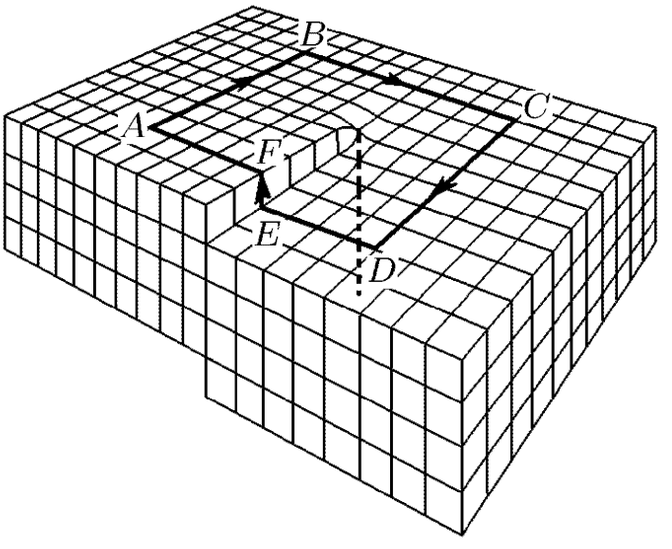


Рис. 178

Дислокации и их движение можно наблюдать с помощью электронного микроскопа. Другой метод основан на травлении кристалла специальными реагентами. В местах выхода дислокаций на поверхность кристалла разрушение его происходит более интенсивно. Благодаря этому возникают ямки травления, делающие дислокации видимыми.

Для характеристики числа дислокаций в теле вводят их плотность. *Плотностью дислокаций* называется число дислокационных линий, пересекающих единичную площадку, мысленно проведенную в теле. Это число меняется примерно от 10^2–10^3 в наиболее совершенных чистых монокристаллах до 10^{11}–10^{12} см$^{-2}$ в сильно деформированных (холоднообработанных) металлах.

3. Дефекты в кристаллах оказывают сильное влияние на их физические свойства (механические, магнитные, электрические и пр.). Рассмотрим, например, деформацию кристалла под действием касательных напряжений. Будем предполагать сначала, что кристалл —

идеальный, т. е. не содержит никаких дефектов. Пусть приложенное касательное напряжение τ параллельно одной из кристаллических плоскостей решетки (рис. 179). Под действием такого напряжения слой CD решетки сместится относительно слоя AB на расстояние x. По-

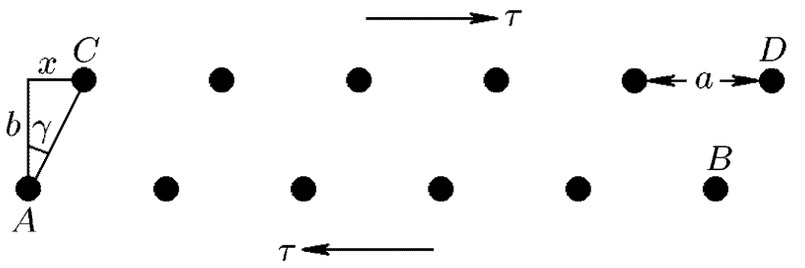


Рис. 179

тенциальная энергия решетки $U(x)$ будет периодической функцией смещения x с периодом a, равным постоянной решетки в направлении смещения. Она минимальна при $x = 0, a, 2a, \ldots$ и максимальна при $x = a/2, 3a/a, 5a/2, \ldots$ Минимумам $U(x)$ соответствуют *устойчивые*, максимумам — *неустойчивые положения равновесия*. Допустим, что на кривой $U = U(x)$ нет никаких других минимумов и максимумов. Если $x < a/2$, то после снятия напряжения решетка вернется в исходное положение равновесия, т. е. деформация будет *упругой*. Не то будет, когда x перейдет через точку максимума $x = a/2$. В этом случае произойдет самопроизвольный переход в ближайшее положение устойчивого равновесия $x = a$. Если при этом напряжение τ не снимать, то за ним последуют переходы в дальнейшие положения устойчивого равновесия $x = 2a$, $x = 3a$ и т. д. Иными словами, деформация станет *пластической*. Таким образом, максимальное смещение, при котором еще не получится пластической деформации, будет $x = a/2$. Соответствующее ему напряжение называется *пределом упругости*, или *пределом текучести*, кристалла. Для оценки предела упругости $\tau_{\text{у}}$ Френкель предположил, что периодическая функция $U(x)$ синусоидальна:

$$U(x) = U_0\left(1 - \cos\frac{2\pi x}{a}\right).$$

Если площадь грани AB взять равной единице, то приложенное напряжение

$$\tau = \frac{dU}{dx} = \frac{2\pi U_0}{a}\sin\frac{2\pi x}{a}.$$

Максимальное значение его, т. е. $2\pi U_0/2$, и будет пределом упругости τ_y. Таким образом,

$$\tau = \tau_{\text{у}}\sin\frac{2\pi x}{a}.$$

При малых x $\tau = \dfrac{2\pi\tau_{\text{у}}}{a}x$. С другой стороны, в этом случае $\tau = G\gamma$, где G — модуль, а γ — угол сдвига. Последний равен $\gamma = x/b$, где b — межплоскостное расстояние, т. е. расстояние между плоскостями AB

и CD. Сравнивая оба выражения, получаем

$$\tau_{\text{у}} = \frac{G}{2\pi}\,\frac{a}{b}. \tag{135.1}$$

В частности, для кубических кристаллов ($a = b$)

$$\tau_{\text{у}} = G/2\pi. \tag{135.2}$$

Если подставить сюда экспериментальные значения модулей сдвига, то для предела упругости наиболее употребительных материалов (металлов и пр.) получатся величины, лежащие в интервале приблизительно от 1000 до 10 000 Н/мм^2. Они примерно на два порядка превосходят наблюдаемые значения. Такое расхождение теории с опытом объясняется тем, что теория не учитывает различные дефекты, всегда содержащиеся в реальном кристалле. Механизм действия дефектов, благодаря их разнообразию и нерегулярному расположению в кристаллах, может быть самым разнообразным. Ограничимся поэтому одним сильно упрощенным примером. Рассмотрим идеализированный дефект, напоминающий краевую дислокацию. Пусть в кристалле имеется

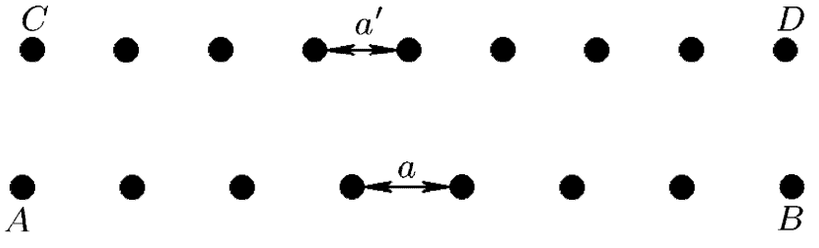


Рис. 180

кристаллическая плоскость, содержащая по сравнению с соседними параллельными плоскостями лишние атомы (рис. 180). Допустим, что n атомов одной плоскости укладываются на том же расстоянии, что и $n + 1$ атомов другой плоскости. (Такой дефект напоминает нониус.) Тогда можно воспользоваться оценкой Френкеля. Только вместо прежнего периода a надо взять больший период, а именно $na = (n + 1)a'$. Это понижает предел упругости в n раз.

4. Совершенно аналогичная картина имеет место в вопросе о *прочности кристаллов на разрыв*. Теоретические оценки прочности приводят к результату, что эта величина для кристаллов может достигать нескольких тысяч — вплоть до 10 тысяч — ньютонов на 1 мм^2. На самом деле в большинстве случаев наблюдаемые значения примерно в 100 и даже в 1000 раз меньше. Расхождение опять объясняется тем, что теория не учитывает наличия дефектов в реальных кристаллах. На прочность кристаллов прежде всего влияют *микроскопические поверхностные и объемные трещинки*, имеющиеся в кристаллах. Один из механизмов влияния трещин состоит в том, что напряжения в окрестности трещин распределяются *неравномерно*. Они максимальны вблизи краев трещин и могут во много раз превосходить средние напряжения в кристалле. Как только эти максимальные напряжения достигают теоретического предела прочности, происходит увеличение

размеров трещинок, а затем и *разрыв кристалла*. А. Ф. Иоффе (1880–1960) экспериментально показал, что при погружении кристаллов поваренной соли в воду ее прочность возрастает с 5 до 1600 Н/мм^2, т. е. до величины, близкой к теоретической. Он объяснил этот результат растворением в воде поверхностного слоя кристалла и ликвидацией в нем микротрещин и других дефектов.

Существуют и другие причины понижения прочности кристаллов на разрыв. Например, как показал П. А. Ребиндер (1898–1972), на прочность материалов большое влияние оказывают *поверхностно-активные вещества*, адсорбированные на поверхности тел и понижающие при этом их поверхностную энергию. По-видимому, частицы адсорбированных веществ распирают зародышевые трещинки, проникая при этом в глубь тела и сильно уменьшая его прочность. В воздухе всегда имеется влага и другие вещества, адсорбирующие на поверхности тел. При удалении воздуха вместе с примесями происходит дезабсорбция, сопровождающаяся повышением прочности на разрыв.

Влияние дефектов на прочность материалов отчетливо проявляется при рассмотрении так называемого «*масштабного фактора*», т. е. зависимости средней прочности образца данного материала от его размера. Для образцов обычных размеров, порядка нескольких сантиметров и больше, эта зависимость практически не наблюдается. Но для тонких нитей она выступает отчетливо. Например, прочность стеклянных нитей с диаметрами 22, 16, 12,5, 8 и 2,5 мкм равна соответственно 220, 1000, 1400, 2000 и 5500 Н/мм^2. Таким образом, при уменьшении диаметра стеклянной нити от 22 до 2,5 мкм прочность повышается примерно в 25 раз. Дело в том, что вероятность встречи «опасного дефекта» (например, трещины), приводящего к разрыву образца, для тонких нитей меньше, чем для толстых.

Прочность материала увеличивается в результате подавления процессов зарождения трещин, а также в результате создания препятствий для их распространения в теле. Поэтому наиболее прочные материалы могут быть получены двумя противоположными способами. Один из них состоит в изготовлении бездефектных (например, нитевидных) кристаллов, где устранены источники внутренних напряжений, на которых могут зарождаться трещины. Другой путь, наоборот, состоит в максимальном искажении правильной структуры кристалла, что затрудняет распространение в теле трещин и пластических деформаций. Техника получения сверхпрочных материалов и сплавов в настоящее время использует только второй способ (*легирование*, т. е. введение в решетку примесей из посторонних атомов, *наклеп*, т. е. сильное пластическое деформирование кристаллической решетки при холодной обработке металлов, *закалка* и пр.)

Именной указатель

Авенариус 391
Авогадро 33, 34, 184, 197, 198, 207, 211–213, 220, 275, 278, 319, 461
Архимед 416
Афанасьева-Эренфест 135, 136

Бернулли 80–84, 183, 431
Бертло 377
Бозе 306–308, 310–312, 316
Бойль 21, 25, 33, 70, 78, 183, 373, 374, 377, 379, 381, 382
Больцман 132, 184, 198, 199, 205, 207, 211, 246, 268–270, 272–276, 278–282, 287, 289, 290, 296, 297, 299, 302, 304, 306, 307, 309–312, 316–318, 334, 346, 374, 375, 447, 448, 524
Борман 333, 334
Борн 320, 333, 334
Браве 503, 504, 508, 510–512, 516
Браун 153, 154, 157, 159, 160, 478
Броун 206
Бюргерс 525

Ван-дер-Ваальс 141, 367–369, 373, 375–382, 386, 392–398, 400, 438, 445, 457, 463, 520
Ван Марум 382
Вант Гофф 484, 486
Вильсон 464
Вроблевский 405

Гальтон 242, 250, 252
Гаусс 243, 250
Гейзенберг 23, 243, 244, 301
Гей-Люссак 25, 33, 68, 69, 184, 274
Гельмгольц 43, 85, 136, 137, 148, 149, 151
Генри 479
Гесс 64, 65
Гиббс 133–137, 148, 149, 297, 305, 491, 493
Глезер 464
Голик 390
Гуи 389

Даламбер 231, 232
Дальтон 34, 184, 191, 206
Дарлинг 412
Дебай 320
Де-Бройль 244
Дезорм 74, 76, 78
Джонс 368
Джоуль 43, 50–52, 57, 67–72, 74, 82, 85, 107, 128, 140–143, 161, 221, 395–401, 403–405
Дикель 352
Дирак 265, 306–308, 310, 312, 316
Дитеричи 374, 376, 377, 380, 381, 401
Дутсон 351
Дьюар 14, 341, 342, 405, 444
Дюлонг 220, 221, 224

Иоффе 528

Кальете 405
Камерлинг-Оннес 377, 405
Каньяр де ля Тур 387, 388
Капица 404, 405, 469
Капплер 212
Карман 320
Карно 85–87, 94–102, 104–106, 110–114, 116, 117, 121, 127, 197, 414, 415, 442, 443
Кельвин (Вильям Томсон) 21, 23, 24, 26, 69, 70, 82, 86, 88–91, 95–98, 100, 112, 113, 116, 117, 120, 135, 140–143, 151, 161, 178–180, 395–401, 403–405, 453, 459, 462
Кирхгоф 447
Клапейрон 34, 35, 68, 70, 71, 73, 75, 94, 107, 197, 198, 315, 343, 366, 369, 396, 400, 440, 442, 443, 445–447, 450, 452, 469, 473, 486–488
Клаузиус 86, 89–91, 95, 96, 111, 113–115, 118–123, 125–128, 130–132, 135, 148, 153, 184, 289, 290, 322, 323, 327, 377, 383, 387, 439, 440, 442, 443, 452, 469, 473, 487, 488
Клеман 74, 76, 78
Клод 403, 404
Клузиус 352
Книппинг 503
Кнудсен 355, 359, 361
Коновалов 493
Копп 221
Крукс 358
Кук 405
Куш 267, 268
Кюри 28, 468

Ламмерт 266, 268
Ландау 516

Лаплас 78, 79, 419–422, 431, 459
Лауэ 503
Лебедев 161
Леннард 368
Ленц 154
Линде 403, 404
Лиувилль 259
Ломоносов 183
Лоренц 274
Лошмидт 334
Лукреций 207

Майер 43, 71, 72, 80, 85, 94, 109, 396, 447
Максвелл 138, 160, 184, 205, 240, 243, 245, 246, 250, 253, 254, 265, 267, 268, 270–274, 279, 280, 297, 304, 324, 325, 341, 342, 351, 361, 362, 381, 383, 385, 439, 442
Мариотт 21, 25, 33, 70, 78, 183, 373, 382
Менделеев 34, 35, 68, 70, 71, 73, 75, 94, 197, 198, 366, 387, 388, 446, 473
Менье 429
Миллер 267, 268

Нернст 313–317, 319, 320, 415, 445
Ньютон 78, 79, 181, 183, 185, 326, 357, 372, 375

Ольшевский 405
Оствальд 87

Пастер 501
Перрен 206, 210, 211, 275–278, 349
Пирсон 405
Планк 88–91, 95, 96, 100, 112, 113, 116, 117, 135, 225, 289, 290, 313, 318, 459
Плато 411, 412
Поль 59, 60, 70, 73, 77, 78, 123, 140, 156, 159, 178, 179, 190, 208, 229, 283, 308, 325, 337, 339, 342, 347, 355, 359, 362, 376, 397, 415, 416, 443, 448, 470
Пти 220, 221, 224
Пуазейль 342, 359
Пуассон 73, 216, 218, 249, 402
Пфеффер 483, 484

Рауль 485–487
Ребиндер 528
Ричардсон 265
Рэлей 404, 410, 411, 429

Сименс 403
Симон 405
Скотт 466
Смолуховский 207, 209
Стирлинг 303, 309
Стокс 208

Томсон Вильям, см. Кельвин

Фарадей 151, 382, 405
Федоров 514
Ферми 265, 306–308, 310, 312, 316
Физо 266
Фик 344
Фихтенгольц 303
Френкель 256, 257, 526, 527
Фридрих 503
Фурье 61, 162

Хемпсон 403, 404

Цартман 267, 268
Цельсий 24, 31, 60, 453

Чепмен 340, 351

Штерн 192, 193, 266, 268

Эйкен 340
Эйлер 175, 205, 420, 432, 433
Эйнштейн 135, 207, 209, 210, 306–308, 310–312, 316, 317, 319, 320, 334, 345, 346, 348
Элдридж 266, 268
Эндрюс 383, 386–388
Энског 351
Эренфест 469, 470

Предметный указатель

Абсолютная термодинамическая температура 98
Абсолютный нуль температуры 23, 100
Агрегатные состояния 314
Адиабата 73
— влажная 474
— сухая 474
Адиабатическая постоянная 73
Адиабатический инвариант 219
— процесс 73
— температурный градиент 471
Азеотропная смесь 495
Анизотропия 515

Базис решетки 503
Базисные (основные) векторы 503
Базисный (основной) параллелепипед 503
Барометрическая формула 270
Ближний порядок 524
Бозе-эйнштейновская конденсация 313
Бозоны 306
Броуновское движение 206, 348
— — вращательное 211
— — , формула Эйнштейна 209

Вакуум 352
Вариантность 492
Вектор Бюргерса 525
— трансляции 503
Вероятностные закономерности 228
Вероятность 229–231, 233, 234
— априорная 231
— математическая 290
— рассеяния 332
— , теорема сложения 232
— , — умножения 232, 233
— термодинамическая 290
— , условие нормировки 233, 239
— условная 233
Вириальные коэффициенты 103, 377
Внутреннее давление 372
Возгонка 435
Волна де-Бройля 244
Время рассеяния атмосферы 280
— — — бесконечно разреженной 282
— релаксации 15
Вырождение полное 312
Вязкость (коэффициент) 334, 337

Газ Ван-дер-Ваальса 373
— и пар 384
Газы вырожденные 244, 316
— идеальные 33
— — , давление 191
— реальные 33
Гальванический элемент 149
— — , электродвижущая сила 150
Гидраты 477
Гравитационные волны 432
Гравитационный эффект 389
Градус 20
Группа симметрии 499
— — пространственная 499
— — точечная 499

Дальний порядок 524
Детандер 403–405
Дефекты кристаллической решетки 523
— — — , вакансии 523
— — — , внедрения 523
— — — макроскопические 523
— — — микроскопические 523
— — — термодинамически равновесные 523
— — — точечные 523
Деформация 526
Джоуль 59
Диаграмма состояния 382, 453, 492
Диаметр молекулы 369
Динамические законы 228
Динамическое отопление 120
— состояние 38
Дислокации 524
— винтовые 524
— краевые 524
Диффузионный поток 344
Диффузия 322
— взаимная 343
— концентрационная 343, 346, 348
Диэлектрическая проницаемость 367
Доска Гальтона 242
Дросселирование газа 82, 141

Естественные грани кристалла 517

Жидкий гелий I 468, 470
— — II 468–470

Жидкость перегретая 385
— переохлажденная 464

Закалка 528
Закон Авогадро 33, 184, 197
— Бойля–Мариотта 21, 33
— Вант Гоффа 484
— возрастания энтропии 128
— Гей-Люссака 25, 33, 184
— Генри 479
— Гесса 64
— Дальтона 34
— Джоуля 70
— кратных отношений 183
— кубов Дебая 320
— постоянного состава 183
— распределения 479
— — Больцмана 269
— — скоростей Максвелла 250
— Рауля 486
— рациональностей граней кристалла 518
— соответственных состояний 380
— Фика 344
Запас теплоты 61
Затвердевание 160
Зеркальная изомерия 501
Зеркальные изомеры 501

Изобара 45
Изображающее пространство 492
Изображающие точки 492
Изотерма 45
— критическая 378
Изотермы газа Ван-дер-Ваальса 377
— реального газа 381
Изотропные тела 297
Изохора 45
Инверсия 498
Интеграл Пуассона 249
Интенсивность пучка 328
Испарение 435, 455
Испытание 229

Калориметр 56
Калориметрическое тело 56
Калория 60
Камера Вильсона 464
Капиллярно-гравитационные волны 430
— —, фазовая скорость 432
Карно машина 95
— теорема 95
— — вторая 114
— цикл 94
Катеноид 429
Квазинейтральный газ 366
Квазиравновесный процесс 44
Квазистатический процесс 44
Кельвин (единица температуры) 24, 98
Кипение 455
Кислородная шкала физическая 34
— — химическая 34
Количество теплоты 57
— —, закон сохранения 61
— — приведенное 123
Компенсирующие процессы (компенсации) 89
Компонент 477, 489
Конвективная устойчивость 470
Конвекция 470
Конденсация 435
— обратная 496
Конденсированные тела 435
Контур Бюргерса 525
Концентрация 477
— массовая 477
— молярная 477
— объемная 477
Коэффициент диффузии 344, 348
— полезного действия 87
— рассеяния 332
— теплопередачи 181
— термодиффузии 351
Кратность уровня (вырожденная) 301
Кривая возгонки 452
— жидкости 493
— испарения 441, 452
— кристаллизации 435
— пара 493
— плавления 452
Кристаллические классы 515
— плоскости 516
— системы 510
Кристаллографические системы координат 516
Кристаллы 519
Критерии термодинамической устойчивости 152
Критическая точка 378, 388, 495
Критический коэффициент 379
— объем 378
Критическое давление 378
— состояние 378
Куб Браве 508

Ламинарное течение адиабатическое 81
Легирование 528

Линия винтовой дислокации 525
— краевой дислокации 525

Макроскопические параметры 39
— — внешние 40
— — внутренние 40
Макросостояние 39, 291
Масштабный фактор 528
Математическая теория теплопроводности 61
Математическое ожидание 236, 239
Международная практическая температурная шкала 31
Мениск 388
Метастабильные состояния 384
Метастабильная жидкость 457
Метод адиабатического размагничивания 406
— Клемана и Дезорма 76
— Клода 404
— термодинамических функций (потенциалов) 138, 144
— циклов 104, 138
Механический эквивалент теплоты 72
Микросостояние 38, 291
Миллеровские индексы 517
Мишень 328
Молекулярная масса 33
— физика 8
Молекулярно-кинетическая теория 8
Моль 33
Монокристаллы 520

Нагреватель 86
Наиболее плотная упаковка 522
Наклеп 528
Начало термодинамики второе 85
— — —, основной постулат 87
— — —, формулировка Клаузиуса 89
— — —, — Планка 88–90
— — —, — Томсона 88–90
— — общее 14
— — первое 43
— — третье 313
Независимые переменные 297
Непрерывность жидкого и газообразного состояния 386
Неравенство Клаузиуса 113, 118
Нулевые колебания 225

Обесценивание энергии 114, 121
Обобщенные координаты 203
— силы 49
— скорости 203
Оболочка адиабатическая 14, 50
— теплонепроводящая 14
— теплопроводящая 16
Общая теория теплоты 8
Однородность 515
Оловянная чума 465
Операции симметрии 498
Осмометр 483
Осмос 483
Осмотический шок 485
Осмотическое давление 483, 484
Основные единицы длины 516
Основные периоды 503
Ось симметрии 499
— — винтовая 513
— — зеркально-поворотная 499, 500
— — n-го порядка 499
Отклонение от среднего значения 237
Относительная частота события 235
Отражение в плоскости и точке 498
Ошибка 237
—, кривая Гаусса 243
—, нормальный закон Гаусса 243
— систематическая 236
— случайная 237
— средняя квадратичная отдельного измерения 237
Пар насыщенный 439
— —, давление 445
— —, теплоемкость 450
— ненасыщенный (перегретый) 436
— пересыщенный 384, 436
Парадокс Гиббса 134
Параллелепипед Браве 508
— приведенный 506
Параллельный перенос 498
Парообразование 435
Парциальное давление 34
Перпетуум мобиле второго рода 87
— — первого рода 59, 87
Пирометр 30
Плавление 435
Плазма 54
Пластинчатое состояние 408
Плоскость зеркального скольжения 513
— симметрии 499, 499
— — зеркально-поворотная 499
Плотность вероятности 239
— потока теплоты 162
Пневматическое огниво 50, 73
Поверхностно-активные вещества 413
Поверхностное натяжение 407–409

Поворот 498
Показатель преломления 367
Полевая частица 328
Поликристаллы 520
Полиморфизм 435
Полиморфные превращения 436
Политропа 74
— , показатель 74
Полупроницаемые перегородки 133, 482
Полярные молекулы 367
Постоянная Авогадро 198
— Больцмана 198
— Планка 225
— Сёзерленда 330
— Фарадея 151
Потенциал Леннарда–Джонса 368
Правило Джоуля и Коппа 221
— Дюлонга и Пти 221
— Максвелла 383, 439, 439
— рычага 385, 480
— фаз 489
Предел упругости (текучести) 526
Преобразование симметрии 498
— тождественное 498
Приведенная температура 380
Приведенное давление 380
Приведенный объем 380
Принцип детального равновесия 256
— Ле Шателье–Брауна 154
— неопределенностей Гейзенберга 243
— суперпозиции температур 174
Прицельное расстояние 329
Пробная частица 328
Пространство скоростей 240
Процесс адиабатический 73
— Джоуля–Томсона 69
— изобарный 45
— изотермический 45
— изохорный 45
— квазиравновесный 44
— квазистатический 44, 91, 92
— Клаузиуса 89
— круговой 48
— — (цикл) Карно 94
— необратимый 91
— обратимый 91
— — в узком смысле 91
— — — широком смысле 91
— политропический 74
— Томсона–Планка 89
Прочность на разрыв 527
Пузырьковая камера 464
Пучок 328
Работа макроскопическая 45, 50
— максимальная 148
— микроскопическая 57
— полезная 149
— элементарная 45
Рабочее тело 86
Равенство Клаузиуса 122
Равновесие динамическое 15
— механическое 16, 434
— тепловое 13, 16
— термическое 13, 16
— термодинамическое 13, 14
— фазовое 436
— химическое 16
Радиометр Крукса 358
Радиометрические силы 358
Радиометрический эффект 358
Разделение 351
— изотопов 352, 354, 358
Разделительная колонка 352
Распределение Бозе–Эйнштейна 308
— Больцмана 269, 296, 312
— Гиббса каноническое 305
— Максвелла–Больцмана 273, 304
— Ферми–Дирака 308
Расстекловывание 465
Раствор 476
— концентрированный 476
— крепкий 476
— насыщенный 477
— ненасыщенный 478
— пересыщенный 478
— слабый (разбавленный) 476
— твердый 481
— — типа внедрения 481
— — — замещения 481
Растворенное вещество 476
Растворимость 478
Растворитель 476
Реакция экзотермическая 64
— эндотермическая 64
Ректификация 494
Релаксация 15
Реперные точки 20, 26
Решетка базоцентрированная 511
— Браве 504
— гранецентрированная 510
— кристаллическая 503
— обратная 506
— объемноцентрированная 510
— примитивная 503
— простая 510, 511

Решетка
— пространственная 503
— сложная 508

Самодиффузия 343
Сверхвысокие температуры 198
Сверхпроводимость 468
Сверхтекучесть 469
Связь гомеополярная (ковалентная) 519
— ионная 519
— металлическая 519
Сегнетоэлектрики 468
Сжижение газов 401
«Сигара» 493
Силы валентные 366
— Ван-дер-Ваальса 367, 520
— дальнодействующие 371
— дипольно-ориентационные 367
— дисперсионные 367
— молекулярные 55, 366
— отталкивания 368
— притяжения 367
— химические 366
Симметрия 498
Система адиабатически изолированная 50
— бинарная 489
— в термостате 305
— двухкомпонентная 489
— дивариантная (бивариантная) 492
— замкнутая 14
— изолированная 14
— инвариантная (нонвариантная) 492
— моновариантная (универиантная) 492
— однокомпонентная 489
— поливариантная 492
— простая 49
Скоростные (изображающие) точки 240
Скорость звука лапласова 79
— — ньютонова 79
— молекул 191
— — наиболее вероятная 252
— — средняя арифметическая 191
— — — квадратичная 191
— убегания 279
Смачивание и несмачивание 417, 418
Событие 229
— достоверное 229
— невозможное 229
— случайное 230
События единственно возможные 230
— независимые 234
— несовместимые 230
— , произведение 230
— противоположные 233
— равновероятностные 230
— , сумма 229
Сольваты 477
Соответственные состояния 380
Соотношение неопределенностей Гейзенберга 301
Соотношения Максвелла 138
— Эренфеста 469
Состояние динамическое 291
— термодинамическое 85
Сосуд Дьюара 14
Спай измерительный 29
— сравнения 29
Среднее значение 236
Средняя длина свободного пробега 323, 324, 327
— кривизна поверхности 420
Статистика Бозе–Эйнштейна 306, 310, 311
— Больцмана 302
— Ферми–Дирака 306, 310
Статистические независимые величины 293
Статистические величины 228
— закономерности 228
Статистический вес 290, 292
Статическая механика 10
Стационарные состояния 301
Степени свободы 203
— — замороженные 224
Степень диссоциации 40
— ионизации 40
Стереоизомеры 501
Столкновения 186
Сублимация 435
Сумма состояний 318
Сфера молекулярного действия 369
— ограждения 323

Температура 13, 14
— абсолютная 23
— — отрицательная 100
— абсолютного кипения 388
— Бойля 374
— вырождения 244
— кинетическая 197
— кипения 455
— критическая 378, 379, 388
— опорная 29

Температура
— смешения 480
— — верхняя 481
— — нижняя 481
— термодинамическая 26, 96, 98
— энергетическая 197
Температурная шкала 19
— — эмпирическая 22, 22
Температурные волны 175–178
— —, коэффициент затухания 177
— —, скорость 176
— точки 20
Температурный коэффициент давления 25, 38
— — объемного расширения 25, 37
— — сопротивления 28
Температуропроводность 163
Теорема Менье 429
— о знаках 157, 160
— — равномерном распределении кинетической энергии 203–205, 258
— Эйлера 420
Теория вероятностей 229
— теплоемкости квантовая твердых тел 221, 317
— — — — — —, формула Эйнштейна 319
— — классическая, газов 212
— — —, твердых тел 220
Тепловая смерть Вселенной 132, 290
— теорема Нернста 313
— функция 63, 136
Тепловое движение 186
— скольжение 357
Тепловой резервуар 61
— эффект реакции 64, 476
Теплоемкость 65
— молярная 66, 221
— удельная 66
Теплообмен 57
Теплообменник 404
Теплопередача 180
Теплопроводность 162, 339
—, теорема единственности 169
Теплород 11
Теплосодержание 63, 136
Теплота возгонки 442
— испарения 442
— образования 64
— — поверхности пленки 414
— плавления 442
— растворения 476
— фазового превращения 442
Термисторы 28
Термическая диффузия (термодиффузия) 350
Термодинамика 8
— аксиоматическая 10
— статическая 10
— техническая 9
— феноменологическая (формальная) 10
— физическая 9
Термодинамическая шкала температур 97, 98
Термодинамические величины интенсивные 139
— — экстенсивные 139
— соотношения 85
Термодинамический потенциал 136, 139
Термодиффузионное отношение 351
Термометр 19
— газовый 22, 26
— жидкостный 27
— платиновый 28
— полупроводниковый 28
— сопротивления 27
— термопарный 29
Термометрическая величина 19
Термометрическое тело 19
Термопара 18
— платинородиевая 30
Термос 14
Термоскоп 17
Термостат 305
Термоэлектрический ток 18
Термоэлектронная эмиссия 265
Термоядерные реакции 199
Течение Кнудсена 359
— молекулярное 359
Точка Кюри 468
Трансляционная симметрия 503
Трансляция 498
Тройная точка 20, 452, 453

Убегающие молекулы 279
Углеродная шкала 34
Узловые линии 517
— плоскости 516
Узлы кристаллической решетки 503
Универсальная газовая постоянная 34
Уравнение адиабаты Пуассона 73, 218
— Бертло 377
— Ван-дер-Ваальса 373
— Гиббса–Гельмгольца 137, 148, 148

Уравнение Дитеричи 376
— Камерлинга-Оннеса 377
— Клапейрона 34
— Клапейрона–Клаузиуса 442
— Клаузиуса 377
— Майера 71
— Менделеева–Клапейрона 34
— состояния 33, 35
— — калориметрическое 68
— — каноническое 137
— — обобщенное 41
— — приведенное 380
— — термическое 68
— теплопроводности 164
Уровни энергии 301
— — кратные (вырожденные) 301
— — простые (невырожденные) 301

Фаза 382, 436
— , условие равновесия 435
Фазовое пространство 300
— равновесие 19
Фазовые превращения 435
— — второго рода 467
— — первого рода 467
— точки 300
Фазовый объем 300
Фён 474
Фермионы 306
Физическая кинетика 10
Флуктуации плотности 299
— энергии 295, 297
— энтропии 298
Флуктуационная гипотеза Больцмана 290
Флуктуация 39, 286, 292
— , среднеквадратичная 293
— , — относительная 293
— , средний квадрат 293
Формула Больцмана 289, 290
— Кнудсена 361
— Лапласа (поверхностное натяжение) 420
— — (скорость звука) 78
— Ньютона (скорость звука) 78
— Планка 318
— Пуазейля 342
— Сёзерленда 330
— Стирлинга 303
— Стокса 208
— Томсона (Кельвина) 459
— Эйнштейна (подвижность частицы) 346
Фотонный газ 193
— — , давление 193
Функции состояния 47
Функция распределения по скоростям 241, 245
Характеристическая температура 225, 226
Химический потенциал 139, 311–313
Холодильная машина 90
Холодильник 86

Центр симметрии 499, 500
Цикл 48
Число степеней свободы 203, 492

Шарик Плато 412
Шкала температур 19
— — идельно-газовая 21, 25
— — Международная практическая 31
— — термическая (Кельвина) 21, 26, 96
— — Цельсия 24

Эвтектика 497
Эвтектическая точка 497
Экспансионный метод 405
Электрическая поляризация 367
Электронвольт 198
Элементы симметрии 499
Энантиоморфизм 513
Энергия внутренняя 43, 52, 53
— нулевая 23, 225, 318
— поверхностная 55
— свободная 136, 139
— связанная 148
— тепловая 43
Энтальпия 63, 139
Энтропия 122, 123, 285
— абсолютная 314
— и вероятность 285
Эффект Джоуля–Томсона 69, 140
— — — дифференциальный 141, 395, 397
— — — — , кривая инверсии 397, 400
— — — — , температура инверсии 397
— — — интегральный 141, 398
— — — — , кривая инверсии 400
— — — — , температура инверсии 400
— — — отрицательный 396
— — — положительный 395
Эффективное сечение 328, 329
Эффузионный поток 353
Эффузия изотермическая 356
— разреженного газа 353
— тепловая 354

Явления переноса 334
Ячейка 503
— фазовая 300

ПРИЛОЖЕНИЕ

ЕДИНИЦЫ И РАЗМЕРНОСТИ ФИЗИЧЕСКИХ ВЕЛИЧИН В СИ

Величина		Единица		
наименование	размерность	наименование	обозначение	связь с основными единицами СИ
Длина	**L**	**метр**	м	**Основная единица**
Площадь	L^2	квадратный метр °гектар	$м^2$ га	$1\ га = 10^4\ м^2$
Объем	L^3	кубический метр °литр	$м^3$ л	$1\ л = 10^{-3}\ м^3$
Плоский угол		радиан	рад	*Дополнительная единица*
Телесный угол	—	стерадиан	ср	*Дополнительная единица*
Время	**T**	**секунда**	**с**	**Основная единица**
Скорость	LT^{-1}	метр в секунду	м/с	
Ускорение	LT^{-2}	метр в секунду в квадрате	$м/с^2$	
Угловая скорость	T^{-1}	радиан в секунду	рад/с	$1\ рад/с = 1\ с^{-1}$
Угловое ускорение	T^{-2}	радиан на секунду в квадрате	$рад/с^2$	$1\ рад/с^2 = 1\ с^{-2}$
Частота периодического процесса	T^{-1}	герц	Гц	$1\ Гц = 1\ с^{-1}$
Частота вращения	T^{-1}	секунда в минус первой степени	$с^{-1}$	
Масса	**M**	**килограмм**	**кг**	**Основная единица**
Плотность	$L^{-3}M$	килограмм на кубический метр	$кг/м^3$	

Величина		Единица		
наименование	размерность	наименование	обозначение	связь с основными единицами СИ
Удельный объем	$\mathrm{L^3M^{-1}}$	килограмм на кубический метр	$\text{м}^3/\text{кг}$	
Массовый расход	$\mathrm{MT^{-1}}$	килограмм в секунду	кг/с	
Объемный расход	$\mathrm{L^3T^{-1}}$	кубический метр в секунду	$\text{м}^3/\text{с}$	
Сила	$\mathrm{LMT^{-2}}$	ньютон	Н	$1\,\text{Н} = 1\,\text{кг}\cdot\text{м}\cdot\text{с}^{-2}$
Давление	$\mathrm{L^{-1}MT^{-2}}$	паскаль	Па	$1\,\text{Па} = 1\,\text{Н}/\text{м}^2 = 1\,\text{м}^{-1}\cdot\text{кг}\cdot\text{с}^{-2}$
Жесткость	$\mathrm{MT^{-2}}$	ньютон на метр	Н/м	$1\,\text{Н}/\text{м} = 1\,\text{кг}\cdot\text{с}^{-2}$
Напряжение (механическое)	$\mathrm{L^{-1}MT^{-2}}$	паскаль	Па	$1\,\text{Па} = 1\,\text{Н}/\text{м}^2 = 1\,\text{м}^{-1}\cdot\text{кг}\cdot\text{с}^{-2}$
Динамическая вязкость	$\mathrm{L^{-1}MT^{-1}}$	паскаль-секунда	Па · с	$1\,\text{Па}\cdot\text{с} = 1\,\text{м}^{-1}\cdot\text{кг}\cdot\text{с}^{-1}$
Кинематическая вязкость	$\mathrm{L^2T^{-1}}$	квадратный метр в секунду	$\text{м}^2/\text{с}$	
Поверхностное натяжение	$\mathrm{MT^{-2}}$	ньютон на метр	Н/м	$1\,\text{Н}/\text{м} = 1\,\text{кг}\cdot\text{с}^{-2}$
Импульс	$\mathrm{LMT^{-1}}$	килограмм-метр в секунду	кг · м/с	
Момент силы	$\mathrm{L^2MT^{-2}}$	ньютон-метр	Н · м	$1\,\text{Н}\cdot\text{м} = 1\,\text{м}^2\cdot\text{кг}\cdot\text{с}^{-2}$
Момент импульса	$\mathrm{L^2MT^{-1}}$	килограмм-метр в квадрате в секунду	$\text{кг}\cdot\text{м}^2/\text{с}$	
Момент инерции	$\mathrm{L^2M}$	килограмм-метр в квадрате	$\text{кг}\cdot\text{м}^2$	
Работа, энергия	$\mathrm{L^2MT^{-2}}$	джоуль	Дж	$1\,\text{Дж} = 1\,\text{Н}\cdot\text{м} = 1\,\text{м}^2\cdot\text{кг}\cdot\text{с}^{-2}$
Мощность, поток энергии	$\mathrm{L^2MT^{-3}}$	ватт	Вт	$1\,\text{Вт} = 1\,\text{Дж}/\text{с} = 1\,\text{м}^2\cdot\text{кг}\cdot\text{с}^{-3}$
Температура (термодинамическая)	$\boldsymbol{\Theta}$	**кельвин**	**К**	**Основная единица**

Величина		Единица		
наименование	размерность	наименование	обозначение	связь с основными единицами СИ
Температурный коэффициент	Θ^{-1}	кельвин в минус первой степени	K^{-1}	
Температурный градиент	$\text{L}^{-1}\Theta$	кельвин на метр	К/м	
Количество вещества	**N**	**моль**	**моль**	**Основная единица**
Молярная масса	MN^{-1}	килограмм на моль	кг/моль	
Молярный объем	L^3N^{-1}	кубический метр на моль	м^3/моль	
Количество теплоты (теплота)	L^2MT^{-2}	джоуль	Дж	$1\ \text{Дж} = 1\ \text{Н}\cdot\text{м} = 1\ \text{м}^2\cdot\text{кг}\cdot\text{с}^{-2}$
Удельная теплота	L^2T^{-2}	джоуль на килограмм	Дж/кг	$1\text{Дж/кг} = 1\text{м}^2\cdot\text{с}^{-2}$
Молярная теплота	$\text{L}^2\text{MT}^{-2}\text{N}^{-1}$	джоуль на моль	Дж/моль	$1\ \text{Дж/моль} = 1\ \text{м}^2\cdot\text{кг}\cdot\text{с}^{-2}\cdot\text{моль}^{-1}$
Теплоемкость, энтропия	$\text{L}^2\text{MT}^{-2}\Theta^{-1}$	джоуль на кельвин	Дж/К	$1\ \text{Дж/К} = 1\ \text{м}^2\cdot\text{кг}\cdot\text{с}^{-2}\cdot\text{К}^{-1}$
Удельная теплоемкость, удельная энтропия	$\text{L}^2\text{T}^{-2}\Theta^{-1}$	джоуль на килограмм-кельвин	Дж/(кг · К)	$1\ \text{Дж/(кг}\cdot\text{К)} = 1\cdot\text{м}^2\cdot\text{с}^{-2}\cdot\text{К}^{-1}$
Молярная теплоемкость, молярная энтропия	$\text{L}^2\text{MT}^{-2}\Theta^{-1}\text{N}^{-1}$	джоуль на моль-кельвин	Дж/(моль·К)	$1\ \text{Дж/(моль}\cdot\text{К)} = 1\ \text{м}^2\cdot\text{кг}\cdot\text{с}^{-2}\cdot\text{К}^{-1}\cdot\text{моль}^{-1}$
Тепловой поток	L^2MT^{-3}	ватт	Вт	$1\ \text{Вт} = 1\ \text{Дж/с} = 1\ \text{м}^2\cdot\text{кг}\cdot\text{с}^{-3}$
Плотность теплового потока	MT^{-3}	ватт на квадратный метр	Вт/м^2	$1\ \text{Вт/м}^2 = \text{кг}\cdot\text{с}^{-3}$
Теплопроводность	$\text{LMT}^{-3}\Theta^{-1}$	ватт на метр-кельвин	Вт/(м · К)	$1\ \text{Вт/(м}\cdot\text{К)} = 1\ \text{м}\cdot\text{кг}\cdot\text{с}^{-3}\cdot\text{К}^{-1}$
Коэффициент теплопередачи	$\text{MT}^{-3}\Theta^{-1}$	ватт на квадратный метр-кельвин	$\text{Вт/(м}^2\cdot\text{К)}$	$1\ \text{Вт/(м}^2\cdot\text{К)} = 1\ \text{кг}\cdot\text{с}^{-3}\cdot\text{К}^{-1}$

Величина		Единица		
наименование	размерность	наименование	обозначение	связь с основными единицами СИ
Концентрация (плотность числа частиц)	L^{-3}	метр в минус третьей степени	м^{-3}	
Молярная концентрация	$\mathrm{L}^{-3}\mathrm{N}$	моль на кубический метр	моль/м^3	
Коэффициент диффузии	$\mathrm{L}^2\mathrm{T}^{-1}$	квадратный метр на секунду	$\text{м}^2/\text{с}$	
Сила электрического тока	**I**	**ампер**	**А**	**Основная единица**
Плотность электрического тока	$\mathrm{L}^{-2}\mathrm{I}$	ампер на квадратный метр	А/м^2	
Количество электричества (электрический заряд)	TI	кулон	Кл	$1\ \text{Кл} = 1\ \text{с} \cdot \text{А}$
Поверхностная плотность электрического заряда	$\mathrm{L}^{-2}\mathrm{TI}$	кулон на квадратный метр	Кл/м^2	$1\ \text{Кл}/\text{м}^2 = 1\ \text{м}^{-2}\text{с} \cdot \text{А}$
Пространственная плотность электрического заряда	$\mathrm{L}^{-3}\mathrm{TI}$	кулон на кубический метр	Кл/м^3	$1\ \text{Кл}/\text{м}^3 = 1\ \text{м}^{-3} \cdot \text{с} \cdot \text{А}$
Электрическое напряжение, электрический потенциал, разность электрических потенциалов, электродвижущая сила	$\mathrm{L}^2\mathrm{MT}^{-3}\mathrm{I}^{-1}$	вольт	В	$1\ \text{В} = 1\ \text{Вт}/\text{А} = 1\ \text{м}^2 \cdot \text{кг} \cdot \text{с}^{-3} \cdot \text{А}^{-1}$
Напряженность электрического поля	$\mathrm{LMT}^{-3}\mathrm{I}^{-1}$	вольт на метр	В/м	$1\ \text{В}/\text{м} = 1\ \text{Вт}/(\text{А} \cdot \text{м}) = 1\ \text{м} \cdot \text{кг} \cdot \text{с}^{-3} \cdot \text{А}^{-1}$
Электрическое сопротивление	$\mathrm{L}^2\mathrm{MT}^{-3}\mathrm{I}^{-2}$	ом	Ом	$1\ \text{Ом} = 1\ \text{В}/\text{А} = 1\ \text{м}^2 \cdot \text{кг} \cdot \text{с}^{-3} \cdot \text{А}^{-2}$
Удельное электрическое сопротивление	$\mathrm{L}^3\mathrm{MT}^{-3}\mathrm{I}^{-2}$	ом · метр	Ом · м	$1\ \text{Ом} \cdot \text{м} = 1\ \text{м}^3 \cdot \text{кг} \cdot c^{-3} \cdot \text{А}^{-2}$

Величина		Единица		
наименование	размерность	наименование	обозначение	связь с основными единицами СИ
Электрическая проводимость	$L^{-2}M^{-1}T^{3}I^{2}$	сименс	См	$1\ См = 1\ Ом^{-1} = 1\ м^{-2} \cdot кг^{-1} \cdot с^{3} \cdot А^{2}$
Удельная электрическая проводимость	$L^{-3}M^{-1}T^{3}I^{2}$	сименс на метр	См/м	$1\ См/м = 1\ Ом^{-1} \cdot м^{-1} = 1\ м^{-3} \cdot кг^{-1} \cdot с^{3} \cdot А^{2}$
Электрическая емкость	$L^{-2}M^{-1}T^{4}I^{2}$	фарад	Ф	$1\ Ф = 1\ Кл/В = 1\ м^{-2} \cdot кг^{-1} \cdot с^{4} \cdot А^{2}$
Электрическая постоянная, абсолютная диэлектрическая проницаемость	$L^{-3}M^{-1}T^{4}I^{2}$	фарад на метр	Ф/м	$1\ Ф/м = 1\ м^{-3} \cdot кг^{-1} \cdot с^{4} \cdot А^{2}$
Поток электрического смещения	TI	кулон	Кл	$1\ Кл = 1\ с \cdot А$
Электрическое смещение	$L^{-2}TI$	кулон на квадратный метр	$Кл/м^{2}$	$1\ Кл/м^{2} = 1\ м^{-2} \cdot с \cdot А$
Магнитный поток (поток магнитной индукции)	$L^{2}MT^{-2}I^{-1}$	вебер	Вб	$1\ Вб = 1\ В \cdot с = 1\ Тл \cdot м^{2} = 1\ м^{2} \cdot кг \cdot с^{-2} \cdot А^{-1}$
Магнитная индукция (плотность магнитного потока)	$MT^{-2}I^{-1}$	тесла	Тл	$1\ Тл = 1\ В \cdot с/м^{2} = 1\ Вб/м^{2} = 1\ кг \cdot с^{-2} \cdot А^{-1}$
Индуктивность	$L^{2}MT^{-2}I^{-2}$	генри	Гн	$1\ Гн = 1\ м^{2} \cdot кг \cdot с^{-2} \cdot А^{-2}$
Магнитная постоянная, абсолютная магнитная проницаемость	$LMT^{-2}I^{-2}$	генри на метр	Гн/м	$1\ Гн/м = 1\ м \cdot кг \cdot с^{-2} \cdot А^{-2}$
Напряженность магнитного поля	$L^{-1}I$	ампер на метр	А/м	
Энергия излучения	$L^{2}MT^{-2}$	джоуль	Дж	$1\ Дж = 1\ м^{2} \cdot кг \cdot с^{-2}$
Мощность излучения (поток излучения)	$L^{2}MT^{-3}$	ватт	Вт	$1\ Вт = 1\ Дж\ /\ с = 1\ м^{2} \cdot кг \cdot с^{-3}$

Величина		Единица		
наименование	размерность	наименование	обозначение	связь с основными единицами СИ
Интенсивность излучения (плотность потока излучения)	MT^{-3}	ватт на квадратный метр	$\text{Вт}/\text{м}^2$	$1\,\text{Вт}/\text{м}^2 = 1\,\text{кг}\cdot\text{с}^{-3}$
Поток частиц	T^{-1}	секунда в минус первой степени	с^{-1}	
Плотность потока частиц	$\mathrm{L}^{-2}\mathrm{T}^{-1}$	секунда в минус первой степени на метр в минус второй степени	$\text{с}^{-1}\cdot\text{м}^{-2}$	
Сила света	**J**	**кандела**	**кд**	**Основная единица**
Световой поток	J	люмен	лм	$1\,\text{лм} = 1\,\text{кд}\cdot\text{ср}$
Световая энергия	TJ	люмен-секунда	$\text{лм}\cdot\text{с}$	$1\,\text{лм}\cdot\text{с} = 1\,\text{с}\cdot\text{кд}\cdot\text{ср}$
Светимость	$\mathrm{L}^{-2}\mathrm{J}$	люмен на квадратный метр	$\text{лм}/\text{м}^2$	$1\,\text{лм}/\text{м}^2 = 1\,\text{м}^{-2}\cdot\text{кд}\cdot\text{ср}$
Освещенность	$\mathrm{L}^{-2}\mathrm{J}$	люкс	лк	$1\,\text{лк} = 1\,\text{лм}/\text{м}^2 = 1\,\text{м}^{-2}\cdot\text{кд}\cdot\text{ср}$
Яркость	$\mathrm{L}^{-2}\mathrm{J}$	кандела на квадратный метр	$\text{кд}/\text{м}^2$	
Оптическая сила	L^{-1}	$^\circ$диоптрия	дптр	$1\,\text{дптр} = 1\,\text{м}^{-1}$
Энергетическая сила (сила излучения)	$\mathrm{L}^2\mathrm{MT}^{-3}$	ватт на стерадиан	Вт/ср	$1\,\text{Вт}/\text{ср} = 1\,\text{м}^2\cdot\text{кг}\cdot\text{с}^{-3}\cdot\text{ср}^{-1}$
Энергетическая светимость (излучательность)	MT^{-3}	ватт на квадратный метр	$\text{Вт}/\text{м}^2$	$1\,\text{Вт}/\text{м}^2 = 1\,\text{кг}\cdot\text{с}^{-3}$
Энергетическая освещенность (облученность)	MT^{-3}	ватт на квадратный метр	$\text{Вт}/\text{м}^2$	$1\,\text{Вт}/\text{м}^2 = 1\,\text{кг}\cdot\text{с}^{-3}$
Энергетическая яркость (лучистость)	MT^{-3}	ватт на стерадиан-квадратный метр	$\text{Вт}/(\text{ср}\cdot\text{м}^2)$	$1\,\text{Вт}/(\text{ср}\cdot\text{м}^2) = 1\cdot\text{кг}\cdot\text{с}^{-3}\cdot\text{ср}^{-1}$

Учебное издание

СИВУХИН Дмитрий Васильевич

ОБЩИЙ КУРС ФИЗИКИ

Том II

ТЕРМОДИНАМИКА И МОЛЕКУЛЯРНАЯ ФИЗИКА

Редактор *Д.А. Миртова*
Оригинал-макет: *Д.В. Горбачев*
Оформление переплета: *А.А. Логунов*

ЛР №071930 от 06.07.99. Подписано в печать 12.05.05.
Формат 60×90/16. Бумага офсетная № 1. Печать офсетная.
Усл. печ. л. 34. Уч.-изд. л. 39,9. Заказ №

Издательская фирма «Физико-математическая литература»
МАИК «Наука/Интерпериодика»
117997, Москва, ул. Профсоюзная, 90
E-mail: fizmat@maik.ru, fmlsale@maik.ru;
http://www.fml.ru

Отпечатано с готовых диапозитивов
в ОАО «Чебоксарская типография № 1»
428019, г. Чебоксары, пр. И. Яковлева, 15

ISBN 5-9221-0601-5

ОБЩИЙ КУРС ФИЗИКИ

Д.В. Сивухин

ЭЛЕКТРИЧЕСТВО

ТОМ III

УДК 530.1(075.8)
ББК 22.2
С 34

Сивухин Д. В. **Общий курс физики.** Учеб. пособие: Для вузов. В 5 т. Т. 3. **Электричество.** — 5-е изд., стер. — М.: ФИЗМАТЛИТ, 2009. — 656 с. — ISBN 978-5-9221-0673-3.

Третий том курса физики, широко известного у нас в стране и за рубежом. Книга написана на основе лекций, которые в течение ряда лет читались автором студентам Московского физико-технического института. Основное внимание уделено выяснению физического смысла и содержания основных законов и понятий науки об электричестве, установлению границ применимости этих законов, развитию у студентов навыков физического мышления и умения ставить и решать конкретные задачи.

Третье издание третьего тома вышло в 1996 г. в двух частях.

Для студентов физических и математических факультетов университетов, физико-технических и инженерно-физических институтов, а также вузов, где физика является основной дисциплиной.

ISBN 978-5-9221-0673-3
© ФИЗМАТЛИТ, 2004, 2006, 2009

ОГЛАВЛЕНИЕ

Предисловие к первому изданию 7

Введение

1. Действие на расстоянии и полевое взаимодействие 9
2. Электрический заряд и напряженность электрического поля . . . 14

Глава I. **Электрическое поле**

3. Закон Кулона. Принцип суперпозиции электростатических полей 17
4. Электрический диполь 22
5. Поток вектора и электростатическая теорема Гаусса 27
6. Применения теоремы Гаусса 31
7. Дифференциальная форма электростатической теоремы Гаусса . 38
8. Математическое дополнение. Формула Гаусса–Остроградского . . 41
9. Теорема Ирншоу 44
10. Электрическое поле в веществе 46
11. Проводники в электрическом поле 50
12. Поляризация диэлектриков 56
13. Теорема Гаусса для диэлектриков 60
14. Граничные условия 61
15. Поляризуемость и диэлектрическая проницаемость 63
16. Поле равномерно поляризованного шара 66
17. Потенциальность электростатического поля 69
18. Электрический потенциал 71
19. Вычисление потенциала по напряженности поля 75
20. Измерение разности потенциалов электрометром. Электрический зонд 78
21. Электрическое поле Земли 82
22. Общая задача математической электростатики 84
23. Метод электрических изображений 86
24. Точечный заряд над плоской поверхностью диэлектрика 91
25. Электрическое поле заряженного проводящего эллипсоида 92
26. Емкость проводников и конденсаторов 97
27. Потенциальные и емкостные коэффициенты 104
28. Электрическая энергия 108
29. Локализация электрической энергии в пространстве 111
30. Взаимная энергия точечных зарядов 117
31. Термодинамика диэлектриков 119
32. Свободная энергия и силы 123
33. Максвелловские натяжения и давления 129
34. Вычисление пондеромоторных сил в общем виде 135
35. Электронная теория поляризации неполярных диэлектриков . . . 140
36. Электронная теория поляризации полярных газообразных диэлектриков 146
37. Пьезоэлектричество 151
38. Пироэлектричество 159
39. Сегнетоэлектричество 161

Глава II. Электрический ток

40. Плотность тока. Закон сохранения электрического заряда 173
41. Закон Ома 175
42. Вывод законов Ома и Джоуля–Ленца 178
43. Сторонние силы. Концентрационный элемент 189
44. Законы Ома и Джоуля–Ленца в интегральной форме 192
45. Правила Кирхгофа 197
46. Стационарные токи в массивных проводниках 200
47. Электролитическая ванна 203
48. Процессы установления тока при зарядке и разрядке конденсатора 206

Глава III. Магнитное поле

49. Силы, действующие в магнитном поле на движущиеся заряды и токи 209
50. Магнитное поле равномерно движущегося заряда. Закон Био и Савара 212
51. Расчет магнитных полей с помощью закона Био и Савара. Системы единиц 216
52. Момент сил, действующих на виток с током в магнитном поле . . 220
53. Теорема Гаусса для магнитных полей 223
54. Дополнительные сведения о телесных углах 224
55. Теорема о циркуляции магнитного поля в вакууме 227
56. Дифференциальная форма теоремы о циркуляции 232
57. Эквивалентность магнитных полей тока и магнитного листка . . . 234
58. Магнитное поле в веществе 242
59. Теорема о циркуляции магнитного поля в веществе 246
60. Граничные условия для векторов **B** и **H** 247
61. Магнитная восприимчивость и магнитная проницаемость 249
62. Работа при перемещении витка с током в постоянном магнитном поле 254
63. Способ Гаусса измерения магнитных полей 255
64. Электромагнитная индукция 256
65. Правило Ленца 260
66. Максвелловская трактовка явления электромагнитной индукции 263
67. Флюксметр и пояс Роговского 267
68. Индуктивность проводов. Явления при замыкании и размыкании тока 271
69. Магнитная энергия токов 276
70. Локализация магнитной энергии в пространстве 279
71. Теорема о сохранении магнитного потока 283
72. Энергия и силы 284
73. Термодинамика магнетиков 289
74. Ферромагнетизм 291
75. Магнитные свойства атомов 296
76. Объяснение диамагнетизма 298
77. Объяснение парамагнетизма 301
78. Гиромагнитные явления 308
79. Формальная теория ферромагнетизма Вейсса 310
80. Сверхпроводники и их магнитные свойства 317

Глава IV. **Уравнения Максвелла**
81. Ток смещения 331
82. Система уравнений Максвелла 336
83. Скорость распространения электромагнитных возмущений 340
84. Энергия и поток энергии 346
85. Международная система единиц (СИ) 354

Глава V. **Движение заряженных частиц в электромагнитных полях**
86. Движение в постоянных и однородных полях 364
87. Дрейф заряженной частицы в неоднородном магнитном поле при наличии слабого электрического поля 368
88. Адиабатический инвариант 375
89. Определение удельного заряда частицы 378
90. Измерение элементарного заряда методом масляных капель 382
91. Электромагнитная масса 386

Глава VI. **Электролиты**
92. Электролиз и электролитическая диссоциация 391
93. Примеры электролиза 395
94. Законы электролиза Фарадея и элементарный заряд 397
95. Скорости ионов и электрическая проводимость электролитов . . . 398
96. Гальванические элементы и аккумуляторы 403

Глава VII. **Электрические токи в металлах, полупроводниках и вакууме**
97. Инерция электронов в металлах 411
98. Явление Холла 414
99. Применения статистики Ферми–Дирака к электронам в металлах 418
100. Металлы и полупроводники 426
101. Термоэлектронная эмиссия 434
102. Электронные лампы и их применения 442
103. Вторичная и автоэлектронная эмиссия 447

Глава VIII. **Электрические явления в контактах**
104. Контактная разность потенциалов 450
105. Термоэлектрический ток 454
106. Явление Пельтье 460
107. Термодинамика термоэлектрических явлений. Явление Томсона . 463
108. Выпрямляющее действие контактов полупроводников 466

Глава IX. **Электрические токи в газах**
109. Ионизация и рекомбинация 470
110. Измерение потенциала ионизации методом электронного удара . 473
111. Измерение слабых токов 476
112. Несамостоятельная проводимость газов 477
113. Измерение коэффициентов рекомбинации 480
114. Измерение подвижностей ионов 484
115. Теория Таунсенда 486
116. Закон Пашена 492

117. Тлеющий разряд 494
118. Искровой разряд 499
119. Коронный разряд 501
120. Дуговой разряд 502
121. Плазма 505

Глава X. **Колебания и волны**
122. Уравнение колебательного контура 513
123. Свободные колебания гармонического осциллятора 518
124. Затухающие колебания 521
125. Баллистический гальванометр 524
126. Векторная диаграмма и комплексные обозначения 525
127. Вынужденные колебания затухающего осциллятора под действием синусоидальной силы 529
128. Вынужденные колебания под действием несинусоидальной силы. Теорема Фурье 536
129. Закон Ома для переменных токов (синусоидально меняющихся во времени) 540
130. Правила Кирхгофа для переменных токов 545
131. Эффективные напряжение и ток 548
132. Процессы установления колебаний 549
133. Автоколебания. Ламповый генератор 560
134. Релаксационные колебания 566
135. Параметрическое возбуждение колебаний 568
136. Трансформатор 572
137. Колебания с двумя степенями свободы 577
138. Волновое уравнение 581
139. Плоские электромагнитные волны 585
140. Стоячие волны 588
141. Поле излучения диполя Герца 592
142. Демонстрационные опыты с электромагнитными волнами 600
143. Волны вдоль проводов 606
144. Свойства быстропеременных токов. Скин-эффект 612
145. Давление и импульс электромагнитных волн 617
146. Принципы радиосвязи 621
147. Распространение электромагнитных волн в волноводах 629

Приложения
Важнейшие формулы электродинамики в системе СИ 640
Фундаментальные физические константы 645
Именной указатель 646
Предметный указатель 648

Предисловие к первому изданию

Учение об электричестве включает три группы вопросов. К первой группе относятся основные понятия и общие принципы, управляющие электрическими и магнитными явлениями; ко второй — электрические и магнитные свойства вещества; к третьей — технические и практические применения электричества.

В настоящем курсе основное внимание уделяется вопросам первой группы, которые излагаются с наибольшей полнотой. Принят индуктивный метод изложения. Основные понятия и принципы устанавливаются путем обобщения опытных фактов, имеющих ограниченную область применимости. Процесс обобщения идет непрерывно и целенаправленно на протяжении всего изложения, завершаясь в середине курса установлением системы уравнений Максвелла. Все последующее изложение строится с учетом этих уравнений.

Электрические и магнитные свойства вещества в общем курсе физики, естественно, не могут быть рассмотрены с той же степенью подробности. Их полное понимание возможно только на основе квантовой механики, а подробное изложение должно производиться в специальных курсах. Несмотря на это, я стремился уделить большое внимание и этой группе вопросов, насколько это возможно сделать в рамках общего курса физики, предназначенного для студентов младших курсов. При этом я широко пользовался методами термодинамики. Без применения термодинамики невозможно полное и ясное изложение не только частных, но и общих вопросов макроскопической электродинамики.

Что касается технических и прочих применений учения об электричестве, то этим вопросам уделено значительно меньше внимания, чем они того заслуживают. Из таких вопросов рассмотрены лишь некоторые, и притом только с принципиальной стороны.

В качестве основной в курсе принята гауссова система единиц. Мотивы, заставившие меня отказаться от Международной системы единиц (СИ), уже приводились в предисловии к первому тому. Более подробно они излагаются в § 85 этого тома. Однако нельзя не считаться с тем, что изложение физики в средней школе и в большинстве технических вузов в настоящее время ведется в системе СИ. Многие книги и журнальные статьи, в особенности прикладного характера, написаны в той же системе. Поэтому в § 85 предлагаемой книги дается подробное изложение принципов построения системы СИ в электродинамике и ее связи с гауссовой системой. Доказывается общее правило, сформулированное в виде табл. 1, позволяющее любую формулу гауссовой системы переписать в системе СИ и наоборот. Кроме того, в Приложении к книге основные формулы электродинамики написаны в системе СИ. Формулу, записанную в тексте книги в гауссовой системе, можно найти под тем же номером в Приложении. Если же соответствующей формулы в Приложении нет, то ее легко получить, пользуясь табл. 1. В табл. 2 приводятся соотношения между единицами СИ и гауссовой системы.

Как и первые два тома, предлагаемый третий том является расширенным курсом лекций, которые я читал на протяжении многих лет (начиная с 1957 г.) студентам второго курса Московского физико-технического института. Поэтому все сказанное в предисловиях к этим двум томам относительно целей и характера изложения относится и к третьему тому.

Многие задачи, включенные в этот том, предлагались студентам Московского физико-технического института на письменных экзаменах и семинарских занятиях. Идеи значительной части этих задач принадлежат преподавателям физики указанного института.

Демонстрации опытов по электричеству, которыми сопровождалось чтение курса, принесли много пользы не только слушателям, но и лектору, читавшему курс. Эти демонстрации готовили и осуществляли лекционные ассистенты: М.И. Маклаков, В.А. Кузнецова, Е.Н. Морозов, В.П. Молчанов. Выражаю им и всем преподавателям, участвовавшим в составлении задач, глубокую благодарность.

Мне приятно также принести глубокую благодарность профессору И.С. Горбаню и сотрудникам руководимой им кафедры экспериментальной физики Киевского государственного университета им. Т.Г. Шевченко, а также профессору С.С. Герштейну за внимательное рецензирование рукописи. Их советы, полные доброжелательности, немало способствовали улучшению предлагаемой книги.

ВВЕДЕНИЕ

§ 1. Действие на расстоянии и полевое взаимодействие

1. Опыт показывает, что между электрически заряженными и намагниченными телами, а также телами, по которым текут электрические токи, действуют силы, называемые *электродинамическими* или *электромагнитными*. Отвлекаясь от деталей, можно сказать, что относительно природы этих сил в науке выдвигались две противоположные точки зрения. Ранее известная из них исходила из представления о *непосредственном действии тел на расстоянии*, без участия каких бы то ни было промежуточных материальных посредников. Более новая точка зрения, принятая в настоящее время, исходит из представления, что взаимодействия передаются с помощью особого материального посредника, называемого *электромагнитным полем*.

2. Теория действия на расстоянии в учении об электрических и магнитных явлениях господствовала примерно до последней четверти XIX века. Основная идея этой теории была заимствована из учения о всемирном тяготении. Громадные успехи небесной механики, основанной на законе всемирного тяготения великого Исаака Ньютона (1643–1727), с одной стороны, и полная неудача как-то объяснить тяготение — с другой, привели многих ученых к представлению, что тяготение, а также электрические и магнитные силы не нуждаются в объяснении, а являются неотъемлемыми, врожденными свойствами материи. По мнению этих ученых, задача теории электричества состояла в том, чтобы установить *элементарные законы* электрических и магнитных сил и на их основе объяснить все электрические и магнитные явления. Под элементарными законами понимали законы, определяющие силы взаимодействия на расстоянии между точечными *электрическими зарядами*, точечными *магнитными полюсами* и *элементами тока*, т. е. между бесконечно короткими участками бесконечно тонких проводов, по которым текут электрические токи. По своему содержанию и форме эти законы напоминали, а часто прямо копировали ньютонов закон всемирного тяготения. Таковы были, например, *законы Кулона*, названные в честь французского физика Ш. Кулона (1736–1806), о взаимодействии электрических зарядов или магнитных полюсов.

В математическом отношении теория действия на расстоянии достигла высокой степени совершенства благодаря трудам Пьера С. Лапласа (1749–1827), Андре М. Ампера (1775–1836), Симеона Д. Пуассона

(1781–1840), Карла Ф. Гаусса (1777–1855), М.В. Остроградского (1801–1862), Грина (1793–1841), Франца Неймана (1798–1895), Карла Неймана (1832–1925), Вильгельма Вебера (1804–1891), Густава Р. Кирхгофа (1824–1887) и многих других математиков и физиков. Теория отличалась формальной простотой и ясностью исходных математических положений, математической строгостью, стройностью и конкретностью. Она совершенно не вводила сомнительных гипотетических представлений относительно физической природы электрических и магнитных сил, а основывалась только на эмпирически прочно установленных фактах и их обобщениях. Количественные выводы теории были прочно обоснованными и достоверными (разумеется, в пределах той области, в которой элементарные законы подтверждены опытом). Не удивительно, что теории действия на расстоянии придерживалось большинство физиков вплоть до последней четверти XIX века. Однако количественное согласие теории с опытом в исследованной ограниченной области явлений не может считаться достаточным доказательством правильности концепции непосредственного действия на расстоянии.

3. Среди физиков XIX века, для которых концепция непосредственного действия на расстоянии была неприемлема, возвышается почти одиноко фигура гениального Майкла Фарадея (1791–1867) — основоположника физической теории электромагнитного поля. Над ним не довлели формальные идеи математиков. Его самобытный ум был свободен от укоренившихся представлений и не мог примириться с мыслью, что тело может производить непосредственное действие в тех местах, в которых оно не находится и которые отделены от него абсолютно пустым пространством. Согласно Фарадею, действие одного тела на другое может либо осуществляться непосредственным соприкосновением, либо передаваться через промежуточную среду.

В случае электромагнитных взаимодействий роль такой среды играл гипотетический мировой эфир, заполняющий все пространство между телами и мельчайшими частицами, из которых они состоят. При электризации и намагничивании тел в окружающем эфире возникают, согласно Фарадею, какие-то изменения, напоминающие упругие деформации и связанные с ними натяжения и давления. Такими натяжениями и давлениями Фарадей и объяснял электромагнитные взаимодействия тел. Центр тяжести с изучения зарядов и токов, являвшихся в теории действия на расстоянии центрами сил, переносился на изучение окружающего пространства. Это пространство с действующими в нем силами называется *электромагнитным полем.*

Яркую характеристику воззрений Фарадея дал Генрих Герц (1857–1894). В докладе, прочитанном в Гейдельберге в 1889 г., он говорил:

«Фарадею говорили, что при электризации тела в него что-то вносят, но он видел, что возникающие изменения обнаруживаются лишь вне тела, а отнюдь не внутри. Фарадея учили, что силы просто перескакивают через пространство, но он видел, какое большое влияние оказывает на эти силы то вещество, которым заполнено это якобы перескакиваемое пространство. Фарадей считал, что электричество существует наверное, но что о его силах спорят. Он видел, однако, насколько осязательно выступают в своих

действиях эти силы, в то время как самого электричества он никак не мог обнаружить. И тогда все обернулось в его представлении. Электрические и магнитные силы стали для него существующими, действительными, осязаемыми, а электричество, магнетизм сделались вещами, о существовании которых можно спорить. Силовые линии, как он называл силы, мыслимые самостоятельно, стояли перед его умственным взором в пространстве как состояния последнего, как напряжения, как вихри, как течения, как многое другое, что и сам он не мог определить, но они стояли там, действуя друг на друга, сдвигая и толкая тела туда и сюда, распространяясь и сообщая друг через друга возбуждение от точки к точке».

4. Применяя изложенные воззрения к конкретным случаям, Фарадей довольствовался преимущественно качественной стороной явлений. Он никогда не пользовался точным языком математических формул. Рассуждения и доказательства Фарадея воспринимались с трудом и даже отвергались его современниками. Зато среди убежденных приверженцев Фарадея был гениальный Джеймс К. Максвелл (1831–1879), в совершенстве владевший математическими методами своего времени. Максвелл облек основные идеи Фарадея в математическую форму. Он обобщил имеющиеся опытные факты и пополнил их новыми. Таким путем в начале 60-х годов XIX века ему удалось сформулировать систему уравнений, в которой в сжатой и точной форме содержатся все количественные законы электромагнитного поля. Установление этих уравнений, пожалуй, является наиболее крупным открытием физики XIX века.

5. Сначала теория Максвелла не получила признания. Это связано главным образом с тем, что вплоть до последней четверти XIX века электродинамика занималась изучением только *постоянных* или *почти постоянных* электрических и магнитных полей. А в этих случаях уравнения Максвелла переходят в уравнения теории действия на расстоянии. Здесь фактические выводы обеих теорий совпадают. Поэтому никакие опыты с постоянными электромагнитными полями не могут ответить на вопрос, какое из двух представлений о силах взаимодействия верно или, точнее, заведомо неверно. Для этого надо было обратиться к изучению *переменных полей.* Максвелл показал, что из его уравнений следует существование *электромагнитных волн*, и вычислил скорость их распространения. Оказалось, что в вакууме эта скорость совпадает со скоростью света (300 000 км/с), т. е. очень велика. Громадный круг явлений воспринимается так, как если бы скорость распространения электромагнитных возмущений была *бесконечна*, т. е. так, как если бы была справедлива теория действия на расстоянии. Электромагнитные волны впервые были получены и экспериментально исследованы в знаменитых опытах Герца в 1887–1888 гг. Их свойства оказались в точности такими, какие предсказывала теория Максвелла. С точки зрения теории действия на расстоянии существование электромагнитных волн абсолютно непонятно. Поэтому после опытов Герца вопрос о характере электродинамических взаимодействий был однозначно решен в пользу теории поля. Громадную роль в деле распространения и развития теории Максвелла сыграло великое изобретение

радио А.С. Поповым (1859–1905), которое в конце концов преобразило науку, технику и саму жизнь человека.

6. Физики XIX века были убеждены, что явления электричества и магнетизма могут быть поняты до конца только тогда, когда они будут сведены к механическим причинам, например к упругим натяжениям, давлениям или каким-то другим механическим изменениям в окружающей среде. Такой средой в теории Фарадея–Максвелла считался *мировой эфир*. Было затрачено много усилий для построения механической теории электрических и магнитных явлений. Сам Максвелл положил этому начало. В первых работах по теории электричества он широко пользовался механическими моделями для представления электромагнитного поля. Однако для представления различных свойств поля потребовались разные модели, противоречащие друг другу. Механические же модели в теории Максвелла сыграли лишь роль лесов строящегося здания. После того как здание построено, леса убираются. Так и в завершенном варианте теории Максвелла, опубликованном им в «Трактате по электричеству и магнетизму» (1873), механические модели совсем не используются. Все усилия по построению непротиворечивой механической теории электрических и магнитных явлений потерпели неудачу. Они убедили физиков последующих поколений в *принципиальной невозможности механической картины мира*. Атомно-молекулярная теория показала, что упругие силы сами являются результатом электрического взаимодействия между электрически заряженными частицами, из которых построены тела. Упругость была сведена к электричеству. После этого программа сведения электрических сил к упругим потеряла всякий смысл. Электрические силы оказались более «простыми» и «понятными», чем силы упругие. Современная физика не связывает с понятием электромагнитного поля никаких «наглядных» картин типа упругих деформаций, натяжений, давлений и пр. Она утверждает лишь, что поле *реально существует* и в этом смысле, наряду с веществом, является *одним из видов материи*. Поле обладает энергией, импульсом и другими физическими свойствами. Посредством полей осуществляются электромагнитные взаимодействия тел. Заряженное тело A возбуждает в окружающем пространстве электрическое поле. Оно проявляется в силе, действующей на другое заряженное тело B, вносимое в это поле. Но поле, возбуждаемое зарядами тела A, реально существует в каждой точке пространства, *даже если в нее не помещено никакое другое тело* B. В этом отличие точек зрения теории поля и теории непосредственного действия на расстоянии. Последняя также пользуется понятием поля. Однако в ней поле выступает не как физическая реальность, а как *вспомогательное математическое понятие*, вводимое лишь для удобства описания электромагнитных взаимодействий. По теории действия на расстоянии не имеет смысла говорить о поле в той или иной точке пространства, пока в него не внесено заряженное тело, на которое действует электромагнитная сила.

7. Первоначальная теория Максвелла не вводила принципиального различия между материальными средами и вакуумом (эфиром). Вакуум рассматривался в ней как *одна из сред*, отличающаяся от других сред только количественно: значениями диэлектрической и магнитной проницаемостей и электрической проводимости. Более глубокую и ясную картину дала *электронная теория*, творцом которой был великий голландский физик Х.А. Лоренц (1853–1928). Она была создана и детально разработана еще до открытия электрона и установления структуры атома. [Электрон был открыт Дж. Дж. Томсоном (1856–1940) в 1897 г., модель атома Эрнеста Резерфорда (1871–1937) появилась в 1911 г., а теория Нильса Бора (1885–1962) — в 1913 г.] На современном языке основную идею электронной теории можно сформулировать следующим образом. Вещество состоит из положительно заряженных атомных ядер и отрицательно заряженных электронов. Для наших целей пока нет необходимости вдаваться в детали строения атомов и атомных ядер. Важно заметить лишь, что вакуум является *универсальной средой*, в которой возбуждается электромагнитное поле. С точки зрения теории электричества всякое вещество следует рассматривать как вакуум, испорченный вкрапленными в него атомными ядрами и электронами. Заряды этих частиц возбуждают электромагнитные поля, накладывающиеся на внешнее поле, в которое внесено вещество. Наложением таких полей и определяется электромагнитное поле в веществе. С этой точки зрения изучение электромагнитного поля в веществе сводится к изучению поля в вакууме. Так мы и поступим в дальнейшем. Сначала изучим электрическое и магнитное поля в вакууме, а затем исследуем, как поле искажается зарядами атомных ядер и электронов вещества. Таким путем электронная теория привела к более глубокому пониманию уравнений Максвелла в веществе. Она явилась рациональной основой для понимания электрических и магнитных свойств вещества с атомистической точки зрения. Сам Лоренц и его последователи, разумеется, пользовались классическими представлениями. Это приводило к трудностям принципиального характера, преодоление которых стало возможным только после того, как электронная теория была перестроена на квантовой основе.

8. *Уравнения Максвелла являются обобщениями опытных фактов. Их доказательство надо искать в сопоставлении с опытом выводимых из них следствий.* Эти уравнения составляют стержень всей электродинамики. Они могут рассматриваться как *основные аксиомы электродинамики*, играющие в ней такую же роль, какую законы Ньютона играют в классической механике. Поэтому изложение электродинамики можно было бы строить *чисто дедуктивно* на основе постулативно установленных уравнений Максвелла. Однако для первоначального изучения такой путь нецелесообразен. Мы глубже проникнем в сущность электродинамики Максвелла, если изберем *индуктивный метод изложения.* Начнем с простейших опытных фактов и явлений, которые правильно описываются не только на языке теории поля, но и на языке

теории непосредственного действия на расстоянии. Постепенным обобщением законов этих явлений мы придем к таким результатам, которые уже не укладываются в рамки теории действия на расстоянии, а могут быть истолкованы только с помощью теории поля. В конце концов мы придем к системе уравнений Максвелла, после чего все последующее изложение будет строиться на их основе.

§ 2. Электрический заряд и напряженность электрического поля

1. Важнейшими понятиями в учении об электричестве являются *электрический заряд* и *напряженность электрического поля.* Качественное представление об этих понятиях дают простейшие опыты по электричеству, известные из элементарного курса физики. Здесь мы остановимся только на *количественной* стороне вопроса. Точное количественное определение заряда и напряженности электрического поля, как и всяких других физических величин, сводится к указанию *принципиального способа их измерения.*

2. Будем предполагать сначала, что электрическое поле не изменяется во времени. Такое поле называется *электростатическим.* Оно может быть возбуждено неподвижными электрическими зарядами. Малое электрически заряженное тельце, настолько малое, что оно практически не вызывает (вследствие индукции) перераспределения электрических зарядов на окружающих телах, может служить *пробным телом* или *пробным зарядом.* Возьмем два пробных заряда и будем последовательно помещать их в одну и ту же точку пространства, и притом так, чтобы оба заряда покоились в соответствующей системе отсчета. Так как поле не меняется во времени, то эти заряды будут подвергаться действию одного и того же поля. Пусть $\mathbf{F}_1$ и $\mathbf{F}_2$ — силы, действующие на эти неподвижные заряды. Обобщением опытных фактов является следующий результат. Силы $\mathbf{F}_1$ и $\mathbf{F}_2$ имеют либо одинаковые, либо прямо противоположные направления, а их отношение не зависит от положения точки, в которой помещаются пробные заряды. Отношение F_1/F_2 является поэтому характеристикой *самих только пробных зарядов*, а не поля, в которое они помещены. Это позволяет характеризовать состояние электризации пробного тела определенным числом q, называемым *электрическим зарядом.* По определению отношение зарядов q_1 и q_2 двух пробных тел равно отношению действующих на них сил F_1 и F_2 при последовательном помещении их в одну и ту же точку поля:

$$\frac{q_1}{q_2} = \frac{F_1}{F_2}. \tag{2.1}$$

При этом предполагается, что силы F_1 и F_2 (а с ними и сравниваемые заряды q_1 и q_2) имеют одинаковые знаки, если эти силы совпадают по направлению, и противоположные знаки, если их направления прямо противоположны.

То обстоятельство, что при введении этого определения заряды предполагались пробными, не лишает его общности. Действительно, каковы бы ни были размеры заряженных тел, всегда можно поместить их в электрическое поле столь удаленных зарядов, что по отношению к этому полю рассматриваемые тела могут считаться пробными.

Заряд какого-либо произвольно взятого тела можно условно принять за единицу. Тогда измерение отношения сил F_1/F_2 дает способ определения величины заряда в «абсолютной мере».

3. *Заряд тела не зависит от выбора (инерциальной) системы отсчета, в которой он измеряется.* Он *инвариантен* относительно перехода от одной инерциальной системы отсчета к другой. Это непосредственно следует из принципа относительности и принятого нами способа сравнения зарядов путем измерения сил, действующих на *неподвижные* заряды. Заряд одинаков во всех инерциальных системах отсчета по той же причине, по которой в них одинакова масса покоя одного и того же тела. Конечно, всякое измерение дает не абсолютное значение заряда, а только *отношение зарядов* рассматриваемых тел. Инвариантны не сами заряды, а *отношения зарядов.* Однако если условиться заряд какого-либо тела считать одинаковым во всех системах отсчета, то это будет справедливо и в отношении всех остальных зарядов. Именно такой смысл мы придаем утверждению об инвариантности заряда относительно выбора системы отсчета.

4. Соотношение (2.1) мы распространяем на случай любых переменных полей. Остается ли поле постоянным или меняется во времени — это не имеет значения. Важно только, чтобы в моменты измерения сил F_1 и F_2 поле, действующее на заряды q_1 и q_2, было одним и тем же и чтобы заряды q_1 и q_2 оставались неподвижными. Конечно, это утверждение также является обобщением опытных фактов.

Сила, действующая на единичный неподвижный пробный электрический заряд, называется напряженностью электрического поля и обозначается через $\mathbf{E}$. Если в соотношении (2.1) взять $q_1 = 1$, то будем иметь $F_1 = E$. Опуская далее индекс 2 и переходя к векторной форме записи, находим силу $\mathbf{F}$, действующую в электрическом поле $\mathbf{E}$ на неподвижный точечный заряд q:

$$\mathbf{F} = q\mathbf{E}. \tag{2.2}$$

Напряженность электрического поля $\mathbf{E}$ есть вектор, так как заряд q является скаляром, а сила $\mathbf{F}$ — вектором (см. т. I, § 7).

5. Указание на неподвижность заряда в определении напряженности электрического поля существенно. Дело в том, что силы, действующие на электрический заряд, зависят не только от электрического, но и от *магнитного поля.* Однако магнитное поле, как показывает опыт, действует только на *движущиеся* заряды и не действует на неподвижные. Используя при определении $\mathbf{E}$ неподвижные заряды, мы исключаем влияние магнитного поля.

Заряд, неподвижный в одной системе отсчета, уже не неподвижен в другой системе, движущейся относительно первой. Полная сила $\mathbf{F}$,

действующая на него (в нерелятивистском приближении), одна и та же в обеих системах отсчета. Но в первой системе, относительно которой заряд покоится, эта сила чисто электрическая, тогда как во второй системе, относительно которой заряд движется, она складывается из электрической и магнитной сил. Если измерить напряженность электрического поля во второй системе с помощью неподвижного в ней пробного заряда, то она окажется иной, чем в первой системе отсчета. Значит, *разделение поля на электрическое и магнитное зависит от выбора системы отсчета.* Электрическое и магнитное поля неразрывно связаны друг с другом. Невозможно разделить весь комплекс явлений на явления чисто электрические и чисто магнитные. При переходе от одной системы отсчета к другой электрическое и магнитное поля определенным образом преобразуются. Законы этого преобразования даются в теории относительности. Поэтому *полная теория электромагнитных явлений должна быть релятивистской.*

6. Фундаментальным свойством электричества является существование его в двух видах — в виде *положительного* и *отрицательного* электричеств. В окружающем нас мире количества этих электричеств в высокой степени одинаковы. Это не удивительно, поскольку одноименные электричества отталкиваются, а разноименные — притягиваются. Другим фундаментальным свойством электричества является *закон сохранения электрического заряда*, являющийся обобщением опытных фактов. Он утверждает, что *полный заряд системы не может измениться, если через ее границу не проходят электрически заряженные частицы.* Это не значит, что сохраняются в отдельности положительный и отрицательный заряды системы. Например, если в системе имеются частицы высоких энергий, то в ней могут рождаться фотоны также высоких энергий. Фотоны могут также вступать в систему через ее границу. Фотоны высокой энергии при взаимодействии с атомами вещества могут породить *пару*, состоящую из отрицательно заряженного электрона и положительно заряженного позитрона. По отношению к такому процессу энергия фотона считается высокой, если она превосходит сумму энергий покоя электрона и позитрона. Возможен обратный процесс, в котором пара электрон–позитрон исчезает (аннигилирует) с испусканием фотонов. Подобного рода процессы меняют количества положительного и отрицательного электричеств в системе. Однако полный заряд ее сохраняется неизменным. Впрочем, в настоящем томе мы будем иметь дело только с частицами «низких» энергий, для которых такие процессы невозможны. Тогда сохраняются не только полный заряд, но также положительный и отрицательный заряды в отдельности.

Глава I

ЭЛЕКТРИЧЕСКОЕ ПОЛЕ

§ 3. Закон Кулона. Принцип суперпозиции электростатических полей

1. Электростатика занимается изучением электрических полей *неподвижных зарядов.* Сначала мы изучим электростатические поля в вакууме. Основной количественный закон электростатики был открыт Кулоном в 1785 г. Он формулируется следующим образом.

Сила взаимодействия F двух точечных зарядов в вакууме направлена вдоль прямой, соединяющей эти заряды, пропорциональна их величинам q_1 и q_2 и обратно пропорциональна квадрату расстояния между ними r_{12}. Она является силой притяжения, если знаки зарядов разные, и силой отталкивания, если эти знаки одинаковы. Математически:

$$F = C\frac{q_1 q_2}{r_{12}^2}, \tag{3.1}$$

где C — числовой коэффициент. Поскольку мы располагаем независимыми способами измерения всех величин F, q, r, опытным путем можно проверить не только обратную пропорциональность силы F квадрату расстояния r, но и пропорциональность ее произведению зарядов q_1 и q_2.

Кулон открыл свой закон, измеряя силы притяжения и отталкивания заряженных шариков с помощью *крутильных весов*, изобретенных им же. (Опыты Кулона общеизвестны и описываются во всех школьных учебниках физики.) Примерно за 11 лет до Кулона тот же закон был получен Г. Кавендишем (1731–1810) из значительно более точных, но косвенных измерений (см. § 6, п. 4). Однако работа Кавендиша не была им опубликована и оставалась неизвестной в течение более 100 лет. Максвелл — первый директор Кавендишской лаборатории — обнаружил в архиве этой лаборатории подготовленную к печати рукопись указанной фундаментальной работы Кавендиша и опубликовал ее в 1879 г.

Точечность зарядов, о взаимодействии которых идет речь в законе Кулона, означает, что *линейные размеры тел, на которых сосредоточены эти заряды, пренебрежимо малы по сравнению с расстоянием между ними.* Закон предполагает, что заряды помещены в *абсолютном вакууме.* Опытная проверка его производится в воздухе. Принципиальный переход к абсолютному вакууму можно осуществить беспредельным уменьшением давления воздуха. Впрочем, в этом нет необходимости, так как влияние воздуха на силы взаимодействия очень мало и в большинстве случаев им можно пренебречь.

2. Числовое значение постоянной C в законе Кулона (3.1) можно выбрать произвольно и приписать этой постоянной любую размерность. Таким путем можно построить бесчисленное множество *систем единиц*. Однако не все они удобны. Наиболее удобной системой при изучении электромагнитного поля в физике является *абсолютная*, или *гауссова*, *система единиц*, которой мы и будем преимущественно пользоваться. Это есть система единиц СГС, дополненная единицами электрических и магнитных единиц. Она является комбинацией двух систем единиц: СГСЭ и СГСМ, т. е. *абсолютной электростатической* и *абсолютной электромагнитной* СГС-*систем единиц*.

Единицы «чисто» электрических величин (заряд, напряженность электрического поля, электрический потенциал, сила электрического тока, сопротивление проводника и пр.) в гауссовой системе совпадают с единицами СГСЭ, а единицы «чисто» магнитных величин (напряженность и индукция магнитного поля, магнитный момент, индуктивности проводов и пр.) — с единицами СГСМ. Система СГСМ будет рассмотрена в § 51. Здесь же мы изложим систему СГСЭ. В этой (и гауссовой) системе единиц коэффициент пропорциональности C в законе Кулона считается *безразмерным* и полагается равным *единице*. Закон Кулона записывается в виде

$$F = \frac{q_1 q_2}{r_{12}^2}. \tag{3.2}$$

Размерность заряда определяется формулой

$$[q] = [F^{1/2} L] = M^{1/2} L^{3/2} T^{-1},$$

а размерность напряженности электрического поля — формулой

$$[E] = [F q^{-1}] = M^{1/2} L^{-1/2} T^{-1}.$$

За единицу заряда принимается величина такого точечного заряда, который действует на такой же точечный заряд в вакууме с силой в одну дину, если расстояние между обоими зарядами равно одному сантиметру. За единицу напряженности электрического поля принимается напряженность такого поля в вакууме, которое действует на единичный точечный заряд с силой в одну дину. Эти единицы не получили специальных названий. Их просто называют *электростатическими* или СГСЭ-единицами заряда и напряженности электрического поля.

Практической единицей заряда является *кулон* (Кл). По точному определению кулон есть десятая доля СГСМ-ед. заряда или приблизительно $2{,}998 \cdot 10^9$ СГСЭ-ед. заряда. В грубых расчетах достаточна точность 1 Кл $= 3 \cdot 10^9$ СГСЭ-ед. заряда. Кулон — очень большая единица. Два точечных заряда в один кулон каждый, удаленные друг от друга на расстояние 1 км, взаимодействовали бы с силой

$$F = \frac{q^2}{r^2} = \frac{9 \cdot 10^{18}}{10^{10}} = 9 \cdot 10^8 \text{ дин} = 9 \cdot 10^3 \text{ Н}.$$

В системе СИ электрическому заряду (или току) приписывается самостоятельная размерность, не зависящая от размерностей длины, массы и времени. Эта система будет изложена в § 85.

3. В векторной форме закон Кулона записывается так:

$$\mathbf{F}_{ik} = \frac{q_i q_k}{r_{ik}^3}\,\mathbf{r}_{ik}, \tag{3.3}$$

где $\mathbf{F}_{ik}$ — сила, действующая на заряд q_k со стороны заряда q_i, а $\mathbf{r}_{ik}$ — радиус-вектор, проведенный от заряда q_i к заряду q_k. Напряженность поля точечного заряда q дается выражением

$$\mathbf{E} = \frac{q}{r^3}\,\mathbf{r}, \tag{3.4}$$

где $\mathbf{r}$ — радиус-вектор, проведенный от заряда q в рассматриваемую «точку наблюдения», к которой относится значение поля $\mathbf{E}$.

Напряженность электрического поля $\mathbf{E}$ нескольких неподвижных точечных зарядов q_1, q_2, ... равна векторной сумме напряженностей полей, которые создавал бы каждый из этих зарядов в отсутствие остальных, т. е.

$$\mathbf{E} = \sum_i \frac{q_i}{r_i^3}\,\mathbf{r}_i, \tag{3.5}$$

где $\mathbf{r}_i$ — радиус-вектор, проведенный из заряда q_i в точку наблюдения. Это положение, являющееся обобщением опытных фактов, называется *принципом суперпозиции электростатических полей*. Не исключена возможность, что принцип суперпозиции нарушается на малых расстояниях порядка размеров атомных ядер (10^{-13} см) и меньше.

Формула (3.5) позволяет в принципе рассчитать напряженность электрического поля любой системы неподвижных зарядов. Если заряды не точечные, то их следует мысленно разделить на малые части, чтобы каждую из них можно было считать точечным зарядом. При непрерывном распределении электричества сумма (3.5) переходит в интеграл.

4. Для наглядного изображения электрических полей широко пользуются *силовыми линиями*. Силовая линия есть математическая линия, направление касательной к которой в каждой точке, через которую она проходит, совпадает с направлением вектора $\mathbf{E}$ в той же точке. За положительное направление силовой линии условились считать направление самого вектора $\mathbf{E}$. При таком соглашении можно сказать, что электрические силовые линии начинаются от положительных зарядов и оканчиваются на отрицательных. В § 5 будет показано, что в пространстве, свободном от электрических зарядов, силовые линии идут гуще там, где поле $\mathbf{E}$ сильнее, и реже там, где оно слабее. Поэтому по густоте силовых линий можно судить и о величине напряженности электрического поля. На рис. 1 изображены силовые линии равномерно заряженных шариков — положительного и отрицательного, а на

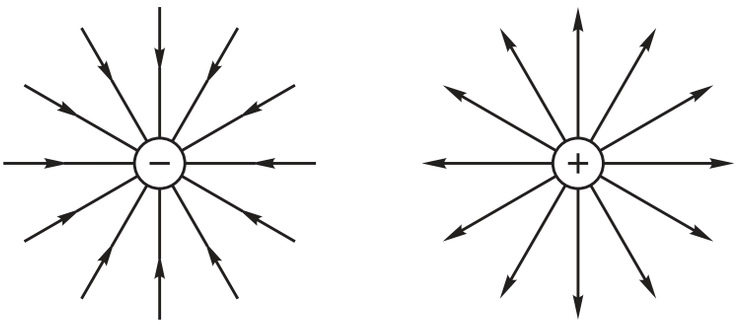


Рис. 1

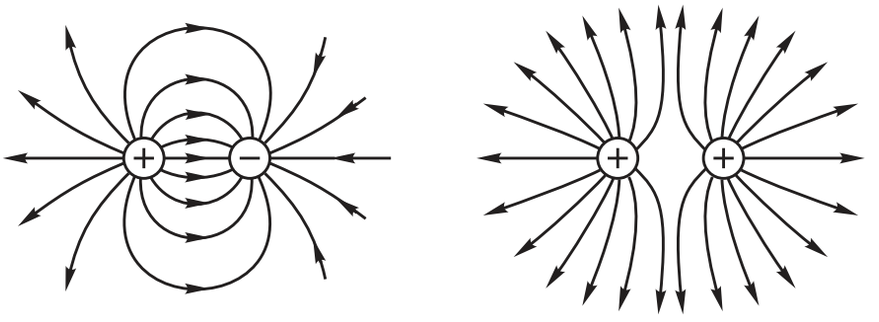


Рис. 2

рис. 2 — двух разноименных и одноименных зарядов равных величин, сосредоточенных на таких шариках.

Для экспериментального воспроизведения силовых линий берут стеклянный сосуд с плоским дном и наливают в него какую-либо изолирующую жидкость, например касторовое масло, глицерин и т. п. В жидкости по возможности равномерно распределяются истертые в порошок кристаллики гипса, асбеста, частицы манной крупы или какие-либо другие продолговатые частицы. В жидкость погружаются металлические электроды. При соединении их с источниками электричества возбуждается электрическое поле. В этом поле частицы электризуются через влияние и, притягиваясь друг к другу разноименно наэлектризованными концами, располагаются в виде цепочек вдоль силовых линий. Картина силовых линий искажается течениями жидкости, вызываемыми силами, действующими на нее в неоднородном электрическом поле. Лучшие результаты получаются по методу, применявшемуся Робертом В. Полем (1884–1976). На стеклянную пластинку наклеиваются электроды из станиоля, между которыми создается электрическое напряжение. Затем на пластинку насыпают, слегка постукивая по ней, продолговатые частицы, на-

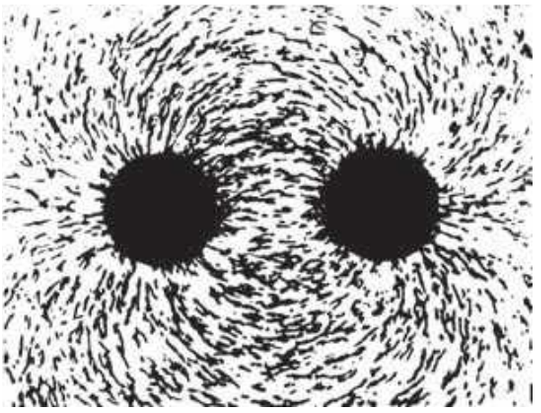


Рис. 3

пример кристаллики гипса. Они располагаются вдоль силовых линий. На рис. 3 изображена полученная таким образом картина силовых линий между двумя разноименно заряженными кружками из станиоля.

Метод силовых линий, конечно, применим для графического представления любых векторных полей. Линии, указывающие направления соответствующего вектора, называются *линиями этого вектора*, или *векторными линиями*. Им дают также специальные названия в зависимости от физического или геометрического смысла рассматриваемого векторного поля. Так, в гидродинамике вводятся *линии тока*, указывающие направление вектора скорости жидкости (см. т. I, § 93). Для представления магнитного поля пользуются *магнитными силовыми линиями* и т. д.

ЗАДАЧИ

1. Два маленьких шарика с массами $m = 0{,}1$ г каждый подвешены на шелковых нитях так, что их поверхности соприкасаются. После сообщения шарикам электрического заряда они оттолкнулись и их центры разошлись на расстояние $D = 6$ см (рис. 4). Определить заряд шарика q, если длина нити $l = 30$ см.

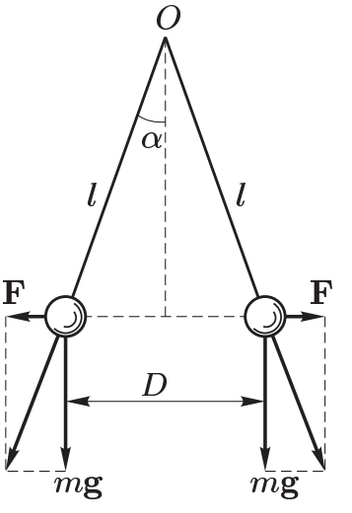


Рис. 4

Решение. В состоянии равновесия $F = mg\,\mathrm{tg}\,\alpha$. По закону Кулона $F = q^2/D^2$. Отсюда

$$q = D\sqrt{mg\,\mathrm{tg}\,\alpha},$$

или для малых углов

$$q = D\sqrt{mg\alpha} = D\sqrt{mgD/2l} = \\ = 18{,}8 \text{ СГСЭ-ед. заряда} = 6{,}3 \cdot 10^{-9} \text{ Кл}. \quad (3.6)$$

Этот пример показывает, что уже ничтожные доли кулона (около одной миллиардной) легко могут быть обнаружены с помощью сравнительно грубого электроскопа.

2. Вычислить напряженность электрического поля равномерно заряженного тонкого диска на его геометрической оси.

Решение. Пусть σ — поверхностная плотность электричества, т. е. количество электричества, приходящееся на единицу площади диска. По условию величина σ одна и та же во всех точках плоскости диска. Из соображений симметрии ясно, что поле E в любой точке A геометрической оси (рис. 5) должно быть направлено вдоль этой оси. Заметив это, возьмем в плоскости диска произвольную бесконечно малую площадку dS с зарядом $dq = \sigma dS$. Напряженность поля, создаваемого этим зарядом в точке A, будет $dE = \sigma dS/r^2$, а его проекция на ось OA $dE_x = \sigma dS\cos\alpha/r^2$, или $dE_x = \sigma d\Omega$, где $d\Omega = dS\cos\alpha/r^2$ — телесный угол, под которым площадка dS видна из точки A. Полное поле E найдется

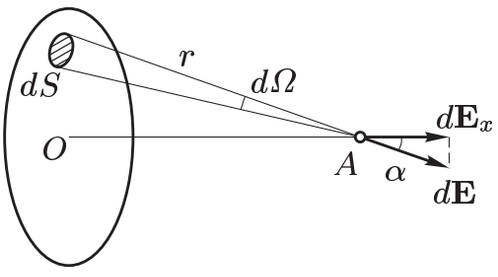


Рис. 5

интегрированием выражения dE_x по телесному углу. Таким путем получаем

$$E = \sigma\Omega, \tag{3.7}$$

где Ω — полный телесный угол, под которым диск виден из точки A. В частности, когда радиус диска бесконечно велик (бесконечная заряженная плоскость),

$$E = 2\pi\sigma. \tag{3.8}$$

3. Один из опытов Кулона, с помощью которого он убедился, что сила притяжения между двумя разноименными точечными зарядами обратно пропорциональна квадрату расстояния между ними, состоял в следующем. В окрестности маленького заряженного шарика подвешивалась на нити небольшая горизонтальная шеллаковая стрелка, на одном конце которой был прикреплен небольшой электрически заряженный кружок из золотой бумаги. Измерялся период малых колебаний стрелки T в зависимости от ее расстояния r до заряженного шарика. Предполагая справедливым закон Кулона, найти зависимость периода колебаний стрелки от указанного расстояния и от других параметров системы. Длина стрелки l очень мала по сравнению с расстоянием r.

Ответ. $T = 2\pi r\sqrt{2I/q_1q_2l}$, где q_1 и q_2 — заряды шарика и кружка, а I — момент инерции стрелки.

§ 4. Электрический диполь

1. Простейшей системой точечных зарядов является *электрический диполь* (двойной полюс). Так называется совокупность равных по величине, но противоположных по знаку двух точечных зарядов — q и $+q$, сдвинутых друг относительно друга на некоторое расстояние (рис. 6). Пусть $\mathbf{l}$ — радиус-вектор, проведенный от отрицательного заряда к положительному. Вектор $\mathbf{p} = q\mathbf{l}$ называется *электрическим моментом диполя* или *дипольным моментом*. Если длина l пренебре-

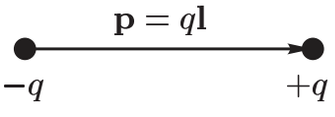


Рис. 6

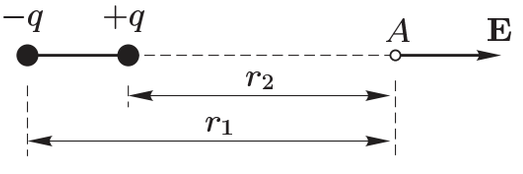


Рис. 7

жимо мала по сравнению с расстоянием от диполя до точки наблюдения, то диполь называется *точечным*. Вычислим электрическое поле точечного диполя. В окончательных формулах, которые мы получим, безразлично (в пределах принятой точности расчета), от какой точки диполя отсчитывается расстояние r до точки наблюдения.

Рассмотрим сначала случай, когда точка наблюдения A лежит на продолжении оси диполя (рис. 7). Напряженность электрического поля в этой точке будет

$$E = q\left(\frac{1}{r_2^2} - \frac{1}{r_1^2}\right) \approx q\frac{d}{dr}\left(\frac{1}{r^2}\right)(r_2 - r_1), \quad \text{или} \quad E = \frac{2ql}{r^3} = \frac{2p}{r^3}.$$

В векторной форме:

$$\mathbf{E} = \frac{2\mathbf{p}}{r^3}. \tag{4.1}$$

Допустим теперь, что точка наблюдения A лежит на перпендикуляре, восстановленном к оси диполя из его центра O (рис. 8). Вектор $\mathbf{E}$ получается геометрическим сложением полей $\mathbf{E}_1$ и $\mathbf{E}_2$, возбуждаемых точечными зарядами $-q$ и $+q$. Как видно из рисунка, вектор $\mathbf{E}$ антипараллелен $\mathbf{p}$, а его длина (для точечного диполя) равна

$$E = E_1\alpha = \frac{q}{r^2}\,\frac{l}{r} = \frac{p}{r^3}.$$

В векторной форме:

$$\mathbf{E} = -\frac{\mathbf{p}}{r^3}. \tag{4.2}$$

Не обязательно, чтобы перпендикуляр AO проходил через центр (точечного) диполя. В принятом приближении формула (4.2) остается верной и тогда, когда за O принята любая точка диполя. Ее можно выбрать даже вне диполя. Важно только, чтобы ее расстояние до центра диполя было пренебрежимо мало по сравнению с расстоянием до точки наблюдения.

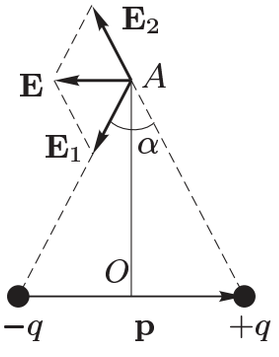


Рис. 8

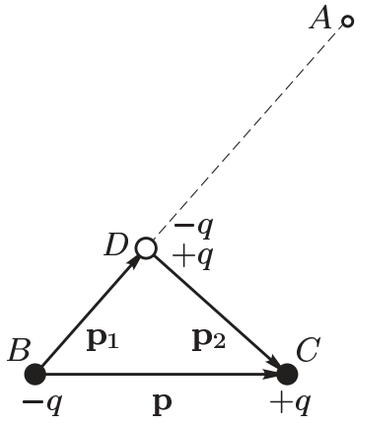


Рис. 9

Общий случай сводится к разобранным частным случаям. Опустим из заряда $+q$ перпендикуляр CD на линию наблюдения BA (рис. 9). Поместим в точке D два точечных заряда: $+q$ и $-q$. Это не изменит поля. Но полученную систему четырех зарядов можно рассматривать как совокупность двух диполей с дипольными моментами $\mathbf{p}_1$ и $\mathbf{p}_2$, изображенными на рис. 9. Вообще, при вычислении напряженности поля, а также сил, действующих на диполь, последний всегда можно заменить системой любого числа диполей, геометрическая сумма моментов которых равна моменту рассматриваемого диполя. Применяя теперь к диполям $\mathbf{p}_1$ и $\mathbf{p}_2$ формулы (4.1) и (4.2), получим

$$\mathbf{E} = \frac{1}{r^3}\,(2\mathbf{p}_1 - \mathbf{p}_2),$$

или, исключая $\mathbf{p}_2$ с помощью соотношения $\mathbf{p}_1 + \mathbf{p}_2 = \mathbf{p}$,

$$\mathbf{E} = \frac{1}{r^3}\,(3\mathbf{p}_1 - \mathbf{p}),$$

или, наконец,

$$\mathbf{E} = \frac{3(pr)}{r^5}\,\mathbf{r} - \frac{\mathbf{p}}{r^3}. \tag{4.3}$$

2. Рассмотрим теперь силы, действующие на диполь в электрическом поле. Если поле однородно, то результирующая сила $\mathbf{F}$ равна нулю, так как силы $\mathbf{F}_1$ и $\mathbf{F}_2$, действующие на отрицательный и положительный заряды диполя, равны по модулю и противоположны по направлению (рис. 10). Момент этих сил $\mathbf{M} = [\,\mathbf{l}\mathbf{F}_2\,] = q[\,\mathbf{l}\mathbf{E}\,]$, или

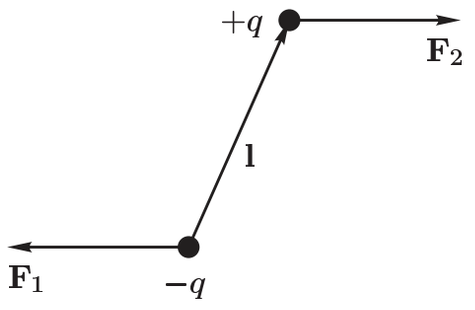


Рис. 10

$$\mathbf{M} = [\,\mathbf{p}\mathbf{E}\,]. \tag{4.4}$$

Момент $\mathbf{M}$ стремится повернуть ось диполя в направлении поля $\mathbf{E}$. Существуют два положения равновесия диполя: когда диполь параллелен электрическому полю и когда он антипараллелен ему. Первое положение устойчиво, второе — неустойчиво. Формула (4.4) верна также для точечного диполя в неоднородном поле.

Если поле неоднородно, то сила $\mathbf{F} = \mathbf{F}_1 + \mathbf{F}_2$, вообще говоря, не обращается в нуль. В этом случае $\mathbf{F} = q(\mathbf{E}_2 - \mathbf{E}_1)$, где $\mathbf{E}_1$ и $\mathbf{E}_2$ — напряженности поля в точках нахождения зарядов $-q$ и $+q$. Для точечного диполя разность $\mathbf{E}_2 - \mathbf{E}_1$ можно приближенно заменить дифференциалом

$$d\mathbf{E} = l_x\frac{\partial \mathbf{E}}{\partial x} + l_y\frac{\partial \mathbf{E}}{\partial y} + l_z\frac{\partial \mathbf{E}}{\partial z}.$$

В этом приближении

$$\mathbf{F} = p_x\frac{\partial \mathbf{E}}{\partial x} + p_y\frac{\partial \mathbf{E}}{\partial y} + p_z\frac{\partial \mathbf{E}}{\partial z}. \tag{4.5}$$

В целях сокращения письма введем оператор «набла», или оператор Гамильтона (1805–1865):

$$\nabla = \mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z}. \tag{4.6}$$

Скалярным умножением $\mathbf{p}$ на ∇ получаем оператор

$$(\mathbf{p}\nabla) = p_x\frac{\partial}{\partial x} + p_y\frac{\partial}{\partial y} + p_z\frac{\partial}{\partial z}.$$

С помощью этого оператора формула (4.5) записывается так:

$$\mathbf{F} = (\mathbf{p}\nabla)\mathbf{E}. \tag{4.7}$$

Смысл этой формулы раскрывается при сравнении ее с выражением (4.5). В частности, если ось X направить вдоль вектора $\mathbf{p}$ ($p_x = p$, $p_y = p_z = 0$), то

$$\mathbf{F} = p\frac{\partial \mathbf{E}}{\partial x}. \tag{4.8}$$

3. Нейтральная система точечных зарядов, занимающая небольшой объем, в первом приближении ведет себя как точечный диполь. Действительно, разделим мысленно заряды системы на более мелкие части, и притом так, чтобы каждому заряду соответствовал равный заряд противоположного знака. Сгруппировав такие заряды попарно, можно рассматривать нашу систему как систему диполей $\mathbf{p}_i$. При вычислении поля на расстояниях, больших по сравнению с размерами системы, такие диполи можно считать точечными. Их можно перенести в одну точку и векторно сложить в один точечный диполь с дипольным моментом $\mathbf{p} = \sum \mathbf{p}_i$. Так же можно поступать при вычислении сил, действующих на систему зарядов во внешнем электрическом поле $\mathbf{E}$. Необходимо только, чтобы размеры системы были настолько малы, чтобы во всех точках занимаемого ею объема внешнее поле $\mathbf{E}$ и его пространственные производные $\partial\mathbf{E}/\partial x$, $\partial\mathbf{E}/\partial y$, $\partial\mathbf{E}/\partial z$ могли с достаточной точностью считаться одинаковыми.

Найдем общее выражение для дипольного момента нейтральной системы зарядов. Предполагая сначала, что система указанным выше способом разделена на пары равных зарядов противоположного знака, напишем

$$\mathbf{p} = \sum q_i^+ \mathbf{l}_i = \sum q_i^+ (\mathbf{r}_i^+ - \mathbf{r}_i^-) = \sum (q_i^+ \mathbf{r}_i^+ + q_i^- \mathbf{r}_i^-),$$

где q_i^+, $\mathbf{r}_i^+$ и q_i^-, $\mathbf{r}_i^-$ — величины положительных и отрицательных зарядов и их радиусы-векторы соответственно. Теперь можно снова соединить мысленно разделенные заряды и вернуться к первоначальной системе их. Тогда получится

$$\mathbf{p} = \sum q_i \mathbf{r}_i, \tag{4.9}$$

где суммирование производится уже по всем зарядам первоначальной системы, как положительным, так и отрицательным.

Для электрически нейтральной системы величина суммы (4.9) не зависит от выбора начала координат. Действительно, при переходе к новой (штрихованной) системе координат с другим началом $\mathbf{r}_i' = \mathbf{r}_i + \mathbf{a}$, где $\mathbf{a}$ — постоянный вектор. Поэтому дипольный момент в новой системе координат будет

$$\mathbf{p}' = \sum q_i \mathbf{r}_i' = \sum q_i \mathbf{r}_i + \mathbf{a} \sum q_i.$$

Последнее слагаемое обращается в нуль из-за электрической нейтральности системы, а потому $\mathbf{p}' = \sum q_i \mathbf{r}_i = \mathbf{p}$.

ЗАДАЧИ

1. Найти силу взаимодействия F между точечным зарядом q и точечным диполем $\mathbf{p}$, если расстояние между ними равно r и дипольный момент $\mathbf{p}$ направлен вдоль соединяющей их прямой.

Ответ. $F = 2qp/r^3$. Диполь будет притягиваться к заряду, если он обращен к нему противоположно заряженным концом, и отталкиваться в противном случае.

2. Найти силу взаимодействия F двух точечных диполей $\mathbf{p}_1$ и $\mathbf{p}_2$, дипольные моменты которых направлены вдоль соединяющей их прямой, а расстояние между диполями равно r.

Ответ. $F = 6p_1p_2/r^4$. Диполи притягиваются, если они обращены друг к другу противоположно заряженными концами, и отталкиваются в противном случае.

3. Найти уравнение силовых линий электрического поля точечного диполя в полярной системе координат.

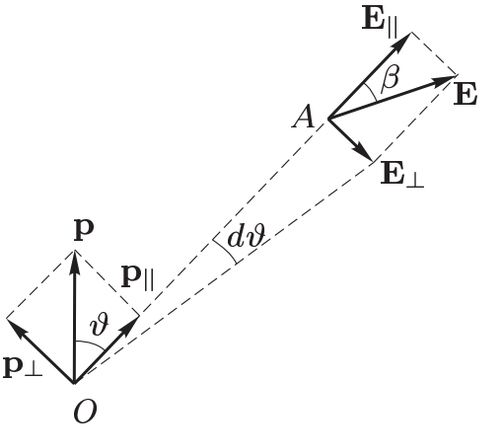


Рис. 11

Решение. Разложим вектор $\mathbf{p}$ (рис. 11) на составляющую $\mathbf{p}_{\parallel}$ вдоль радиуса $\mathbf{r}$ и составляющую $\mathbf{p}_{\perp}$, к нему перпендикулярную. Соответствующие им поля в точке наблюдения A будут

$$\mathbf{E}_{\parallel} = \frac{2\mathbf{p}_{\parallel}}{r^3}, \quad \mathbf{E}_{\perp} = -\frac{\mathbf{p}_{\perp}}{r^3}.$$

Угол β между радиусом $\mathbf{r}$ и электрической силовой линией определится формулой

$$\operatorname{tg}\beta = \frac{E_{\perp}}{E_{\parallel}} = \frac{p_{\perp}}{2p_{\parallel}} = \frac{1}{2}\operatorname{tg}\vartheta.$$

Проекция бесконечно малого участка силовой линии на направление вектора $\mathbf{p}_{\perp}$ может быть, с одной стороны, представлена как $dr\operatorname{tg}\beta = (dr/2)\operatorname{tg}\vartheta$, с другой стороны, как $rd\vartheta$. Поэтому

$$\frac{dr}{2}\operatorname{tg}\vartheta = r\,d\vartheta.$$

Интегрируя это уравнение, получаем искомое уравнение электрической силовой линии:

$$r = r_0\sin^2\vartheta.$$

Постоянная r_0 имеет смысл длины радиуса-вектора $\mathbf{r}$ в экваториальной плоскости, т. е. при $\vartheta = \pi/2$.

4. Возможны ли круговые движения с постоянной скоростью точечного электрического заряда вокруг неподвижного точечного электрического диполя?

Ответ. Да, возможны, и притом на любом расстоянии заряда от диполя. Плоскость круговой орбиты заряда перпендикулярна к оси диполя. Угол α между направлением дипольного момента и радиусом-вектором, проведенным от диполя к движущемуся заряду, определяется выражением $\cos\alpha = \mp\sqrt{1/3}$, где минус относится к положительному заряду, а плюс — к отрицательному.

§ 5. Поток вектора и электростатическая теорема Гаусса

1. Понятие *потока вектора* является одним из важнейших понятий векторного анализа. Оно используется при формулировке важнейших свойств электрического, магнитного и других векторных полей. Первоначально это понятие было введено в гидродинамике. Возьмем в поле скоростей жидкости малую площадку S, перпендикулярную к вектору скорости жидкости $\mathbf{v}$ (рис. 12). Объем жидкости, протекающей через эту площадку за время dt, равен $vS\,dt$. Если площадка наклонена

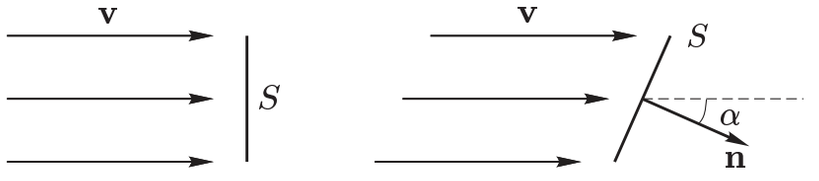


Рис. 12

к потоку, то соответствующий объем будет $vS\cos\alpha\,dt$, где α — угол между вектором скорости $\mathbf{v}$ и нормалью $\mathbf{n}$ к площадке S. Объем жидкости, протекающей через площадку S в единицу времени, получится делением этого выражения на dt. Он равен $vS\cos\alpha$, т. е. скалярному произведению $(\mathbf{vS})$ вектора скорости $\mathbf{v}$ на вектор площадки $\mathbf{S} = S\mathbf{n}$. Единичный вектор $\mathbf{n}$ нормали к площадке S можно провести в двух прямо противоположных направлениях. Одно из них условно принимается за положительное. В этом направлении и проводится нормаль $\mathbf{n}$. Та сторона площадки, из которой исходит нормаль $\mathbf{n}$, называется *внешней*, а та, в которую нормаль $\mathbf{n}$ входит, — *внутренней*. Если поверхность S не бесконечно мала, то при вычислении объема протекающей жидкости ее надо разбить на бесконечно малые площадки dS, а затем вычислить интеграл $\int(\mathbf{v}\,d\mathbf{S})$ по всей поверхности S.

Выражения типа $(\mathbf{v}\,d\mathbf{S})$ или $\int(\mathbf{v}\,d\mathbf{S})$ встречаются в самых разнообразных вопросах физики и математики. Эти выражения имеют смысл независимо от конкретной физической природы вектора $\mathbf{v}$. Они называются *потоком вектора* $\mathbf{v}$ через бесконечно малую площадку dS или конечную поверхность S соответственно. Так, интеграл $\Phi = \int \mathbf{E}\,d\mathbf{S}$ называют потоком вектора напряженности электрического поля $\mathbf{E}$, хотя с этим понятием и не связано никакое реальное течение.

Допустим, что вектор $\mathbf{E}$ представляется геометрической суммой

$$\mathbf{E} = \sum \mathbf{E}_i.$$

Умножив это соотношение скалярно на dS и проинтегрировав, получим

$$\Phi = \sum \Phi_i, \tag{5.1}$$

где Φ_1, Φ_2, ... — потоки векторов $\mathbf{E}_1$, $\mathbf{E}_2$ ... через ту же самую поверхность. Таким образом, из того факта, что векторы складываются

геометрически, следует, что их потоки через одну и ту же поверхность складываются *алгебраически*.

2. Перейдем теперь к доказательству важнейшей теоремы электростатики — *теоремы Гаусса*. Она определяет поток вектора напряженности электрического поля через произвольную замкнутую поверхность S. За положительную нормаль к поверхности S примем *внешнюю нормаль*, т. е. нормаль, направленную наружу (рис. 13). Предположим

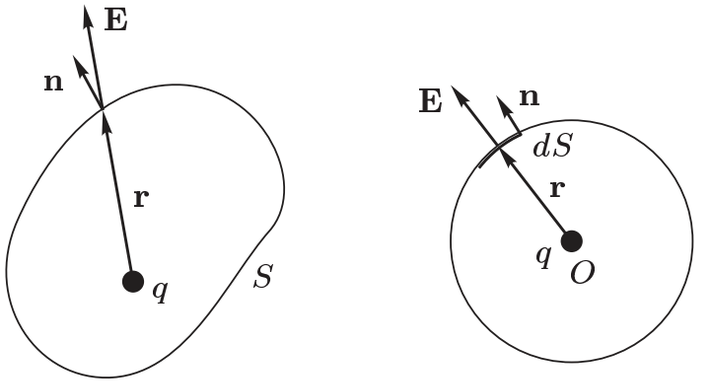


Рис. 13

сначала, что электрическое поле создается единственным точечным зарядом q. На поверхности S поле определяется выражением

$$\mathbf{E} = \frac{q\mathbf{r}}{r^3}. \tag{5.2}$$

Рассмотрим сначала простейший случай, когда поверхность S является сферой, а точечный заряд q помещен в ее центре. Поток вектора $\mathbf{E}$ через элементарную площадку сферы равен

$$d\Phi = (\mathbf{En})\, dS = \frac{q\, dS}{r^2},$$

а поток через всю сферу $\Phi = qS/r^2$. Так как поверхность сферы S равна $4\pi r^2$, то

$$\Phi = 4\pi q. \tag{5.3}$$

Покажем теперь, что результат (5.3) не зависит от формы поверхности S, окружающей заряд q. Возьмем произвольную элементарную площадку dS с установленным на ней положительным направлением нормали $\mathbf{n}$ (рис. 14). Поток вектора $\mathbf{E}$ через эту площадку будет

$$d\Phi = (\mathbf{En})\, dS = E\, dS \cos\alpha = E\, dS_r,$$

где dS_r — проекция площадки $d\mathbf{S}$ на плоскость, перпендикулярную к радиусу $\mathbf{r}$. Используя выражение (5.2), получим $d\Phi = q\, dS_r/r^2$. Величина dS_r/r^2 есть телесный угол $d\Omega$, под которым из точки нахождения заряда q видна площадка dS_r, а следовательно, и площадка dS. Условимся считать его положительным, если площадка dS обращена к q внутренней стороной, и отрицательным в противоположном случае.

Тогда $d\Phi = q\,d\Omega$. Поток Φ через произвольную (вообще говоря, незамкнутую) конечную поверхность S найдется интегрированием этого выражения по $d\Omega$. Так как заряд q не зависит от положения площадки dS, то $\Phi = q\int d\Omega$, или

$$\Phi = q\Omega, \tag{5.4}$$

где Ω — телесный угол, под которым из точки нахождения заряда q видна поверхность S.

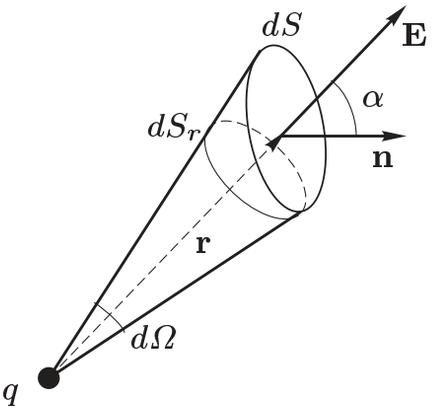


Рис. 14

Если поверхность S замкнутая, то следует различать два случая.

Случай 1. Заряд q лежит внутри пространства, окруженного поверхностью S (рис. 15 *а*). В этом случае телесный угол $d\Omega$ охватывает все направления в пространстве, т. е. равен 4π, а потому формула (5.4) переходит в (5.3). Не имеет значения, сколько раз прямая, исходящая из q, пересекает поверхность S. Допустим, например, что пересечение происходит три раза (рис. 15 *б*). Абсолютные значения телесных углов, под которыми видны элементарные площадки dS_1, dS_2, dS_3, одинаковы. Однако

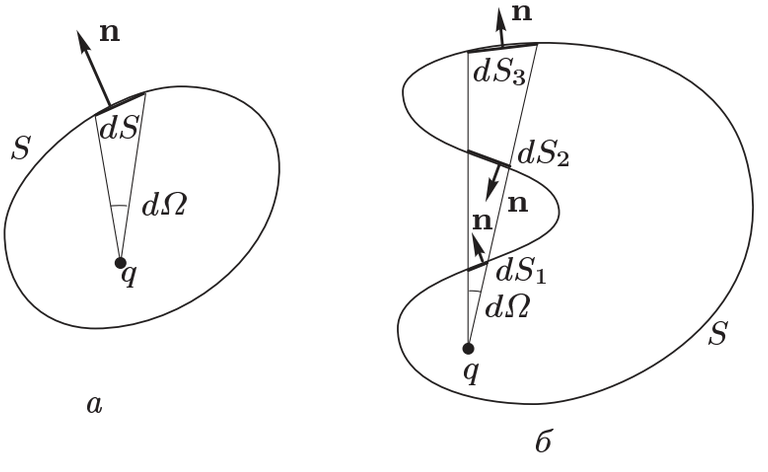


Рис. 15

площадка dS_3 обращена к q внутренней, а dS_2 — внешней сторонами. Сумма телесных углов, под которыми видны эти две площадки, равна нулю. Остается только телесный угол $d\Omega$, под которым видна площадка dS_1. И так будет всегда, когда число пересечений нечетное, т. е. в тех случаях, когда поверхность S окружает заряд q. Нечетное число пересечений при вычислении потока сводится к одному пересечению.

Случай 2. Заряд q лежит вне пространства, окруженного поверхностью S (рис. 16). В этом случае прямая, исходящая из заряда q, либо совсем не пересекает замкнутую поверхность S, либо пересекает ее четное число раз. Поэтому полный телесный угол Ω, а с ним и поток Φ равны нулю.

Случай, когда точечный заряд q лежит точно на поверхности S, физического смысла не имеет. Точечный заряд есть идеализация, пользоваться которой допустимо только в тех случаях, когда линейные размеры заряженного тела малы по сравнению с расстояниями, на которых рассматривается поле этого тела. Если же заряд лежит на поверхности, то точки последней вблизи самого заряда этому условию не удовлетворяют.

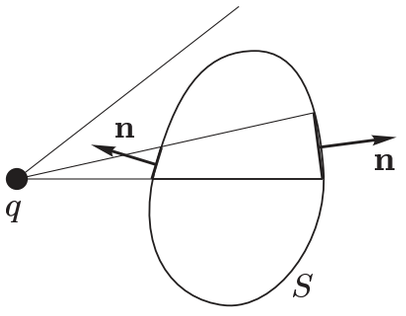


Рис. 16

Допустим, что поле $\mathbf{E}$ является суперпозицией полей $\mathbf{E}_1, \mathbf{E}_2, \ldots$ точечных зарядов $q_1, q_2, \ldots$ По теореме, доказанной выше, поток вектора $\mathbf{E}$ равен сумме потоков векторов $\mathbf{E}_1, \mathbf{E}_2, \ldots$ Если заряд q_i окружен замкнутой поверхностью S, то его поток через эту поверхность будет $4\pi q_i$. Если же заряд лежит во внешнем пространстве по отношению к поверхности S, то его поток равен нулю. В результате получается следующее фундаментальное соотношение:

$$\Phi \equiv \oint (\mathbf{E}\, d\mathbf{S}) = 4\pi q, \tag{5.5}$$

называемое *электростатической теоремой Гаусса*. Здесь q — алгебраическая сумма всех зарядов, окруженных замкнутой поверхностью S. Заряды, расположенные во внешнем пространстве по отношению к этой поверхности, на величину потока не влияют.

При доказательстве предполагалось, что все заряды точечные. Но это ограничение легко снять, так как всякий заряд можно мысленно разделить на малые части, каждая из которых может рассматриваться как точечный заряд.

3. Возьмем какой-либо произвольный замкнутый контур L и через каждую точку его проведем электрическую силовую линию (рис. 17). Эти линии образуют трубчатую поверхность, называемую *силовой трубкой*. Рассмотрим произвольное поперечное сечение трубки поверхностью S. Положительную нормаль к S проведем в ту же сторону, в какую направлены силовые линии. Пусть Φ — поток вектора $\mathbf{E}$ через сечение S. Мы утверждаем, что если внутри трубки нет электрических зарядов, то поток Φ остается одним и тем же по всей длине трубки. Для доказательства возьмем другое поперечное сечение трубки S'. Рассмотрим замкнутую поверхность, ограниченную сечениями S и S' и боковой поверхностью трубки. Применим к ней теорему Гаусса. Поток через боковую поверхность равен нулю, так как на поверхности трубки вектор $\mathbf{E}$ касается этой поверхности. Поток через основание S численно равен Φ, но противоположен по знаку, так

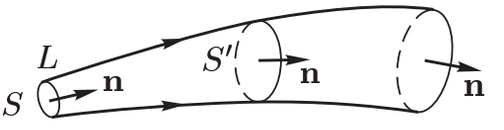


Рис. 17

как внешняя нормаль к замкнутой поверхности направлена противоположно $\mathbf{n}$. Напротив, поток через основание S' равен $+\Phi'$. Полный поток через рассматриваемую замкнутую поверхность будет $\Phi' - \Phi$. По теореме Гаусса тот же поток равен нулю, так как внутри силовой трубки нет электрических зарядов. Таким образом, $\Phi' = \Phi$. В частности, если трубка бесконечно узкая, а сечения S и S' нормальны к ней, то

$$ES = E'S'.$$

Получается полная аналогия с течением несжимаемой жидкости. В тех местах, где трубка у́же, поле $\mathbf{E}$ сильнее. В тех местах, где она шире, поле $\mathbf{E}$ слабее. Следовательно, по густоте силовых линий можно судить о напряженности электрического поля (см. § 3, п. 4).

4. Теорема Гаусса есть следствие закона Кулона. Последний по своей форме не отличается от закона всемирного тяготения Ньютона. И тут и там сила взаимодействия меняется обратно пропорционально квадрату расстояния. Поэтому теорема Гаусса справедлива также для гравитационных полей. Роль заряда играет гравитационная масса (умноженная на гравитационную постоянную). Различие состоит только в том, что электрические заряды могут быть и положительными и отрицательными, тогда как гравитационные массы всегда положительны.

§ 6. Применения теоремы Гаусса

Для расчета электрических полей произвольной системы зарядов теоремы Гаусса недостаточно. Это видно уже из того, что теорема Гаусса есть *скалярное соотношение*. А одного скалярного уравнения мало для определения трех неизвестных — составляющих E_x, E_y, E_z вектора $\mathbf{E}$. Необходима известная симметрия задачи, чтобы последняя свелась к решению одного скалярного уравнения. В таких случаях (но и то далеко не всегда) теорема Гаусса может оказаться достаточной для вычисления вектора $\mathbf{E}$. Приведем примеры.

1. Электростатическое поле бесконечной равномерно заряженной плоскости. Поверхностная плотность электричества σ на заряженной плоскости, по предположению, постоянна. Ввиду симметрии вектор $\mathbf{E}$ должен быть перпендикулярен к этой плоскости. Он направлен от плоскости, если она заряжена положительно, и к плоскости, если ее заряд отрицателен. Ввиду той же симметрии длина вектора $\mathbf{E}$ может зависеть только от расстояния до заряженной плоскости, но не может зависеть от того, по какую сторону от нее находится точка наблюдения. Заметив это, построим цилиндр с основаниями, симметрично расположенными по разные стороны плоскости, и с образующими, перпендикулярными к ней (рис. 18). Если S — площадь каждого из оснований, то поток вектора $\mathbf{E}$ через одно основание будет ES, а через оба основания $2ES$. Поток через боковую поверхность цилиндра равен нулю, так как на ней векторы $\mathbf{E}$ и $\mathbf{n}$ взаимно перпендикулярны. Поток через всю поверхность цилиндра будет равен поэтому $\Phi = 2ES$.

По теореме Гаусса тот же поток можно представить в виде $\Phi = 4\pi q = 4\pi\sigma S$. Сравнивая оба выражения, находим

$$E = 2\pi\sigma. \tag{6.1}$$

Такой же результат был получен ранее непосредственно из закона Кулона (см. задачу 2 к § 3).

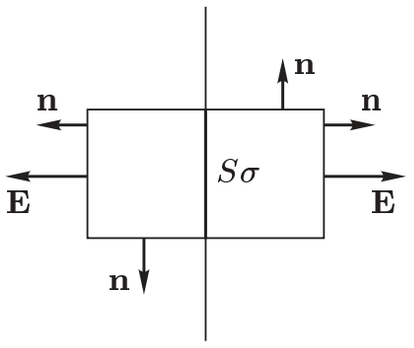


Рис. 18

Напряженность поля бесконечной заряженной плоскости, таким образом, не зависит от расстояния до нее. Плоскость может считаться бесконечной, если расстояние от нее пренебрежимо мало по сравнению с ее размерами. Только на таких расстояниях E не зависит от расстояния до плоскости. На больших расстояниях формула (6.1) неприменима — напряженность поля убывает с расстоянием. Если расстояния порядка размеров самой плоскости, то величина и направление поля в пространстве меняются очень сложно. На расстояниях, очень больших по сравнению с размерами плоскости, заряженная плоскость действует как точечный заряд — поле убывает обратно пропорционально квадрату расстояния.

Отметим еще, что по разные стороны плоскости векторы $\mathbf{E}$ одинаковы по модулю, но противоположны по направлению. Поэтому при переходе через заряженную плоскость напряженность электрического поля меняется скачком на величину $4\pi\sigma$.

2. Поле равномерно заряженной бесконечной плоскопараллельной пластинки. Пусть $2a$ — толщина пластинки, а ρ — объемная плотность электричества внутри пластинки. По предположению величина ρ постоянна. Начало координат O поместим в средней плоскости пластинки, а ось x направим перпендикулярно к ней (рис. 19). Рассуждая, как в предыдущей задаче, найдем

$$E = \begin{cases} 4\pi\rho x & \text{внутри пластинки,} \\ 4\pi\rho a & \text{вне пластинки.} \end{cases} \tag{6.2}$$

Будем беспредельно уменьшать толщину пластинки, одновременно увеличивая плотность электричества ρ, чтобы величина ρa оставалась постоянной. В пределе получится бесконечная равномерно заряженная плоскость с поверхностной плотностью электричества $\sigma = 2\rho a$, а формула (6.2) перейдет в ранее полученную формулу (6.1).

3. Поле шара, равномерно заряженного по поверхности и объему. Ввиду шаровой симметрии вектор $\mathbf{E}$ параллелен или антипараллелен радиусу-вектору $\mathbf{r}$, проведенному из центра шара в точку наблюдения, а его длина E может зависеть только от расстояния r. Заметив это, проведем вне шара концентрическую с ним сферу S радиуса r (рис. 20 *а*). Поток вектора $\mathbf{E}$ через эту сферу $4\pi r^2 E$ по теореме Гаусса равен $4\pi q$, а потому для напряженности поля вне

шара получаем

$$E = \frac{q}{r^2}, \tag{6.3}$$

независимо от того, заряжен ли шар по объему или по поверхности. Таким образом, *равномерно заряженный шар создает во внешнем пространстве такое поле, как если бы весь заряд был сосредоточен в его центре.* Этот результат остается справедливым при *любом сферически симметричном* распределении заряда по объему шара. Разумеется, он верен и для поля тяготения.

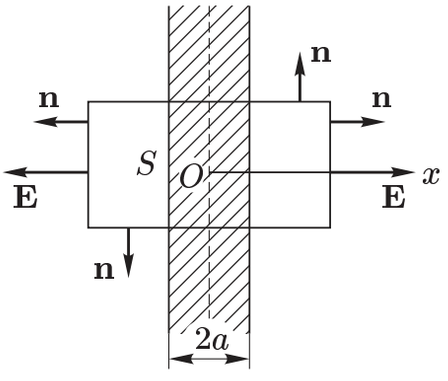


Рис. 19

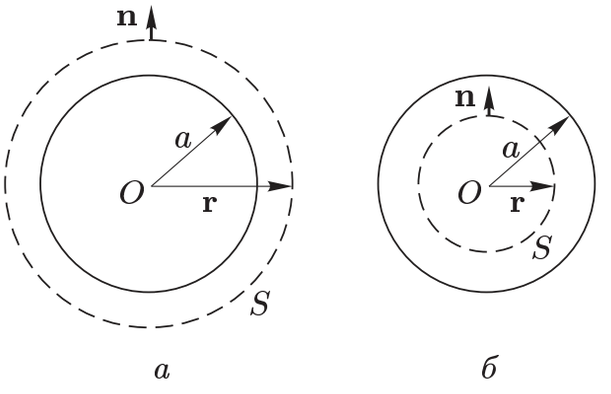


Рис. 20

Когда радиус шара пренебрежимо мал по сравнению с расстоянием r, мы получаем кулоново поле точечного заряда. Нельзя, однако, сказать, что закон Кулона — следствие теоремы Гаусса. Он получается из нее при дополнительном предположении, что поле неподвижного точечного заряда радиально и обладает шаровой симметрией.

Совершенно так же вычисляется поле внутри шара (рис. 20 *б*). Оно определяется выражением

$$E = \frac{q'}{r^2}, \tag{6.4}$$

где q' — заряд, ограниченный сферой радиуса r. Если заряд равномерно распределен по объему шара, то $q' = q(r/a)^3$. В этом случае

$$E = \frac{qr}{a^3} = \frac{4\pi}{3}\,\rho r. \tag{6.5}$$

Если же заряд равномерно распределен по поверхности шара, то $q' = 0$, а потому также $E = 0$. Таким образом, *электрическое поле внутри сферической полости, равномерно заряженной по поверхности, равно нулю.* Результат остается справедливым и для случая, когда внутри сферической полости зарядов нет, а внешние заряды распределены сферически симметрично.

4. Уясним последний результат непосредственно с помощью закона Кулона. Пусть поверхность сферы равномерно заряжена электричеством. Через произвольную точку A окружаемой ею полости проведем пучок лучей, вырезающий из сферы бесконечно малые площадки s_1

и s_2 (рис. 21). Проекции этих площадок s_1' и s_2' на плоскость, перпендикулярную к оси пучка, пропорциональны квадратам расстояний r_1 и r_2. То же справедливо для самих площадок s_1 и s_2 и находящихся на них зарядов q_1 и q_2. Действительно, если через ось пучка и центр сферы O провести плоскость (плоскость рисунка), то углы α_1 и α_2 будут равны между собой и, кроме того, $s_1' = s_1 \sin\alpha_1$, $s_2' = s_2 \sin\alpha_2$. Отсюда и следует наше утверждение. Из него получаем $q_1/r_1^2 = q_2/r_2^2$. Значит, кулоновы электрические поля, возбуждаемые в точке A зарядами q_1 и q_2, равны по модулю и противоположны по направлению. Это справедливо для каждой пары зарядов типа q_1 и q_2, на которые можно мысленно разбить всю поверхность заряженной сферы. Поэтому полное электрическое поле должно обращаться в нуль в каждой точке сферической полости.

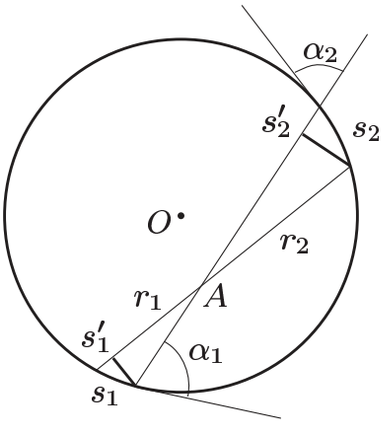


Рис. 21

Мы видим, что теорема об отсутствии электрического поля внутри сферической полости, равномерно заряженной по ее поверхности, является следствием закона Кулона — обратной пропорциональности напряженности поля квадрату расстояния. Если бы поле менялось с расстоянием иначе, то теорема была бы неверна. Поэтому, как заметил Дж. Пристли (1733–1804), экспериментальная проверка этой теоремы может служить подтверждением самого закона Кулона. Такая косвенная проверка может быть выполнена с гораздо большей точностью, чем прямое измерение сил взаимодействия точечных зарядов. Впервые она была произведена в 1774 г. Кавендишем, а затем с большей точностью повторена Максвеллом в 1879 г. Кавендиш и Максвелл представили результаты своих измерений в следующей форме. Если принять, что сила взаимодействия точечных зарядов $F \sim 1/r^n$, то n может отличаться от 2 по измерениям Кавендиша не более чем на $1/50$, а по измерениям Максвелла — не более чем на $1/20\,000$. Опыт с современными средствами измерений был повторен Плимптоном и Лаутоном в 1936 г. Они нашли, что n может отличаться от 2 не более чем на 10^{-9}. Конечно, зависимость типа $F \sim 1/r^n$ может и не соблюдаться. Реальная проблема состоит в нахождении расстояний, с которых начинает нарушаться закон Кулона, а также в установлении истинного закона сил, которым следует заменить закон Кулона на таких расстояниях. В настоящее время закон Кулона подтвержден до расстояний сантиметра или десятков сантиметров с относительной точностью порядка 10^{-9}.

Может показаться, что для достижения столь высокой точности с не меньшей точностью должна быть отшлифована и поверхность заряжаемой сферы, т. е. постоянство радиуса сферы должно быть выдержано с относительной точностью 10^{-9}. При радиусе сферы в 10 см это составляет 10^{-8} см — размер одного атома. С не меньшей точностью, казалось бы, должна оставаться постоянной и плотность электричества на поверхности сферы, а для этого необходимо, чтобы сфера абсолютно

не подвергалась электрическому влиянию со стороны окружающих тел. Ни то, ни другое сделать, конечно, невозможно. Однако заботиться об этом и не надо. Форма полости и наружной поверхности тела не играет роли. В § 11 и 22 будет показано, что поле в полости, окруженной проводящей оболочкой любой формы, строго равно нулю, если только внутри полости нет заряженных тел. Это было бы не так, если бы закон Кулона нарушался.

Есть две области, в которых априори можно ожидать нарушений закона Кулона. Это, во-первых, область расстояний, меньших 10^{-14} см, где нет уверенности в применимости электромагнитной теории вообще. Это, во-вторых, область очень больших расстояний, начиная с географических и больше. Здесь мы также не располагаем непосредственными экспериментальными подтверждениями закона Кулона. Однако если бы закон Кулона нарушался при больших расстояниях, то, согласно современной квантовой электродинамике, квант света (фотон) обладал бы отличной от нуля массой покоя. А это означало бы, что скорость электромагнитных волн в вакууме должна зависеть от длины волны. Отсутствие такой зависимости позволяет сделать вывод, что закон Кулона верен до расстояний, по крайней мере, в несколько километров с относительной точностью 10^{-6}.

5. Поле бесконечной прямой линии и бесконечно длинного цилиндра. Поле бесконечной прямой линии, равномерно заряженной электричеством, направлено радиально — к линии или от нее, в зависимости от знака заряда. На расстоянии r от линии это поле определяется формулой

$$E = 2\varkappa/r, \tag{6.6}$$

где $\varkappa$ — линейная плотность заряда, т. е. заряд, приходящийся на единицу длины линии. Той же формулой определяется поле бесконечно длинного кругового цилиндра, равномерно заряженного по объему или по поверхности, если точка наблюдения находится вне цилиндра. Если цилиндр полый и равномерно заряжен по поверхности, то поле внутри него равно нулю. Если же цилиндр равномерно заряжен по объему, то

$$E = 2\pi\rho r. \tag{6.7}$$

Эти результаты нетрудно получить из теоремы Гаусса.

6. Рассмотрим теперь какую-то поверхность S, заряженную электричеством (рис. 22). Полупространство по одну сторону этой поверхности обозначим индексом *1*, а по другую — индексом *2*. Поверхностная плотность электричества σ может меняться вдоль поверхности S произвольно. Возьмем бесконечно малый цилиндр, основания которого расположены по разные стороны от S. Высота цилиндра должна быть бесконечно мала по сравнению с линейными размерами его оснований. Если площадь основания ΔS, то внутри цилиндра находится электрический заряд $q = \sigma\Delta S$. Сумма потоков вектора $\mathbf{E}$ через основания цилиндра будет $(E_{n_1} + E_{n_2})\Delta S$, поток через боковую поверхность пренебрежимо мал. Приравнивая последнее выражение величине $4\pi\sigma\Delta S$,

получаем

$$E_{n_1} + E_{n_2} = 4\pi\sigma. \tag{6.8}$$

Здесь $\mathbf{n}_1$ означает внешнюю нормаль к поверхности S для полупространства *1*, а $\mathbf{n}_2$ — для полупространства *2*. Формуле (6.8) можно придать другой вид, проведя к поверхности S единую нормаль $\mathbf{n}$. Направим ее от полупространства *1* к полупространству *2*. Тогда

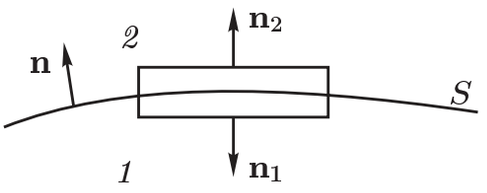


Рис. 22

$$E_{2n} - E_{1n} = 4\pi\sigma. \tag{6.9}$$

Таким образом, при переходе через заряженную поверхность нормальная составляющая вектора $\mathbf{E}$ претерпевает разрыв, равный $4\pi\sigma$.

Происхождение скачка нормальной составляющей вектора $\mathbf{E}$ полезно уяснить с другой точки зрения. Полное электрическое поле в любой точке пространства складывается из внутреннего поля $\mathbf{E}_{\text{внутр}}$, т. е. поля, создаваемого зарядами самой площадки ΔS, и внешнего поля $\mathbf{E}_{\text{внеш}}$, т. е. поля, возбуждаемого всеми остальными зарядами. При пересечении площадки ΔS внешнее поле меняется непрерывно. Сама же площадка на бесконечно близких расстояниях от нее ведет себя как бесконечная заряженная плоскость. Создаваемое ею поле $\mathbf{E}_{\text{внутр}}$ нормально к площадке и равно $2\pi\sigma$. Однако направления этого поля по разные стороны площадки противоположны. По одну сторону оно увеличивает, а по другую уменьшает нормальную составляющую полного поля, т. е.

$$\mathbf{E}_1 = \mathbf{E}_{\text{внеш}} + 2\pi\sigma\mathbf{n}_1, \quad \mathbf{E}_2 = \mathbf{E}_{\text{внеш}} + 2\pi\sigma\mathbf{n}_2. \tag{6.10}$$

Таким образом, скачок претерпевает только *внутреннее поле*, тогда как внешнее меняется непрерывно. А так как внутреннее поле не имеет тангенциальной составляющей, то тангенциальная составляющая полного поля меняется также непрерывно:

$$E_{1t} = E_{2t}. \tag{6.11}$$

7. Из формул (6.10) получаем также

$$\mathbf{E}_{\text{внеш}} = \frac{1}{2}(\mathbf{E}_1 + \mathbf{E}_2). \tag{6.12}$$

Пользуясь этим выражением, легко рассчитать электрические силы, действующие на заряженную поверхность. На площадку ΔS, очевидно, могут действовать только *внешние заряды*, а не заряды самой площадки. Электрическая сила, отнесенная к единице площади площадки ΔS, будет

$$\mathbf{f} = \sigma\mathbf{E}_{\text{внеш}} = \frac{\sigma}{2}(\mathbf{E}_1 + \mathbf{E}_2). \tag{6.13}$$

Исключив σ с помощью формулы (6.9), преобразуем этот результат к виду

$$\mathbf{f} = \frac{1}{8\pi}\,(E_{2n} - E_{1n})(\mathbf{E}_1 + \mathbf{E}_2). \tag{6.14}$$

В частности, если поле нормально к заряженной поверхности, то

$$\mathbf{f} = \frac{1}{8\pi}\,(E_2^2 - E_1^2)\mathbf{n}. \tag{6.15}$$

Отсюда видно, что на единицу площади заряженной поверхности действуют силы натяжения $E_1^2/(8\pi)$ и $E_2^2/(8\pi)$, которые тянут ее наружу в противоположных направлениях.

8. Поле двух параллельных разноименно и равномерно заряженных плоскостей. Если плотности зарядов на обеих плоскостях одинаковы по модулю, то будут одинаковы, но противоположно направлены и создаваемые ими поля. Между плоскостями направления полей совпадают, и при их сложении получается поле

$$E = 4\pi\sigma \tag{6.16}$$

(рис. 23). Во внешнем пространстве направления полей противоположны, и результирующее поле равно нулю. К тем же результатам можно прийти с помощью общего соотношения (6.9).

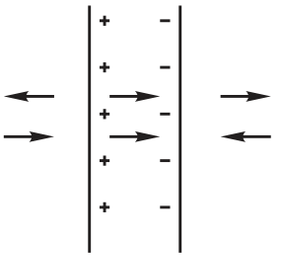


Рис. 23

ЗАДАЧИ

1. Две бесконечные плоскопараллельные металлические пластинки помещены в вакууме параллельно друг другу (рис. 24). Полный заряд на единицу площади (т. е. сумма зарядов на обеих поверхностях пластинки) равен q_1 для первой пластинки и q_2 для второй. Определить поверхностные плотности электрических зарядов на пластинках, а также напряженность электрического поля между ними и во внешнем пространстве.

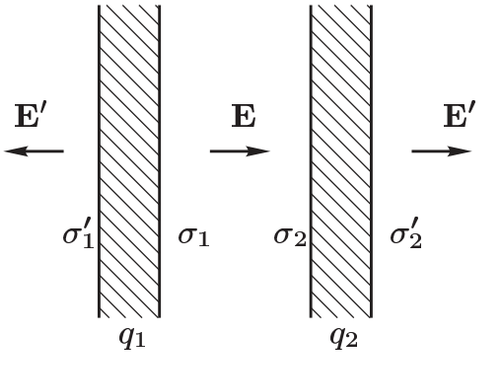


Рис. 24

Ответ. $\sigma_1 = -\sigma_2 = \dfrac{1}{2}\,(q_1 - q_2)$; $\sigma_1' = \sigma_2' = \dfrac{1}{2}\,(q_1 + q_2)$; $E = 2\pi(q_1 - q_2)$; $E' = 2\pi(q_1 + q_2)$.

2. Проводящая сфера радиуса R составлена из двух полусфер. Определить силу F, с которой отталкиваются эти полусферы, если полный заряд сферы равен Q.

Решение. Согласно формуле (6.15) на единицу поверхности сферы действует выталкивающая сила $\mathbf{f} = Q^2\mathbf{n}/(8\pi R^4)$. Отсюда интегрированием легко получить

$$F = Q^2/(8R^2).$$

3. Как изменится ответ в предыдущей задаче, если в центре сферы поместить дополнительно точечный заряд q? Сферу считать полой и бесконечно тонкой.

Ответ. $F = Q(Q + 2q)/(8R^2)$.

4. Длинный проводящий цилиндр радиуса R разрезан вдоль продольной оси. Определить силу отталкивания F, действующую на единицу длины каждого полуцилиндра, если на единицу длины цилиндра приходится заряд $\varkappa$.

Ответ. $F = \varkappa^2/(\pi R)$.

5. Как изменится ответ в предыдущей задаче, если вдоль оси цилиндра поместить дополнительно тонкую заряженную нить, на единицу длины которой приходится заряд $\varkappa_0$? Цилиндр считать полым, а его стенки — бесконечно тонкими.

Ответ. $F = \varkappa(\varkappa + 2\varkappa_0)/(\pi R)$.

§ 7. Дифференциальная форма электростатической теоремы Гаусса

1. Соотношение (5.5) выражает теорему Гаусса *в интегральной форме*. Придадим теперь этой теореме *дифференциальную форму*. Назовем *объемной плотностью электричества* ρ количество электричества, отнесенное к единице объема. Тогда заряд в элементе объема dV представится выражением $dq = \rho\, dV$. Будем предполагать, что величина ρ является непрерывной функцией пространственных координат. Представление о непрерывном распределении электричества в пространстве является такой же идеализацией, как и представление о непрерывном распределении вещества. Такими представлениями широко пользуются в макроскопической физике.

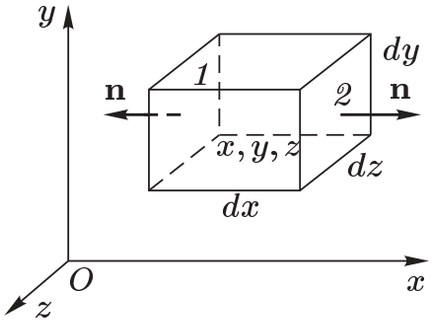


Рис. 25

Возьмем в пространстве бесконечно малый прямоугольный параллелепипед со сторонами dx, dy, dz, параллельными координатным осям прямоугольной системы координат (рис. 25). На грани *1* внешняя нормаль направлена в отрицательную сторону оси x. Поэтому поток вектора $\mathbf{E}$ через эту грань будет $-E_x(x)\, dy\, dz$. На противоположной грани *2*, наоборот, направление внешней нормали совпадает с положительным направлением оси x, и для потока через эту грань следует писать $+E_x(x + dx)\, dy\, dz$. Сумма обоих потоков будет

$$[E_x(x + dx) - E_x(x)]\, dy\, dz = \frac{\partial E_x}{\partial x}\, dx\, dy\, dz = \frac{\partial E_x}{\partial x}\, dV,$$

где $dV \equiv dx\, dy\, dz$ — объем параллелепипеда. Аналогично найдутся потоки через две пары остальных граней. Полный поток через всю поверхность параллелепипеда:

$$d\Phi = \operatorname{div} \mathbf{E}\, dV, \tag{7.1}$$

где введено обозначение

$$\operatorname{div}\mathbf{E} \equiv \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z}. \tag{7.2}$$

По теореме Гаусса тот же поток равен $4\pi q = 4\pi\rho\, dV$. Приравнивая оба выражения, получим

$$\operatorname{div}\mathbf{E} = 4\pi\rho. \tag{7.3}$$

Эта формула и выражает электростатическую теорему Гаусса в дифференциальной форме.

Величина, определяемая выражением (7.2), сохраняет смысл независимо от конкретной физической или геометрической природы вектора **E**. Она называется *дивергенцией* вектора **E**. С дивергенцией приходится встречаться в самых разнообразных вопросах математики и физики, чем и оправдывается введение этого математического понятия.

2. Теорема Гаусса в дифференциальной форме (7.3) является следствием той же теоремы в интегральной форме (5.5). Обращая порядок рассуждений, легко убедиться, что из дифференциальной формы теоремы Гаусса можно получить интегральную. Обе формы математически эквивалентны. Правда, дифференциальная форма имеет смысл лишь в том случае, когда электричество распределено в пространстве с конечной плотностью ρ. Если ρ обращается в бесконечность в отдельных точках, на линиях или поверхностях, то дифференциальная форма становится неприменимой [1)], тогда как интегральная форма (5.5) применима и в таких случаях. В этом смысле интегральная форма обладает большей математической общностью, чем дифференциальная. Однако разрывные распределения электричества с бесконечно большими значениями ρ являются математическими абстракциями и в физике должны рассматриваться как предельные случаи непрерывных распределений с всюду конечными значениями ρ. Если это иметь в виду, то можно утверждать, что *интегральная и дифференциальная формы теоремы Гаусса полностью эквивалентны.*

3. В электростатике теорема Гаусса является не более как одним из следствий закона Кулона. Но мы не можем ограничиться электростатикой. Наша задача значительно шире. Мы должны путем обобщения опытных фактов отыскать общие уравнения и законы, применимые не только в электростатике, но и во всей электродинамике. В качестве руководящего принципа при отыскании таких законов можно выставить требование, чтобы они были законами *теории поля*, исключающими мгновенное действие на расстоянии. Закон Кулона этому требованию не удовлетворяет. Он может быть справедлив только в электростатике. Однако следствия, выводимые из него, могут иметь и более широкую область применимости. К числу таких

[1)] С использованием так называемых *обобщенных функций* дифференциальную форму теоремы Гаусса можно распространить и на эти случаи.

следствий и относится теорема Гаусса. Она не противоречит теории поля с ее представлением о конечной скорости распространения взаимодействий. Записанная в дифференциальной форме, теорема Гаусса не содержит никаких намеков на дальнодействующий характер сил. Она является *локальной теоремой*, т. е. связывает различные физические величины (ρ и $\operatorname{div}\mathbf{E}$) в одной и той же точке пространства. Законы теории поля не обязательно должны быть локальными. Однако все локальные законы совместимы с основным представлением этой теории о передаче взаимодействий посредством полей. С другой стороны, закон Кулона только достаточен, но не необходим для доказательства теоремы Гаусса. Поэтому естественно ввести гипотезу, что *теорема Гаусса верна не только в электростатике, но и в электродинамике*, имеющей дело с переменными во времени электромагнитными полями. Верна эта гипотеза или нет — на этот вопрос может дать ответ только опыт. Вся совокупность опытных фактов говорит в пользу этой гипотезы. Поэтому она и была принята в физике. Тем самым уравнение (7.3) и математически эквивалентное ему уравнение (5.5) перестают быть скромными следствиями закона Кулона, а возводятся в ранг *основных постулатов* теории электричества. Они входят в систему основных уравнений Максвелла.

Теорема Гаусса в интегральной форме (5.5) устанавливает связь между физическими величинами в сколь угодно удаленных точках пространства в один и тот же момент времени. Поэтому может показаться, что ее справедливость связана с предположением о мгновенном действии на расстоянии. Возможность представления теоремы Гаусса в дифференциальной форме показывает, что это не так. Формула (5.5) утверждает только, что с электрическим зарядом q всегда связано какое-то электрическое поле. Поле неограниченно долго покоящегося заряда кулоново на любых расстояниях. Но если заряд движется, то это уже не так. Например, ускоренно движущийся заряд излучает электромагнитные волны. Однако поток вектора $\mathbf{E}$ через любую замкнутую поверхность, окружающую одни и те же заряды, не зависит от формы и положения этой поверхности, а также от характера движения зарядов.

ЗАДАЧА

Получить формулы (6.1)–(6.3), (6.5)–(6.7), пользуясь теоремой Гаусса в дифференциальной форме.

Р е ш е н и е. В качестве примера получим формулу (6.5). Ввиду шаровой симметрии

$$\mathbf{E} = E(r)\frac{\mathbf{r}}{r},$$

или в координатной форме

$$E_x = E(r)\frac{x}{r}, \quad E_y = E(r)\frac{y}{r}, \quad E_z = E(r)\frac{z}{r}.$$

Дифференцируя E_x и учитывая, что $\partial r/\partial x = x/r$ (последнее получается дифференцированием равенства $r^2 = x^2 + y^2 + z^2$), находим

$$\frac{\partial E_x}{\partial x} = \frac{dE}{dr}\,\frac{x^2}{r^2} - \frac{E}{r^3}\,x^2 + \frac{E}{r}.$$

Написав аналогичные соотношения для производных $\partial E_y/\partial y$ и $\partial E_z/\partial z$ и сложив их, получим

$$\operatorname{div}\mathbf{E} = \frac{dE}{dr} + \frac{2E}{r} = \frac{1}{r^2}\,\frac{d}{dr}\,(r^2 E). \tag{7.4}$$

Внутри шара

$$\frac{1}{r^2}\,\frac{d}{dr}\,(r^2 E) = 4\pi\rho,$$

откуда

$$E = \frac{4\pi}{3}\,\rho r + \frac{C}{r^2}.$$

Постоянная C должна равняться нулю, так как напряженность поля E в центре шара конечна, как это ясно из физических соображений. Аналогично найдем выражение для напряженности поля вне шара.

Обратим внимание на формулу (7.4). Она дает выражение для дивергенции любого вектора, когда этот вектор направлен радиально и зависит только от расстояния до начала координат (сферическая симметрия).

§ 8. Математическое дополнение. Формула Гаусса–Остроградского

1. В различных вопросах теории электричества и других разделов теоретической физики часто применяется математическая формула, с помощью которой поток вектора через произвольную замкнутую поверхность выражается через объемный интеграл по области, ограниченной этой поверхностью. Приведем вывод этой формулы, хотя в большинстве случаев мы и будем избегать пользоваться ею. Пусть $f(x, y, z)$ — некоторая функция, а S — замкнутая поверхность, ограничивающая объем V (рис. 26). На отрезке *12*, параллельном оси x, f является функцией одного аргумента x. Интегрируя вдоль этого отрезка, получим

$$\int\limits_{12} \frac{\partial f}{\partial x}\,dx = f_2 - f_1,$$

где f_1 и f_2 — значения функции f на концах рассматриваемого отрезка. Построим теперь бесконечно узкий цилиндр, одной из образующих которого является отрезок *12*. Пусть $d\sigma$ — площадь его поперечного сечения (величина существенно положительная). Умножим предыдущее соотношение на $d\sigma$. Так как $d\sigma\,dx$ есть элементарный объем dV, заштрихованный

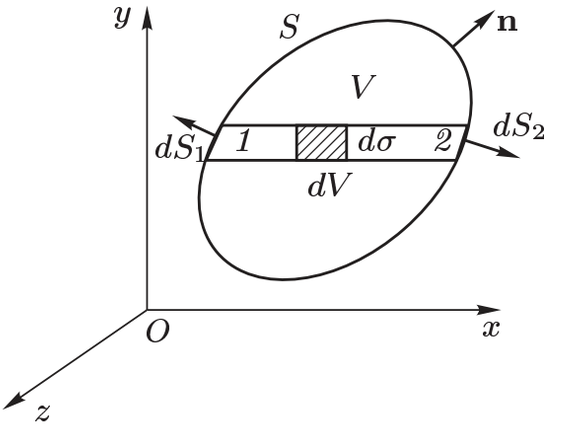


Рис. 26

на рисунке, то в результате получится

$$\int\limits_{\Delta V} \frac{\partial f}{\partial x}\, dV = d\sigma(f_2 - f_1),$$

где ΔV — часть объема V, вырезаемого из него поверхностью цилиндра. Пусть dS_1 и dS_2 — элементарные площадки, вырезаемые тем же цилиндром на поверхности S, а $\mathbf{n}_1$ и $\mathbf{n}_2$ — единичные нормали к ним, проведенные наружу от поверхности S. Тогда $d\sigma = dS_2 \cdot n_{2x} = -dS_1 \cdot n_{1x}$, а потому

$$\int\limits_{\Delta V} \frac{\partial f}{\partial x}\, dV = f_1 n_{1x}\, dS_1 + f_2 n_{2x}\, dS_2,$$

или, короче,

$$\int\limits_{\Delta V} \frac{\partial f}{\partial x}\, dV = \int\limits_{dS_1 + dS_2} f n_x\, dS,$$

где поверхностный интеграл распространен по сумме площадок dS_1 и dS_2. Весь объем V можно разделить на элементарные цилиндры рассматриваемого вида и написать для каждого из них такие же соотношения. Суммируя эти соотношения, получим

$$\int\limits_{V} \frac{\partial f}{\partial x}\, dV = \oint\limits_{S} f n_x\, dS. \tag{8.1}$$

Интеграл слева распространен по всему объему V, справа — по поверхности S, ограничивающей этот объем. Аналогичные соотношения можно написать для осей y и z.

Возьмем теперь произвольный вектор $\mathbf{A}$ и применим к его компонентам соотношение (8.1). Получим

$$\int\limits_{V} \frac{\partial A_x}{\partial x}\, dV = \oint\limits_{S} A_x n_x\, dS,$$

и аналогично для компонент A_y и A_z. Складывая эти соотношения, найдем

$$\int\limits_{V} \left(\frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} \right) dV = \oint\limits_{S} (A_x n_x + A_y n_y + A_z n_z)\, dS,$$

или

$$\int\limits_{V} \operatorname{div} \mathbf{A}\, dV = \oint\limits_{S} (\mathbf{An})\, dS. \tag{8.2}$$

Это и есть *формула Гаусса–Остроградского*. Ее можно также записать в виде

$$\int\limits_{V} \operatorname{div} \mathbf{A}\, dV = \oint\limits_{S} (\mathbf{A}\, d\mathbf{S}). \tag{8.3}$$

2. Если объем V бесконечно мал, то величина $\operatorname{div} \mathbf{A}$ внутри него может считаться постоянной. Вынося ее за знак интеграла и переходя к пределу $V \to 0$, получим

$$\operatorname{div} \mathbf{A} = \lim_{V \to 0} \frac{1}{V} \oint\limits_{S} (\mathbf{A}\, d\mathbf{S}). \tag{8.4}$$

Предельный переход надо понимать в том смысле, что область V должна стягиваться в точку, т. е. размеры этой области должны беспредельно уменьшаться по всем направлениям. Наши рассуждения показывают, что величина, стоящая в правой части (8.4), не зависит от формы поверхности S, стягиваемой в точку. Поэтому выражение (8.4) можно было бы принять за исходное определение дивергенции, как это часто и делается. Такое определение обладает тем преимуществом, что оно *инвариантно*, т. е. никак не связано с выбором системы координат.

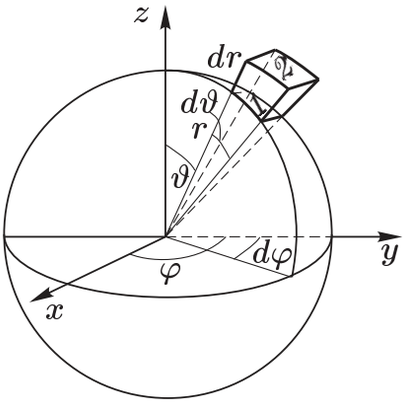


Рис. 27

На формуле (8.4) основан наиболее простой и общий способ вычисления дивергенции в различных системах координат. Вычислим, например, $\operatorname{div}\mathbf{A}$ в сферической системе координат r, ϑ, φ (рис. 27). Рассмотрим бесконечно малый «кубик», ограниченный плоскостями $r = \text{const}$, $\vartheta = \text{const}$, $\varphi = \text{const}$. Сумма потоков вектора $\mathbf{A}$ через противоположные грани кубика *1* и *2* будет

$$d_1\Phi = -A_r(r)dS_1 + A_r(r+dr)dS_2 = \frac{\partial}{\partial r}\left(A_r dS\right)dr.$$

Подставляя сюда $dS = r^2 \sin\vartheta\, d\vartheta\, d\varphi$, получим

$$d_1\Phi = \frac{\partial}{\partial r}\left(r^2 A_r\right)\sin\vartheta\, d\vartheta\, d\varphi\, dr = \frac{1}{r^2}\frac{\partial}{\partial\vartheta}\left(r^2 A_r\right)dV,$$

где $dV = r^2 \sin\vartheta\, dr\, d\vartheta\, d\varphi$ — объем рассматриваемого «кубика». Аналогично находятся потоки через остальные две пары противоположных граней, а затем и выражение для дивергенции:

$$\operatorname{div}\mathbf{A} = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 A_r\right) + \frac{1}{r\sin\vartheta}\frac{\partial}{\partial\vartheta}\left(\sin\vartheta A_\vartheta\right) + \frac{1}{r\sin\vartheta}\frac{\partial A_\varphi}{\partial\varphi}. \tag{8.5}$$

Если вектор $\mathbf{A}$ направлен по $\mathbf{r}$ и зависит только от расстояния r, то это выражение переходит в (7.4).

Приведем другой пример на применение формулы Гаусса–Остроградского. Согласно (5.5) $\oint \mathbf{E}\, d\mathbf{S} = 4\pi q$, или для бесконечно малого объема $\oint \mathbf{E}\, d\mathbf{S} = 4\pi\rho V$. Подставив это значение в (8.4), получим $\operatorname{div}\mathbf{E} = 4\pi\rho$, т. е. теорему Гаусса в дифференциальной форме.

ЗАДАЧА

Электрическое поле в электростатике всегда перпендикулярно к поверхности проводника (см. § 11, п. 4). Пользуясь этим, доказать, что вблизи искривленной поверхности заряженного проводника электрическое поле удовлетворяет соотношению

$$\frac{\partial E}{\partial n} = -E\left(\frac{1}{R_1} + \frac{1}{R_2}\right), \tag{8.6}$$

где производная берется по направлению внешней нормали к поверхности проводника, а R_1 и R_2 — главные радиусы кривизны этой поверхности (они считаются положительными для выпуклых и отрицательными для вогнутых сечений поверхности).

§ 9. Теорема Ирншоу

1. Для равновесия системы точечных электрических зарядов необходимо и достаточно, чтобы сила, действующая на каждый заряд, обращалась в нуль. Примером, где это условие соблюдается, может служить система двух одинаковых точечных зарядов q и q, посередине между которыми помещен заряд противоположного знака $-q/4$ (рис. 28). Другие примеры приводятся в задаче к этому параграфу. На вопрос об устойчивости такого равновесия дает ответ *теорема Ирншоу.* Согласно этой теореме *всякая равновесная конфигурация покоящихся точечных электрических зарядов неустойчива, если на них, кроме кулоновских сил притяжения и отталкивания, никакие другие силы не действуют.*

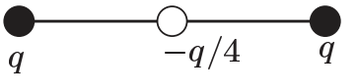


Рис. 28

Теорема Ирншоу является следствием теоремы Гаусса. Допустим, что какая-то система неподвижных точечных зарядов находится в устойчивом равновесии. Рассмотрим произвольный заряд q этой системы, находящейся в равновесии в положении A (рис. 29). Предположим для определенности, что заряд q положителен (в случае отрицательного заряда доказательство аналогично). Если заряд q сместится в бесконечно близкую точку A', то ввиду предположенной устойчивости равновесия должна возникнуть сила, направленная к точке A и стремящаяся вернуть заряд снова в ту же точку. Пусть $\mathbf{E}$ — электрическое поле, создаваемое всеми зарядами, за исключением заряда q. В точке A' оно должно быть направлено к A, каково бы ни было направление смещения AA'. Окружим заряд q произвольной замкнутой поверхностью S и притом такой, чтобы все прочие заряды были расположены во внешнем пространстве по отношению к этой поверхности. На поверхности S поле $\mathbf{E}$ направлено к точке A, а потому поток вектора $\mathbf{E}$ через поверхность S отрицателен. Но это противоречит теореме Гаусса. Последняя требует, чтобы указанный поток был равен нулю, поскольку он создается зарядами, расположенными вне S. Получившееся противоречие и доказывает теорему Ирншоу.

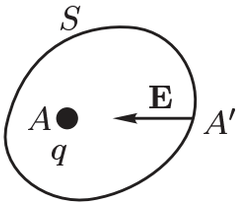


Рис. 29

Если помимо электрических сил в системе действуют какие-либо другие силы, то равновесие может оказаться устойчивым. Возьмем, например, три одинаковых и одинаково заряженных шарика. Два из них закрепим на концах цилиндрической трубки из изолятора, а третий поместим посередине между ними (рис. 30). Пусть средний шарик может скользить внутри трубки без трения. Тогда в среднем положении он будет находиться в устойчивом равновесии.

2. Теорема Гаусса справедлива и для гравитационных полей. Однако в этом случае все силы являются *силами притяжения.* Равновесных

конфигураций материальных точек, притягивающихся по закону всемирного тяготения Ньютона, не существует, и теорема Ирншоу лишена содержания. Это очевидно для системы из двух материальных точек. Если в некоторый момент времени обе точки находятся в покое, то силы притяжения между ними приведут их в движение, пока обе точки не сольются в одном месте. Рассмотрим теперь произвольное число материальных точек (рис. 31). В любой момент времени в системе можно найти такую крайнюю точку A, что все прочие материальные

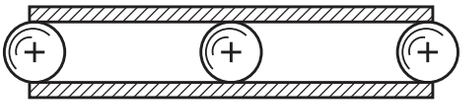


Рис. 30

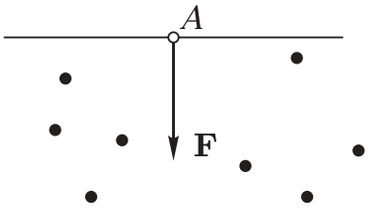


Рис. 31

точки окажутся расположенными по одну сторону какой-то плоскости, проходящей через A. Сила $\mathbf{F}$, действующая на A, будет направлена в сторону пространства, в котором расположены эти материальные точки. Если точка A в рассматриваемый момент времени находится в покое, то под действием силы $\mathbf{F}$ она придет в движение, т. е. равновесие будет нарушено. Возможность равновесных конфигураций неподвижных электрических зарядов связана с тем, что имеются заряды обоих знаков, встречаются как силы притяжения, так и силы отталкивания.

3. Атомы химических элементов состоят из положительно заряженных атомных ядер и окружающих их отрицательно заряженных электронов. Размеры атомов порядка 10^{-8} см. Размеры атомных ядер и электронов примерно в 10^5 раз меньше, т. е. пренебрежимо малы по сравнению с размерами атомов. Таким образом, атом с большой точностью может рассматриваться как система, состоящая из точечных электрически заряженных частиц. К такой системе применима теорема Ирншоу. Атомы, несомненно, являются устойчивыми системами. Поэтому электроны и атомные ядра, из которых они состоят, не могут находиться в покое. Для объяснения *устойчивости* атомных систем в классической физике была введена *планетарная модель атома.* По этой модели электроны вращаются вокруг ядра подобно планетам, вращающимся вокруг Солнца. Однако планетарная модель атома оказалась также неустойчивой. Электрон, вращающийся вокруг ядра, движется ускоренно. А по законам классической электродинамики *ускоренно движущийся заряд излучает электромагнитные волны.* Непрерывно растрачивая энергию на излучение, вращающийся электрон в конце концов должен был бы упасть на ядро. Классическая физика оказалась неспособной объяснить устойчивость атома. Объяснение было дано только квантовой механикой.

ЗАДАЧА

Одинаковые (по величине и по знаку) точечные заряды помещены в вершинах правильного 1) треугольника; 2) четырехугольника; 3) шестиугольника. Какой заряд q противоположного знака надо поместить в центре системы, чтобы она находилась в равновесии?

Ответ. 1) $Q = -q/\sqrt{3}$; 2) $Q = -(1/2)(\sqrt{2} + 1/2)q$; 3) $Q = -(5/4 + 1/\sqrt{3})q$.

§ 10. Электрическое поле в веществе

1. Размеры атомных ядер и электронов примерно в сто тысяч раз меньше размеров самих атомов. На долю заряженных частиц приходится ничтожная (примерно 10^{-15}) часть занимаемого телом пространства. Весь остальной объем тела составляет вакуум. Атомные ядра и электроны возбуждают в нем электромагнитные поля. Поле в промежутках между атомами и электронами, а также внутри этих частиц необычайно сложно меняется в пространстве и во времени. Такое поле называется *микроскопическим*, или, короче, *микрополем*. Столь же сложно меняется плотность распределения электричества. Она очень велика внутри атомных ядер и электронов и обращается в нуль в промежутках между ними. Такая плотность также называется *микроскопической*, или *микроплотностью*. Микроскопические величины обозначаются посредством $K_{\text{микро}}$, $\rho_{\text{микро}}$ и т. п. Их нельзя измерить путем внесения в вещество пробного заряда. Наименьшим зарядом является элементарный заряд e (заряд электрона). А такой заряд существенно исказил бы микрополе и распределение электричества в атомной системе. Таким образом, введение $\mathbf{E}_{\text{микро}}$, $\rho_{\text{микро}}$ и прочих микроскопических величин встречает определенную трудность принципиального порядка. Можно поставить под сомнение принципиальную возможность самого описания поля с помощью микроскопических величин типа $\mathbf{E}_{\text{микро}}$, $\rho_{\text{микро}}$ и т. п. Тем не менее классическая физика допускает такую возможность. Х.А. Лоренц показал, как, исходя из представления о микрополе, можно прийти к уравнениям для описания *макроскопических процессов* в телах. Такой подход к макроскопическим уравнениям электродинамики принят и в настоящем руководстве. Разумеется, справедливость макроскопических уравнений электродинамики еще не означает, что полностью верна и та микрокартина, которая была положена в основу вывода этих уравнений.

2. Задание микроскопических величин в каждой точке пространства и в каждый момент времени дало бы наиболее детальное классическое описание поля. Однако практически (а может быть, и принципиально) оно неосуществимо. Для многих целей достаточно более простое и несравненно более грубое описание, которым пользуется макроскопическая электродинамика. Она отвлекается от атомистического строения электричества и связанных с ним мелкомасштабных изменений поля, происходящих на ядерных и атомных расстояниях. Она принимает во внимание только изменения поля на макроскопических

расстояниях. Она оперирует со сглаженными полями и распределениями электричества, плавно меняющимися в пространстве и во времени. Такие поля называются *средними*, или *макроскопическими полями* (короче, *макрополями*). Напряженность электрического макрополя будем обозначать посредством $\mathbf{E}_{\text{микро}}$, или, короче, $\mathbf{E}$.

Описание поля в веществе посредством величин типа $\mathbf{E}_{\text{микро}}$, $\rho_{\text{микро}}$ и т. д. аналогично детальному механическому описанию движения вещества, в котором указывается положение и скорость каждой молекулы и составляющих ее частиц в любой момент времени. Описание с помощью макроскопических величин, напротив, аналогично гидродинамическому рассмотрению движения жидкости как сплошной среды. При таком рассмотрении распределение вещества в пространстве характеризуется его объемной плотностью, а движение — скоростью гидродинамического потока $\mathbf{v}$ как непрерывными функциями времени и пространственных координат. Молекулярные силы учитываются также суммарно — посредством внутренних давлений и касательных напряжений, возникающих при движении жидкости.

Дадим теперь более точное количественное определение макроскопического поля $\mathbf{E}$. Под $\mathbf{E}$ мы будем понимать микрополе $\mathbf{E}_{\text{микро}}$, усредненное по физически бесконечно малым объемам пространства. Чтобы вычислить макроскопическое поле $\mathbf{E}$ в какой-либо точке пространства, надо взять физически бесконечно малый объем V, внутри которого находится эта точка, проинтегрировать вектор $\mathbf{E}_{\text{микро}}$ по этому объему и значение интеграла разделить на объем V:

$$\mathbf{E} = \frac{1}{V} \int\limits_V \mathbf{E}_{\text{микро}}\, dV. \tag{10.1}$$

Так же определяется макроскопическая плотность $\overline{\rho} = \langle \rho_{\text{микро}} \rangle$ и любая другая макроскопическая величина. Результат вычисления практически не должен зависеть от величины и формы объема V. Для этого необходимо, чтобы внутри объема V содержалось еще очень много атомов. В то же время объем V должен быть настолько малым, чтобы с ним, а также с любыми линейными размерами его можно было обращаться, как с математическими дифференциалами. Объемы V, удовлетворяющие обоим этим условиям, и называются *физически бесконечно малыми.* Усреднение по таким объемам в смысле операции (10.1) сглаживает все нерегулярные и быстро меняющиеся вариации микрополя на расстояниях порядка атомных, но сохраняет плавные изменения его на макроскопических расстояниях.

3. Уравнения макроскопического поля могут быть получены из уравнений для микроскопического поля. Если те и другие уравнения представить в дифференциальной форме, то возникает вопрос: как связаны между собой производные обоих полей? На этот вопрос отвечает математическая формула

$$\frac{\partial \langle A \rangle}{\partial x} = \left\langle \frac{\partial A}{\partial x} \right\rangle. \tag{10.2}$$

Она утверждает, что усреднение и дифференцирование по координате можно переставлять местами. То же справедливо и для дифференцирования по времени. Для доказательства запишем формулу типа (10.1) более подробно:

$$\langle A(\mathbf{r})\rangle = \frac{1}{V}\int A(\mathbf{r}')\,dV'. \tag{10.3}$$

В соседней точке

$$\langle A(\mathbf{r}+\Delta\mathbf{r})\rangle = \frac{1}{V}\int A(\mathbf{r}'+\Delta\mathbf{r}')\,dV'.$$

Поскольку результат усреднения не зависит от величины и формы области интегрирования, последнюю мы выбрали одинаковой в обоих случаях. Элементы объема также можно выбрать одинаковыми. Тогда

$$\langle A(\mathbf{r}+\Delta\mathbf{r})\rangle - \langle A(\mathbf{r})\rangle = \frac{1}{V}\int\limits_V [A(\mathbf{r}'+\Delta\mathbf{r}) - A(\mathbf{r}')]\,dV'.$$

Выберем теперь вектор $\Delta\mathbf{r}$ так, чтобы он был параллелен оси X, т. е. положим $\Delta\mathbf{r} = \mathbf{i}\Delta x$. Разделив последнее соотношение на Δx и перейдя к пределу $\Delta x \to 0$, получим

$$\frac{\partial\langle A\rangle}{\partial x} = \frac{1}{V}\int\limits_V \frac{\partial A(\mathbf{r}')}{\partial x'}\,dV',$$

а это и есть формула (10.2). Аналогично, дифференцируя выражение (10.3) по времени как параметру, найдем

$$\frac{\partial\langle A\rangle}{\partial t} = \left\langle \frac{\partial A}{\partial t}\right\rangle. \tag{10.4}$$

Примем теперь, что для микроскопического поля справедлива теорема Гаусса в виде

$$\operatorname{div}\mathbf{E}_{\text{микро}} = 4\pi\rho_{\text{микро}}. \tag{10.5}$$

Усредняя это соотношение и принимая во внимание, что $\langle\operatorname{div}\mathbf{E}_{\text{микро}}\rangle = \operatorname{div}\langle\mathbf{E}_{\text{микро}}\rangle \equiv \operatorname{div}\mathbf{E}$, получим

$$\operatorname{div}\mathbf{E} = 4\pi\overline{\rho}, \tag{10.6}$$

где $\overline{\rho}$ — средняя (макроскопическая) плотность электричества в веществе, т. е. $\overline{\rho} = \langle\rho_{\text{микро}}\rangle$. Это — *дифференциальная форма теоремы Гаусса в веществе.* Она справедлива не только в электростатике, но и во всей макроскопической электродинамике.

4. При рассмотрении электрических явлений в веществе очень важно иметь правильное представление о порядке величин сил, действующих между протонами и электронами. Эти силы очень велики по сравнению с гравитационными силами притяжения между теми же частицами. Вычислим, например, отношение силы электрического

отталкивания двух протонов F_e к силе их гравитационного притяжения F_g. Заряд протона $e = 4{,}8 \cdot 10^{-10}$ СГСЭ-ед. заряда, масса $m_\text{p} = 1{,}67 \cdot 10^{-24}$ г. Используя эти данные, найдем

$$\frac{F_e}{F_g} = \frac{e^2}{Gm_\text{p}^2} = 1{,}24 \cdot 10^{36},$$

где $G = 6{,}67 \cdot 10^{-8}$ дин $\cdot$ см^2/г^2 — гравитационная постоянная. Значение этого отношения не зависит от расстояния между взаимодействующими частицами, так как обе силы F_e и F_g меняются обратно пропорционально квадрату этого расстояния. Масса электрона $m_\text{e} = 9{,}11 \cdot 10^{-28}$ г, т. е. в 1836 раз меньше массы протона. Поэтому для электронов отношение рассматриваемых сил в 1836^2 раз больше и составляет $4{,}17 \cdot 10^{42}$.

Что было бы, если бы Земля потеряла все свои электроны? Число протонов в Земле практически равно числу нейтронов. Поэтому каждый протон отталкивался бы от Земли с силой, превышающей его собственный вес приблизительно в $(1{,}24/2) \cdot 10^{36} = 0{,}62 \cdot 10^{36}$ раз. Эта сила равна весу груза в миллион тонн! Столь чудовищно большие силы не проявляются только потому, что в обычных условиях тела *электрически нейтральны.* Положительные заряды атомных ядер почти полностью скомпенсированы отрицательными зарядами электронов. При электризации тел нарушения такой компенсации ничтожны. Допустим, например, что шарику с радиусом $a = 1$ см сообщен заряд $q = 100$ СГСЭ-ед. $= (1/3) \cdot 10^{-7}$ Кл. Это довольно большой заряд. (Напряженность поля у поверхности шарика составит $E = q/a^2 = 100$ СГСЭ-ед. $= 30\,000$ В/см. А при разности потенциалов 30 000 В между металлическими шариками такого размера в сухом воздухе проскакивает электрическая искра, если расстояние между ними 1 см и шарики находятся в воздухе при нормальных давлении и температуре.) Если заряд q положителен, то для такой электризации от шарика надо отнять $n_\text{e} = q/e \approx 2 \cdot 10^{11}$ электронов. Пусть масса шарика $M = 30$ г. Тогда сумма содержащихся в нем протонов и нейтронов будет

$$N_\text{p} + N_\text{n} = \frac{M}{m_\text{p}} = \frac{30}{1{,}67 \cdot 10^{-24}} \approx 2 \cdot 10^{25}.$$

Превышение протонов над электронами ничтожно и составляет всего $n_\text{e}/n_\text{p} \sim 10^{-14}$ от общего числа протонов. Уменьшение массы шарика из-за электризации составляет $n_\text{e} m_\text{e} \approx 2 \cdot 10^{-16}$ г, т. е. примерно 10^{-17} массы самого шарика. Такое уменьшение массы не может быть обнаружено даже на самых чувствительных весах. Допустим теперь, что при электризации электроны были удалены от поверхностного слоя шарика. Оценим его толщину δ. Общее число протонов и нейтронов в слое будет n_e, а масса слоя $\Delta M \sim n_\text{e} m_\text{p} \sim 10^{-13}$ г. Так как $\Delta M/M = 3\delta/a$, то получаем $\delta \sim 10^{-15}$ см.

5. При внесении тела в электрическое поле легкие электроны испытывают смещения против поля. Смещения атомных ядер по сравнению с ними пренебрежимо малы. Происходит частичное разделение положительных и отрицательных зарядов. В отдельных местах тела появляются макроскопические заряды различных знаков. Это явление называется электрическим смещением, а появившиеся в результате разделения заряды — *индукционными зарядами*. К возникновению индукционных зарядов и сводится влияние вещества на электрическое поле. Индукционные заряды создают дополнительное электрическое поле, накладывающееся на поле первичных зарядов. Если известны все первичные и индукционные заряды, то при вычислении полного электрического поля можно «забыть» о наличии вещества. Полное поле найдется суперпозицией кулоновых полей, возбуждаемых в вакууме всеми первичными и индукционными зарядами.

§ 11. Проводники в электрическом поле

1. Смещения электрических зарядов в металлах и изоляторах носят совершенно различный характер. В металлах имеются *свободные электроны*, которые в пределах тела могут перемещаться на какие угодно расстояния. Поэтому индукционные заряды, возникающие в электрическом поле на противоположных концах тела, могут быть механически отделены друг от друга. Возьмем два металлических цилиндра A и B, установленных на изолирующих подставках и соединенных с электроскопами (рис. 32). Сблизим их до соприкосновения. Если к ним поднести заряженный шар C, то стрелки обоих электроскопов отклонятся. При удалении шара C отклонение пропадает. Раздвинем теперь цилиндры A и B в присутствии влияющего тела C, а затем тело C удалим. Электрические заряды на A и B, а также на стержнях и стрелках электроскопов сохранятся. Если шар C был заряжен положительно, то на A окажется отрицательный, а на B — положительный заряды. В этом можно убедиться, взяв потертую о кожу стеклянную палочку. Если этой палочкой коснуться цилиндра A, то отклонение стрелки электроскопа A уменьшится. Если же коснуться цилиндра B, то стрелка электроскопа B отклонится еще больше.

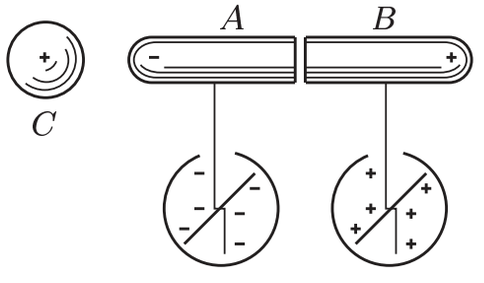


Рис. 32

2. Если бы внутри однородного проводника существовало макроскопическое электрическое поле, то оно привело бы в движение свободные электроны. В проводнике возник бы электрический ток, и равновесие электричества было бы невозможно. Для равновесия необходимо, чтобы макроскопическое поле $\mathbf{E}$ обращалось в нуль во всех точках

внутри проводника [1]. Будет равна нулю также дивергенция вектора $\mathbf{E}$, а с ней, в силу теоремы Гаусса (10.6), и величина $\overline{\rho}$. Таким образом, *при равновесии объемная плотность электричества внутри однородного проводника равна нулю. Электричество может располагаться только на поверхности, а не внутри проводника.* Как было выяснено в предыдущем параграфе (п. 4), толщина поверхностного слоя, в котором нарушается электрическая нейтральность вещества, настолько мала, что в макроскопической электростатике ее можно совсем не принимать во внимание. Можно пользоваться математической идеализацией, в которой электричество ведет себя как сплошная жидкость, располагающаяся на границе проводника как на геометрической поверхности.

Электрические заряды располагаются по поверхности проводника потому, что между ними действуют кулоновские силы притяжения и отталкивания. Допустим, что внутри проводника возникли электрические заряды. Согласно теореме Ирншоу никакая статическая конфигурация их внутри проводника не может быть устойчивой. Притяжение между разноименными зарядами приведет к их сближению и нейтрализации, а отталкивание одноименных зарядов — к тому, что они разойдутся как можно дальше и сосредоточатся на поверхности тела. Приведенное рассуждение показывает также, что поверхностная плотность электричества будет максимальна на наиболее удаленных выступающих частях проводника, обладающих максимальной кривизной, например на остриях. Это видно также из формулы (8.6), согласно которой поле $\mathbf{E}$ должно очень резко меняться вблизи острия.

3. Распределение электричества на поверхности проводника можно исследовать с помощью *пробного шарика*, т. е. маленького металлического шарика, насаженного на изолирующую ручку. Возьмем металлическое тело, форма которого изображена на рис. 33, установленное на изолирующей подставке. Зарядив его, коснемся пробным шариком острия A, а затем поднесем шарик к электроскопу. Стрелка электроскопа отклонится. Если то же самое повторить, касаясь пробным шариком боковой поверхности тела, то отклонение стрелки будет меньше. Если же коснуться впадины B, то стрелка совсем не отклонится. Это показывает, что плотность электричества максимальна у A и минимальна у B. Возьмем гибкую металлическую сетку, к которой по обе стороны приклеены легкие бумажные листочки (рис. 34). Установим ее на изолирующих подставках, а затем зарядим электричеством. Если сетка плоская, то листочки расходятся одинаково с обеих сторон. При изгибании сетки листочки с выпуклой стороны расходятся еще больше, а с вогнутой — спадают.

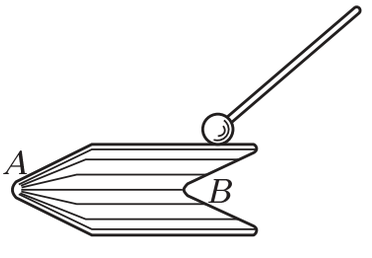


Рис. 33

[1] В неоднородном проводнике это не обязательно из-за действия сил неэлектростатического происхождения.

На остриях заряженного проводника поверхностная плотность электричества может быть настолько большой, что электричество начинает с них стекать. В сильном и сильно неоднородном поле вблизи острия молекулы воздуха заряжаются через влияние и притягиваются к нему. Коснувшись острия, молекулы заряжаются одноименно с ним и отталкиваются. Сила отталкивания превосходит ранее действовавшую силу притяжения, так как она действует на заряженные молекулы, а сила притяжения — на нейтральные. По этой причине заряженные молекулы будут удаляться от острия с бóльшими скоростями, чем они приближались к нему. Возникает поток заряженных частиц воздуха, направленный от острия, называемый *электрическим ветром.* Им можно погасить зажженную свечу.

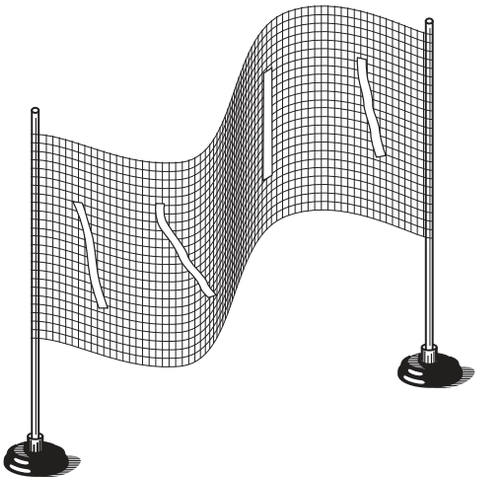


Рис. 34

В демонстрационном приборе — *колесе Франклина* (1706–1790) — электрический ветер, стекающий с остриев, приводит во вращение легкий крест из металлических проволок (рис. 35). Прибор действует по тому же принципу, что и сегнерово колесо. В другом демонстрационном опыте электрический ветер с остриев (рис. 36) приводит во вращение

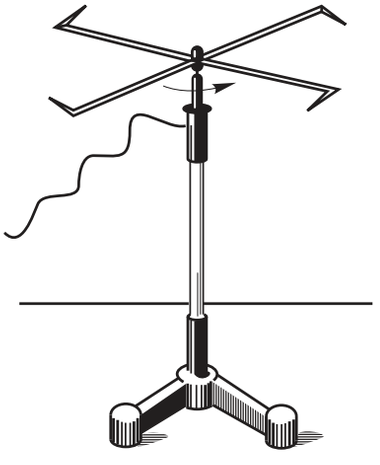


Рис. 35

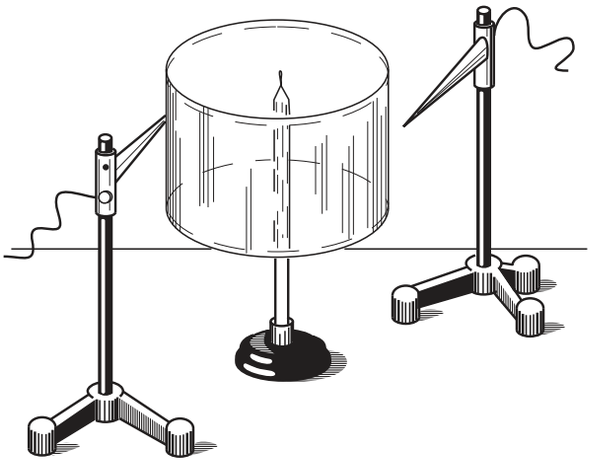


Рис. 36

легкий цилиндр, насаженный на вертикальную ось (электрический ротор).

В сильных электрических полях механизм утечки зарядов с острия становится более сложным. Вблизи острия молекулы воздуха ионизуются сильным электрическим полем. Воздух становится проводником электричества. Возникает сильный электрический ток, направленный к острию или от него (*электрический пробой*). Этот ток снимает электрические заряды с острия. На этом основано действие громоотвода.

4. При переходе через любую границу раздела сред тангенциальная составляющая электрического поля меняется непрерывно (см. § 6, п. 6). В электростатике электрическое поле внутри проводника обращается в нуль. Отсюда следует, что во внешнем пространстве поле $\mathbf{E}$ должно быть перпендикулярно к поверхности проводника. Если бы это было не так, то под действием касательной составляющей поля $\mathbf{E}$ заряды пришли бы в движение по поверхности проводника, т. е. равновесие их было бы невозможно. Силовые линии, таким образом, нормальны к поверхности проводника и оканчиваются на ней, не проникая внутрь проводника (рис. 37).

Напряженность электрического поля вблизи поверхности проводника легко найти по формуле (6.9). Полагая в ней $E_{1n} = 0$ (внутри проводника), $E_{2n} = E$ (вне проводника), получим $E = 4\pi\sigma$, или в векторной форме

$$\mathbf{E} = 4\pi\sigma\mathbf{n} \tag{11.1}$$

(нормаль $\mathbf{n}$ проведена наружу от поверхности проводника). Электрическая сила $\mathbf{f}$, действующая на единицу площади заряженной поверхности проводника, согласно общим формулам (6.13) и (6.15), будет

$$\mathbf{f} = \frac{\sigma}{2}\,\mathbf{E} = 2\pi\sigma^2\mathbf{n} = \frac{E^2}{8\pi}\,\mathbf{n}. \tag{11.2}$$

Она всегда направлена наружу, т. е. стремится удалить электричество с поверхности проводника.

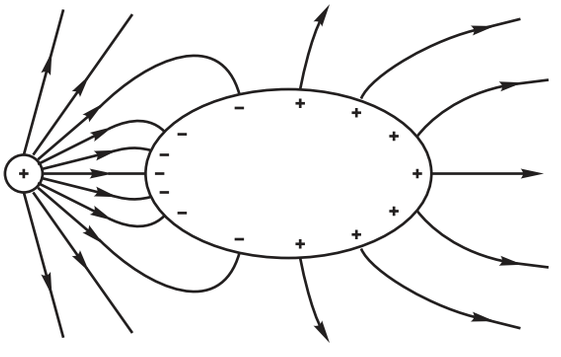


Рис. 37

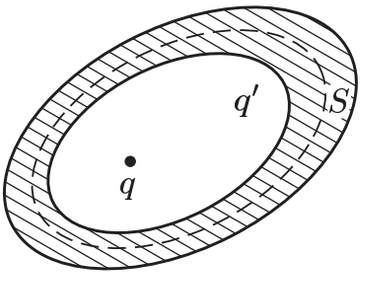


Рис. 38

5. Докажем теперь несколько положений, совокупность которых называется *теоремой Фарадея*. Пусть в однородном проводнике имеется полость, внутрь которой внесены электрические заряды (рис. 38). Проведем замкнутую поверхность S, окружающую полость и целиком проходящую в проводнике. Так как напряженность электрического поля на поверхности S равна нулю, то будет равен нулю и полный заряд, окруженный этой поверхностью. Таким образом, *сумма индуцированных зарядов на внутренней поверхности проводящей оболочки равна и противоположна по знаку сумме зарядов, окруженных этой оболочкой.*

При равновесии индуцированные заряды q' располагаются по внутренней поверхности проводящей оболочки таким образом, чтобы

полностью скомпенсировать внутри этой оболочки кулоновское поле зарядов q, окружаемых ею. Такая компенсация должна иметь место не только в стенках проводящей оболочки, но и во всем внешнем пространстве. Чтобы убедиться в этом, достаточно представить, что все внешнее пространство заполнено электрически нейтральной проводящей средой. Поле в ней при равновесии равно нулю. Но такая среда не будет оказывать никакого влияния на электрическое поле, поскольку положительные и отрицательные заряды в ней скомпенсированы в каждой точке пространства. Поэтому если среду удалить, оставив только проводящую оболочку, то от этого поле нигде не изменится. В частности, оно останется равным нулю во всем пространстве, из которого была удалена среда. Значит, *кулоновское поле зарядов, окруженных проводящей оболочкой, и зарядов, индуцированных на ее внутренней поверхности, равно нулю во всем внешнем пространстве.*

Допустим теперь, что все заряды находятся во внешнем пространстве. Если проводящее тело сплошное, то в нем электрического поля нет. Удалим из тела часть (электрически нейтрального) вещества. От этого, как выяснено выше, поле нигде не изменится, а равновесие электричества не нарушится. Зато в теле образуется полость. Таким образом, *если в полости нет электрических зарядов, то электрическое поле в ней равно нулю. Внешние заряды не создают в полости никакого электрического поля.* Чтобы предохранить какие-либо тела, например измерительные приборы, от влияния внешних электрических полей, их окружают проводящей оболочкой (*электростатическая защита*).

Мы видим, что замкнутая проводящая оболочка разделяет все пространство на внутреннюю и внешнюю части, в электрическом отношении совершенно не зависящие друг от друга. Например, при любом перемещении зарядов внутри (вне) оболочки во внешнем (внутреннем) пространстве не будет происходить никаких изменений. В частности, если все внутренние заряды привести в контакт со стенками оболочки, то произойдет их нейтрализация индуцированными зарядами. При этом внутреннее поле исчезнет, но наружное поле и распределение электричества на наружной поверхности тела останутся неизменными.

6. Все эти результаты были получены Фарадеем экспериментально. Опишем некоторые демонстрационные опыты, иллюстрирующие их. Возьмем *«цилиндр Фарадея»*, т. е. длинный металлический сосуд цилиндрической формы, открытый сверху. Насадим его на стержень электроскопа (рис. 39). Затем внесем в цилиндр заряженный шарик на изолирующей ручке. Стрелка электроскопа отклонится. Если шарик погружен достаточно глубоко, то при любых перемещениях его в цилиндре угол отклонения стрелки электроскопа не изменяется. Он не изменяется и тогда, когда заряженным шариком касаются внутренней поверхности стенки цилиндра. При этом шарик оказывается незаряженным. В этом легко убедиться, извлекая шарик из цилиндра и касаясь им другого незаряженного электроскопа. Весь заряд с шарика переходит к цилиндру и располагается на его наружной поверхности. В идеально поставленном опыте отверстие цилиндра Фарадея после

внесения в него шарика должно закрываться металлической крышкой. Но если цилиндр длинный, то опыт хорошо удается и без крышки.

Основываясь на опытах такого типа, Фарадей указал способ, с помощью которого заряд проводящего тела можно полностью передать другому проводящему телу. Для этого во втором теле надо сделать полость и внести в нее первое (заряженное) тело. При соприкосновении вносимого тела с внутренней поверхностью полости заряд от него полностью переходит ко второму телу. Внесенное тело можно

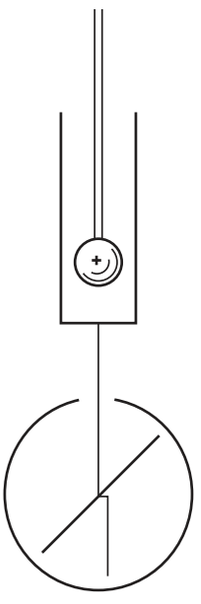


Рис. 39

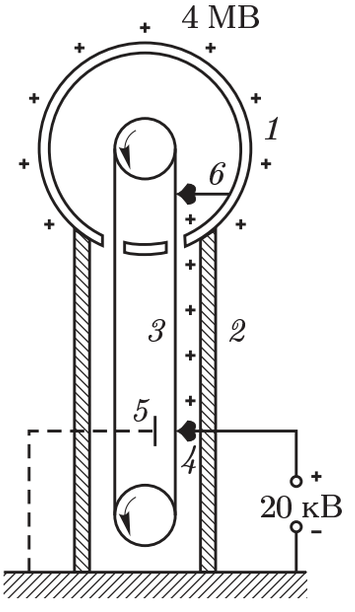


Рис. 40

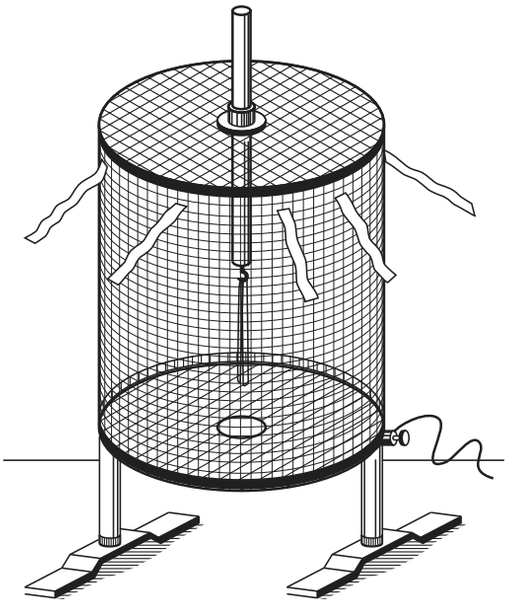


Рис. 41

извлечь из полости и зарядить снова. Внося его в полость второй раз, можно опять передать заряд второму телу. Такую операцию можно повторить многократно и сообщить полому телу теоретически сколь угодно большой заряд. Практически величина заряда лимитируется утечкой электричества из-за ионизации окружающего воздуха. По этому принципу работает электростатический генератор Ван-де-Граафа (1901–1967).

Генератор Ван-де-Граафа состоит из полого металлического шара *1* диаметром в несколько метров, укрепленного на изолирующей колонне *2* (рис. 40). Движущаяся бесконечная лента *3* из прорезиненной ткани заряжается от источника напряжения с помощью системы остриев *4*. С обратной стороны ленты против остриев помещена заземленная пластина *5*, усиливающая стекание зарядов с остриев *4* на ленту *3*. Другая система остриев *6* снимает заряды с ленты и передает их полому шару. Обычные генераторы позволяют получать напряжения до 2–5 миллионов вольт. Они применяются для ускорения электронов и ионов.

7. Опишем теперь демонстрационный опыт с «*клеткой Фарадея*». Она представляет собой цилиндр, дно и крышка которого сделаны из листового металла, а боковые стенки — из металлической сетки, чтобы сквозь них было видно, что делается внутри цилиндра (рис. 41).

К боковой поверхности цилиндра снаружи приклеены бумажные полоски. Вдоль оси клетки может перемещаться металлический стержень, все время находящийся в электрическом контакте с крышкой. Стержень оканчивается легкими бумажными листочками, играющими роль электроскопа. Они могут выдвигаться наружу через отверстие в дне клетки. Клетка устанавливается на изолирующих подставках. Если клетку зарядить электричеством, то наружные листочки расходятся, а листочки внутри остаются неподвижными. Если их выдвинуть наружу, то они также расходятся. Фарадей произвел аналогичный опыт (1836 г.) в большом масштабе, поместившись сам с электроскопом внутри металлической клетки, подвешенной к потолку на шелковых веревках. К клетке подносились металлические тела, заряженные от электростатической машины. Между ними и клеткой проскакивали сильные электрические искры, сообщавшие клетке большой электрический заряд. Никаких отклонений листочков электроскопа внутри клетки не наблюдалось, и никаких особых ощущений Фарадей не испытывал. Физики, работающие на высоковольтных ускорителях Ван-де-Граафа, автоматически повторяют опыт Фарадея. Они располагаются с измерительными приборами внутри шара работающего ускорителя на миллионы вольт и уверены, что находятся под надежной защитой теоремы Фарадея.

§ 12. Поляризация диэлектриков

1. Диэлектрики являются непроводниками электричества. Как и в металлах, в них также могут возбуждаться индукционные заряды. Поднесем, например, к шарику заряженного электроскопа C электрически нейтральное тело из диэлектрика AB (рис. 42). Угол отклонения стрелки электроскопа уменьшается. Дело в том, что заряд шарика C возбуждает на конце диэлектрика B индукционные заряды того же, а на конце A — противоположного знака. Эти заряды оттягивают часть зарядов со стрелки и стержня электроскопа на шарик, с чем и связано уменьшение угла отклонения стрелки.

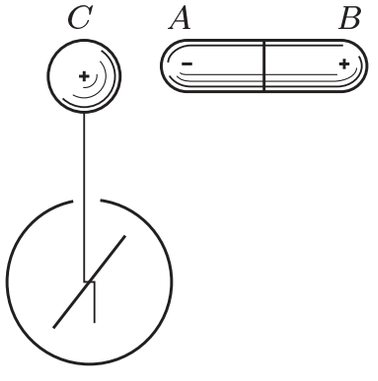


Рис. 42

Попытаемся разделить индукционные заряды, возникшие на диэлектрике. Для этого воспользуемся тем же приемом, который применялся в случае металлов (см. рис. 32). Пусть диэлектрик состоит из двух половин A и B, соприкасающихся между собой. Если в присутствии заряженного электроскопа эти части разъединить, а затем убрать или разрядить электроскоп, то обе они окажутся незаряженными. Это показывает, что заряды в диэлектрике лишены той свободы передвижения, какая свойственна электронам в металлах.

2. Заряды в диэлектрике могут смещаться из своих положений равновесия лишь на малые расстояния, порядка атомных. Допустим, например, что диэлектрик состоит из электрически нейтральных молекул. Под действием приложенного электрического поля центр тяжести электронов в молекуле немного смещается относительно центра тяжести атомных ядер. Молекулы становятся электрическими диполями, ориентированными положительно заряженными концами в направлении электрического поля $\mathbf{E}$. В этом случае говорят, что диэлектрик *поляризован*, а само смещение положительных и отрицательных зарядов диэлектрика в разные стороны называют *электрической поляризацией*. На схематическом рис. 43 диэлектрик изображен в виде прямоугольного параллелепипеда, а молекулы — в виде шариков. Положительно заряженная половина молекулы закрашена в черный цвет, отрицательно заряженная оставлена светлой. Мы видим, что на конце AB параллелепипеда $ABDC$ выступают нескомпенсированные отрицательные, а на конце CD — положительные поверхностные заряды. Это и есть индукционные заряды, появляющиеся в результате поляризации диэлектрика. Их называют *поляризационными*, или *связанными*, *зарядами*. Последним термином хотят подчеркнуть, что свобода перемещения связанных зарядов ограничена. Они могут смещаться лишь внутри электрически нейтральных молекул. В объеме диэлектрика происходит компенсация положительных и отрицательных зарядов молекул, и никаких макроскопических поляризационных зарядов не появляется. Однако это справедливо только тогда, когда поляризация диэлектрика *однородна*, т. е. когда все молекулы диэлектрика поляризованы и ориентированы одинаково. Если же поляризация неоднородна, то компенсации нет, и в диэлектрике могут появиться объемные поляризационные (связанные) заряды.

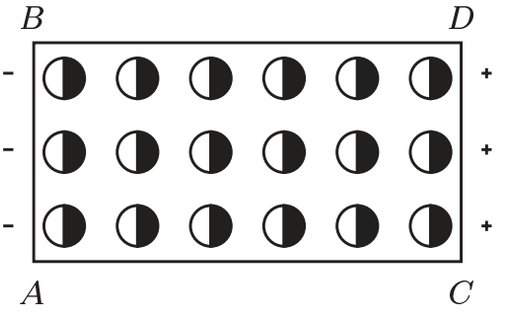


Рис. 43

Помимо электрически нейтральных молекул в диэлектрике могут существовать положительно или отрицательно заряженные *ионы*. Избыток ионов того или иного знака в какой-либо части диэлектрика означает наличие в этой части нескомпенсированных макроскопических зарядов. Такие заряды называются *свободными*. Они возникают в диэлектрике, например при электризации трением. К свободным зарядам относятся также все заряды, находящиеся на проводниках.

3. Механизм поляризации диэлектрика может быть и иным. Существуют диэлектрики, молекулы которых обладают дипольными моментами уже в отсутствие электрического поля. Такие молекулы называются *полярными*. Если поля нет, то полярные молекулы совершают хаотические тепловые движения и ориентированы совершенно беспорядочно. При наложении электрического поля дипольные моменты

молекул ориентируются преимущественно в направлении поля. А это означает, что диэлектрик становится поляризованным.

Наконец, существуют диэлектрические кристаллы (например, кристаллы NaCl), построенные из ионов противоположного знака. Такие кристаллы называются *ионными*. Ионный кристалл состоит из двух кристаллических решеток, вдвинутых одна в другую. Одна решетка построена из положительных, другая — из отрицательных ионов. В этом случае уже нельзя говорить о молекулах или атомах в кристалле. Кристалл в целом должен рассматриваться как одна гигантская молекула. При наложении электрического поля решетка положительных ионов сдвигается в одну, а отрицательных — в противоположную сторону. В этом и состоит электрическая поляризация ионных кристаллов. Существуют ионные кристаллы, поляризованные даже в отсутствие внешнего электрического поля.

4. Конкретное строение диэлектрика и механизм его поляризации для наших ближайших целей не имеют значения. Существенно лишь, что поляризация диэлектрика сопровождается появлением на нем нескомпенсированных макроскопических зарядов. Мы можем довольствоваться грубой моделью, в которой положительное и отрицательное электричества рассматриваются как непрерывные жидкости, равномерно перемешанные друг с другом. При поляризации диэлектрика происходит смещение одной жидкости относительно другой. Существенно, что такие смещения в обычных условиях ничтожны даже по сравнению с размерами атомов. Это связано с тем, что внешние поля, действующие на диэлектрик, очень слабы, если их сравнивать с внутренними электрическими полями атомов и молекул. Так, на электрон в атоме водорода действует электрическое поле ядра $E = e/r^2 \sim$ $\sim 10^7$ СГСЭ-ед. $\approx 10^{11}$ В/м, громадное по сравнению с обычными макроскопическими полями.

5. Для количественного описания поляризации диэлектрика пользуются *вектором поляризации*. Так называется дипольный момент единицы объема диэлектрика, возникающий при его поляризации. Возьмем кусок однородного изотропного диэлектрика, имеющий форму косого параллелепипеда (рис. 44). Поместим его в однородное электри-

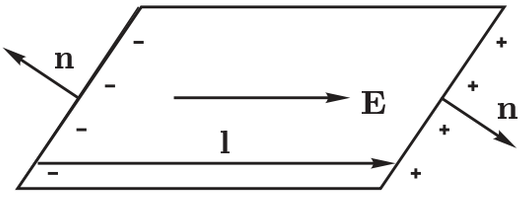


Рис. 44

ческое поле, направленное параллельно боковым ребрам. На основаниях параллелепипеда появятся поляризационные заряды с поверхностной плотностью $\sigma_{\text{пол}}$. На боковых гранях поляризационных зарядов не возникнет, так как смещение зарядов внутри диэлектрика происходит

параллельно этим граням. Если S — площадь основания параллелепипеда, то диэлектрик приобретет дипольный момент $\sigma_{\text{пол}} S\mathbf{l}$, где $\mathbf{l}$ — вектор, проведенный от отрицательного основания параллелепипеда к положительному параллельно боковым ребрам. Вектор поляризации диэлектрика будет

$$\mathbf{P} = \frac{\sigma_{\text{пол}} S}{V}\,\mathbf{l}, \tag{12.1}$$

где V — объем параллелепипеда. Эта формула справедлива и для однородного анизотропного диэлектрика (кристалла), когда направления векторов $\mathbf{E}$ и $\mathbf{P}$ могут и не совпадать.

Пусть $\mathbf{n}$ — единичный вектор внешней нормали к основанию параллелепипеда, заряженному положительно. Тогда $V = S(\mathbf{ln})$. Подставив это значение в формулу (12.1) и умножив ее скалярно на $\mathbf{n}$, найдем

$$\sigma_{\text{пол}} = (\mathbf{Pn}) = P_n. \tag{12.2}$$

В частности, если параллелепипед прямоугольный, то $\sigma_{\text{пол}} = P$. Формула (12.2) была выведена применительно к положительно заряженному основанию. Но она верна и для отрицательно заряженного основания, так как на нем внешняя нормаль $\mathbf{n}$ направлена в противоположную сторону, а потому проекция P_n отрицательна. Формула справедлива и на боковой поверхности параллелепипеда, так как на ней, как мы видели, $\sigma_{\text{пол}} = 0$, что согласуется с формулой (12.2). Таким образом, формула (12.2) справедлива в общем случае.

Формула (12.2) показывает, что нормальная составляющая P_n представляет по величине количество электричества, смещаемое при поляризации через единичную площадку в направлении нормали $\mathbf{n}$ к ней. Эта интерпретация применима и в случае неоднородной поляризации, т. е. такой, при которой вектор $\mathbf{P}$ меняется от точки к точке. Чтобы убедиться в этом, достаточно мысленно разделить диэлектрик на малые объемы, в пределах каждого из которых поляризация может считаться однородной.

6. Как сказано выше, при неоднородной поляризации поляризационные заряды могут появляться не только на поверхности, но и в объеме диэлектрика. Вычислим теперь плотность объемных поляризационных зарядов. Выделим мысленно в диэлектрике произвольный объем V, ограниченный замкнутой поверхностью S (рис. 45). Заряд, смещенный при поляризации через площадку dS в отрицательном направлении нормали $\mathbf{n}$, согласно формуле (12.2) равен $-P_n\,dS$. Через всю поверхность S внутрь объема V при поляризации поступает поляризационный заряд

$$q_{\text{пол}} = -\oint P_n\,dS = -\oint (\mathbf{P}\,d\mathbf{S}). \tag{12.3}$$

Если поляризация однородна, то $q_{\text{пол}} = 0$.

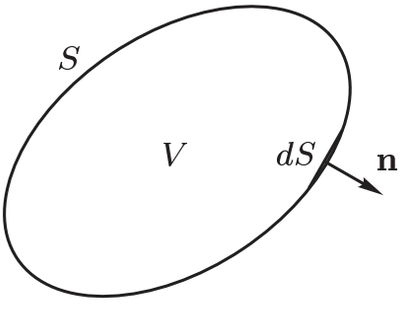


Рис. 45

§ 13. Теорема Гаусса для диэлектриков

1. Как выяснено в § 11, влияние диэлектрика на электрическое поле сводится к действию поляризационных зарядов. Поэтому к диэлектрикам можно применить соотношение (5.5), добавив при этом к свободным зарядам q поляризационные заряды $q_{\text{пол}}$:

$$\oint E_n \, dS = 4\pi(q + q_{\text{пол}}). \tag{13.1}$$

Подставив сюда значение $q_{\text{пол}}$ из формулы (12.3), получим

$$\oint (E_n + 4nP_n) \, dS = 4nq. \tag{13.2}$$

Введем новый вектор

$$\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P}, \tag{13.3}$$

называемый *вектором электрической индукции*, или *электрического смещения*. Тогда

$$\oint D_n \, dS = 4\pi q. \tag{13.4}$$

Это и есть *теорема Гаусса для электрического поля в диэлектрике*. Мы видим, что поток вектора $\mathbf{D}$ через замкнутую поверхность определяется *только свободными зарядами*. Этим и оправдывается введение вектора $\mathbf{D}$. В вакууме векторы $\mathbf{D}$ и $\mathbf{E}$ совпадают.

2. В дифференциальной форме соотношение (13.4) имеет вид

$$\operatorname{div}\mathbf{D} = 4\pi\rho, \tag{13.5}$$

где ρ — объемная плотность *свободных* зарядов. Нелишне напомнить, что теоремы (13.4) и (13.5) справедливы не только в электростатике. Они постулируются также для переменных во времени полей. Эти теоремы входят как составные части в систему фундаментальных электродинамических уравнений Максвелла.

Подставив в (13.5) выражение (13.3), получим

$$\operatorname{div}\mathbf{E} = 4\pi(\rho - \operatorname{div}\mathbf{P}).$$

Но для той же величины можно написать

$$\operatorname{div}\mathbf{E} = 4\pi(\rho + \rho_{\text{пол}}).$$

Следовательно,

$$\rho_{\text{пол}} = -\operatorname{div}\mathbf{P}. \tag{13.6}$$

3. Теорема Гаусса для вектора электрического смещения в диэлектрике имеет такой же вид, как и для напряженности электрического поля в вакууме. Поэтому все математические соотношения, полученные из нее для вакуума, сохраняют силу и для однородного диэлектрика. Надо только вектор $\mathbf{E}$ заменить вектором $\mathbf{D}$. Таким путем из формул

(6.1)–(6.3) и (6.5) получаем, например,

$$D = 2\pi\sigma, \tag{13.7}$$

$$D = \begin{cases} 4\pi\rho x & \text{внутри пластинки,} \\ 4\pi\rho a & \text{вне пластинки,} \end{cases} \tag{13.8}$$

$$D = \begin{cases} (4\pi/3)\rho r & \text{внутри шара,} \\ (4\pi/3)\rho a^3/r^2 & \text{вне шара.} \end{cases} \tag{13.9}$$

Электрическое смещение точечного заряда в однородном диэлектрике определяется выражением

$$\mathbf{D} = q\mathbf{r}/r^3. \tag{13.10}$$

§ 14. Граничные условия

1. Из соотношения (5.5) мы получили граничное условие (6.9), которому должны удовлетворять нормальные составляющие вектора $\mathbf{E}$ на заряженной поверхности. Поступая совершенно так же, из теоремы Гаусса для диэлектриков (13.4) получаем следующее условие на границе раздела двух диэлектриков:

$$D_{2n} - D_{1n} = 4\pi\sigma, \tag{14.1}$$

где σ — поверхностная плотность свободных зарядов на этой границе.

Отличие формулы (14.1) от аналогичной формулы (6.9) обусловлено влиянием поляризационных зарядов, появляющихся на границе диэлектриков. Поверхностная плотность поляризационных зарядов равна $\sigma_{\text{пол}} = P_{1n} - P_{2n}$ (рис. 46). Учитывая ее, получаем $E_{2n} - E_{1n} = 4\pi(\sigma + \sigma_{\text{пол}})$, или

$$(E_{2n} + 4\pi P_{2n}) - (E_{1n} + 4\pi P_{1n}) = 4\pi\sigma, \tag{14.2}$$

а эта формула тождественна с (14.1).

В частности, вектор электрического смещения в диэлектрике на границе с проводником определяется выражением

$$\mathbf{D} = 4\pi\sigma\mathbf{n}. \tag{14.3}$$

Здесь единичная нормаль $\mathbf{n}$ проведена от металла к диэлектрику.

Если на границе раздела нет свободных зарядов, то

$$D_{1n} = D_{2n}. \tag{14.4}$$

Таким образом, при переходе через незаряженную границу двух диэлектриков нормальная составляющая вектора $\mathbf{D}$ остается непрерывной. Что касается вектора $\mathbf{E}$, то на любой границе остаются непрерывными его тангенциальные составляющие:

$$E_{1t} = E_{2t}. \tag{14.5}$$

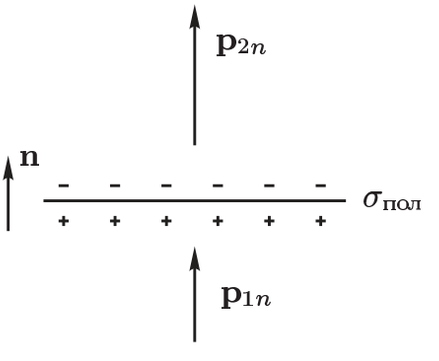


Рис. 46

Это утверждение доказывается так же, как и для поля в вакууме (см. § 6, а также § 17).

2. Пользуясь граничными условиями (14.4) и (14.5), можно указать принципиальный способ измерения векторов $\mathbf{E}$ и $\mathbf{D}$ в диэлектрике. Обычный метод измерения $\mathbf{E}$ по силе, действующей на пробный заряд, годится для поля в вакууме, а к веществу применим не всегда. Дело в том, что выражение для действующей силы $\mathbf{F} = q\mathbf{E}$ справедливо в вакууме, а в веществе в лучшем случае является приближенным. Кроме того, внесение пробного заряда в вещество может оказаться просто невозможным, например при необходимости измерить напряженность или электрическое смещение поля в твердом диэлектрике. Единственный *принципиальный способ* измерения векторов $\mathbf{E}$ и $\mathbf{D}$ внутри среды, пригодный во всех случаях, состоит в том, чтобы в теле сделать *полость* и внести в нее пробный заряд. Однако измеренное таким путем поле, вообще говоря, не будет совпадать ни с вектором $\mathbf{E}$, ни с вектором $\mathbf{D}$. Результат зависит от *формы полости.* Только для полостей специальной формы измерение непосредственно дает $\mathbf{E}$ и $\mathbf{D}$. Рассмотрим два случая.

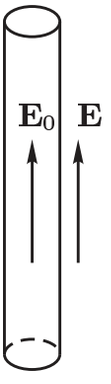


Рис. 47

Случай 1. Полость имеет форму очень длинного и тонкого цилиндрического канала, параллельного полю $\mathbf{E}$ (рис. 47). Количество вещества внутри канала бесконечно мало. Его удаление из этого канала меняет электрическое поле в окружающем диэлектрике бесконечно мало. На концах канала появляются лишь поляризационные заряды, влияние которых на электрическое поле вдали от этих концов пренебрежимо мало. Из соображений симметрии следует, что поле в канале $\mathbf{E}_0$ параллельно наружному полю $\mathbf{E}$. Поэтому из граничного условия (14.5) получаем $\mathbf{E} = \mathbf{E}_0$. Таким образом, измерение $\mathbf{E}$ сводится к измерению $\mathbf{E}_0$.

Случай 2. Полость имеет форму бесконечно короткого цилиндра с основаниями, перпендикулярными к вектору $\mathbf{D}$ (рис. 48). Как и в предыдущем случае, удаление вещества из такой бесконечно малой полости сказывается на поле в окружающем диэлектрике также бесконечно мало. Действительно, на границах полости появляются поляризационные заряды противоположных знаков. Вне полости поля этих зарядов почти полностью компенсируют друг друга. Однако внутри полости поля поляризационных зарядов усиливают друг друга, что существенно меняет поле в полости. Внутри полости электрическое поле $\mathbf{E}_0$, ввиду симметрии, перпендикулярно к ее основаниям. В полости напряженность и индукция поля совпадают ($\mathbf{E}_0 = \mathbf{D}_0$). Поэтому из граничного условия (14.4) получаем $\mathbf{D} = \mathbf{E}_0$. Измерение $\mathbf{D}$ сводится к измерению $\mathbf{E}_0$.

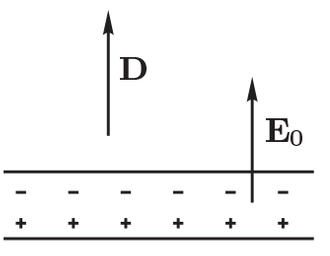


Рис. 48

§ 15. Поляризуемость и диэлектрическая проницаемость

1. Одним из фундаментальных уравнений электростатики является теорема Гаусса (13.4) или (13.5). Второе фундаментальное уравнение электростатики будет сформулировано в § 17 при введении понятия потенциала. В вакууме, где поле характеризуется одним только вектором $\mathbf{E}$, этих уравнений достаточно. Они образуют *полную систему* уравнений электростатики. В веществе к вектору $\mathbf{E}$ надо добавить еще один вектор ($\mathbf{P}$ или $\mathbf{D}$). Поэтому уравнения электростатики надо дополнить еще одним векторным уравнением. Принципиальный способ получения такого уравнения содержится в самом определении вектора поляризации $\mathbf{P}$. Если известна атомная структура вещества, то в принципе можно рассчитать смещения электронов и атомных ядер, которые они получают при внесении вещества в электрическое поле. После этого можно вычислить вектор $\mathbf{P}$ и тем самым получить недостающее уравнение. Ясно, что в зависимости от конкретных условий таким путем должны получаться весьма разнообразные и сложные соотношения. Универсальной связи между векторами $\mathbf{P}$ и $\mathbf{E}$, пригодной для всех веществ, не существует. Здесь мы не можем идти по указанному пути. Мы получим недостающее уравнение, опираясь на опыт.

2. Опыт показывает, что для обширного класса диэлектриков и широкого круга явлений связь между векторами $\mathbf{P}$ и $\mathbf{E}$ *линейна* и *однородна.* Такая закономерность объясняется тем, что напряженности макроскопических электрических полей обычно очень малы по сравнению с напряженностями микрополей внутри атомов и молекул (см. § 12, п. 4). Если среда изотропна, то векторы $\mathbf{P}$ и $\mathbf{E}$ коллинеарны и можно написать

$$\mathbf{P} = \alpha \mathbf{E}, \tag{15.1}$$

где α — безразмерный коэффициент, называемый *поляризуемостью диэлектрика.* Он зависит от плотности и температуры диэлектрика. Связь между $\mathbf{D}$ и $\mathbf{E}$ записывается в виде

$$\mathbf{D} = \varepsilon \mathbf{E}. \tag{15.2}$$

Новая безразмерная величина

$$\varepsilon = 1 + 4\pi\alpha \tag{15.3}$$

называется *диэлектрической проницаемостью диэлектрика.* Этой величиной обычно и характеризуются индивидуальные свойства диэлектриков. Для вакуума $\alpha = 0$, $\varepsilon = 1$.

3. В кристаллах направления векторов $\mathbf{P}$ и $\mathbf{E}$ не совпадают. Соотношение (15.1) заменяется более общей линейной однородной зависимостью:

$$\begin{aligned} P_x &= \alpha_{xx} E_x + \alpha_{xy} E_y + \alpha_{xz} E_z, \\ P_y &= \alpha_{yx} E_x + \alpha_{yy} E_y + \alpha_{yz} E_z, \\ P_z &= \alpha_{zx} E_x + \alpha_{zy} E_y + \alpha_{zz} E_z, \end{aligned}$$

или сокращенно:

$$P_i = \sum_j \alpha_{ij} E_j \quad (i, j = x, y, z). \tag{15.4}$$

Здесь α_{ij} — безразмерные коэффициенты, зависящие от выбора координатных осей. Совокупность этих девяти коэффициентов называется *тензором поляризуемости диэлектрика.* Аналогично,

$$D_i = \sum_j \varepsilon_{ij} E_j \quad (i, j = x, y, z), \tag{15.5}$$

где ε_{ij} — новые безразмерные постоянные, образующие *тензор диэлектрической проницаемости вещества.* Они связаны с коэффициентами α_{ij} соотношениями

$$\varepsilon_{ij} = \delta_{ij} + 4\pi\alpha_{ij}, \tag{15.6}$$

где δ_{ij} — единичный тензор, определяемый условиями: $\delta_{ij} = 1$ при $i = j$, $\delta_{ij} = 0$ при $i \neq j$. Пользуясь понятием энергии как функции состояния, находим, что α_{ij} и ε_{ij} симметричны, т. е.

$$\alpha_{ij} = \alpha_{ji}, \quad \varepsilon_{ij} = \varepsilon_{ji}. \tag{15.7}$$

4. Все приведенные приближенные соотношения, несмотря на их важность, не относятся к числу фундаментальных соотношений электростатики и электродинамики. Область применимости их ограничена. Существуют диэлектрики, к которым они неприменимы. Мы указывали уже, что ионные кристаллы могут быть поляризованы даже в отсутствие внешнего электрического поля. Другим примером тел, обладающих тем же свойством, могут служить *электреты.* Эти диэлектрики подобны постоянным магнитам. Они длительно сохраняют состояние поляризации и благодаря этому создают электрическое поле в окружающем пространстве. Электрет можно получить, поместив в сильное электрическое поле расплавленный диэлектрик, состоящий из полярных молекул. Подходящим веществом может служить смесь воска и смолы, помещаемая в электрическое поле порядка 10^6 В/м. После застывания диэлектрика поле выключается. Застывший диэлектрик сохраняет поляризацию в течение нескольких часов и даже суток. Поляризация в конце концов исчезает благодаря медленным релаксационным процессам, идущим внутри диэлектрика. Существуют электреты, время жизни которых составляет многие годы. Приближенно поведение электретов и аналогичных им диэлектриков в электрическом поле можно описать соотношением вида

$$\mathbf{P} = \mathbf{P}_0 + \alpha\mathbf{E}, \tag{15.8}$$

где величины $\mathbf{P}_0$ и α от напряженности поля $\mathbf{E}$ не зависят.

5. Опишем еще демонстрационный опыт с лейденской банкой. Удобно взять банку, наружной обкладкой которой служит длинная металлическая трубка. В нее вставляется диэлектрическая трубка из кварца, а в последнюю — внутренняя обкладка, представляющая собой металлический стержень на изолирующей ручке. Заземлив наружную обкладку, банку заряжают от электростатической машины, затем отсоединяют от нее и разбирают. Внутренний стержень вытягивается за изолирующую

ручку, вынимается кварцевая трубка, и обе металлические обкладки приводятся в контакт друг с другом. Теперь на обкладках зарядов нет. Если банку собрать снова, то она опять окажется заряженной. Такую операцию можно повторить многократно, и всякий раз после сборки банка оказывается заряженной. Это доказывает, что кварцевая трубка поляризована даже тогда, когда она не окружена заряженными обкладками. Поляризованный кварц создает в окружающем пространстве электрическое поле, которое и индуцирует электрические заряды на обкладках собранной банки.

Еще более сложные явления наблюдаются в так называемых сегнетоэлектриках (см. § 39). Здесь связь между векторами $\mathbf{P}$ и $\mathbf{E}$ нелинейна и зависит от предшествующей истории изменения поля.

Простейший случай, когда применимы формулы (15.1) и (15.2), является довольно распространенным и практически наиболее важным. В дальнейшем, если не оговорено противное, мы будем иметь дело именно с таким случаем.

6. Рассмотрим теперь поведение силовых линий при прохождении через границу раздела двух диэлектриков (рис. 49). Если на границе раздела нет свободных зарядов, то должны выполняться граничные условия

$$E_{1t} = E_{2t}, \quad \varepsilon_1 E_{1n} = \varepsilon_2 E_{2n}.$$

Если ввести углы β_1 и β_2 между силовыми линиями и нормалью к границе раздела, то эти условия можно записать в виде

$$E_1 \sin\beta_1 = E_2 \sin\beta_2,$$

$$\varepsilon_1 E_1 \cos\beta_1 = \varepsilon_2 E_2 \cos\beta_2.$$

Из них получаем

$$\frac{\operatorname{tg}\beta_1}{\operatorname{tg}\beta_2} = \frac{\varepsilon_1}{\varepsilon_2}. \qquad (15.9)$$

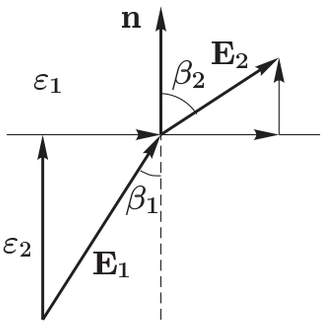


Рис. 49

Отсюда видно, что при переходе через границу раздела двух диэлектриков силовые линии испытывают *преломление*. При переходе из диэлектрика с меньшей ε в диэлектрик с большей ε угол β увеличивается, т. е. силовая линия удаляется от нормали к границе раздела. С этим связана концентрация (сгущение) силовых линий в диэлектрике с большей диэлектрической проницаемостью. Примером может служить диэлектрическая пластинка, внесенная в однородное электрическое поле (рис. 50).

Если полый диэлектрик с большой диэлектрической проницаемостью внести в электрическое поле, то из-за преломления силовые линии сконцентрируются преимущественно в стенках диэлектрика (рис. 51). Внутри полости они расположатся редко. Это значит, что поле в полости будет ослаблено по сравнению с наружным полем. Полость

внутри диэлектрика, таким образом, обладает *экранирующим действием.* В этом отношении она ведет себя аналогично полости в металле.

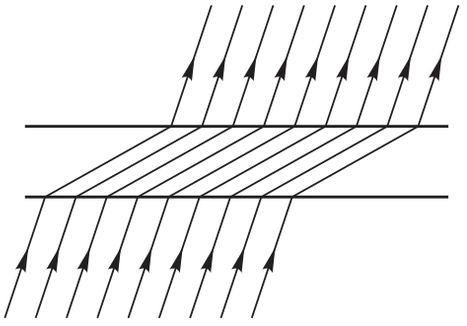


Рис. 50

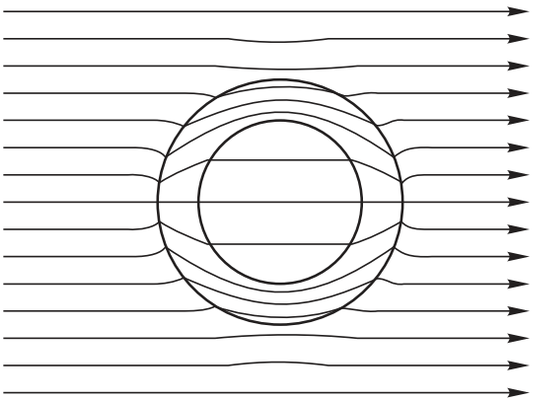


Рис. 51

Однако, в отличие от металла, экранирование диэлектриком не полное. Чем больше диэлектрическая проницаемость, тем сильнее экранирующее действие.

§ 16. Поле равномерно поляризованного шара

1. До поляризации в шаре была однородная смесь положительного и отрицательного электричеств с объемными плотностями $+\rho$ и $-\rho$. Сдвинем все положительные заряды относительно отрицательных на одно и то же расстояние $\delta\mathbf{l}$. (На рис. 52 смещение $\delta\mathbf{l}$ сильно преувеличено. В практически важных случаях оно мало даже по сравнению с атомными размерами.) Шар равномерно поляризуется, причем вектор поляризации будет $\mathbf{P} = \rho\delta\mathbf{l}$. Мы видим, что поле $\mathbf{E}$ равномерно поляризованного шара есть векторная сумма полей двух равномерно и разноименно заряженных шаров, немного смещенных друг относительно друга. Рассмотрим два случая.

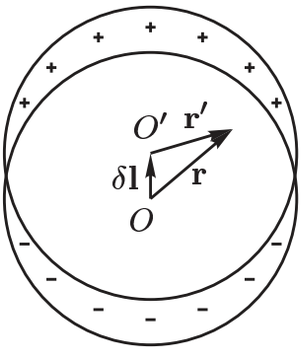


Рис. 52

Случай 1. Поле внутри равномерно поляризованного шара. Пусть O и O' — центры отрицательно и положительно заряженных шаров, а $\mathbf{r}$ и $\mathbf{r}'$ — радиусы-векторы, проведенные из этих центров. Согласно формуле (6.5) поля этих шаров равны соответственно

$$\mathbf{E}_1 = -\frac{4\pi}{3}\,\rho\mathbf{r}, \quad \mathbf{E}_2 = -\frac{4\pi}{3}\,\rho\mathbf{r}',$$

а их геометрическая сумма

$$\mathbf{E}^{(i)} = \frac{4\pi}{3}\,\rho(\mathbf{r}' - \mathbf{r}) = -\frac{4\pi}{3}\,\rho\delta\mathbf{l},$$

или

$$\mathbf{E}^{(i)} = -\frac{4\pi}{3}\,\mathbf{P}. \tag{16.1}$$

Случай 2. Поле равномерно поляризованного шара во внешнем пространстве. Пусть q — заряд положительного шара. Каждый шар возбуждает во внешнем пространстве такое поле, как если бы весь заряд был сосредоточен в центре шара. Поэтому поле равномерно поляризованного шара во внешнем пространстве будет совпадать с полем точечного диполя с дипольным моментом $\mathbf{p} = q\delta\mathbf{l} = V\mathbf{P}$, где V — объем шара. Следовательно, вне шара

$$\mathbf{E}^{(e)} = V\left[\frac{3(Pr)}{r^5}\,\mathbf{r} - \frac{\mathbf{P}}{r^3}\right].$$

Чтобы найти $\mathbf{E}^{(e)}$ на границе шара, следует положить $V = (4\pi/3)r^3$. Это дает

$$\mathbf{E}^{(e)} = 4\pi(\mathbf{Pn})\,\mathbf{n} - \frac{4\pi}{3}\,\mathbf{P}, \tag{16.2}$$

где $\mathbf{n}$ — единичный вектор внешней нормали к поверхности шара.

2. Равномерную поляризацию шара можно получить, поместив его во внешнее однородное электрическое поле $\mathbf{E}_0$. Для доказательства достаточно убедиться, что при этом будут удовлетворены условия в бесконечности и граничные условия на поверхности шара. Последние требуют, чтобы по разные стороны поверхности шара были одинаковы касательные составляющие векторов $\mathbf{E}$ и нормальные составляющие векторов $\mathbf{D}$. Полное поле $\mathbf{E}$ слагается из внешнего поля $\mathbf{E}_0$ и поля поляризованного шара. На бесконечности полное поле должно переходить в $\mathbf{E}_0$. Это условие, очевидно, удовлетворяется, так как на бесконечности поле поляризованного шара исчезает, поскольку оно убывает обратно пропорционально кубу расстояния. Внутри шара $\mathbf{E} = \mathbf{E}_0 + \mathbf{E}^{(i)}$, вне шара $\mathbf{E} = \mathbf{E}_0 + \mathbf{E}^{(e)}$. Касательные составляющие обоих полей на поверхности шара одинаковы, как это видно из выражений (16.1) и (16.2). Вне шара электрическое смещение совпадает с напряженностью поля, т. е. равно $\mathbf{E}_0 + \mathbf{E}^{(e)}$, внутри шара оно будет $\mathbf{E}_0 + \mathbf{E}^{(i)} + 4\pi\mathbf{P}$. С учетом (16.1) и (16.2) отсюда получаем на поверхности шара

$$\mathbf{D}^{(e)} = \mathbf{E}_0 + 4\pi(\mathbf{Pn})\,\mathbf{n} - \frac{4\pi}{3}\,\mathbf{P}, \quad \mathbf{D}^{(i)} = \mathbf{E}_0 - \frac{4\pi}{3}\,\mathbf{P} + 4\pi\mathbf{P}.$$

Отсюда видно, что нормальные составляющие этих векторов одинаковы. Значит, граничные условия удовлетворены, чем и завершается доказательство.

Полное поле внутри шара, как следует из формулы (16.1), определяется выражением

$$\mathbf{E} = \mathbf{E}_0 - \frac{4\pi}{3}\,\mathbf{P}. \tag{16.3}$$

Вектор $\mathbf{P}$ отсюда можно исключить, используя выражение $\mathbf{P} = \alpha\mathbf{E}$. Таким образом, получаем соотношение между внутренним и внешним полями:

$$\left(1 + \frac{4\pi}{3}\,\alpha\right)\mathbf{E} = \mathbf{E}_0, \tag{16.4}$$

или

$$\mathbf{E} = \frac{3}{\varepsilon + 2}\,\mathbf{E}_0. \tag{16.5}$$

Вектор поляризации внутри шара будет

$$\mathbf{P} = \alpha\mathbf{E} = \frac{3\alpha}{\varepsilon+2}\,\mathbf{E}_0 = \frac{3}{4\pi}\,\frac{\varepsilon-1}{\varepsilon+2}\,\mathbf{E}_0.$$

В результате во внешнем однородном поле $\mathbf{E}_0$ шар радиуса a приобретет дипольный момент $\mathbf{p} = V\mathbf{P}$, или

$$\mathbf{p} = a^3\,\frac{\varepsilon-1}{\varepsilon+2}\,\mathbf{E}_0. \tag{16.6}$$

3. Рассчитаем теперь напряженность поля $\mathbf{E}'$ в сферической полости, вырезанной внутри равномерно поляризованного диэлектрика в предположении, что поляризация вне полости всюду однородна. Тогда и внешнее поле в диэлектрике $\mathbf{E}$ будет также однородно. Если полость заполнить тем же равномерно поляризованным диэлектриком, то к полю в полости E' добавится поле равномерно поляризованного шара $-(4\pi/3)\mathbf{P}$. В результате должно получиться поле $\mathbf{E}$, т. е. $\mathbf{E}' - (4\pi/3)\mathbf{P} = \mathbf{E}$. Отсюда

$$\mathbf{E}' = \mathbf{E} + \frac{4\pi}{3}\,\mathbf{P}. \tag{16.7}$$

Исключив вектор $\mathbf{P}$, найдем

$$\mathbf{E}' = \frac{\varepsilon+2}{3}\,\mathbf{E}. \tag{16.8}$$

ЗАДАЧИ

1. Найти приближенное выражение для силы, действующей в неоднородном электрическом поле на маленькие диэлектрический и металлический шарики радиуса a.

Р е ш е н и е. Если бы внешнее поле $\mathbf{E}_0$ было однородно, то шарик приобрел бы дипольный момент, определяемый выражением (16.6). Тот же результат приближенно справедлив и в неоднородном поле. Используя его и формулу (4.8), находим искомую силу:

$$F = a^3\,\frac{\varepsilon-1}{\varepsilon+2}\,\frac{\partial}{\partial x}\Big(\frac{E_0^2}{2}\Big),$$

причем ось X мы направили вдоль вектора $\mathbf{E}_0$. Полагая $\varepsilon = \infty$, получаем формулу для проводящего шарика:

$$F = \frac{a^3}{2}\,\frac{\partial E_0^2}{\partial x}.$$

Сила $\mathbf{F}$ направлена в сторону возрастания поля $\mathbf{E}_0$. Силами такого рода объясняется первое явление, с которого началось изучение электричества: притяжение наэлектризованными телами легких тел.

2. Как меняется с расстоянием r сила взаимодействия $\mathbf{F}$ между двумя маленькими шариками, из которых один заряжен, а другой не заряжен?

О т в е т. $F \sim 1/r^5$.

3. В шаре, равномерно заряженном электричеством с объемной плотностью ρ, сделана сферическая полость, центр которой O' смещен относительно центра шара O на расстояние R. Определить электрическое поле внутри полости.

Ответ. $\mathbf{E} = (4/3)\pi\rho\mathbf{R}$, где $\mathbf{R} = \overrightarrow{OO'}$. Поле однородно.

Указание. Заполнить мысленно полость электричествами противоположных знаков с плотностями $+\rho$ и $-\rho$. Тогда поле в полости можно рассматривать как суперпозицию полей двух равномерно и противоположно заряженных шаров. См. аналогичную задачу в т. I (§ 55, задача 7).

4. В неограниченной диэлектрической однородной жидкости с диэлектрической проницаемостью ε помещен однородный шар с той же диэлектрической проницаемостью, равномерно заряженный электричеством с объемной плотностью ρ. В шаре сделана сферическая полость, куда помещен меньший шар радиуса a из того же материала, также равномерно заряженный с объемной плотностью ρ электричеством того же знака. Зазор между поверхностью малого шара и стенками полости пренебрежимо мал. Определить силу $\mathbf{F}$, действующую на меньший шар, зная расстояние между центрами обоих шаров.

Решение. Поле большого шара в его полости однородно и равно $4\pi\rho\mathbf{R}/(3\varepsilon)$, где $\mathbf{R} = \overrightarrow{OC}$ — вектор, проведенный от центра большого шара O к центру малого шара C. Умножив это поле на заряд малого шара, найдем

$$\mathbf{F} = (4\pi\rho)^2 a^3 \mathbf{R}/(9\varepsilon).$$

§ 17. Потенциальность электростатического поля

1. Неподвижный точечный заряд q возбуждает в вакууме электрическое поле $\mathbf{E} = \dfrac{Q}{r^3}\,\mathbf{r}$. Пусть в этом поле перемещается другой точечный заряд q, переходя из начального положения *1* в конечное положение *2* вдоль произвольной кривой *12* (рис. 53). Работа, совершаемая силами поля при таком перемещении, выражается криволинейным интегралом

$$A_{12} = \int\limits_{12} q(\mathbf{E}\,d\mathbf{r}) = qQ \int\limits_{12} \frac{\mathbf{r}\,d\mathbf{r}}{r^3}.$$

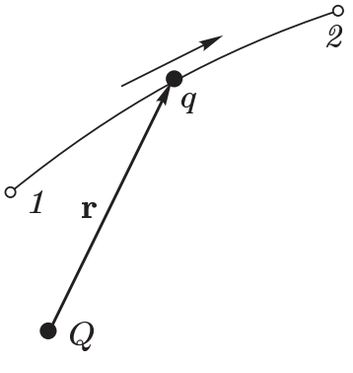


Рис. 53

Но $\mathbf{r}\,d\mathbf{r} = r\,dr$, в чем легко убедиться, дифференцируя тождество $\mathbf{r}^2 = r^2$. Поэтому криволинейный интеграл сводится к определенному:

$$A_{12} = qQ \int\limits_{r_1}^{r_2} \frac{dr}{r^2} = qQ\left(\frac{1}{r_1} - \frac{1}{r_2}\right). \qquad (17.1)$$

Таким образом, при любом выборе начальной и конечной точек *1* и *2* работа A_{12} не зависит от формы пути, а определяется только положениями этих точек. Силовые поля, удовлетворяющие такому условию, называются *потенциальными* или *консервативными* (см. т. I, § 24). Следовательно, *электростатическое поле точечного заряда есть поле потенциальное.*

Доказанное справедливо для электрического поля любой системы неподвижных точечных зарядов. Это непосредственно следует из принципа суперпозиции электрических полей и из известной теоремы

механики, согласно которой работа результирующей силы равна сумме работ составляющих сил.

В общем случае любую систему зарядов можно мысленно разделить на достаточно малые части, каждая из которых может рассматриваться как точечный заряд. В число таких зарядов должны быть включены и индукционные заряды на проводниках и диэлектриках. Поэтому *всякое электростатическое поле, независимо от того, создается оно в вакууме или в веществе, является полем потенциальным.* Это было бы очевидно для микрополя $\mathbf{E}_{\text{микро}}$, если бы возбуждающие его заряды были неподвижны. Макроскопическое поле $\mathbf{E}_{\text{макро}}$ было бы также потенциально, так как оно получается из потенциального поля $\mathbf{E}_{\text{микро}}$ путем его усреднения. Однако электроны и атомные ядра движутся, а электрические микрополя не потенциальны. Поэтому уравнения макроскопической электростатики в общем случае нельзя получить из уравнений электростатики для микрополей. Нужны уравнения микрополей для движущихся зарядов. Таковыми являются уравнения электронной теории Лоренца. Но мы не будем обосновывать электростатику с помощью уравнений электронной теории Лоренца. В конце концов, макроскопические уравнения Максвелла устанавливаются постулативно. А из этих уравнений, как будет видно из дальнейшего, непосредственно следует, что электростатическое макрополе потенциально.

2. Допустим, что в электростатическом поле заряд переносится из точки *1* в точку *2* сначала по пути *132*, а затем по пути *142* (рис. 54).

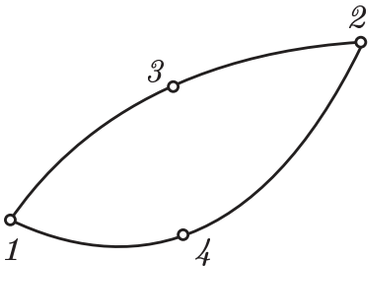


Рис. 54

В обоих случаях работы сил поля одинаковы: $A_{132} = A_{142}$. Если заряд переносится по замкнутому пути *13241*, то на участке *241* работа изменит знак: $A_{241} = -A_{142}$, а потому $A_{132} + A_{241} = A_{13241} = 0$. Значит, при перемещении заряда по любому замкнутому пути работа в электростатическом поле равна нулю. Если перемещаемый заряд единичный, то работа сводится к криволинейному интегралу $\oint \mathbf{E}\, d\mathbf{s}$. Такой интеграл называется *циркуляцией вектора* $\mathbf{E}$ по соответствующему замкнутому контуру. Таким образом, для любого замкнутого контура

$$\oint \mathbf{E}\, d\mathbf{s} = 0. \tag{17.2}$$

Это приводит к другому определению потенциальности поля, эквивалентному данному выше. *Векторное поле* $\mathbf{E}$ *называется потенциальным, если циркуляция вектора* $\mathbf{E}$ *по любому замкнутому контуру равна нулю.*

Уравнение (17.2) есть второе из фундаментальных уравнений электростатики, о которых говорилось в § 15.

3. Из уравнения (17.2) следует, что *линии электростатического поля не могут быть замкнутыми.* Для доказательства допустим противное. Пусть силовая линия замкнута. Возьмем ее в качестве контура интегрирования C. При обходе этого контура в положительном направлении силовой линии подынтегральное выражение в интеграле $\oint \mathbf{E}\, d\mathbf{s}$, а с ним и самый интеграл существенно положительны. Это противоречит уравнению (17.2), что и доказывает наше утверждение.

Невозможны также квазизамкнутые линии потенциального поля. Так мы называем силовые линии, обладающие следующим свойством. Силовая линия, выйдя из любой точки A, извивается и возвращается в сколь угодно малую окрестность той же точки, никогда, однако, не проходя точно через A. Для доказательства опять предположим противное. Пусть A и B — бесконечно близкие точки, через которые проходит силовая линия. Замкнем ее бесконечно малым отрезком, соединяющим эти точки. Так как интеграл $\int \mathbf{E}\, d\mathbf{s}$ вдоль этого отрезка также бесконечно мал, то для циркуляции $\oint \mathbf{E}\, d\mathbf{s}$ вдоль образовавшегося замкнутого контура получилась бы величина, отличная от нуля. А это невозможно.

4. С помощью формулы (17.2) можно строго доказать граничное условие (14.5) для вектора $\mathbf{E}$. Пусть $ABCD$ — бесконечно малый прямоугольный контур, стороны которого AD и BC проходят по разные стороны границы раздела двух сред (рис. 55). Применим к нему формулу (17.2). Боковые стороны AB и CD возьмем бесконечно короткими по сравнению с AD и BC. Тогда можно пренебречь вкладом в циркуляцию, вносимым этими сторонами, и написать

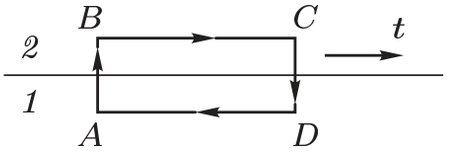


Рис. 55

$$\oint\limits_{ABCD} \mathbf{E}\, d\mathbf{s} = (E_{2t} - E_{1t})\, l,$$

где l — длина стороны BC или равной ей стороны AD. Из обращения в нуль этой циркуляции получаем

$$E_{1t} = E_{2t}. \tag{17.3}$$

§ 18. Электрический потенциал

1. Для потенциальных полей можно ввести понятие *потенциала* или, точнее, *разности потенциалов. Разностью потенциалов* $\varphi_1 - \varphi_2$ *между точками 1 и 2 называется работа, совершаемая силами поля при перемещении единичного положительного заряда по произвольному пути из точки 1 в точку 2.* Такое определение имеет смысл потому, что для потенциальных полей эта работа *не зависит от формы пути*, а определяется только положениями начальной и конечной точек его. Потенциалу какой-либо произвольной точки поля O можно условно приписать любое значение φ_0. Тогда потенциалы всех прочих

точек поля определятся однозначно. Если изменить значение φ_0, то потенциалы в точке O и во всех других точках изменятся на одну и ту же постоянную. Таким образом, *потенциал определен с точностью до аддитивной постоянной.* Значение этой постоянной не играет роли, так как физические явления зависят только от напряженностей электрических полей. Электрические же поля связаны не с абсолютными значениями потенциалов, а с их *разностями* между различными точками пространства. От значения аддитивной постоянной эти поля не зависят. В теоретической физике за нулевой потенциал удобно принимать потенциал бесконечно удаленной точки пространства. На практике за нулевой потенциал обычно принимают потенциал Земли.

Работа сил электростатического поля при перемещении заряда q по произвольному пути из начальной точки *1* в конечную точку *2* определяется выражением

$$A_{12} = q(\varphi_1 - \varphi_2). \tag{18.1}$$

В гауссовой и СГСЭ-системах единиц за единицу потенциала принимается разность потенциалов между такими двумя точками, что при перемещении электростатической единицы электричества из одной точки в другую электрическое поле совершает работу в один эрг. Эта единица не получила специального названия. Практической единицей потенциала является *вольт*. Вольт есть разность потенциалов между такими точками, когда при перемещении одного кулона электричества из одной точки в другую электрическое поле совершает работу в один джоуль. Приближенно

$$1\ \text{В} = \frac{1\ \text{Дж}}{1\ \text{Кл}} = \frac{10^7\ \text{эрг}}{3 \cdot 10^9\ \text{СГСЭ-ед. заряда}} = \frac{1}{300}\ \text{СГСЭ-ед. потенциала}.$$

2. Найдем связь потенциала с напряженностью электрического поля. Пусть *1* и *2* — бесконечно близкие точки, расположенные на оси x, так что $x_2 - x_1 = dx$. Работа при перемещении единицы заряда из точки *1* в точку *2* будет $E_x\,dx$. Та же работа равна $\varphi_1 - \varphi_2 = -d\varphi$. Приравнивая оба выражения, получим $d\varphi = -E_x\,dx$. Аналогичное рассуждение применимо для осей y и z. В результате получаются три соотношения:

$$E_x = -\frac{\partial\varphi}{\partial x}, \quad E_y = -\frac{\partial\varphi}{\partial y}, \quad E_z = -\frac{\partial\varphi}{\partial z}. \tag{18.2}$$

Их можно объединить в одну векторную формулу:

$$\mathbf{E} = -\left(\frac{\partial\varphi}{\partial x}\,\mathbf{i} + \frac{\partial\varphi}{\partial y}\,\mathbf{j} + \frac{\partial\varphi}{\partial z}\,\mathbf{k}\right). \tag{18.3}$$

Так как $\mathbf{E}$ есть вектор, то и выражение, стоящее в скобках, есть также вектор. Он называется *градиентом скаляра* φ и обозначается $\operatorname{grad}\varphi$, или $\nabla\varphi$. Его можно рассматривать как произведение символического

вектора или оператора (4.6) на скаляр φ. Таким образом, по определению

$$\operatorname{grad}\varphi \equiv \nabla\varphi = \frac{\partial\varphi}{\partial x}\,\mathbf{i} + \frac{\partial\varphi}{\partial y}\,\mathbf{j} + \frac{\partial\varphi}{\partial z}\,\mathbf{k}. \tag{18.4}$$

Теперь формулу (18.3) можно записать короче, а именно:

$$E = -\operatorname{grad}\varphi = -\nabla\varphi. \tag{18.5}$$

Произвольное векторное поле $\mathbf{E}(x, y, z)$ характеризуется тремя скалярными функциями — $E_x(x, y, z)$, $E_y(x, y, z)$, $E_z(x, y, z)$. Потенциальность накладывает на поле столь сильное ограничение, что для его характеристики достаточно *одной* скалярной функции, а именно потенциала $\varphi(x, y, z)$. Зная эту функцию, можно вычислить напряженность поля по формуле (18.3) или (18.5). Формулы (18.3) и (18.5) с особой отчетливостью показывают несущественность аддитивной постоянной в выражении для потенциала: при дифференцировании аддитивная постоянная выпадает и не влияет на результат. Те же формулы показывают, что напряженность поля имеет размерность потенциала, деленного на длину. На практике напряженность электрического поля часто выражают в вольтах на сантиметр или в вольтах на метр. Приближенно

$$1\ \text{В/см} \approx \frac{1}{300}\ \text{СГСЭ-единиц}, \quad 1\ \text{В/м} \approx \frac{1}{30\,000}\ \text{СГСЭ-единиц}.$$

3. Для выяснения геометрического смысла градиента введем понятие *эквипотенциальных поверхностей*, или *поверхностей равного потенциала*. Как показывает само название, эквипотенциальная поверхность есть такая поверхность, на которой потенциал остается постоянным. Он может меняться только при переходе от одной эквипотенциальной поверхности к другой. Возьмем на эквипотенциальной поверхности произвольную точку O и введем локальную систему координат с началом в этой точке (рис. 56). Ось z направим по нормали $\mathbf{n}$ к эквипотенциальной поверхности в сторону возрастания потенциала φ. То же направление примем за положительное направление нормали $\mathbf{n}$. Координатная плоскость xy, очевидно, совместится с касательной плоскостью к эквипотенциальной поверхности. Тогда в точке O $\partial\varphi/\partial x = \partial\varphi/\partial y = 0$. Кроме того, $\mathbf{k} = \mathbf{n}$, $\partial\varphi/\partial z \equiv \partial\varphi/\partial n$. Формула (18.4) переходит в

$$\operatorname{grad}\varphi = \frac{\partial\varphi}{\partial n}\,\mathbf{n}. \tag{18.6}$$

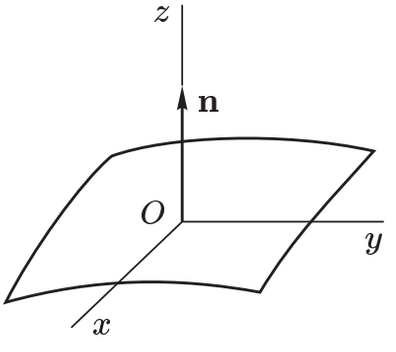


Рис. 56

Функция φ возрастает наиболее быстро в направлении $\mathbf{n}$. Поэтому можно дать следующее определение. *Градиент функции $\varphi(x, y, z)$ есть вектор, направленный в сторону максимального возрастания этой функции, а его длина равна производной функции φ в том же направлении.* Преимущество этого определения состоит в том, что оно

носит *инвариантный характер*, т. е. никак не связано с выбором какой бы то ни было системы координат.

Проведем теперь в каком-либо направлении единичный вектор $\mathbf{s}$. Проекция вектора $\mathbf{A} \equiv \operatorname{grad}\varphi$ на это направление будет $A_s = (\mathbf{As}) = (\mathbf{s}\operatorname{grad}\varphi)$. Но ту же величину можно представить в виде производной $A_s = \partial\varphi/\partial s$, в чем легко убедиться, проведя в направлении $\mathbf{s}$ координатную ось и воспользовавшись одной из формул (18.2). Таким образом, получается соотношение

$$\frac{\partial\varphi}{\partial s} = (\mathbf{s}\operatorname{grad}\varphi). \tag{18.7}$$

Производная функции φ в каком-либо направлении равна проекции вектора градиента этой функции на то же направление. Ясно, что эта производная максимальна, когда вектор $\mathbf{s}$ направлен вдоль $\operatorname{grad}\varphi$, т. е. по нормали к эквипотенциальной поверхности.

4. Вектор $\mathbf{E}$ направлен противоположно вектору градиента потенциала φ. Электрические силовые линии являются, таким образом, линиями, вдоль которых потенциал φ изменяется *наиболее быстро.* Они *нормальны* к эквипотенциальным поверхностям. Эквипотенциальные поверхности могут служить поэтому для наглядного изображения картины поля. Обычно их чертят так, что при переходе от одной эквипотенциальной поверхности к соседней потенциал получает *одно и то же* приращение $\Delta\varphi$. Чем меньше выбрано $\Delta\varphi$, тем детальнее будет представлено распределение потенциала в пространстве, а с ним и картина электростатического поля. Для большей наглядности чертят также силовые линии, ортогональные к семейству поверхностей равного потенциала. Там, где (при постоянном $\Delta\varphi$) соседние эквипотенциальные поверхности наиболее близко подходят друг к другу, напряженность электрического поля максимальна. Наоборот, в местах, где расстояния между ними велики, будет мала и напряженность поля $\mathbf{E}$. Поверхность проводника есть одна из эквипотенциальных поверхностей, и силовые линии должны подходить к ней нормально. Внутри проводника $E = 0$, а потому потенциал φ должен иметь одно и то же значение во всех точках проводника. Здесь эквипотенциальная поверхность вырождается в эквипотенциальный объем.

ЗАДАЧИ

1. Доказать формулу

$$\operatorname{div}(\varphi\mathbf{A}) = (\mathbf{A}\operatorname{grad}\varphi) + \varphi\operatorname{div}\mathbf{A}. \tag{18.8}$$

2. Доказать формулу

$$\operatorname{grad}(\mathbf{ar}) = \mathbf{a}, \tag{18.9}$$

где $\mathbf{a}$ — постоянный вектор.

3. Из трех концентрических бесконечно тонких металлических сфер с радиусами $R_1 < R_2 < R_3$, находящихся в вакууме, крайние заземлены, а средней сообщен электрический заряд Q. Найти напряженность электрического поля во всем пространстве.

Ответ. Электрическое поле радиально и определяется выражениями

$$E=\begin{cases}0, & \text{если } r<R_1 \text{ или } R_3<r<\infty,\\ \dfrac{R_1(R_2-R_3)}{R_2(R_3-R_1)}\,\dfrac{Q}{r^2}, & \text{если } R_1<r<R_2,\\ \dfrac{R_3(R_2-R_1)}{R_2(R_3-R_1)}\,\dfrac{Q}{r^2}, & \text{если } R_2<r<R_3,\end{cases}$$

где r — расстояние от центра сфер.

4. Из трех параллельных металлических пластинок A, B и C (рис. 57) крайние A и B неподвижны и соединены с гальванической батареей, поддерживающей разность потенциалов V между ними постоянной. Средняя пластинка C сначала находится в контакте с верхней пластинкой A. Затем с помощью изолирующей ручки она перемещается по направлению к нижней пластинке. Пренебрегая краевыми эффектами, найти напряженности полей E_1 и E_2 в зазорах между пластинками в зависимости от переменного расстояния x.

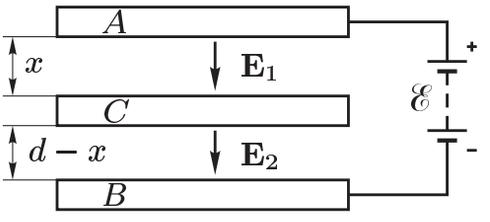


Рис. 57

Ответ. $E_1=Vx/d^2$, $E_2=V(x+d)/d^2$.

5. Внутри плоского воздушного конденсатора, обкладки которого соединены между собой, помещена поляризованная пластинка из электрета толщины h. Вектор поляризации пластинки $\mathbf{P}$ перпендикулярен к ее боковым граням. Пренебрегая зависимостью поляризованности электрета $\mathbf{P}$ от электрического поля, определить напряженность и электрическое смещение поля внутри и вне пластинки, если расстояние между обкладками конденсатора AB равно d.

Ответ. В воздушном зазоре $E=D=4\pi hP/d$. Внутри пластинки $E=-4\pi(d-h)P/d$, $D=4\pi hP/d$.

§ 19. Вычисление потенциала по напряженности поля

Если известен потенциал $\varphi(x,y,z)$, то напряженность электрического поля можно вычислить его *дифференцированием* по координатам. Обратная задача вычисления потенциала по напряженности поля решается *интегрированием*. Исходными являются формулы (18.2) или им аналогичные. Рассмотрим простейшие примеры на вычисление потенциала.

1. Потенциал поля точечного заряда q в однородном диэлектрике. Электрическое смещение поля определяется выражением (13.10). Из него получаем

$$E=E_r=\frac{q}{\varepsilon r^2}=-\frac{d\varphi}{dr},$$

или после интегрирования

$$\varphi=-\frac{q}{\varepsilon r}+\text{const}.$$

В качестве постоянной интегрирования следует взять нуль, чтобы при $r = \infty$ потенциал φ обратился в нуль. Тогда

$$\varphi = \frac{q}{\varepsilon r}. \tag{19.1}$$

2. Потенциал поля системы точечных зарядов в однородном диэлектрике. На основании принципа суперпозиции из (19.1) получаем

$$\varphi = \frac{1}{\varepsilon} \sum \frac{q_i}{r_i}, \tag{19.2}$$

где r_i — расстояние i-го заряда до точки наблюдения. Суммирование ведется по всем зарядам.

3. Потенциал непрерывно распределенных электрических зарядов. Рассматривая заряды элементов объема и поверхности как точечные и применяя формулу (19.2), для потенциала φ в однородном диэлектрике можно написать

$$\varphi(\mathbf{r}) = \frac{1}{\varepsilon} \int \frac{\rho(\mathbf{r}')\, dV'}{|\mathbf{r} - \mathbf{r}'|} + \frac{1}{\varepsilon} \int \frac{\sigma(\mathbf{r}')\, dS'}{|\mathbf{r} - \mathbf{r}'|}. \tag{19.3}$$

Здесь dV' и dS' — элементы объема и заряженных поверхностей с центрами в точке $\mathbf{r}'$. Интегрирование ведется по всем объемным и поверхностным зарядам. Если диэлектрик неоднороден, то интегрирование надо распространить и на поляризационные заряды. Включение таких зарядов автоматически учитывает влияние среды, и величину ε вводить не надо:

$$\varphi(\mathbf{r}) = \int \frac{\rho(\mathbf{r}') + \rho_{\text{пол}}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}\, dV' + \int \frac{\sigma(\mathbf{r}') + \sigma_{\text{пол}}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}\, dS'. \tag{19.4}$$

Формула (19.3) является частным случаем этой формулы. Влияние поляризационных зарядов в ней учитывается посредством ε.

4. Потенциал бесконечной равномерно заряженной плоскости в однородном диэлектрике. В этом случае

$$\varphi = \begin{cases} -2\pi\sigma x/\varepsilon + C & \text{при } x > 0, \\ +2\pi\sigma x/\varepsilon + C & \text{при } x < 0, \end{cases} \tag{19.5}$$

Начало координат помещено на заряженной плоскости, ось X направлена перпендикулярно к ней. Постоянная C одна и та же в обоих выражениях, так как при переходе через заряженную плоскость потенциал должен изменяться непрерывно.

Никаким выбором постоянной C нельзя добиться обращения потенциала в нуль в бесконечности. Это связано с тем, что в рассматриваемом случае в бесконечности имеются не только поля, но и сами заряды. Для плоскости конечных размеров выражениями (19.5) можно пользоваться только при таких x, которые малы по сравнению с размерами плоскости. При x порядка размеров плоскости выражение для φ

становится очень сложным. На очень больших расстояниях плоскость ведет себя как точечный заряд. Разумеется, для конечной плоскости постоянную C в формулах (19.5) всегда можно выбрать так, чтобы в бесконечности потенциал обратился в нуль. Однако для вычисления C надо знать выражение для потенциала не только вблизи плоскости, но и на любых расстояниях от нее.

ЗАДАЧИ

Всюду предполагается, что диэлектрик однороден.

1. Вычислить потенциал поля точечного диполя.

Ответ.

$$\varphi = \frac{(\mathbf{pr})}{\varepsilon r^3}. \tag{19.6}$$

2. Дифференцированием (19.6) вычислить напряженность поля точечного диполя.

3. Вычислить потенциал поля шара радиуса a, равномерно заряженного по объему.

Ответ. Вне шара φ определяется формулой (19.1). Внутри шара

$$\varphi = \frac{2\pi\rho}{3\varepsilon}\,(3a^2 - r^2). \tag{19.7}$$

4. Вычислить потенциал поля сферы радиуса a, равномерно заряженной по поверхности.

Ответ. Вне сферы φ определяется формулой (19.1). Внутри сферы

$$\varphi = \frac{q}{\varepsilon a} = \text{const}. \tag{19.8}$$

5. Вычислить потенциал поля равномерно заряженной бесконечной плоскопараллельной пластинки толщины $2a$.

Ответ. Внутри пластинки

$$\varphi = -\frac{2\pi\rho}{\varepsilon}\,(x^2 + a^2) + C, \tag{19.9}$$

вне пластинки

$$\varphi = \begin{cases} -4\pi\rho a x/\varepsilon + C & \text{при } x > a, \\ +4\pi\rho a x/\varepsilon + C & \text{при } x < a. \end{cases} \tag{19.10}$$

Начало координат помещено в средней плоскости пластинки, ось X направлена нормально к ней.

6. Вычислить потенциал поля бесконечно длинной и бесконечно тонкой прямолинейной нити, равномерно заряженной электричеством с линейной плотностью $\varkappa$.

Ответ.

$$\varphi = -\frac{2\varkappa}{\varepsilon}\,\ln r + C, \tag{19.11}$$

где r — расстояние до нити.

7. Вычислить потенциал поля бесконечно длинного цилиндра радиуса a, равномерно заряженного по объему.

Ответ.

$$\varphi = \begin{cases} \dfrac{\pi}{\varepsilon}\,\rho(a^2 - r^2) + C & \text{при } r \leqslant a, \\ -\dfrac{2\pi a^2 \rho}{\varepsilon}\,\ln\dfrac{r}{a} + C & \text{при } r \geqslant a. \end{cases} \tag{19.12}$$

8. Вычислить потенциал поля бесконечно длинного цилиндра радиуса a, равномерно заряженного по поверхности.

Ответ.

$$\varphi = \begin{cases} C & \text{при } r \leqslant a, \\ -\dfrac{2\pi a\sigma}{\varepsilon}\,\ln\dfrac{r}{a} + C & \text{при } r \geqslant a. \end{cases} \tag{19.13}$$

9. Показать, что эквипотенциальными поверхностями двух параллельных бесконечно длинных прямых, равномерно заряженных электричествами противоположных знаков, являются круговые цилиндры, оси которых параллельны рассматриваемым линиям и лежат с ними в одной плоскости.

Указание. Уравнение эквипотенциальной поверхности имеет вид $r_1/r_2 = \mathrm{const}$, где r_1 и r_2 — расстояния до рассматриваемых прямых. Записав это уравнение в координатах, нетрудно убедиться, что эквипотенциальными поверхностями будут круговые цилиндры.

§ 20. Измерение разности потенциалов электрометром. Электрический зонд

1. Электроскоп, листочки или стрелка которого окружены металлической оболочкой, может служить *электрометром*, т. е. прибором для измерения разности потенциалов между двумя проводниками. Один из проводников соединяют с шариком электрометра, а другой — с оболочкой (рис. 58). Стрелка электрометра примет потенциал тела A, оболочка — потенциал тела B. Возникнет электрическое поле с силовыми линиями, идущими от оболочки к стрелке или обратно. Угол отклонения стрелки определяется напряженностью и конфигурацией этого поля. Но поле внутри замкнутой металлической оболочки совершенно не зависит от наружного поля. Оно однозначно определяется разностью потенциалов между оболочкой и стрелкой. Следовательно, угол отклонения стрелки может служить мерой разности потенциалов между телами A и B. Электрометр можно проградуировать, чтобы сразу отсчитывать разность потенциалов в вольтах. Обычно в качестве тела B берут Землю, т. е. оболочку электрометра заземляют. Тогда электрометр покажет потенциал тела A относительно Земли.

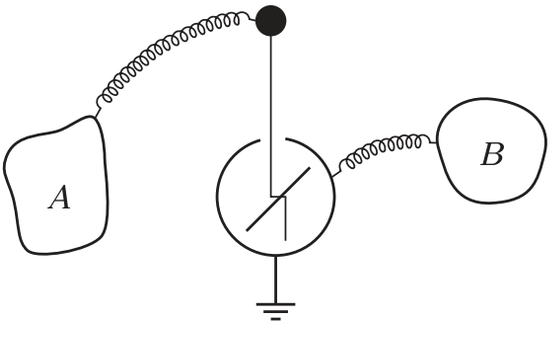


Рис. 58

2. В принципе безразлично, что заземлять: оболочку или шарик. От этого зависит только направление силовых линий, но не их конфигурация и значение напряженности поля внутри оболочки. Угол отклонения стрелки в обоих случаях будет одинаков. Установим электрометр на диэлектрике и заземлим его оболочку. Затем с помощью наэлектризованной палочки зарядим шарик. Стрелка отклонится. Заземлим теперь шарик, а оболочку изолируем. Той же палочкой электризуем оболочку, стрелка отклоняется так же. Но стрелка электрометра защищена от влияния внешних электрических зарядов, так как она окружена металлической оболочкой. Почему же при электризации оболочки стрелка отклоняется? Дело в том, что ее защита *не полная.* Оболочка не замкнута. В ней есть отверстие, через которое проходит металлический стержень, соединяющий стрелку с шариком электрометра. Электричество, сообщенное наружной поверхности оболочки, индуцирует заряды на шарике, которые по стержню переходят на стрелку электрометра. Заряженные стрелка и стержень оттягивают часть электричества с наружной поверхности оболочки на внутреннюю. В результате появляется электрическое поле не только снаружи, но и внутри оболочки. Мы видим, что электрометр только тогда может служить измерительным прибором, когда его стрелка защищена от внешних электрических полей не полностью. Но связь стрелки с внешними телами должна быть слабой. Для этого отверстие в оболочке, шарик электрометра и наружная часть стержня, соединяющего шарик со стрелкой, должны быть малыми. Иначе на этих частях электрометра могли бы возникать заметные заряды, индуцированные посторонними внешними телами. Переходя на стрелку, такие заряды вносили бы заметные искажения в измеряемую разность потенциалов. По той же причине провода, соединяющие тела A и B с шариком и оболочкой электрометра, должны быть тонкими.

3. Поучителен также следующий опыт. Пробный шарик на изолирующей ручке соединен гибким проводом с шариком электрометра, оболочка которого заземлена (рис. 59). При перемещении пробного шарика по поверхности заряженного металлического тела C показание электрометра не меняется. Это и понятно. Электрометр измеряет разность потенциалов между шариками. Потенциал же пробного шарика не зависит от того, какой точки проводящего тела C он касается. В аналогичном опыте, описанном в § 11 (см. рис. 33), электрометр не чувствовал заряда, снятого с впадины B. Отсюда не следует, что в точке B поверхностная плотность электричества равна нулю. Можно лишь сказать, что прежний опыт недостаточно чувствителен, чтобы ее обнаружить.

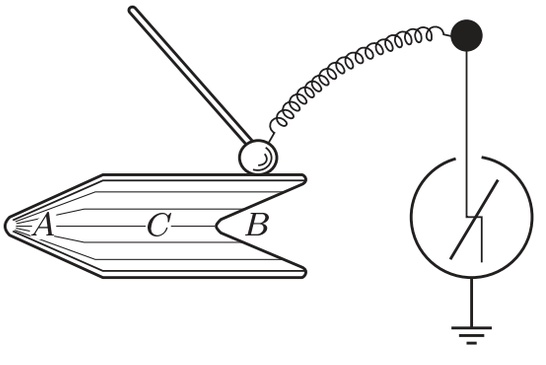


Рис. 59

Новый опыт позволяет определенно утверждать, что плотность электричества у B хотя и мала, но не равна нулю. Если бы это было не так, то стрелка электрометра не могла бы отклоняться, когда незаряженным пробным шариком касаются проводника C в точке B. Заряд, перешедший с тела C на пробный шарик и электрометр, не зависит от положения точки касания. От этого может зависеть только время зарядки электрометра. Оно хотя и очень мало, но всегда конечно. Когда пробный шарик касается острия A, зарядка происходит всего быстрее; когда он касается впадины B, электрометр заряжается медленнее.

4. Для измерения потенциала в различных точках жидкого или газообразного диэлектрика можно пользоваться методом *электрического зонда*. Зонд представляет собой малое металлическое тельце (кончик проволочки, выступающий из диэлектрической трубочки, шарик или диск), соединенное проволокой с шариком электрометра, оболочка которого заземлена. Он вводится в ту точку диэлектрика, потенциал которой надо измерить. Очевидно, электрометр укажет разность потенциалов между стрелкой и оболочкой или, что то же самое, между зондом и Землей. Однако зонд, соединенный с электрометром, вообще говоря, существенно меняет потенциал той точки пространства, в которую он вносится. Причиной этого являются индукционные заряды, появляющиеся на зонде и шарике электрометра, с которым он соединен.

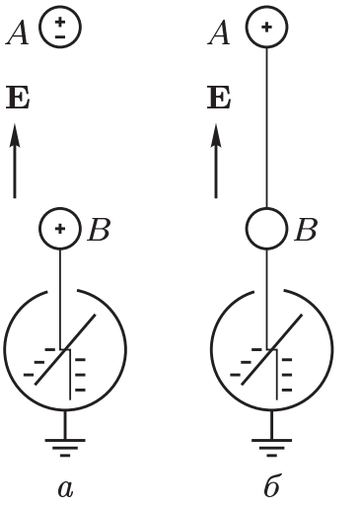


Рис. 60

Допустим сначала, что маленький зонд A изолирован (рис. 60 *а*). При введении в электрическое поле зонд поляризуется. Положительные и отрицательные заряды немного смещаются в противоположных направлениях. Так как эти смещения малы, то внесение зонда исказит поле в окрестности точки A очень мало. На шарике и стрелке электрометра также возникают индукционные заряды противоположных знаков. Соединим теперь тонкой проволокой зонд A с шариком электрометра B (рис. 60 *б*). Тогда на зонде A останется электричество только одного знака, электричество противоположного знака перейдет на электрометр. Так как количество перешедшего электричества очень мало, то потенциал шарика B и стержня электрометра при этом практически не изменится. Не то будет с потенциалом зонда A. Геометрические размеры зонда малы. Небольшой заряд, ушедший с зонда, существенно меняет его потенциал. Зонд A принимает потенциал шарика и стержня электрометра. Таким образом, соединение зонда с электрометром практически не сказывается на угле отклонения стрелки. Показание электрометра не имеет никакого отношения к потенциалу, существовавшему в точке A до внесения в нее зонда.

Для возможности измерения потенциала необходимо, чтобы при внесении зонда в точку A зонд и соединенный с ним шарик электромет-

ра приняли потенциал, существовавший в точке A до внесения зонда. Этого можно достигнуть, убирая индукционные заряды, образующиеся на зонде. В *капельном зонде* телом A служит малое ведерко, наполненное проводящей жидкостью. В дне ведерка имеется малое отверстие. Капли жидкости, уходящие через это отверстие, уносят индукционные заряды, образующиеся на зонде. При этом заряды противоположного знака переходят с зонда на стержень и стрелку электрометра. Благодаря этому угол отклонения стрелки изменяется. В установившемся состоянии, когда зонд оказывается незаряженным, потенциал зонда становится равным потенциалу окружающего пространства. А так как зонд соединен проводником с шариком электрометра B, то потенциал последнего будет таким же. Электрометр покажет тот потенциал, который надо измерить.

Удаление индукционных зарядов с зонда можно осуществить и другими способами. Например, зондом может служить кончик металлической проволоки, выступающий из диэлектрической трубочки, играющей роль газовой горелки (*пламенный зонд*). Благодаря высокой температуре пламени окружающий воздух немного ионизуется и становится проводящим. Образующиеся ионы снимают индукционные заряды с зонда и уносятся потоком окружающего газа. Аналогичная идея используется в *радиоактивном зонде*. На зонд наносится небольшое количество радиоактивного вещества, которое и создает нужную ионизацию окружающего газа. В том и другом случае на зонде устанав-

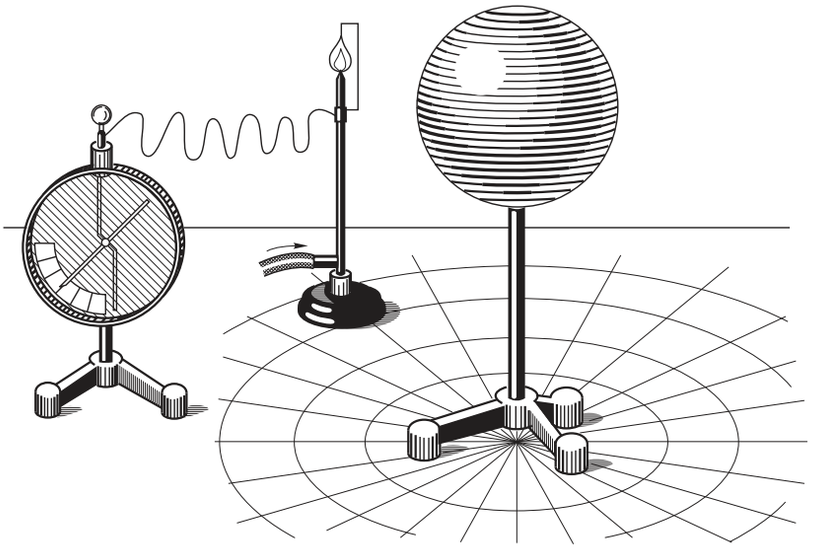


Рис. 61

ливается потенциал, равный потенциалу окружающего пространства до внесения в него зонда. То же самое происходит в проводящих жидкостях (электролитах), в которые погружены заряженные электроды. Здесь удаление индуцированных зарядов с зонда обеспечивается электрической проводимостью самой жидкости.

5. Следующий демонстрационный опыт может служить для иллюстрации действия зонда. Холодный пламенный зонд помещается вблизи

большого изолированного металлического шара (рис. 61). Пока зонд холодный, стрелка удаленного от него электрометра не отклоняется, если даже шар сильно зарядить. Но отклонение стрелки становится большим, если зажечь светильный газ, вытекающий из трубочки зонда. При приближении электрометра к шару отклонение стрелки увеличивается, при удалении — уменьшается. Стрелка стоит на месте при перемещении зонда по сферической эквипотенциальной поверхности, концентрической с поверхностью шара.

§ 21. Электрическое поле Земли

1. Электрическим зондом можно измерять и напряженность поля в диэлектриках, например газах. Надо измерить потенциал φ в различных точках пространства, а затем вычислить его градиент. Проще и точнее, однако, напряженность поля измеряется непосредственно. Один из способов состоит в следующем. Две параллельные металлические пластины (воздушный конденсатор) устанавливаются перпендикулярно к направлению измеряемого электрического поля (рис. 62). Пластины соединяются между собой через баллистический гальванометр, с помощью которого можно измерять величину заряда, прошедшего через цепь. Так как пластины соединены, то их потенциалы одинаковы, а напряженность электрического поля между ними равна нулю. Поле существует только снаружи пластин и связано с поверхностной плотностью электричества соотношением $E = 4\pi\sigma$ (в газах различием между $\mathbf{D}$ и $\mathbf{E}$ можно пренебречь). Если зазор между пластинами мал по сравнению с их размерами, то можно также пренебречь влиянием краев пластин, считая величину σ постоянной. Пластины быстро поворачивают на 90° в вертикальное положение. При этом через баллистический гальванометр проходит заряд $q = S\sigma$, где S — площадь пластины. Измерив q, можно вычислить σ, а затем и напряженность поля E.

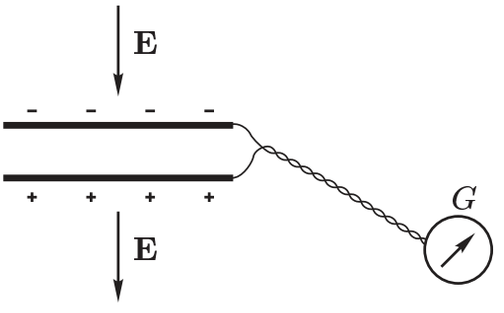


Рис. 62

2. Подобные измерения показали, что земной шар заряжен отрицательно. Земное электрическое поле меняется во времени. Эти изменения могут быть регулярными (суточными и годичными) и нерегулярными. В среднем напряженность поля у самой поверхности Земли составляет 130 В/м. Между двумя уровнями, отстоящими друг от друга на рост человека, существует разность потенциалов примерно 200 В. Человек не чувствует этой разности потенциалов и его не поражает ток потому, что сам он является хорошим проводником электричества. Как и всякий проводник, тело человека сильно искажает электрическое поле. Электрические силовые линии подходят к поверхности тела человека нормально, а эквипотенциальные поверхности огибают его совершенно так же, как они огибают металлический предмет. Все точки тела человека находятся под одним и тем же потенциалом.

Зная напряженность поля вблизи земной поверхности, нетрудно подсчитать, что полный отрицательный заряд Земли составляет около $6 \cdot 10^5$ Кл. На высоте 1 км напряженность земного поля падает примерно до 40 В/м. На высоте 10 км поле не превышает нескольких вольт на метр. На высоте 50 км и больше земное поле едва заметно. Большая часть падения потенциала приходится на малые высоты. Полная разность потенциалов

между поверхностью Земли и верхними слоями атмосферы составляет почти 400 000 В. Эти данные показывают, что земная атмосфера заряжена положительно.

3. Воздух обладает электрической проводимостью. Она обусловлена ионами, образующимися в результате ионизации молекул и атомов космическими лучами Благодаря проводимости в атмосфере текут токи, стремящиеся разрядить Землю. Средняя плотность тока порядка 10^{-6} мкА/м^2. Полный электрический ток, достигающий земной поверхности, равен приблизительно 1800 А. Заряд Земли, как указывалось выше, равен примерно $6 \cdot 10^5$ Кл. Используя эти данные, нетрудно подсчитать, что заряд Земли должен был бы убывать в два раза примерно через каждые 4 минуты Земля должна была бы полностью потерять свой заряд менее чем за полчаса. Если этого не происходит, то существуют какие-то процессы, непрерывно заряжающие Землю отрицательным электричеством и поддерживающие ее средний заряд неизменным. Долгое время оставалось загадкой, что это за процессы. В настоящее время, по-видимому, можно считать установленным, что заряд Земли поддерживается грозовой деятельностью атмосферы.

Гроза является одним из величественных явлений природы. Хотя электрическая природа молнии давно установлена, в учении об атмосферном электричестве еще очень много неисследованного и неясного. Далеко не ясен даже механизм возникновения электрических зарядов в атмосфере. Однако, по-видимому, можно считать установленным, что заряд земного шара поддерживается молниями, проскакивающими между атмосферой и поверхностью Земли. С каждым ударом молнии на Землю низвергается в среднем 20–30 Кл отрицательного электричества. Ричард Фейнман (1918–1988) в своих известных лекциях по физике приводит следующие аргументы в подтверждение этой точки зрения. При ясной погоде над поверхностью моря систематически измерялся электрический ток, идущий на Землю. Оказалось, что этот ток, если его усреднить по земной поверхности, испытывает суточные вариации. Ток в максимуме примерно на 15 % больше среднего суточного тока, причем максимум приходится примерно на 19 часов по лондонскому времени. Такие же вариации испытывает и напряженность земного электрического поля. Самое интересное здесь то, что момент максимума *один и тот же* для всех точек земного шара *не по местному времени*, *а по единому* (например, лондонскому) *времени*. Это не так уж удивительно, потому что верхние слои атмосферы (ионосфера) сильно ионизованы и обладают высокой электрической проводимостью. Высокой электрической проводимостью обладает и морская вода. Имеются, таким образом, две хорошо проводящие сферические оболочки, между которыми и возбуждается электрическое поле Земли, как между обкладками сферического конденсатора. В самих оболочках из-за их высокой электрической проводимости не может возникнуть сколько-нибудь существенной разности потенциалов. С другой стороны, установлено, что максимум грозовой деятельности, усредненный по всей поверхности земного шара, приходится также на 19 часов по лондонскому времени. Это подтверждает гипотезу о наличии связи между грозовой деятельностью атмосферы и электрическим полем Земли.

ЗАДАЧА

Земля непрерывно облучается космическими лучами высокой энергии, приходящими из пространства вне Солнечной системы. Космические лучи в основном состоят из протонов, средняя энергия $\mathscr{E}$ которых составляет несколько миллиардов электронвольт. Интенсивность I потока протонов, достигающих земной атмосферы, равна примерно одному протону на квад-

ратный сантиметр в секунду. Оценить время, необходимое для того, чтобы протоны космических лучей повысили потенциал Земли настолько, чтобы они уже не могли попадать на поверхность Земли из-за электрического отталкивания. Объяснить, почему и после этого времени протоны космических лучей продолжают достигать земной поверхности.

Ответ. $t = \mathscr{E}/(4\pi RIe^2) \approx 10^6$ с ≈ 10 сут (R — радиус Земли, e — заряд протона; числовой ответ получен при $\mathscr{E}/e \sim 10^9$ В). В действительности потенциал Земли не может достигать такого значения, так как наряду с приходом протонов существует обратный процесс, в котором земная атмосфера теряет положительные заряды в виде протонов и положительных ионов, уходящих в космическое пространство под действием возникшего электрического поля.

§ 22. Общая задача математической электростатики

1. Если известен потенциал φ как функция пространственных координат, то его дифференцированием можно вычислить напряженность электрического поля по формуле (18.5). Зная диэлектрическую проницаемость, можно затем определить вектор электрического смещения $\mathbf{D} = \varepsilon\mathbf{E}$ и по теореме Гаусса (13.5) найти объемную плотность свободных зарядов ρ. Поверхностная плотность свободных зарядов найдется по скачку нормальной составляющей вектора $\mathbf{D}$ из соотношения (14.1).

Наоборот, если известны плотности свободных и связанных зарядов, то интегрированием можно вычислить потенциал по формуле (19.3) или (19.4), а затем найти и все остальные величины. Как правило, интегралы (19.3) или (19.4) не берутся аналитически, но они всегда могут быть найдены численно.

Реальные задачи, к которым приводит электростатика, гораздо сложнее. Дело в том, что связанные заряды, а также распределение свободного электричества по поверхности проводников не бывают известными, а сами подлежат определению. Общая задача математической электростатики формулируется следующим образом.

В диэлектрической среде заданы расположение и форма всех проводников. Известна диэлектрическая проницаемость среды ε между проводниками и объемная плотность свободных электрических зарядов во всех точках диэлектриков. Кроме того, известны: а) либо потенциалы всех проводников, б) либо заряды всех проводников, в) либо заряды некоторых проводников и потенциалы всех остальных проводников. Требуется определить напряженность электрического поля во всех точках пространства и распределение электричества по поверхностям проводников.

2. Задача сводится к нахождению потенциала φ как функции пространственных координат x, y, z. Найдем дифференциальное уравнение, которому должна удовлетворять эта функция. Для этого теорему Гаусса (13.5) запишем в виде $\operatorname{div}(\varepsilon\mathbf{E}) = 4\pi\rho$ и подставим в нее выражение для $\mathbf{E}$ из формулы (18.5). Получим

$$\operatorname{div}(\varepsilon \operatorname{grad}\varphi) = -4\pi\rho, \tag{22.1}$$

или в координатной форме

$$\frac{\partial}{\partial x}\left(\varepsilon\frac{\partial\varphi}{\partial x}\right)+\frac{\partial}{\partial y}\left(\varepsilon\frac{\partial\varphi}{\partial y}\right)+\frac{\partial}{\partial z}\left(\varepsilon\frac{\partial\varphi}{\partial z}\right)=-4\pi\rho. \tag{22.1a}$$

Если диэлектрик однороден (ε не зависит от координат), то

$$\operatorname{div}\operatorname{grad}\varphi=-\frac{4\pi\rho}{\varepsilon}, \tag{22.2}$$

или

$$\frac{\partial^2\varphi}{\partial x^2}+\frac{\partial^2\varphi}{\partial y^2}+\frac{\partial^2\varphi}{\partial z^2}=-\frac{4\pi\rho}{\varepsilon}. \tag{22.2a}$$

Введем так называемый *оператор Лапласа*, или *лапласиан*:

$$\Delta\equiv\nabla^2\equiv\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}. \tag{22.3}$$

Тогда

$$\frac{\partial^2\varphi}{\partial x^2}+\frac{\partial^2\varphi}{\partial y^2}+\frac{\partial^2\varphi}{\partial z^2}\equiv\Delta\varphi\equiv\nabla^2\varphi, \tag{22.4}$$

и уравнение (22.2а) запишется в более кратком виде:

$$\Delta\varphi=-\frac{4\pi\rho}{\varepsilon}. \tag{22.5}$$

Это уравнение называется *уравнением Пуассона*. При отсутствии свободных зарядов ($\rho=0$) оно переходит в *уравнение Лапласа*:

$$\Delta\varphi=0. \tag{22.6}$$

Общая электростатическая задача сводится к нахождению решения дифференциального уравнения (22.1), удовлетворяющего всем условиям, перечисленным выше. Можно показать, что такая задача не может иметь более одного решения [1]. Для того чтобы не повторять дважды одни и те же математические выкладки, мы отложим доказательство этой *теоремы единственности* до § 29. Нахождение самого решения, вообще говоря, задача очень сложная. Аналитические решения известны лишь для немногих частных случаев. Однако если удалось угадать функцию φ, удовлетворяющую всем условиям задачи, то можно утверждать, что она и будет искомым (единственным) решением задачи. В этом и состоит значение теоремы единственности.

Не всегда требуется задавать заряды или потенциалы тел *во всем пространстве*. Если требуется найти электрическое поле в полости, окруженной проводящей оболочкой, то это достаточно сделать только для тел, заключенных *внутри самой полости*. Наоборот, если требуется найти внешнее поле, то надо задать заряды или потенциалы только тел *вне проводящей оболочки*, а также общий заряд или потенциал на внешней поверхности этой оболочки. Теорема единственности распространяется и на эти случаи.

[1] Если заданы только заряды всех проводников, то потенциал φ определяется с точностью до несущественной аддитивной постоянной.

Приведем два примера на применение теоремы единственности.

3. Пусть в вакууме распределено электричество с объемной плотностью ρ. Потенциал электрического поля φ_0 будет удовлетворять уравнению $\Delta\varphi_0 = -4\pi\rho$. Заполним все пространство однородным диэлектриком с диэлектрической проницаемостью ε, оставляя свободные заряды на прежних местах. Тогда потенциал изменится и будет удовлетворять уравнению (22.5). Сравнив оба уравнения и применив теорему единственности, найдем $\varphi = \varphi_0/\varepsilon$. Таким образом, заполнение пространства однородным диэлектриком уменьшает потенциал, а с ним и напряженность электрического поля в ε раз. Результат остается верным и при наличии поверхностных зарядов.

4. В качестве второго примера рассмотрим теорему Фарадея (§ 11, п. 5). Часть результатов, относящихся к этой теореме, была получена нами интуитивно на основе физических соображений. Теорема единственности позволяет обосновать их строго. Пусть в полости, окруженной проводящей оболочкой, нет электрических зарядов. Потенциал внутри полости удовлетворяет уравнению Лапласа (22.6). На стенках полости он должен принимать какое-то постоянное значение C. Решение уравнения (22.6), удовлетворяющее этому условию, можно указать сразу. Это есть $\varphi(x, y, z) = C$. По теореме единственности других решений быть не может. Таким образом, внутри полости $\mathbf{E} = -\operatorname{grad}\varphi = 0$, что и доказывает теорему Фарадея.

§ 23. Метод электрических изображений

1. В ближайших четырех параграфах рассматриваются некоторые задачи на вычисление электростатических полей. Все они решаются *искусственными методами.* Начнем с *метода электрических изображений.*

Допустим, что в пространстве имеется несколько точечных электрических зарядов. Пусть S — какая-либо эквипотенциальная поверхность, разделяющая все пространство на два полупространства: I и I' (рис. 63). Обозначим через q_1, q_2, ... точечные заряды в полупространстве I, а через q_1', q_2', ... — в полупространстве I'. Точечные заряды можно рассматривать как предельные случаи малых проводящих тел, например металлических шариков. К ним применима теорема единственности. Заданием величины и расположения зарядов q_1, q_2, ..., а также потенциала поверхности S поле в полупространстве I определяется однозначно. Аналогичное утверждение справедливо для полупространства I'. Поэтому если поверхность S сделать проводящей, то поле во всем пространстве не претерпит никаких изменений. Однако поля в полупространствах I и I' сделаются теперь совершенно независимыми друг от друга. В результате мы получаем решение сразу двух задач, вполне аналогичных друг другу. Одна из них состоит в следующем.

В полупространстве I по одну сторону от поверхности проводящего тела S находятся точечные заряды $q_1, q_2, \ldots$ Найти электрическое поле

в этом полупространстве. Конечно, это поле векторно складывается из полей зарядов q_1, q_2, ... и зарядов, индуцированных на поверхности тела S. Однако в силу теоремы единственности, поле индуцированных зарядов в полупространстве I эквивалентно полю, создаваемому зарядами q_1', q_2', ... При вычислении искомого поля проводящее тело можно убрать и заменить его точечными зарядами q_1', q_2', ... Совокупность этих последних зарядов называется *электрическим изображением* зарядов q_1, q_2, ... в поверхности S. Таким образом, задача об электрическом поле зарядов, расположенных по одну сторону от проводящей поверхности, сводится к отысканию электрических изображений этих зарядов в этой поверхности.

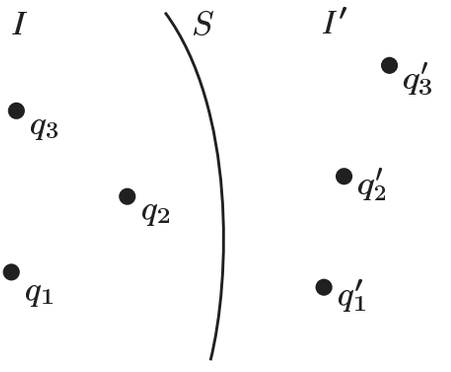


Рис. 63

Приведем два примера на метод электрических изображений.

2. Точечный заряд q над бесконечной плоской поверхностью проводника. Электрическим изображением заряда q в плоскости AB будет заряд противоположного знака $q' = -q$, расположенный по другую сторону плоскости AB на таком же расстоянии от нее, что и заряд q (рис. 64). Действительно, потенциал поля точечных зарядов q и q' в какой-либо точке C над поверхностью проводника будет

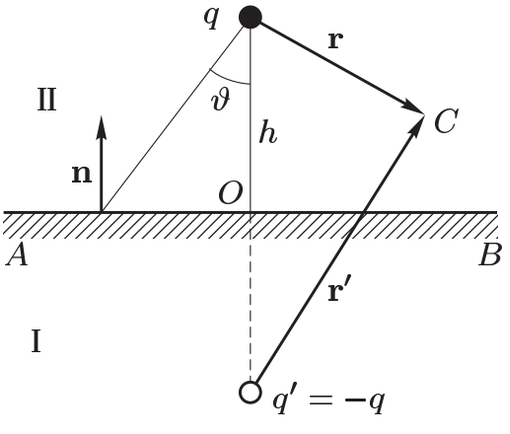


Рис. 64

$$\varphi = q\left(\frac{1}{r} - \frac{1}{r'}\right). \qquad (23.1)$$

Он обращается в нуль на плоскости AB, а потому эта плоскость является эквипотенциальной. Формула (23.1) и определяет потенциал поля в верхнем полупространстве I. В нижнем полупространстве II, заполненном проводящей средой, поле, разумеется, равно нулю. Таким образом, заряд q индуцирует на плоскости AB такие заряды, которые создают в верхнем полупространстве Iтакое же поле, что и вспомогательный точечный заряд $q' = -q$. Отсюда следует, что индуцированные заряды притягивают заряд q с той же силой, что и вспомогательный точечный заряд $q' = -q$, т. е. с силой $F = q^2/(2h)^2$, где h — расстояние между зарядом q и плоскостью AB. Поэтому эта сила называется *силой электрического изображения*. В нижнем полупространстве IIиндуцированные заряды компенсируют поле заряда q.

Поверхностная плотность электричества найдется по формуле

$$\sigma = \frac{1}{4\pi} E_n = -\frac{1}{4\pi} \frac{\partial \varphi}{\partial n}.$$

Простое вычисление дает

$$\sigma = -\frac{q}{2\pi h^2}\cos^3\vartheta. \tag{23.2}$$

Полный индуцированный заряд на бесконечной плоскости AB равен и противоположен по знаку заряду q. В этом легко убедиться непосредственным интегрированием выражения (23.2) по плоскости AB. Еще проще воспользоваться теоремой Гаусса. Окружим заряд q и индуцированные заряды бесконечно удаленной сферой с центром в точке O. На полусфере, проходящей внутри проводящей среды, поле и его поток равны нулю. На полусфере, проходящей в вакууме, поле совпадает с полем точечного диполя, а потому обратно пропорционально кубу радиуса. Сама поверхность полусферы возрастает пропорционально квадрату радиуса. Поток вектора $\mathbf{E}$ через нее в пределе $r \to \infty$ обращается в нуль. По теореме Гаусса должен обращаться в нуль и полный заряд, окруженный сферой. Но этот заряд равен $q + q_{\text{инд}}$, где $q_{\text{инд}}$ — полный индуцированный заряд на плоскости AB. Значит, $q_{\text{инд}} = -q$.

3. Точечный заряд q вблизи проводящей сферы (рис. 65). Допустим, что сфера S радиуса a заземлена, т. е. потенциал ее равен нулю. Величина точечного заряда q и его расстояние до центра сферы $R = OA$ заданы. Этими условиями решение электростатической задачи определяется однозначно. Поле внутри сферы, по теореме Фарадея, равно нулю. Найдем поле вне сферы. Выберем на прямой OA такую точку C, чтобы треугольник OBC был подобен треугольнику OBA. Поместим в этой точке вспомогательный точечный заряд q'. Если b и b' — длины отрезков BA и BC, то потенциал зарядов q и q' в точке B будет $(q/b + q'/b')$. Он обратится в нуль, если

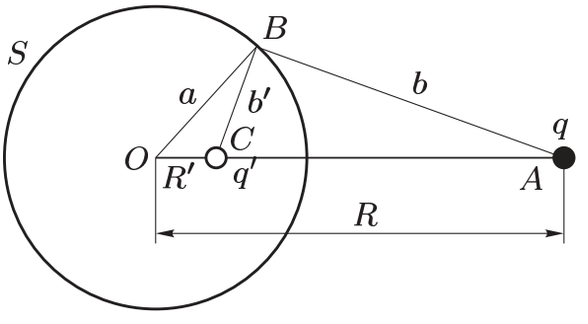


Рис. 65

$$q' = -\frac{b'}{b}\,q = -\frac{a}{R}\,q. \tag{23.3}$$

Мы видим, что величина q' не зависит от положения точки B на сфере S. Следовательно, потенциал, создаваемый зарядами q и q', обращается в нуль во всех точках сферы S, т. е. q' является электрическим изображением заряда q в сфере S. Вне сферы на расстояниях r и r' от зарядов q и q' потенциал определяется выражением

$$\varphi = \frac{q}{r} + \frac{q'}{r'}. \tag{23.4}$$

Общий заряд $q_{\text{инд}}$, индуцированный на сфере S, равен по абсолютной величине и совпадает по знаку с зарядом q'. Для доказательства

возьмем произвольную замкнутую поверхность Σ, окружающую сферу S, но не окружающую заряд q. На поверхности Σ поле $\mathbf{E}$ совпадает с полем точечных зарядов q и q', из которых q лежит вне Σ. Поэтому поток Φ этого поля через поверхность Σ будет $\Phi = 4\pi q'$. По теореме Гаусса тот же поток равен $\Phi = 4\pi q_{\text{инд}}$. Следовательно, $q_{\text{инд}} = q'$.

Если потенциал сферы S равен φ_0, то для решения задачи надо ввести еще один фиктивный заряд $q_0 = a\varphi_0$, поместив его в центре O сферы S. Поле во внешнем пространстве представится суперпозицией полей трех зарядов: q, q', q_0. Действительно, потенциал зарядов q и q' на сфере S равен нулю. Потенциал на S создается только зарядом q_0, т. е. равен $q_0/a = \varphi_0$.

Заметим, наконец, что электрические заряды q и q' обладают *свойством взаимности.* Оно заключается в следующем. Если q' является электрическим изображением заряда q, то, и обратно, заряд q является электрическим изображением заряда q'. Это замечание позволяет распространить изложенный метод на случай, когда точечный заряд внесен внутрь сферической полости, сделанной в проводящей среде.

4. Допустим теперь, что сфера S изолирована и задан ее заряд q_0. Нетрудно заметить, что в этом случае для определения поля во внешнем пространстве к зарядам q и q' надо добавить третий (фиктивный) заряд $q_0 - q'$, поместив его в центре сферы O. Рассмотрим специально частный случай, когда $q_0 = 0$. Индукционные заряды возбуждают во внешнем пространстве такое же поле, что и диполь длины R' с дипольным моментом $p = -q'R'$, который направлен по полю $\mathbf{E}$, создаваемому зарядом q. Будем неограниченно удалять заряд q, одновременно увеличивая его так, чтобы поле $\mathbf{E}$ в центре сферы оставалось неизменным. В пределе получится однородное электрическое поле, в которое внесен проводящий шар. При этом $p = -q'R' = a^3 q/R^2 = a^3 E$, или в векторной форме

$$\mathbf{p} = a^3 \mathbf{E}. \tag{23.5}$$

Этот результат уже был получен в § 16.

ЗАДАЧИ

1. Определить силу притяжения между точечным зарядом q и металлическим шаром (см. рис. 65). Рассмотреть два случая: 1) шар заземлен, 2) шар изолирован, а полный заряд его равен нулю.

Ответ. 1) $F = \dfrac{Ra}{(R^2 - a^2)^2}\, q^2$; 2) $F = \left(\dfrac{Ra}{(R^2 - a^2)^2} - \dfrac{a}{R^3}\right) q^2$.

2. В условиях предыдущей задачи найти работу A, которую надо совершить, чтобы точечный заряд q удалить в бесконечность.

Ответ. 1) $\dfrac{aq^2}{2(R^2 - a^2)}$; 2) $\dfrac{a^3 q^2}{2R^2(R^2 - a^2)}$.

3. Внутри сферической незаряженной проводящей оболочки в точке A на расстоянии $OA = a$ от ее центра помещен точечный заряд q (рис. 66). Радиус внутренней поверхности оболочки равен r, а внешней R. Найти: 1) поверхностную плотность индуцированных электрических зарядов на внешней поверхности оболочки; 2) потенциал оболочки, принимая за нуль потенциал

бесконечно удаленной точки; 3) поверхностную плотность индуцированных зарядов в точках B и C внутренней поверхности оболочки.

Ответ. 1) $\sigma = \dfrac{q}{4\pi R^2}$; 2) $\varphi = \dfrac{q}{R}$;

3) $\sigma_B = \dfrac{q}{4\pi(r-a)^2}\left(1+\dfrac{a}{r}\right)$, $\sigma_C = \dfrac{q}{4\pi(r+a)^2}\left(1-\dfrac{a}{r}\right)$.

4. Найти силу, действующую на точечный заряд q, помещенный на биссектрисе прямого двугранного угла между двумя проводящими плоскостями (рис. 67). Расстояние между зарядом q и вершиной двугранного угла равно a.

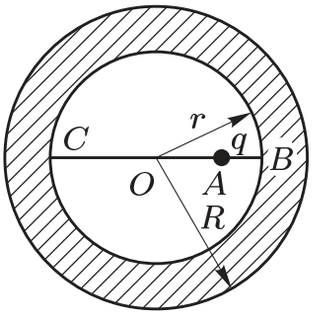


Рис. 66

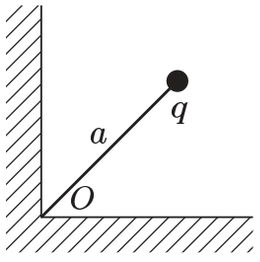


Рис. 67

Ответ. $F = \dfrac{q^2}{4a^2}\,(2\sqrt{2}-1)$. Сила F направлена к вершине двугранного угла O.

5. Точечный заряд q находится между двумя металлическими плоскостями, образующими между собой угол 60° (рис. 68). Найти предел, к которому стремится напряженность электрического поля $\mathbf{E}$, когда точка наблюдения приближается к ребру O, все время оставаясь между металлическими плоскостями. Как изменится результат, если заряд будет не точечным?

Решение. Нетрудно убедиться, что электрическим изображением заряда q относительно поверхности AOB будет совокупность пяти зарядов: q_1, q_2, q_3, q_4, q_5. Поле этих зарядов и заряда q в точке O равно нулю. Результат не изменится, если заряд q будет не точечным.

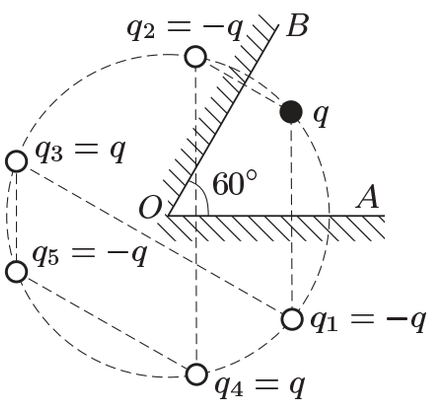


Рис. 68

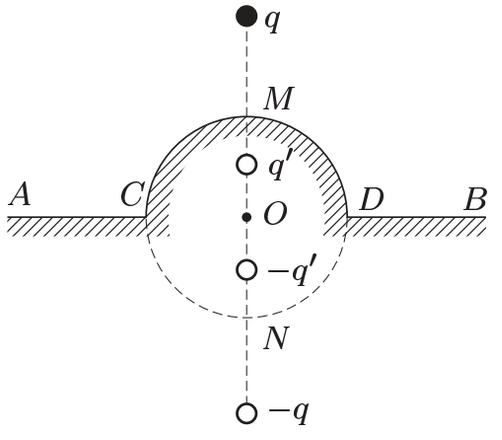


Рис. 69

6. На бесконечной плоской поверхности проводника имеется сферический бугор CMD, центр которого O лежит в той же плоскости (рис. 69). На перпендикуляре OM вне проводника расположен точечный заряд q. Найти электрическое поле во всем пространстве.

Решение. Введем электрические изображения в сфере и плоскости, как указано на рис. 69. Сгруппируем заряды попарно: 1) q с $-q$; 2) q' с $-q'$. Каждая пара в плоскости $ACDB$ создает нулевой потенциал. Сгруппируем

теперь те же заряды по-другому: 1) q с q'; 2) $-q$ с $-q'$. При такой группировке каждая пара будет создавать нулевой потенциал на сфере $CMDN$. Ясно поэтому, что потенциал четырех зарядов q, $-q$, q', $-q'$ обращается в нуль на поверхности $ACMDB$. Следовательно, поле этих зарядов в верхнем полупространстве будет тождественно с полем, которое требуется рассчитать.

7. Найти силу притяжения F между точечным электрическим диполем и бесконечной металлической пластинкой, если момент диполя $\mathbf{p}$ перпендикулярен к плоскостям пластинки, а расстояние его до ближайшей поверхности пластинки равно h. Определить также работу A_{12}, которую надо затратить, чтобы удалить диполь от пластинки с расстояния h_1 до h_2.

Ответ. $F = \dfrac{3p^2}{8h^4}$, $A_{12} = \dfrac{p^2}{8}\left(\dfrac{1}{h_1^3} - \dfrac{1}{h_2^3}\right)$. Обратим внимание, что перемещение диполя сопровождается перемещением индуцированных им зарядов. Однако это перемещение происходит перпендикулярно к силовым линиям и поэтому не сопровождается дополнительной работой.

§ 24. Точечный заряд над плоской поверхностью диэлектрика

Пусть два однородных диэлектрика с диэлектрическими проницаемостями ε_1 и ε_2 граничат друг с другом вдоль плоскости MN (рис. 70). В точке A первого диэлектрика помещен точечный заряд q. Найдем электрическое поле в каждом из диэлектриков. В окрестности точки A поле должно стремиться к бесконечности, как кулоново поле точечного заряда q. Поэтому поле в первом диэлектрике должно содержать слагаемое $qr/(e_1 r^3)$. К нему надо добавить поле поляризационных зарядов, возникших на границе раздела диэлектриков. Введем предположение, оправдываемое последующими вычислениями, что поле поляризационных зарядов в первом диэлектрике эквивалентно полю какого-то точечного заряда q', помещенного в точке A', зеркально симметричной с A относительно границы раздела. Тогда для поля в первом диэлектрике можно написать

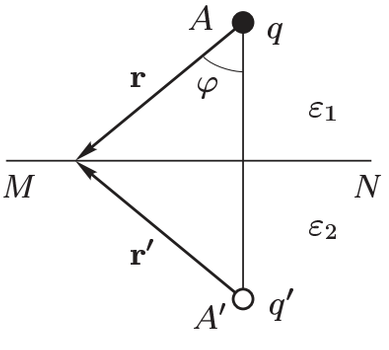


Рис. 70

$$\mathbf{E}_1 = \frac{q}{\varepsilon_1 r^3}\,\mathbf{r} + \frac{q'}{\varepsilon_1 r'^3}\,\mathbf{r}',$$

где $\mathbf{r}$ и $\mathbf{r}'$ — радиусы-векторы, проведенные из зарядов q и q' в рассматриваемую точку. Введем второе предположение, также оправдываемое последующими вычислениями, что поле во втором диэлектрике представляется выражением

$$\mathbf{E}_2 = \frac{q''}{\varepsilon_2 r^3}\,\mathbf{r},$$

причем второй (фиктивный) заряд q'' совмещен пространственно с зарядом q (на рис. 70 он не изображен). Теперь необходимо выражения для $\mathbf{E}_1$ и $\mathbf{E}_2$ «сшить», чтобы на границе раздела диэлектриков удовлетворялись граничные условия: непрерывность касательных компонент вектора $\mathbf{E}$ и нормальных компонент вектора $\mathbf{D}$. Первое условие имеет вид

$$\frac{q}{\varepsilon_1}\sin\varphi + \frac{q'}{\varepsilon_1}\sin\varphi = \frac{q''}{\varepsilon_2}\sin\varphi,$$

а второе —

$$q\cos\varphi - q'\cos\varphi = q''\cos\varphi.$$

Существенно, что угол φ выпадает из обоих уравнений. Поэтому если q' и q'' определить из этих уравнений, то граничные условия будут удовлетворены во всех точках границы раздела. Таким путем находим

$$\mathbf{E}_1 = \frac{q}{\varepsilon_1 r^3}\,\mathbf{r} - \frac{1}{\varepsilon_1}\,\frac{\varepsilon_2-\varepsilon_1}{\varepsilon_1+\varepsilon_2}\,\frac{q}{r'^3}\,\mathbf{r}', \quad \mathbf{E}_2 = \frac{2}{\varepsilon_1+\varepsilon_2}\,\frac{q}{r^3}\,\mathbf{r}. \tag{24.1}$$

Эти выражения удовлетворяют всем условиям задачи и, в силу теоремы единственности, дают ее решение. При $\varepsilon_2 \to \infty$ они переходят в соответствующие выражения для поля точечного заряда над проводящей плоскостью.

ЗАДАЧА

Какая сила действует на точечный заряд q вблизи плоской границы раздела двух диэлектриков, если заряд помещен в первом диэлектрике?

Ответ. $F = \dfrac{1}{4\varepsilon_1}\,\dfrac{\varepsilon_2-\varepsilon_1}{\varepsilon_2+\varepsilon_1}\,\dfrac{q^2}{h^2}$, где h — расстояние заряда от границы раздела. Заряд притягивается к плоскости, если $\varepsilon_2 > \varepsilon_1$ и отталкивается, если $\varepsilon_2 < \varepsilon_1$.

§ 25. Электрическое поле заряженного проводящего эллипсоида

1. Возьмем сферический слой между двумя концентрическими сферами, равномерно заряженный по всем объему. Электрическое поле в полости, ограниченной таким слоем, как известно, равно нулю. Хотя этот результат является непосредственным следствием теоремы Гаусса (см. § 6), для последующего важно получить его непосредственно из закона Кулона. С этой целью через произвольную точку M (рис. 71) внутри сферической полости, ограниченной рассматриваемым слоем, проведем бесконечно узкий конус прямых. Он вырежет из слоя два бесконечно малых элемента объема dV_1 и dV_2, заштрихованных на рисунке. Проведем далее через точки A, B, A', B' участки сферических поверхностей с общим центром в точке M, изображенные на рисунке штриховыми линиями. Между ними тот же конус прямых вырежет элементы объема $d\tau_1$ и $d\tau_2$ (не обозначенные на рисунке). Очевидно, $d\tau_1/d\tau_2 = r_1^2/r_2^2$, где r_1 и r_2 — расстояния рассматриваемых элементов объема от точки M. Кроме того, $d\tau_1 = dV_1$, $d\tau_2 = dV_2$, а потому $dq_1/dq_2 = r_1^2/r_2^2$, где dq_1 и dq_2 — заряды элементов объема dV_1 и dV_2 соответственно. На основании закона Кулона можно утверждать, что поля зарядов dq_1 и dq_2 в точке M равны и противоположно направлены. Проведя через точку M конусы прямых во всевозможных направлениях, можно разделить весь заряженный сферический слой на пары аналогичных бесконечно малых зарядов, электрические поля которых в точке M взаимно уничтожают друг друга. Отсюда следует, что электрическое поле будет равно нулю во всех точках полости, так как точку M можно взять произвольно.

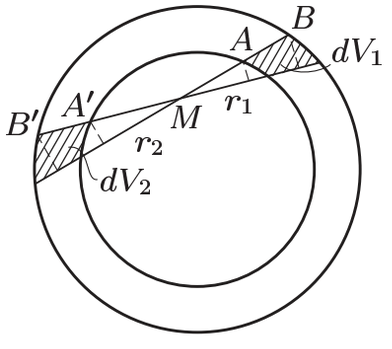


Рис. 71

2. Предполагая, что электрические заряды в сферическом слое неподвижно закреплены, будем равномерно сжимать или растягивать всю фигуру

на рис. 71 последовательно вдоль трех взаимно перпендикулярных направлений. При такой деформации сферический слой перейдет в равномерно заряженный эллипсоидальный слой, ограниченный двумя подобными и подобно расположенными концентрическими эллипсоидами. Каждая прямая перейдет в прямую, равномерно сжатую или растянутую. Величины зарядов dq_1 и dq_2, разумеется, останутся без изменения. Элементы объема dV_1 и dV_2, а также расстояния r_1 и r_2 изменятся, но их отношения останутся прежними. В частности, $dq_1/dq_2 = r_1^2/r_2^2$. Значит, сохранится и компенсация полей в каждой точке внутри полученной эллипсоидальной полости. Отсюда следует, что *электрическое поле в полости, ограниченной равномерно заряженным эллипсоидальным слоем между двумя подобными и подобно расположенными концентрическими эллипсоидами, равно нулю.*

3. Будем теперь уменьшать толщину эллипсоидального слоя беспредельно, сохраняя его общий заряд q неизменным. В пределе получится *поверхностное распределение электричества*, при котором поле внутри эллипсоидальной полости всюду равно нулю. Если поверхность полости сделать проводящей, то равновесие электричества на ней не нарушится. Значит, в результате такого предельного перехода получится равновесное распределение электричества по поверхности проводящего эллипсоида. Выполним теперь указанный предельный переход и найдем величину поверхностной плотности электричества σ на поверхности эллипсоида. Проведем из центра эллипсоидального слоя O бесконечно узкий конус прямых, вырезающий из этого слоя объем dV, а из его поверхности — элементарную площадку dS (рис. 72). Очевидно, $dV = dS\,dN$, где dN — толщина слоя в рассматриваемом месте. Если ρ — объемная плотность электричества в слое, то заряд элемента объема dV будет $dq = \rho\,dS\,dN$. Если $dN \to 0$, а общий заряд эллипсоидального слоя сохраняется неизменным, то получается поверхностное распределение электричества с поверхностной плотностью

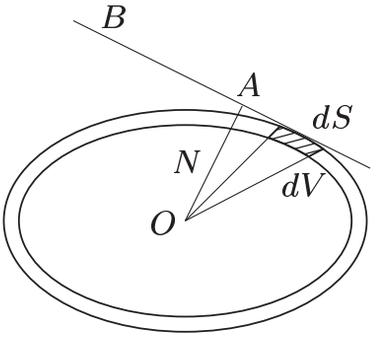


Рис. 72

$$\sigma = \frac{dq}{dS} = \rho\,dN.$$

Проведем в рассматриваемом месте эллипсоидальной поверхности касательную плоскость AB и опустим на нее перпендикуляр OA, длину которого обозначим через N. При неизменном направлении перпендикуляра OA объемы подобных и подобно расположенных эллипсоидов, очевидно, пропорциональны N^3. Если V — объем эллипсоида, то $V \sim N^3$, а потому

$$\frac{\Delta V}{V} = \frac{3\,dN}{N},$$

где ΔV — объем эллипсоидального слоя. Исключив dN и воспользовавшись соотношением $q = \rho\Delta V$, получим

$$\sigma = \frac{qN}{3V}. \tag{25.1}$$

Пусть уравнение поверхности эллипсоида имеет вид

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1. \tag{25.2}$$

Тогда, как известно из аналитической геометрии,

$$N = \left(\frac{x^2}{a^4} + \frac{y^2}{b^4} + \frac{z^2}{c^4}\right)^{-1/2}.$$

Кроме того, $V = (4\pi/3)abc$. Поэтому формула (25.1) переходит в

$$\sigma = \frac{q}{4\pi abc}\left(\frac{x^2}{a^4} + \frac{y^2}{b^4} + \frac{z^2}{c^4}\right)^{-1/2}. \tag{25.3}$$

Из рассуждений настоящего параграфа непосредственно следует, что *равновесное распределение электричества на поверхности проводящего эллипсоида остается равновесным при любом равномерном растяжении или сжатии его.* Пользуясь этим, можно было бы найти равновесное распределение электричества по поверхности эллипсоида путем равномерного растяжения или сжатия заряженного проводящего шара. Такой путь является наиболее коротким. Разумеется, он также приводит к формуле (25.3).

Зная распределение электричества по поверхности эллипсоида, можно путем интегрирования по этой поверхности найти потенциал, а затем и напряженность электрического поля в любой точке пространства. Такой путь, однако, приводит к громоздким вычислениям. Значительно проще другой, искусственный прием (см. задачи 3 и 4 к этому параграфу).

ЗАДАЧИ

1. Найти поверхностную плотность электричества на бесконечно тонкой проводящей эллиптической пластинке, получающейся равномерным сжатием трехосного эллипсоида в направлении оси Z.

Решение. Пренебрегая в (25.3) первыми двумя членами в скобках, а затем воспользовавшись формулой (25.2), получаем

$$\sigma = \frac{q}{4\pi ab}\left(\frac{z^2}{c^2}\right)^{-1/2} = \frac{q}{4\pi ab}\left(1 - \frac{x^2}{a^2} - \frac{y^2}{b^2}\right)^{-1/2}. \tag{25.4}$$

2. Заряженный проводящий эллипсоид мысленно разделен на части равноотстоящими плоскостями, перпендикулярными к одной из его главных осей. Показать, что, каково бы ни было число таких частей, величины их зарядов будут всегда одинаковы. В частности, если эллипсоид является вытянутым и бесконечно тонким, то электричество распределится по его длине равномерно.

3. Найти условие, при выполнении которого поверхности семейства $\lambda(x, y, z) = \text{const}$ могут быть эквипотенциальными.

Решение. По условию потенциал φ постоянен, если постоянна функция λ. Значит, он должен быть функцией только λ: $\varphi = \varphi(\lambda)$. Дифференцируя φ по координате x, получаем

$$\frac{\partial\varphi}{\partial x} = \varphi'\frac{\partial\lambda}{\partial x}, \quad \frac{\partial^2\varphi}{\partial x^2} = \varphi''\left(\frac{\partial\lambda}{\partial x}\right)^2 + \varphi'\frac{\partial^2\lambda}{\partial x^2}.$$

Написав аналогичные соотношения для y и z и складывая их, найдем

$$\Delta\varphi = \varphi''(\operatorname{grad}\lambda)^2 + \varphi'\Delta\lambda.$$

С учетом уравнения Лапласа $\Delta\varphi = 0$ отсюда следует

$$\frac{\Delta\lambda}{(\operatorname{grad}\lambda)^2} = -\frac{\varphi''(\lambda)}{\varphi'(\lambda)}.$$

Таким образом, чтобы уравнение $\lambda(x, y, z) = \text{const}$ представляло семейство эквипотенциальных поверхностей, необходимо, чтобы левая часть последнего соотношения была функцией только λ:

$$\Delta\lambda \frac{\Delta\lambda}{(\operatorname{grad}\lambda)^2} = \Phi(\lambda). \tag{25.5}$$

Если функция $\Phi(\lambda)$ известна, то потенциал φ можно найти интегрированием уравнения

$$\frac{\varphi''(\lambda)}{\varphi'(\lambda)} = -\Phi(\lambda), \tag{25.6}$$

что сводится к выполнению двух квадратур.

4. Найти потенциал электрического поля заряженного проводящего эллипсоида.

Р е ш е н и е. Пусть $\lambda(x, y, z)$ — неявная функция, определяемая уравнением

$$\frac{x^2}{a^2+\lambda} + \frac{y^2}{b^2+\lambda} + \frac{z^2}{c^2+\lambda} = 1 \tag{25.7}$$

($a \geqslant b \geqslant c > 0$). Уравнение $\lambda(x, y, z) = \text{const}$ представляет семейство софокусных поверхностей второго порядка. При $c^2+\lambda > 0$ это есть семейство эллипсоидов; при $b^2 + \lambda > 0$, $c^2 + \lambda < 0$ — однополостных гиперболоидов; при $a^2 + \lambda > 0$, $b^2 + \lambda < 0$, $c^2 + \lambda < 0$ — двухполостных гиперболоидов. Во всех случаях уравнение $\lambda(x, y, z) = \text{const}$ может быть семейством эквипотенциальных поверхностей. Для доказательства достаточно убедиться, что выполняется условие (25.5). Введем обозначение

$$F_n = \frac{x^2}{(a^2+\lambda)^n} + \frac{y^2}{(b^2+\lambda)^n} + \frac{z^2}{(c^2+\lambda)^n}.$$

Тогда дифференцированием (25.7) по x получим

$$\frac{2x}{a^2+\lambda} - F_2 \frac{\partial\lambda}{\partial x} = 0,$$

откуда

$$\frac{\partial\lambda}{\partial x} = \frac{2x}{(a^2+\lambda)F_2}.$$

Вторичное дифференцирование дает

$$\frac{\partial^2\lambda}{\partial x^2} = \frac{2}{(a^2+\lambda)F_2} - \frac{2x}{(a^2+\lambda)^2 F_2}\frac{\partial\lambda}{\partial x} - \frac{2x}{(a^2+\lambda)F_2^2}\left[\frac{2x}{(a^2+\lambda)^2} - 2F_3\frac{\partial\lambda}{\partial x}\right] =$$

$$= \frac{2}{(a^2+\lambda)F_2} - \frac{8x^2}{(a^2+\lambda)^3 F_2^2} + \frac{8F_3x^2}{(a^2+\lambda)^2 F_2^3}.$$

Написав аналогичные выражения для производных по y и z и складывая их, находим

$$(\operatorname{grad}\lambda)^2 = \frac{4}{F_2},$$

$$\Delta\lambda = \frac{2}{F_2}\left(\frac{1}{a^2+\lambda} + \frac{1}{b^2+\lambda} + \frac{1}{c^2+\lambda}\right).$$

Следовательно, условие (25.5) выполняется, причем

$$\Phi(\lambda) = \frac{1}{2}\left(\frac{1}{a^2+\lambda} + \frac{1}{b^2+\lambda} + \frac{1}{c^2+\lambda}\right). \tag{25.8}$$

Задача свелась к интегрированию уравнения

$$\frac{\varphi''(\lambda)}{\varphi'(\lambda)} = -\frac{1}{2}\left(\frac{1}{a^2+\lambda} + \frac{1}{b^2+\lambda} + \frac{1}{c^2+\lambda}\right). \tag{25.9}$$

Допустим теперь, что все знаменатели в правой части уравнения (25.9) положительны. Тогда эквипотенциальными поверхностями (25.7) будут софокусные эллипсоиды. При $\lambda = 0$ получается эллипсоид (25.2). Таким образом, если потенциал $\varphi(x, y, z)$ найти из уравнения (25.9), то на поверхности эллипсоида (25.2) он обратится в постоянную и, следовательно, будет давать решение рассматриваемой задачи. Интегрируя уравнение (25.9), получаем

$$\varphi'(\lambda) = A[(a^2+\lambda)(b^2+\lambda)(c^2+\lambda)]^{-1/2},$$

где A — постоянная интегрирования. Интегрируя вторично и принимая за нуль потенциал бесконечно удаленной точки, находим

$$\varphi(\lambda) = A\int_{\infty}^{\lambda}[(a^2+\lambda)(b^2+\lambda)(c^2+\lambda)]^{-1/2}\,d\lambda.$$

Постоянная A определится из условия, что на больших расстояниях функция φ должна переходить в потенциал точечного заряда $\varphi = q/r$. При больших λ эллипсоид (25.7) переходит в шар радиуса $r = \sqrt{\lambda}$, а предыдущий интеграл — в

$$\varphi(\lambda) = A\int_{\infty}^{\lambda}\lambda^{-3/2}\,d\lambda = -2A\lambda^{-1/2} = -\frac{2A}{r}.$$

Значит,

$$\varphi = \frac{q}{2}\int_{\lambda}^{\infty}[(a^2+\lambda)(b^2+\lambda)(c^2+\lambda)]^{-1/2}\,d\lambda. \tag{25.10}$$

Входящий сюда интеграл является *эллиптическим.* Если эллипсоид является эллипсоидом вращения, то интегрирование выполняется в элементарных функциях. Для *вытянутого эллипсоида вращения* ($b = c$)

$$\varphi = \frac{q}{\sqrt{a^2-b^2}}\ln\frac{\sqrt{b^2+\lambda}}{\sqrt{a^2+\lambda}-\sqrt{a^2-b^2}}. \tag{25.11}$$

Введем обозначения: $A = \sqrt{a^2+\lambda}$, $f = \sqrt{a^2-b^2}$. Величина $2A$ есть *большая ось* эквипотенциального эллипсоида, проходящего через точку наблюдения, а $2f$ — расстояние между *фокусами* рассматриваемого семейства *софокусных эллипсоидов.* Выражение (25.11) можно преобразовать к виду

$$\varphi = \frac{q}{2f}\ln\frac{A+f}{A-f}. \tag{25.11a}$$

Для *сплюснутого эллипсоида вращения* ($a = b$)

$$\varphi = \frac{q}{\sqrt{a^2-c^2}}\operatorname{arctg}\frac{\sqrt{a^2-c^2}}{\sqrt{\lambda+c^2}}. \tag{25.12}$$

Введем длину малой оси эквипотенциального эллипсоида, проходящего через точку наблюдения: $2B = 2\sqrt{c^2+\lambda}$, а также расстояние между фокусами

$2f = 2\sqrt{a^2 - c^2}$. Тогда

$$\varphi = \frac{q}{f} \operatorname{arctg} \frac{f}{B}. \tag{25.12а}$$

5. Бесконечно тонкая диэлектрическая палочка равномерно заряжена электричеством с постоянной линейной плотностью. Показать, что эквипотенциальными поверхностями поля такой палочки будут софокусные эллипсоиды, фокусы которых находятся на ее концах.

6. Бесконечно тонкая круглая диэлектрическая пластинка радиуса a заряжена электричеством с поверхностной плотностью

$$\sigma = \frac{q}{2\pi a\sqrt{a^2 - r^2}},$$

где r — расстояние от центра пластинки. Показать, что эквипотенциальными поверхностями поля пластинки будут софокусные сплюснутые эллипсоиды вращения с фокальной линией, расположенной по окружности пластинки.

§ 26. Емкость проводников и конденсаторов

1. Рассмотрим заряженный уединенный проводник, погруженный в неподвижный диэлектрик. Потенциал создаваемого им электрического поля на бесконечности условимся считать равным нулю. Если удвоить заряд проводника, то его потенциал также удвоится. Вообще, между зарядом проводника q и его потенциалом φ существует прямая пропорциональность:

$$q = C\varphi. \tag{26.1}$$

Коэффициент C зависит только от размеров и формы проводника, а также от диэлектрической проницаемости окружающего диэлектрика и ее распределения в пространстве. Он называется *емкостью уединенного проводника.* Например, для шара радиуса a в однородном диэлектрике $\varphi = q/(\varepsilon a)$, а потому

$$C = \varepsilon a. \tag{26.2}$$

2. Более важным является понятие *емкости конденсатора.* Всякий конденсатор состоит из двух металлических обкладок, отделенных одна от другой слоем диэлектрика. Пусть обкладками конденсатора являются две замкнутые металлические оболочки: наружная и внутренняя, причем внутренняя целиком окружена наружной. Тогда поле между обкладками совершенно не будет зависеть от внешних электрических полей. Заряды на поверхностях обкладок, обращенных одна к другой, по теореме Фарадея, равны по величине и противоположны по знаку. В реальном конденсаторе, поскольку его обкладки не являются полностью замкнутыми, это верно только приближенно, хотя и с большой точностью. Практическая независимость внутреннего поля конденсатора от внешнего поля здесь достигается тем, что обкладки располагаются очень близко одна от другой. В этом случае заряды будут почти целиком сосредоточены на внутренних поверхностях обкладок, т. е. поверхностях, обращенных друг к другу. Если q — заряд одной

из обкладок (для определенности положительной), а $\varphi \equiv (\varphi_1 - \varphi_2)$ — разность потенциалов между обкладками, то

$$q = C(\varphi_1 - \varphi_2). \tag{26.3}$$

Постоянная C зависит только от размеров и устройства конденсатора. Она называется *емкостью конденсатора.*

Возьмем два конденсатора. В одном пространство между обкладками заполнено однородным диэлектриком с диэлектрической проницаемостью ε, в другом между обкладками — вакуум (такой конденсатор обычно называют *воздушным*, что не вполне точно). В остальных отношениях оба конденсатора тождественны. При одних и тех же зарядах разность потенциалов между обкладками первого конденсатора будет в ε раз меньше, чем между обкладками второго (см. § 22, п. 3). Следовательно, емкость C конденсатора с диэлектриком будет в ε раз больше емкости C_0 воздушного конденсатора:

$$C = \varepsilon C_0. \tag{26.4}$$

3. Во всякой системе единиц за единицу емкости принимают емкость такого конденсатора, который единичным зарядом заряжается до разности потенциалов, равной единице. В гауссовой и СГСЭ-системах, как это видно из формулы (26.2), единичной емкостью обладает уединенный шарик в вакууме, если радиус его равен 1 см. Размерность емкости в этих системах единиц совпадает с размерностью длины. Поэтому указанная единица емкости называется *сантиметром.* Практической единицей емкости является *фарад.* Фарад есть емкость такого конденсатора, который одним кулоном электричества заряжается до разности потенциалов в один вольт. Очевидно,

$$1\ \Ф = \frac{1\ \text{Кл}}{1\ \text{В}} = \frac{3 \cdot 10^9\ \text{СГСЭ-ед. заряда}}{\frac{1}{300}\ \text{СГСЭ-ед. потенциала}} = 9 \cdot 10^{11}\ \text{см}.$$

Емкостью в один фарад обладает уединенный шар в вакууме с радиусом $9 \cdot 10^{11}$ см $= 9 \cdot 10^6$ км. Это очень большая емкость. На практике применяется *микрофарад*, равный миллионной доле фарада, а также *пикофарад*, который в миллион раз меньше микрофарада. Емкостью в один пикофарад обладает шарик в вакууме, если его радиус равен 0,9 см.

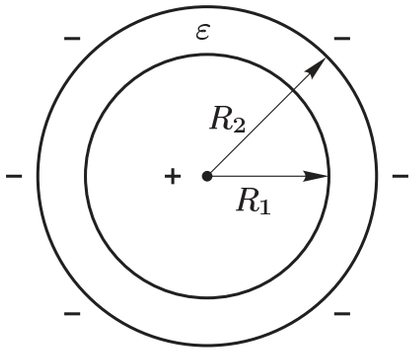


Рис. 73

4. Емкость шарового конденсатора. Обкладками конденсатора являются две сферы: внутренняя с радиусом R_1 и внешняя с радиусом R_2 (рис. 73). Разность потенциалов между ними

$$\varphi = \varphi_1 - \varphi_2 = \frac{q}{\varepsilon}\left(\frac{1}{R_1} - \frac{1}{R_2}\right).$$

Емкость конденсатора

$$C = \frac{\varepsilon}{1/R_1 - 1/R_2} = \varepsilon \frac{R_1 R_2}{R_2 - R_1}. \tag{26.5}$$

Если толщина зазора между обкладками $d = R_2 - R_1$ мала по сравнению с радиусами R_1 и R_2, то площади обкладок почти одинаковы и приближенно равны $S \approx 4\pi R_1^2 \approx 4\pi R_2^2 \approx 4\pi R_1 R_2$. Тогда

$$C = \frac{\varepsilon S}{4\pi d}. \tag{26.6}$$

Эта формула, как и следовало ожидать, совпадает с формулой для емкости плоского конденсатора, которая выводится ниже.

5. Емкость плоского конденсатора. Поле между обкладками конденсатора почти всюду однородно (рис. 74). Однородность поля нарушается только вблизи краев конденсатора. Такими «*краевыми эффектами*» при вычислении емкости конденсатора мы пренебрежем. Это можно делать, когда расстояние d между обкладками очень мало по сравнению с их линейными размерами. Если σ — поверхностная плотность электричества на положительной обкладке, а S — площадь последней, то $q = \sigma S$. Напряженность поля $E = 4\pi\sigma/\varepsilon$, разность потенциалов $\varphi_1 - \varphi_2 = Ed = 4\pi\sigma d/\varepsilon$, емкость конденсатора

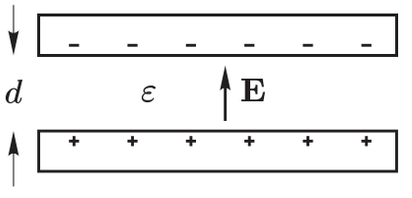


Рис. 74

$$C = \frac{\varepsilon S}{4\pi d}.$$

Краевые эффекты вносят некоторую поправку к этой формуле, но ее вычислением мы заниматься не будем.

Возьмем воздушный конденсатор, состоящий из двух пластин — неподвижной A и подвижной B (рис. 75). Пластину B заземлим, пластину A хорошо изолируем и соединим с шариком электрометра. Зарядим пластину A и будем перемещать пластину B. При увеличении расстояния между пластинами стрелка электрометра отклоняется сильнее, при уменьшении — слабее. Дело в том, что при разведении пластин емкость конденсатора уменьшается. Заряд q пластины A практически не меняется, так как на электрометр переходит пренебрежимо малое количество электричества. Поэтому разность потенциалов $\varphi = q/C$ возрастает, что и показывает электрометр. Наоборот, при сближении пластин емкость возрастает, а разность потенциалов уменьшается.

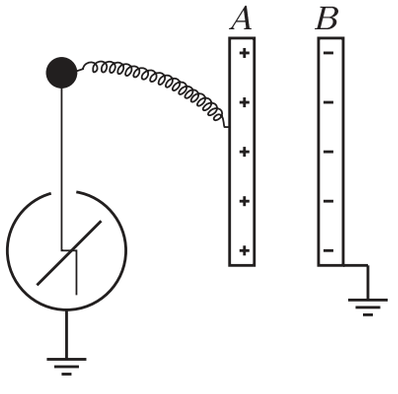


Рис. 75

Явление можно также объяснить непосредственно с помощью закона Кулона. При соединении заряженной пластины A с электрометром часть заряда уходит с нее на электрометр. Оставшийся заряд удерживается электрическим полем пластины B. Когда B приближается к A, поле возрастает. Благодаря этому часть зарядов с электрометра оттягивается к пластине A, и угол отклонения стрелки уменьшается. Из приведенного объяснения видно, что явление все же обусловлено влиянием краевых эффектов. Напряженность поля между бесконечными пластинами не зависит от расстояния между ними. Но для конечных пластин, в особенности вблизи их краев, это не так. При приближении конечной пластины B на A увеличиваются индукционные заряды противоположного, а на шарике и стрелке электрометра — того же знаков. Такое перемещение зарядов меняет заряд конденсатора и напряженность электрического поля между его пластинами пренебрежимо мало. Но его достаточно, чтобы сильно изменить заряд на электрометре и соответствующую ему разность потенциалов.

Если, оставляя расстояние между пластинами A и B неизменным, ввести между ними лист из диэлектрика, то электрометр покажет меньшую разность потенциалов. Такой опыт был поставлен Фарадеем. На нем Фарадей впервые установил влияние *промежуточной среды* на электрическое поле между наэлектризованными телами. Результат опыта, очевидно, объясняется возрастанием емкости конденсатора при введении диэлектрика между пластинами. Можно также сказать, что на диэлектрике появляются поляризационные заряды, оттягивающие часть электричества с электрометра на пластины конденсатора. Металлический лист в таком опыте действует как диэлектрик с бесконечно большим значением ε. Его введение эквивалентно уменьшению зазора d между пластинами конденсатора.

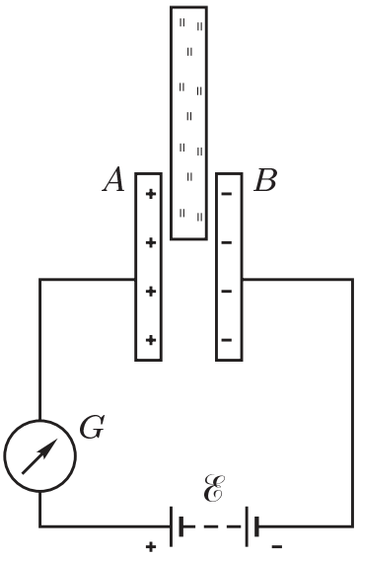


Рис. 76

Опыт можно осуществить в другой постановке (рис. 76). Пластины воздушного конденсатора через баллистический гальванометр подсоединяют к полюсам гальванической батареи, поддерживающей постоянной разность потенциалов между ними. Если между пластинами ввести лист из диэлектрика, то емкость и заряд конденсатора должны возрастать. По цепи пройдет кратковременный электрический ток, и баллистический гальванометр «даст отброс».

6. Емкость цилиндрического конденсатора. Цилиндрический конденсатор состоит из коаксиальных цилиндрических обкладок, разделенных слоем диэлектрика (рис. 77). Примером такого конденсатора может служить лейденская банка. Пусть a и b — радиусы внутренней и наружной обкладок, а l — длина конденсатора. Если

пренебречь краевыми эффектами, то

$$C = \frac{\varepsilon l}{2\ln(b/a)}. \tag{26.7}$$

Когда зазор между обкладками конденсатора $d = b - a$ мал по сравнению с a и b, эта формула, как легко убедиться, переходит в (26.6).

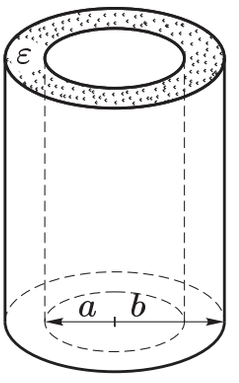


Рис. 77

7. Емкость двух параллельных прямых проволок. Пусть l — длина каждой проволоки, a и b — их радиусы, $2h$ — расстояние между ними. Предполагая, что $l \gg 2h \gg a$, $2h \gg b$, можно воспользоваться формулой (19.11) и написать

$$\varphi = -2q\ln\frac{r_1}{\varepsilon l} + 2q\ln\frac{r_2}{\varepsilon l} + \text{const},$$

где r_1 и r_2 — расстояния точки наблюдения от осей проволок. Полагая сначала $r_1 = a$, $r_2 = 2h$, а затем $r_1 = 2h$, $r_2 = b$, находим потенциалы проволок φ_1, φ_2 и разность потенциалов $\varphi_1 - \varphi_2$. В результате получаем для емкости формулу

$$C = \frac{\varepsilon l}{2\ln\dfrac{4h^2}{ab}}. \tag{26.8}$$

В частности, при $a = b$

$$C = \frac{\varepsilon l}{4\ln(2h/a)}. \tag{26.9}$$

Из последней формулы методом зеркальных изображений легко получить выражение для емкости цилиндрической прямой проволоки, подвешенной над заземленной бесконечной проводящей плоскостью (телеграфный провод над земной поверхностью). Если проволока параллельна плоскости, а расстояние между ними равно h, то емкость будет

$$C = \frac{\varepsilon l}{2\ln(2h/a)}, \tag{26.10}$$

т. е. вдвое больше, чем в предыдущем случае.

8. Конденсаторы часто соединяют в батареи. Соединение может быть *параллельным* (рис. 78) или *последовательным* (рис. 79). Применяют также *комбинированное соединение.* Ограничимся ради простоты случаем двух конденсаторов. При параллельном соединении разности потенциалов между обкладками обоих конденсаторов одинаковы, а заряды обкладок складываются: $q = q_1 + q_2$. Делением на общую разность потенциалов $\varphi = \varphi_1 - \varphi_2$ находим емкость батареи:

$$C = C_1 + C_2. \tag{26.11}$$

При последовательном соединении средние пластины, соединенные между собой, электризуются через влияние, а потому их заряды равны

и противоположны по знаку. Таким образом, заряды на обоих конденсаторах одинаковы. Разности потенциалов складываются:

$$\varphi_1 - \varphi_3 = (\varphi_1 - \varphi_2) + (\varphi_2 - \varphi_3).$$

А так как

$$\varphi_1 - \varphi_3 = \frac{q}{C}, \quad \varphi_1 - \varphi_2 = \frac{q}{C_1}, \quad \varphi_2 - \varphi_3 = \frac{q}{C_2},$$

то отсюда получаем

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2}. \tag{26.12}$$

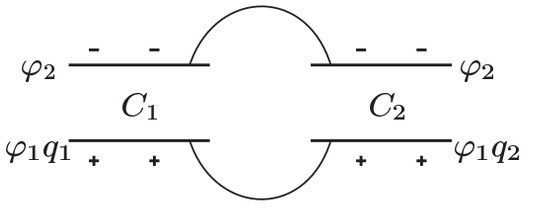


Рис. 78

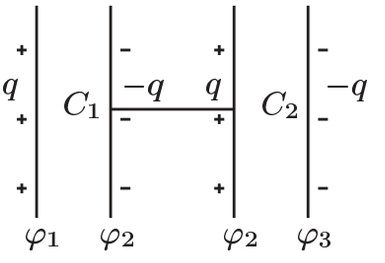


Рис. 79

Обобщение формул (26.11) и (26.12) на случай нескольких конденсаторов тривиально. Параллельное соединение применяется для увеличения емкости конденсатора. Последовательное применяют тогда, когда во избежание пробоя большую разность потенциалов требуется распределить между несколькими конденсаторами.

9. Применим формулу (26.12) к расчету емкости *слоистого плоского конденсатора.* Конденсатор состоит из двух параллельных металлических обкладок, разделенных плоскими слоями из диэлектриков с различными диэлектрическими проницаемостями (рис. 80). Представим себе, что между слоями диэлектриков введены бесконечно тонкие металлические листы. От этого заряды на обкладках конденсатора и напряженности полей в слоях диэлектрика не изменятся. Не изменится и разность потенциалов между обкладками, а с ней и емкость конденсатора. Однако введение металлических листов превращает слоистый конденсатор в батарею последовательно соединенных конденсаторов. Применяя к ней формулы (26.6) и (26.12), получим

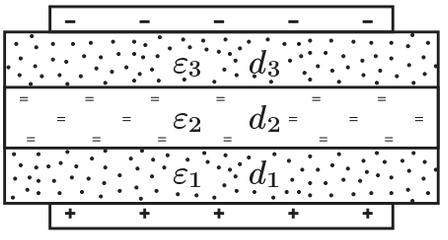


Рис. 80

$$\frac{1}{C} = \frac{4\pi}{S}\left(\frac{d_1}{\varepsilon_1} + \frac{d_2}{\varepsilon_2} + \dots\right), \tag{26.13}$$

где d_1, d_2, ... — толщины диэлектрических слоев, а ε_1, ε_2, ... — их диэлектрические проницаемости.

ЗАДАЧИ

1. Три конденсатора с емкостями $C_1 = 2$ мкФ, $C_2 = 2$ мкФ, $C_3 = 4$ мкФ и допустимыми напряжениями $V_1 = 1000$ В, $V_2 = 450$ В, $V_3 = 250$ В соединены в батарею. При каком соединении конденсаторов можно получить наибольшее напряжение? Чему равно это напряжение и соответствующая емкость батареи?

Ответ. При последовательном; $V = 1125$ В; $C = 0{,}8$ мкФ.

2. Определить емкость единицы длины двух параллельных бесконечно длинных круговых цилиндров, радиусы которых равны a и b, а расстояние между осями l.

Решение. Возьмем две бесконечно длинные параллельные прямые A и A', равномерно заряженные электричествами противоположных знаков (рис. 81). Эквипотенциальными поверхностями будут круговые цилиндры (см. задачу 9 к § 19). Пусть S и S' — два из них, расположенные вне друг

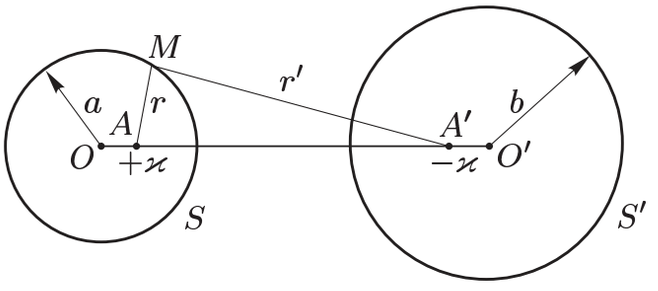


Рис. 81

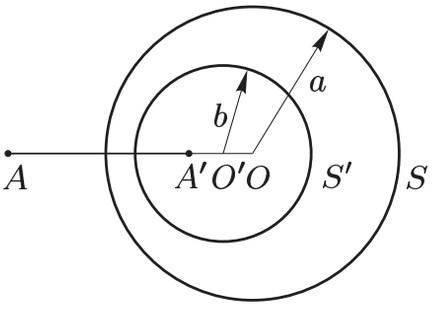


Рис. 82

друга. Положение осей цилиндров O и O' определим условиями $OA \cdot OA' = a^2$, $O'A' \cdot O'A = b^2$ (см. § 23, п. 3). Потенциалы поверхностей S и S' будут

$$\varphi = \frac{2\varkappa}{\varepsilon} \ln \frac{a}{d}, \quad \varphi' = -\frac{2\varkappa'}{\varepsilon} \ln \frac{b}{d'},$$

где $d = OA$, $d' = O'A'$. Если заряды $\varkappa$ и $\varkappa'$ перейдут с линий A и A' на поверхности цилиндров S и S', то поле внутри цилиндров обратится в нуль, а во внешнем пространстве не изменится. Получится распределение зарядов, отвечающее условиям задачи. Вычислив разность потенциалов $\varphi - \varphi'$, найдем емкость на единицу длины

$$C = \frac{\varepsilon}{2 \ln (ab/dd')}. \tag{26.14}$$

Для определения d и d' имеем уравнения

$$d(l - d') = a^2, \quad d'(l - d) = b^2,$$

из которых находим

$$d = \frac{l^2 + a^2 - b^2 - \sqrt{(l^2 + a^2 - b^2)^2 - 4a^2 l^2}}{2l},$$
$$d' = \frac{l^2 + b^2 - a^2 - \sqrt{(l^2 + b^2 - a^2)^2 - 4b^2 l^2}}{2l}. \tag{26.15}$$

Рассмотрим теперь случай, когда цилиндр S' целиком лежит внутри цилиндра S (рис. 82). Все рассуждения и вычисления в этом случае остаются без изменений. Окончательная формула (26.14) также остается неизменной.

Только для определения $d = OA$ и $d' = O'A'$ получаются уравнения

$$d'(d-l) = b^2, \quad d(d'+l) = a^2 \quad (a > b > l).$$

Из них находим

$$d = \frac{a^2 - b^2 - l^2 + \sqrt{(a^2-b^2-l^2)^2 - 4a^2l^2}}{2l},$$
$$d' = \frac{a^2 - b^2 - l^2 - \sqrt{(a^2-b^2-l^2)^2 - 4a^2l^2}}{2l}. \tag{26.16}$$

3. Показать, что для тонких цилиндров формула (26.14) переходит в формулу (26.8), а для коаксиальных цилиндров — в формулу (26.7).

4. Найти выражение для емкости эллипсоида.

Ответ.

$$\frac{1}{C} = \frac{1}{2\varepsilon} \int_0^\infty [(a^2+\lambda)(b^2+\lambda)(c^2+\lambda)]^{-1/2}\, d\lambda. \tag{26.17}$$

Для вытянутого эллипсоида вращения:

$$\frac{1}{C} = \frac{1}{2ae\varepsilon} \ln \frac{1+e}{1-e}, \tag{26.18}$$

для сплюснутого:

$$\frac{1}{C} = \frac{1}{2ae\varepsilon} \operatorname{arctg} \frac{e}{\sqrt{1-e^2}}, \tag{26.19}$$

где e — эксцентриситет эллипсоида.

§ 27. Потенциальные и емкостные коэффициенты

1. Рассмотрим произвольную систему n заряженных неподвижных проводников, пространство между которыми заполнено неподвижным диэлектриком — однородным или неоднородным. Такую систему иногда называют *сложным конденсатором.* Будем предполагать, что свободных зарядов в диэлектрике нет. Докажем, что при этих условиях потенциалы проводников будут *линейными однородными функциями* их зарядов. При этом, как обычно, потенциал поля в бесконечности принимается равным нулю.

Предположим сначала, что все проводники не заряжены. Сообщим затем одному только i-му проводнику заряд, равный единице. Этим однозначно определится электрическое поле $\mathbf{E}_i(\mathbf{r})$ во всем пространстве и соответствующие ему потенциал $V_i(\mathbf{r})$ и индукция $\mathbf{D}_i = \varepsilon(\mathbf{r})\mathbf{E}_i(\mathbf{r})$. По теореме Гаусса поток вектора $\mathbf{D}_i$ через поверхность i-го проводника будет равен 4π, а через поверхности остальных проводников — нулю. Значение потенциала $V_i(\mathbf{r})$ в месте нахождения j-го проводника обозначим V_{ji}. Коэффициенты V_{ji} зависят только от формы и расположения проводников, а также от диэлектрической проницаемости диэлектрика между ними. Они называются *потенциальными коэффициентами.* Ввиду линейности и однородности уравнений электростатики произвольная линейная комбинация векторов $\mathbf{E}_i(\mathbf{r})$ и $\mathbf{D}_i(\mathbf{r})$ с постоянными

коэффициентами q_i, т. е.

$$\mathbf{E}(\mathbf{r}) = \sum_{i=1}^{n} q_i \mathbf{E}_i(\mathbf{r}), \tag{27.1}$$

$$\mathbf{D}(\mathbf{r}) = \varepsilon \mathbf{E}(\mathbf{r}) = \sum_{i=1}^{n} q_i \mathbf{D}_i(\mathbf{r}), \tag{27.2}$$

удовлетворяет этим уравнениям. Действительно, вектор $\mathbf{E}(\mathbf{r})$ потенциальный, так как все поля $\mathbf{E}_i(\mathbf{r})$ потенциальны. В диэлектрике вектор $\mathbf{D}$ удовлетворяет уравнению $\operatorname{div}\mathbf{D} = 0$, так как $\operatorname{div}\mathbf{D}_i = 0$. Наконец, в проводниках $\mathbf{E} = 0$. Таким образом, выражения $\mathbf{E}$ и $\mathbf{D}$ могут рассматриваться как напряженность и индукция какого-то электростатического поля. Заряды, возбуждающие такое поле, не могут находиться внутри диэлектрика, так как $4\pi\rho = \operatorname{div}\mathbf{D} = 0$. Остается выяснить физический смысл постоянных коэффициентов q_i, введенных выше чисто формально. Для этого замечаем, что заряд на поверхности i-го проводника согласно теореме Гаусса равен

$$Q_i = \frac{1}{4\pi} \oint\limits_{S_i} (\mathbf{D}\, d\mathbf{S}) = \frac{1}{4\pi} \sum_j q_j \oint\limits_{S_i} (\mathbf{D}_j\, d\mathbf{S}) = \frac{q_i}{4\pi} \oint\limits_{S_i} (\mathbf{D}_i\, d\mathbf{S}) = q_i.$$

На основании теоремы единственности можно поэтому сказать, что выражение (27.1) определяет электростатическое поле системы n проводников, заряды которых равны соответственно $q_1, q_2, \ldots, q_n$. Потенциал поля (27.1) определяется выражением

$$\varphi(\mathbf{r}) = \sum_{j=1}^{n} q_j V_j(\mathbf{r}). \tag{27.3}$$

Поместив точку $\mathbf{r}$ на поверхности i-го проводника, находим его потенциал

$$\varphi_i = \sum_{j=1}^{n} V_{ij} q_j. \tag{27.4}$$

Разрешив эти уравнения относительно q_i, получим

$$q_i = \sum_{j=1}^{n} C_{ij} \varphi_j. \tag{27.5}$$

Постоянные C_{ij} называются *емкостными коэффициентами* [1]. Как и потенциальные коэффициенты, они определяются только величиной,

[1] В старой литературе коэффициенты C_{ij} назывались *коэффициентами емкости*, когда $i = j$, и *коэффициентами индукции*, когда $i \neq j$. Мы не следуем этой терминологии.

формой и расположением проводников, а также диэлектрической проницаемостью промежуточной среды. В § 28 будет доказано, что емкостные, а следовательно, и потенциальные коэффициенты симметричны, т. е. $C_{ij} = C_{ji}$, $V_{ij} = V_{ji}$.

Таким образом, *заряды проводников являются линейными однородными функциями их потенциалов, а потенциалы — линейными однородными функциями зарядов.*

Если диэлектрик между проводниками однороден, то все емкостные коэффициенты C_{ik} будут пропорциональны его диэлектрической проницаемости ε.

Для конденсатора число проводников (обкладок) равно двум. В этом случае

$$q_1 = C_{11}\varphi_1 + C_{12}\varphi_2, \quad q_2 = C_{21}\varphi_1 + C_{22}\varphi_2,$$

причем $q_2 = -q_1$. Решив эти уравнения относительно φ_1 и φ_2, найдем разность потенциалов $\varphi_1 - \varphi_2$ и емкость конденсатора:

$$C = \frac{C_{11}C_{22} - C_{12}C_{21}}{C_{11} + C_{22} + C_{12} + C_{21}}. \tag{27.6}$$

2. Отметим некоторые свойства потенциальных и емкостных коэффициентов. Мы не будем приводить их строгие математические доказательства, а ограничимся интуитивными, но достаточно убедительными физическими соображениями.

Все потенциальные коэффициенты V_{ij} положительны. Действительно, если j-му проводнику сообщить положительный заряд q, а остальные проводники оставить незаряженными, то интуитивно ясно, что потенциал во всех точках пространства будет положителен. При этом $\varphi_i = V_{ij}q$. Из положительности φ_i следует $V_{ij} > 0$.

Интуитивно также ясно, что максимальным будет потенциал проводника, которому сообщен заряд, т. е. $\varphi_j > \varphi_i$. Отсюда следует, что $V_{jj} > V_{ij}$ $(i \neq j)$.

Емкостные коэффициенты C_{ij} с одинаковыми индексами положительны, а с разными — отрицательны. Действительно, заземлим все проводники, за исключением i-го. Тогда $q_i = C_{ii}\varphi_i$. Величины q_i и φ_i будут иметь одинаковые знаки, а потому должно быть $C_{ii} > 0$. Теперь заземлим все проводники, за исключением i-го и j-го, i-му проводнику сообщим положительный заряд q_i а j-й проводник оставим незаряженным. Тогда потенциалы φ_i и φ_j будут положительны, причем $q_j = 0 = C_{ji}\varphi_i + C_{jj}\varphi_i = 0$. Это равенство может соблюдаться только при условии $C_{ji} < 0$.

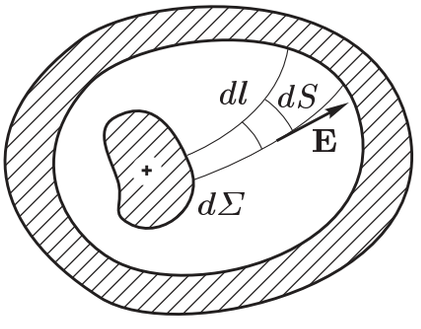


Рис. 83

Покажем, наконец, что $\sum_{j=1}^{n} C_{ij} \geqslant 0$. Для этого предположим, что i-й проводник, которому сообщен положительный заряд q_i, окружен со всех сторон заземленной замкнутой проводящей оболочкой произвольной формы (рис. 83). По теореме Фарадея на оболочке индуцируется отрицательный заряд $q' = -q_i$. Удалим в бесконечность некоторые части оболочки, чтобы из оставшихся частей образовалось $n - 1$ заземленных проводников с зарядами $q_1, q_2, \ldots, q_{i-1}, q_{i+1}, \ldots, q_n$. Сумма этих зарядов по абсолютной величине будет меньше q_i. С учетом

знаков можно написать $\sum\limits_{j=1}^{n} q_j \geqslant 0$. Подставив сюда $q_j = C_{ji}\varphi_i$ и приняв во внимание, что $\varphi_i > 0$, получим требуемый результат.

ЗАДАЧИ

1. К положительно заряженному уединенному проводнику подносится незаряженный изолированный проводник. Показать, что при этом потенциалы обоих проводников будут увеличиваться, а разность потенциалов между ними уменьшаться. Рассмотреть также случай, когда первый проводник был заряжен отрицательно.

2. Проводник заряжается от электрофора путем повторяющихся поднесений к пластинке, которая после каждого поднесения снова заряжается от того же электрофора до заряда Q. Пусть q_1 — заряд на проводнике после первой операции. Определить заряд q на проводнике после очень большого числа операций.

Р е ш е н и е. При поднесении проводника к пластинке общий заряд в определенном отношении распределяется между этими телами. При первом поднесении проводник получает заряд q_1, на пластинке остается $Q - q_1$. Если операция зарядки повторена многократно, то при последующих соприкосновениях проводника с пластинкой его заряд практически уже меняться не будет. Заряд пластинки также не будет меняться и будет равен Q, поскольку пластинка заряжается от электрофора. Искомый заряд q определится из пропорции

$$\frac{q}{Q} = \frac{q_1}{Q - q_1}.$$

3. Три одинаковых изолированных металлических шара расположены в вершинах равностороннего треугольника. Проволочкой, подключенной к удаленному заряженному проводнику, потенциал которого неизвестен, но поддерживается постоянным, по очереди касаются каждого из шаров. Заряды на первых двух шарах оказались после этого равными q_1 и q_2. Найти заряд на третьем шаре q_3.

Р е ш е н и е. В силу симметрии имеем $V_{11} = V_{22} = V_{33} = A$, $V_{12} = V_{21} = V_{23} = B$. При зарядке первого шара он получает потенциал $\varphi_1 = Aq_1$. При зарядке остальных двух шаров потенциал первого шара меняется, но его значения для решения не нужны. При зарядке второго шара его потенциал становится равным также $\varphi_1 = Aq_2 + Bq_1$. Аналогично для третьего шара: $\varphi_1 = Aq_3 + B(q_1 + q_2)$. Таким образом,

$$Aq_1 = Aq_2 + Bq_1 = Aq_3 + B(q_1 + q_2).$$

Отсюда

$$q_3 = \frac{q_2^2}{q_1}.$$

4. Четыре одинаковых изолированных металлических шара расположены в вершинах правильного тетраэдра. Проволочкой, подключенной к удаленному заряженному проводнику, потенциал которого неизвестен, но поддерживается постоянным, по очереди касаются каждого из шаров. Заряды на первых двух шарах оказались после этого равными q_1 и q_2. Найти заряды на двух остальных шарах.

О т в е т. $q_3 = q_2^2/q_1$, $q_4 = q_2^3/q_1^2$.

§ 28. Электрическая энергия

1. Электрическая энергия, как и всякая другая энергия, зависит только от *состояния системы*, но не зависит от способа, каким система была приведена в это состояние. Вычислим сначала электрическую энергию заряженного конденсатора. Она однозначно определяется зарядами его обкладок или разностью потенциалов между ними. Способ зарядки на величину энергии не влияет. Применим такой способ зарядки, чтобы вычисления были максимально просты. Если конденсатор не заряжен, то на каждой из его обкладок имеется смесь одинаковых количеств положительного и отрицательного электричеств. Будем переносить положительное электричество бесконечно малыми порциями dq с отрицательной обкладки на положительную. Для переноса заряда dq необходимо совершить работу против электрического поля:

$$\delta A^{\text{внеш}} = \varphi\, dq,$$

где φ — мгновенное значение разности потенциалов между обкладками. Работа самого конденсатора будет такой же по величине, но противоположной по знаку:

$$\delta A = -\varphi\, dq. \tag{28.1}$$

Зарядка конденсатора может сопровождаться выделением или поглощением тепла, а также изменением плотности диэлектрика. Однако в большинстве случаев эти эффекты незначительны, и мы временно оставим их без внимания. Тогда работа $\delta A^{\text{внеш}}$ целиком пойдет на увеличение электрической энергии конденсатора W, т. е.

$$dW = \varphi\, dq = \frac{q\, dq}{C}. \tag{28.2}$$

Если, как мы предположили, температура и плотность диэлектрика при зарядке не изменяются, то не будет изменяться также диэлектрическая проницаемость ε, а с ней и емкость конденсатора C. Поэтому интегрированием предыдущего выражения находим

$$W = \frac{q^2}{2C} = \frac{1}{2}\, q\varphi = \frac{1}{2}\, C\varphi^2. \tag{28.3}$$

2. Проведем теперь вычисление электрической энергии в общем виде. Рассмотрим несколько тел с зарядами q_1, q_2, ... и потенциалами φ_1, φ_2, ..., пространство между которыми заполнено неподвижным диэлектриком (однородным или неоднородным). Под q_1, q_2, ... здесь следует понимать только *свободные*, а не поляризационные заряды. Тела, на которых они находятся, могут быть как проводниками, так и диэлектриками. Проводники могут быть любых размеров. Размеры диэлектрических тел должны быть настолько малы, чтобы потенциал каждого тела мог с достаточной точностью считаться одинаковым во всех его точках. Этого всегда можно достигнуть, мысленно разделяя диэлектрик на достаточно малые части и считая каждую из них за отдельное диэлектрическое тело. Примем за начальное такое состояние,

в котором все тела не заряжены. Будем переносить электричество из бесконечности на эти тела бесконечно малыми порциями. Рассуждая так же, как в случае конденсатора, найдем

$$W = \int \sum_i \varphi_i' \, dq_i',$$

где суммирование ведется по всем заряженным телам. Штрихи над φ_i и q_i поставлены для того, чтобы указать, что эти величины *переменные*, т. е. меняются во время зарядки. Интеграл легко вычислить, используя то обстоятельство, что его значение *не зависит от способа зарядки.* Пусть q_i и φ_i — заряд и потенциал i-го тела в конечном состоянии. Осуществим зарядку так, чтобы в любой момент времени переменные заряды q_i' были пропорциональны их конечным значениям q_i:

$$q_i' = kq_i,$$

где k — переменная величина, одинаковая для всех зарядов q_i. Во время зарядки она возрастает от начального значения $k = 0$ до конечного $k = 1$. Поскольку связь между $\mathbf{D}$ и $\mathbf{E}$ предполагается линейной ($\mathbf{D} = \varepsilon\mathbf{E}$), увеличение всех зарядов в несколько раз ведет к увеличению всех потенциалов в такое же число раз. На этом основании можно написать

$$\varphi_i' = k\varphi_i.$$

Единственной переменной, определяющей при зарядке мгновенные значения зарядов и потенциалов, стала величина k. Ее мы и примем за переменную интегрирования. Очевидно, $dq_i' = q_i \, dk$, и, следовательно,

$$W = \sum \varphi_i q_i \int_0^1 k \, dk.$$

Выполнив интегрирование, получим

$$W = \frac{1}{2} \sum \varphi_i q_i. \tag{28.4}$$

Ограничение, касающееся размеров диэлектриков, введенное при получении формулы (28.4), можно снять, если записать эту формулу в виде

$$W = \frac{1}{2} \int \varphi\rho \, dV + \frac{1}{2} \int \varphi\sigma \, dS, \tag{28.5}$$

где ρ — объемная, а σ — поверхностная плотности (свободного) электричества. В таком виде формула справедлива при *любом* распределении проводящих и диэлектрических сред в пространстве. Интегрирование должно проводиться по всем свободным зарядам.

3. Рассмотрим теперь специально случай, когда все заряды находятся на проводниках. Тогда выражение для W надо взять в виде (28.4). Размеры проводников и их распределение в пространстве могут быть какими угодно. В частном случае конденсатора число проводников равно двум, а их заряды равны и противоположны по знаку. В этом

случае формула (28.4) переходит в (28.3). Пусть теперь число проводников произвольно. Если заряды на проводниках получат бесконечно малые приращения δq_i, то изменятся и потенциалы φ_i. Электрическая энергия изменится на

$$\delta W = \frac{1}{2}\sum \varphi_i\,\delta q_i + \frac{1}{2}\sum q_i\,\delta\varphi_i.$$

Но та же величина равна $\delta A^{\text{внеш}} = \sum \varphi_i\,\delta q_i$. Приравнивая оба выражения, находим

$$\sum \varphi_i\,\delta q_i = \sum q_i\,\delta\varphi_i. \tag{28.6}$$

С помощью этого соотношения легко доказать симметрию емкостных и потенциальных коэффициентов C_{ij} и V_{ij}. Для упрощения доказательства допустим, что заряженными являются только i-й и j-й проводники, причем потенциал φ_j поддерживается постоянным. Тогда

$$\varphi_i\,\delta q_i + \varphi_j\,\delta q_j = q_i\,\delta\varphi_i. \tag{28.6а}$$

Далее,

$$q_i = C_{ii}\varphi_i + C_{ij}\varphi_j, \quad q_j = C_{ji}\varphi_i + C_{jj}\varphi_j.$$

Ввиду постоянства φ_j

$$\delta q_i = C_{ii}\,\delta\varphi_i, \quad \delta q_j = C_{ji}\,\delta\varphi_j.$$

Подставляя эти значения в соотношение (28.6а), получим

$$C_{ij} = C_{ji}. \tag{28.7}$$

Отсюда следует также, что $V_{ij} = V_{ji}$. Этот результат можно, конечно, получить непосредственно из соотношения (28.6).

4. Учтем теперь, что поляризация диэлектрика, возникающая при возбуждении в нем электрического поля, может сопровождаться *изменением температуры* диэлектрика и появлением в нем *механических сил*, и *упругих напряжений.* Вследствие этого во время зарядки тел диэлектрическая проницаемость ε может измениться, поскольку она зависит от температуры и плотности диэлектрика. Это обстоятельство не сказывается на выражении для элементарной работы (28.1). Но последующее интегрирование этого выражения было проведено в предположении *постоянства* ε. Ясно, что таким путем нельзя получить выражение для внутренней электрической энергии системы. Однако выражения (28.3)–(28.5) определяют не менее важную физическую величину, а именно *свободную энергию системы*, точнее, ту часть ее, которая связана с электризацией тел. Чтобы показать это, переведем систему в конечное состояние в два этапа. Сначала в отсутствие электрического поля деформируем все тела и подведем к ним такое количество тепла, чтобы получилось то же распределение температуры и плотности, что и в конечном состоянии. Приращение свободной энергии в таком процессе обозначим $\Psi_{\text{упр}}$. Оно, очевидно, от электрического поля не зависит. Затем бесконечно малыми порциями будем

подводить к телам системы электричество, сохраняя неизменными температуру и плотность в каждой точке пространства. Диэлектрическая проницаемость среды изменяться не будет. Следовательно, на этом этапе применим способ вычисления работы, которым мы пользовались при выводе формул (28.3) и (28.5). Но, как известно из термодинамики, внешняя работа при изотермическом квазистатическом процессе идет на приращение свободной энергии системы. Значит, приращение свободной энергии на втором этапе процесса будет $\Psi_{\text{эл}} = W$. Если свободную энергию в начальном состоянии принять за нуль, то для полной свободной энергии системы в конечном состоянии можно написать

$$\Psi = \Psi_{\text{упр}} + W. \tag{28.8}$$

Первое слагаемое в правой части дает *упругую* часть свободной энергии, второе — *электрическую.*

§ 29. Локализация электрической энергии в пространстве

1. В предыдущем параграфе работу по электризации диэлектрика и его электрическую энергию мы выразили через заряды и потенциалы наэлектризованных тел. Но те же величины (работу и энергию) можно выразить через векторы $\mathbf{E}$ и $\mathbf{D}$, что более адекватно соответствует духу теории поля.

Начнем с вычисления элементарной работы $\delta^* A^{\text{внеш}}$, которую производят внешние силы при квазистатическом процессе электризации диэлектрика. При этом мы будем предполагать, что во время электризации объем диэлектрика остается неизменным, а сам диэлектрик — неподвижным. Это освобождает нас от необходимости учета работы упругих и вязких сил. Вся работа внешних сил затрачивается на отделение положительного электричества от отрицательного, и эта работа в конечном результате приводит к определенному распределению электричества в пространстве.

Возьмем в изотропном диэлектрике две бесконечно малые плоские площадки AB и CD, перпендикулярные к электрическому полю E (рис. 83). Расстояние между ними l предполагается бесконечно малым высшего порядка по сравнению с их линейными размерами. При этом условии площадки AB и CD могут рассматриваться как бесконечные плоскости, и краевые эффекты на них можно не учитывать. Пусть внешние силы, направленные против поля $\mathbf{E}$, переносят с площадки AB на площадку CD электрический заряд $\delta q = S\,\delta\sigma$, где S — площадь одной площадки, а $\delta\sigma$ — приращение поверхностной плотности электричества на площадке CD. Работа внешних сил против поля E при таком переносе равна

$$\delta^* A^{\text{внеш}} = \delta q E l = S l E\,\delta\sigma = V E\,\delta\sigma,$$

где V — бесконечно малый объем между AB и CD. В результате такого переноса, как это следует из теоремы Гаусса, электрическое поле

изменяется только между площадками, а вне их всюду остается неизменным. Длина вектора $\mathbf{D}$ между площадками изменяется на величину $dD = 4\pi\,\delta\sigma$, так что для $\delta^* A^{\text{внеш}}$ получается

$$\delta^* A^{\text{внеш}} = \frac{1}{4\pi}\, E\, dD. \tag{29.1}$$

При выводе этой формулы было использовано предположение об одинаковой направленности векторов $\mathbf{E}$ и $\mathbf{D}$, что в общем случае имеет место только для *изотропных сред.* В *анизотропных диэлектриках* (*кристаллах*) направления векторов $\mathbf{E}$ и $\mathbf{D}$, вообще говоря, различны. Но легко видеть, что в этом случае роль величины D играет проекция вектора $\mathbf{D}$ на направление поля $\mathbf{E}$. Поэтому формула (29.1) обобщается и принимает вид

$$\delta^* A^{\text{внеш}} = \frac{1}{4\pi}\, \mathbf{E}\, d\mathbf{D}. \tag{29.2}$$

2. Формула (29.2) справедлива для бесконечно малого объема V любой формы. Для доказательства достаточно разбить произвольный объем V бесконечно близкими параллельными плоскостями, заменить поверхность, ограничивающую объем V ступенчатой поверхностью, как это изображено на рис. 83, а затем перейти к пределу, неограниченно уменьшая расстояние между соседними плоскостями. Для каждого объемчика между такими плоскостями справедлива формула (29.2). Суммируя выражение (29.2) по всем объемчикам, мы получим выражение того же типа, в котором, однако, под V следует понимать уже полный объем выделенной бесконечно малой части диэлектрика. Существенно, что при указанном способе переноса зарядов, как это следует из теоремы Гаусса, электрическое поле меняется только внутри объема V, но остается неизменным вне этого объема.

Заметим еще, что заряды, переносимые с одной плоскости на другую, соседнюю, не обязательно должны быть одинаковы. Если они различны, то, вообще говоря, на рассматриваемых плоскостях будет происходить накопление (или убыль) электричества. При переходе к пределу, когда расстояние между соседними плоскостями беспредельно уменьшается, поверхностное распределение электричества на плоскостях перейдет в объемное. Отсюда следует, что формула (29.2) справедлива не только при отсутствии, но и при наличии в диэлектрике объемных зарядов. Впрочем, это и так очевидно, так как при выводе этой формулы никаких ограничений на распределение электричества по объему пространства не накладывалось.

Работа внешних сил, отнесенная к единице объема, найдется делением выражения (29.2) на V. Для указания на бесконечную малость этой величины мы будем пользоваться символом δ, в отличие от символа δ^*, который применялся для аналогичной работы в бесконечно малом объеме V. Таким образом, для работы, производимой внешними силами в единице объема неподвижного диэлектрика при бесконечно малом

смещении электрических зарядов, мы получаем выражение

$$\delta A^{\text{внеш}} = \frac{1}{4\pi}\,\mathbf{E}\,d\mathbf{D}. \tag{29.3}$$

Существенно отметить *универсальность* формулы (29.3), т. е. ее применимость для любых диэлектриков, поскольку никакая конкретная связь между векторами $\mathbf{E}$ и $\mathbf{D}$ при ее выводе не использовалась.

3. Допустим теперь, что кроме работы по электризации диэлектрика никаких других работ над диэлектриком не производится. При этом, как уже указывалось в предыдущем параграфе, может выделяться тепло. Но мы пока отвлечемся от этого осложняющего обстоятельства и учтем его только в следующем параграфе. Тогда формула (29.3) будет определять элементарное приращение электрической энергии в единице объема пространства W:

$$dW = \frac{1}{4\pi}\,\mathbf{E}\,d\mathbf{D}. \tag{29.4}$$

Если известна зависимость вектора $\mathbf{D}$ от вектора $\mathbf{E}$, то сама электрическая энергия W в единице объема найдется интегрированием этого выражения по всему процессу:

$$W = \int dW = \frac{1}{4\pi}\int \mathbf{E}\,d\mathbf{D}. \tag{29.5}$$

В частности, если $\mathbf{D} = \varepsilon\mathbf{E}$ и во время процесса электризации ε остается постоянной, то

$$W = \frac{\varepsilon}{4\pi}\int \mathbf{E}\,d\mathbf{E} = \frac{\varepsilon}{8\pi}\,\mathbf{E}^2 = \frac{1}{8\pi}\,\mathbf{E}\mathbf{D}. \tag{29.6}$$

Если при этом ε не зависит от температуры, то эта формула определяет *внутреннюю электрическую энергию.* Если же ε зависит от температуры, то при выводе формулы (29.6) надо предполагать, что температура диэлектрика постоянна. В этом случае формулой (29.6) определяется *свободная электрическая энергия* в единице объема. Все это находится в соответствии с тем, что было сказано по поводу формулы (28.5) в п. 4 предыдущего параграфа.

4. Аналогичное вычисление энергии можно провести и для кристаллов. В этом случае вместо скаляра ε надо пользоваться тензором диэлектрической проницаемости ε_{ij}: зависимость между компонентами векторов $\mathbf{D}$ и $\mathbf{E}$ определяется формулой (15.5). При этом надо воспользоваться симметрией тензора ε_{ij}, т. е. соотношением $\varepsilon_{ij} = \varepsilon_{ji}$ $(i, j = x, y, z)$. Об этой симметрии упоминалось в § 15 без доказательства. Приведем теперь это доказательство.

В кристаллах выражение (29.4) определяет дифференциал внутренней электрической энергии (когда тензор ε_{ij} не зависит от температуры) в единице объема. Когда ε_{ij} является функцией температуры, то то же выражение есть дифференциал свободной электрической энергии. Какой из этих двух случаев осуществляется — безразлично. Существенно только то, что величина W является *функцией состояния*,

а следовательно, дифференциал dW — *полным дифференциалом.* Для него можно написать

$$4\pi\, dW = \mathbf{E}\, d\mathbf{D} = \sum_i E_i\, dD_i = \sum_i E_i\, d \sum_j \varepsilon_{ij} E_j = \sum_i \sum_j \varepsilon_{ij} E_i\, dE_j. \tag{29.7}$$

Чтобы это выражение было полным дифференциалом, необходимо и достаточно, чтобы

$$\frac{\partial(\varepsilon_{ij} E_i)}{\partial E_i} = \frac{\partial(\varepsilon_{ji} E_j)}{\partial E_j}.$$

Так как тензор ε_{ij} не зависит от напряженности поля, то отсюда получается

$$\varepsilon_{ij} = \varepsilon_{ji}. \tag{29.8}$$

Распространим теперь формулу (29.6) на случай кристаллических сред. Для этого следует проинтегрировать выражение (29.4). Прежде всего заметим, что значение суммы не зависит от того, какими буквами обозначены индексы суммирования. Пользуясь этим, в выражении (29.7) заменим индекс i на j, а индекс j — на i:

$$4\pi\, dW = \sum_j \sum_i \varepsilon_{ij} E_j\, dE_i,$$

или в силу (29.8)

$$4\pi\, dW = \sum_j \sum_i \varepsilon_{ij} E_j\, dE_i = \sum_i \sum_j \varepsilon_{ij} E_j\, dE_i.$$

Сложив это с (29.7), получим

$$8\pi\, dW = \sum_i \sum_j \varepsilon_{ij} (E_i\, dE_j + E_j\, dE_i) = \sum_i \sum_j \varepsilon_{ij}\, d(E_i E_j).$$

Ввиду независимости тензора ε_{ij} от напряженности поля, интегрирование легко выполняется и дает

$$W = \frac{1}{8\pi} \sum_i \sum_j \varepsilon_{ij}\, dE_j = \frac{1}{8\pi} \sum_i E_i D_i = \frac{1}{8\pi} (\mathbf{ED}).$$

Таким образом, и для кристаллов мы приходим к прежней формуле (29.6).

5. Полная электрическая энергия (внутренняя или свободная) выражается формулой

$$W = \frac{1}{8\pi} \int \mathbf{ED}\, dV = \frac{1}{8\pi} \int \varepsilon \mathbf{E}^2\, dV = \frac{1}{8\pi} \int \frac{\mathbf{D}^2}{\varepsilon}\, dV, \tag{29.9}$$

где интегрирование распространяется по всему бесконечному пространству. Для той же величины ранее была получена формула (28.5). С чисто математической точки зрения формулы (28.5) и (28.9) эквивалентны и отличаются одна от другой только по форме.

Однако за формальным различием этих формул стоит и различие в физической интерпретации электрической энергии. Уравнение (29.9)

выражает электрическую энергию в виде предела бесконечного множества слагаемых, каждое из которых равно $(\varepsilon/8\pi)\,\mathbf{E}^2\,dV$ и относится к определенному элементу объема dV. Это можно понимать в том смысле, что *носителем электрической энергии является электрическое поле*, причем энергия поля *локализована* в пространстве так, что в каждой единице объема содержится энергия W, определяемая формулой (29.8), в которой $\mathbf{E}$ означает напряженность электрического поля в данном элементе объема. Величина W называется *объемной плотностью электрической энергии.* Напротив, выражение (28.5) может быть формально истолковано как *потенциальная энергия взаимодействия электрических зарядов*, и притом *взаимодействия на расстоянии.* Такое истолкование исключает представление о локализации энергии в определенных участках пространства. В нем акцент ставится на электрические заряды, в частности на заряды проводников, находящихся в диэлектрике. При первой же интерпретации центр тяжести переносится в диэлектрики, окружающие эти проводники. Внутри проводников $\mathbf{E} = 0$, и в них совсем не локализована электрическая энергия. Она целиком локализована в диэлектриках.

Какому же из этих двух представлений об электрической энергии следует отдать предпочтение? В рамках электростатики *принципиально невозможно* указать ни одного опыта, который позволил бы сделать выбор между ними. Дело в том, что в электростатике электрическое поле неотделимо от зарядов, являющихся его источниками. Величиной и расположением зарядов *однозначно* определяется электростатическое поле. Обратно, заданием поля во всем пространстве *также однозначно* определяется плотность электрических зарядов. Не так обстоит дело в случае переменных полей. Переменные электромагнитные поля могут существовать *самостоятельно*, *независимо от возбудивших их электрических зарядов.* Заряды могут нейтрализоваться, а поле, которое они возбудили, может продолжать существовать в виде электромагнитных волн, которым присущ определенный запас энергии. Эта энергия не может быть представлена как потенциальная энергия зарядов, взаимодействующих на расстоянии, поскольку самих зарядов уже нет. Формула (28.5) теряет смысл. Но формула (29.9), а также выражения (29.6) сохраняют смысл и для переменных электромагнитных полей.

Представляют ли они в этом случае электрическую энергию и ее плотность — этот вопрос требует особого исследования. Во всяком случае нельзя выдвинуть каких-либо возражений против возможности представления электромагнитной энергии через напряженности электрического и магнитного полей. Сомнения могут относиться только к конкретным формулам, с помощью которых производится такое представление. Следовательно, если электростатику рассматривать как предельный случай электродинамики, то *даже в электростатике* следует отдать предпочтение теории поля с ее представлением о локализации электрической энергии в пространстве.

6. Заметим в заключение, что математическая эквивалентность выражений (28.5) и (29.9) для статических полей может быть использо-

вана для доказательства единственности решения электростатической задачи, сформулированной в п. 1 § 22. Действительно, предположим, что задача допускает несколько решений. Возьмем два из них: 1) $\mathbf{E}_1 = -\operatorname{grad}\varphi_i$, $\mathbf{D}_1 = \varepsilon\mathbf{E}_1$; 2) $\mathbf{E}_2 = -\operatorname{grad}\varphi_2$, $\mathbf{D}_2 = \varepsilon\mathbf{E}_2$. Оба решения могут отличаться значениями плотности связанных электрических зарядов. Но плотности свободных электрических зарядов ρ_1 и ρ_2 должны быть одинаковы, так как они заданы. Если во втором решении изменить знаки всех свободных и поляризационных зарядов, то очевидно возникшее при этом поле будет равно $\mathbf{E}_2' = -\mathbf{E}_2$, так как свободными и поляризационными зарядами электростатическое поле определяется однозначно. В силу принципа суперпозиции вектор

$$\mathbf{E} \equiv \mathbf{E}_1 + \mathbf{E}_2' = \mathbf{E}_1 - \mathbf{E}_2$$

будет представлять также какое-то электростатическое поле. Этому полю соответствует объемная плотность свободных зарядов $\rho = \rho_1 - \rho_2$ и потенциал $\varphi = \varphi_1 - \varphi_2$. Потенциал φ не может обращаться в тождественный нуль, так как по нашему предположению соответствующее поле $\mathbf{E}$ во всех точках пространства не обращается в нуль. Ввиду эквивалентности формул (28.5) и (29.9) можно написать

$$\int \frac{\varepsilon\mathbf{E}^2}{8\pi}\,dV = \frac{1}{2}\int \varphi\rho\,dV + \frac{1}{2}\int \varphi\sigma\,dS.$$

Первый интеграл в правой части равен нулю, так как $\rho = 0$. Поверхностный интеграл достаточно взять только по поверхностям проводников, так как на всякой поверхности, проходящей в диэлектрике, $\sigma \equiv \sigma_1 - \sigma_2 = 0$. Поскольку потенциал каждого проводника $\varphi^{(i)}$ постоянен, поверхностный интеграл можно представить в виде

$$\int \varphi\sigma\,dS = \sum_i \varphi_i \oint\limits_{S^{(i)}} \sigma\,dS = \sum_i \varphi^{(i)} q^{(i)},$$

где $q_i = q_1^{(i)} - q_2^{(i)}$ — полный заряд i-то проводника. Если задан потенциал i-го проводника, то $\varphi^{(i)} = 0$, а если задан заряд, то $q^{(i)} = 0$. В обоих случаях $\varphi^{(i)} q^{(i)} = 0$. Таким образом,

$$\int \frac{\varepsilon\mathbf{E}^2}{8\pi}\,dV = 0.$$

Ввиду положительности ε отсюда следует, что $\mathbf{E}^2 = 0$. Следовательно, $\mathbf{E} = \mathbf{E}_1 - \mathbf{E}_2 = 0$, т. е. введенное нами предположение неверно. Это и доказывает единственность решения электростатической задачи.

ЗАДАЧИ

1. Вычислить электрическую энергию шара радиуса a, заряд которого q равномерно распределен по его поверхности.

Ответ. $W = q^2/(2a)$.

2. То же для шара, заряд которого равномерно распределен по его объему.

Ответ. $W = 3q^2/(5a)$.

3. Доказать математическую эквивалентность выражений (28.5) и (29.9), используя математическую теорему Гаусса–Остроградского.

Р е ш е н и е. Не нарушая общности, можно сначала считать, что поверхностных зарядов нет. Тогда векторы $\mathbf{E}$ и $\mathbf{D}$ можно дифференцировать. Подставляя в формулу (29.9) $\mathbf{E} = -\operatorname{grad}\varphi$ и воспользовавшись тождеством $\operatorname{div}(\varphi\mathbf{D}) = \varphi \operatorname{div}\mathbf{D} + \mathbf{D}\operatorname{grad}\varphi$, а также соотношением (13.5), получим

$$W = -\frac{1}{8\pi}\int \operatorname{div}(\varphi\mathbf{D})\,dV + \frac{1}{2}\int \varphi\rho\,dV,$$

где интегрирование производится по всему бесконечному пространству. Первый интеграл в правой части этого равенства равен нулю. Действительно, предполагая, что все заряды находятся в конечной области пространства, окружим ее сферой S большого радиуса r и воспользуемся формулой Гаусса–Остроградского

$$\int\limits_V \operatorname{div}(\varphi\mathbf{D})\,dV = \oint\limits_S \varphi(\mathbf{D}\,d\mathbf{S}),$$

где V означает объем, ограниченный сферой S. Асимптотически при $r \to \infty$ вектор $\mathbf{D}$ на сфере S меняется с радиусом r так же, как в случае точечного заряда, т. е. пропорционально $1/r^2$, а потенциал убывает пропорционально $1/r$. Следовательно, $\varphi\mathbf{D}$ будет убывать асимптотически как $1/r^3$. Поверхность же S возрастает пропорционально r^2. Значит, поверхностный интеграл в правой части предыдущего равенства асимптотически убывает как $1/r$ и в пределе при $r \to 0$ (т. е. при переходе к бесконечному пространству) обратится в нуль. Итак,

$$W = \frac{1}{2}\int \varphi\rho\,dV.$$

Введя элемент заряда $dq = \rho\,dV$, перепишем это равенство в виде

$$W = \frac{1}{2}\int \varphi\,dq.$$

После этого отпадает необходимость в специальном исследовании поверхностных зарядов. Достаточно заметить, что заряженную поверхность можно рассматривать как предельный случай тонкого слоя, заряженного по объему. К такому слою последнее выражение применимо. Разделив элементы заряда на объемные $\rho\,dV$ и поверхностные $\sigma\,dS$, получим

$$W = \frac{1}{2}\int \varphi\rho\,dV + \frac{1}{2}\int \varphi\sigma\,dS,$$

т. е. формулу (28.5).

§ 30. Взаимная энергия точечных зарядов

1. Пусть точечные заряды q_1 и q_2 находятся в вакууме на бесконечном расстоянии друг от друга. Чтобы их сблизить до расстояния r_{12}, надо затратить работу $q_1 q_2/r_{12}$. Потенциальная энергия взаимодействия зарядов будет

$$U = \frac{q_1 q_2}{r_{12}}. \tag{30.1}$$

Для нескольких точечных зарядов

$$U = \frac{1}{2}\sum\sum_{i\neq k}\frac{q_i q_k}{r_{ik}}. \tag{30.2}$$

Коэффициент $1/2$ поставлен потому, что при суммировании потенциальная энергия каждой пары зарядов учитывается дважды: в виде слагаемого $q_i q_k/r_{ik}$ и в виде равного ему слагаемого $q_k q_i/r_{ki}$. Формулу (30.2) можно представить в виде

$$U = \frac{1}{2}\sum \varphi_i q_i, \tag{30.3}$$

где φ_i — потенциал в точке нахождения i-го заряда, создаваемый всеми остальными зарядами:

$$\varphi_i = \sum_{k\neq i}\frac{q_k}{r_{ik}}. \tag{30.4}$$

2. По внешнему виду формула (30.3) совпадает с аналогичной формулой (28.4) для электрической энергии заряженных проводников. На самом деле между обеими формулами имеется глубокое различие. Это видно уже из того, что выражение (28.4) может быть преобразовано в объемный интеграл (29.9), который всегда положителен. Выражение (30.3) не допускает такого преобразования, так как оно может быть и положительным и отрицательным. Например, оно отрицательно для двух точечных зарядов противоположных знаков. Каждый заряд q_i, взятый в отдельности, обладает электрической энергией. Она называется *собственной энергией* заряда q_i, и представляет собой энергию взаимного отталкивания бесконечно малых частей, на которые его можно мысленно разбить. Эта энергия учитывалась при выводе формулы (28.4), но не учитывалась при выводе формулы (30.3). При получении формулы (30.3) каждый заряд q_i рассматривался как нечто целое и неизменное. Учитывалась только работа, затрачиваемая на *сближение* таких неизменных зарядов, но не на их *образование.* Напротив, при выводе формулы (28.4) учитывалась также работа, затрачиваемая на *образование* зарядов q_i путем конденсации их из бесконечно малых порций электричества, переносимых из бесконечности. В соответствии с этим формула (28.4) определяет *полную электрическую энергию системы зарядов*, а формула (30.3) — только их *взаимную потенциальную энергию.* В формуле (28.4) φ_i означает потенциал проводника, создаваемый всеми зарядами, а в формуле (30.3) — всеми зарядами, за исключением i-го.

3. Для лучшего уяснения вопроса рассмотрим два бесконечно малых шарика неизменных размеров. Пусть сначала шарики не заряжены, бесконечно далеко находятся друг от друга, а электричество распределено по всему бесконечному пространству с бесконечно малой плотностью. Соберем все электричество на шариках. Так как расстояние между ними бесконечно велико, то они не будут оказывать никакого

влияния друг на друга. Вся работа пойдет на увеличение собственных энергий шариков. Эти энергии будут равны соответственно

$$W_1 = \frac{1}{8\pi}\int E_1^2\,dV, \quad W_2 = \frac{1}{8\pi}\int E_2^2\,dV.$$

Затем сблизим заряженные шарики, для чего потребуется совершить работу $U = q_1 q_2/r_{12}$ (расстояние между шариками r_{12} должно быть очень велико по сравнению с их размерами). Полная электрическая энергия шариков будет

$$W = \frac{1}{8\pi}\int E_1^2\,dV + \frac{1}{8\pi}\int E_2^2\,dV + U.$$

Но ту же энергию можно выразить иначе. Пока шарики не заряжены, сблизим их до расстояния r_{12}, а затем будем собирать на них электричество. Потребуется работа

$$W = \frac{1}{8\pi}\int(\mathbf{E}_1+\mathbf{E}_2)^2\,dV = \frac{1}{8\pi}\int \mathbf{E}_1^2\,dV + \frac{1}{8\pi}\int \mathbf{E}_2^2\,dV + \frac{1}{4\pi}\int(\mathbf{E}_1\mathbf{E}_2)\,dV.$$

Сравнивая оба выражения, находим

$$U = \frac{1}{4\pi}\int(\mathbf{E}_1\mathbf{E}_2)\,dV. \tag{30.5}$$

§ 31. Термодинамика диэлектриков

1. Применим к процессу поляризации диэлектриков начала термодинамики. Будем предполагать, что диэлектрики *изотропны* как в отсутствие, так и при наличии электрического поля. К таким диэлектрикам относятся, например, жидкости и газы. Выделим мысленно достаточно малую часть диэлектрика, которая с достаточной точностью может считаться однородной. С той же точностью могут рассматриваться как однородные давление $\mathscr{P}$, а также напряженность электрического поля внутри этой части. Первое начало термодинамики для выделенной части запишем в виде

$$\delta A^{\text{внеш}} + \delta Q = dU, \tag{31.1}$$

где δQ — количество сообщенной теплоты, а dU — приращение внутренней энергии. Работа над диэлектриком $\delta A^{\text{внеш}}$ слагается из двух частей. Первая часть $\mathscr{P}\,dV$ есть работа, производимая внешним давлением. Вторая идет на электризацию диэлектрика и представляется выражением (29.3). Влияние слагаемого $\mathscr{P}\,dV$ было подробно исследовано во втором томе нашего курса. Поэтому в целях сокращения изложения мы его здесь опустим. Таким образом, мы предполагаем, что поляризация диэлектрика либо сопровождается незначительными изменениями его объема, либо этот объем поддерживается постоянным. Кроме того, не теряя общности, мы можем относить все величины к единице объема диэлектрика, т. е. положить $V = 1$. В результате получим

$$\delta Q = dU - \frac{1}{4\pi}\,E\,dD. \tag{31.2}$$

2. Введем энтропию системы S, абсолютную температуру T, а также термодинамические функции: свободную энергию

$$\Psi = U - TS, \tag{31.3}$$

термодинамический потенциал

$$\Phi = \Psi - \frac{1}{4\pi} ED \tag{31.4}$$

и энтальпию

$$I = U - \frac{1}{4\pi} ED. \tag{31.5}$$

Для квазистатических процессов $\delta Q = T\,dS$ и уравнение (31.2) запишется в виде

$$dU = T\,dS + \frac{1}{4\pi} E\,dD. \tag{31.6}$$

Используя это соотношение и взяв дифференциалы от выражений (31.3)–(31.5), найдем

$$d\Psi = -S\,dT + \frac{1}{4\pi} E\,dD, \tag{31.7}$$

$$d\Phi = -S\,dT - \frac{1}{4\pi} D\,dE, \tag{31.8}$$

$$dI = T\,dS - \frac{1}{4\pi} D\,dE. \tag{31.9}$$

Уравнения (31.6)–(31.9) являются *основными уравнениями термодинамики диэлектриков.* Для получения из них конкретных выводов их необходимо дополнить «*уравнением состояния*». Роль такого уравнения играет соотношение вида

$$D = f(E, T, \tau), \tag{31.10}$$

где τ — плотность вещества диэлектрика. Такое уравнение не может быть выведено чисто термодинамическими методами. Оно должно быть заимствовано либо из опыта, либо из молекулярной теории поляризации диэлектриков.

3. Считая, что векторы $\mathbf{D}$ и $\mathbf{E}$ связаны уравнением состояния (31.10), и интегрируя выражение (31.7) при постоянных T и τ, получим

$$\Psi = \frac{1}{4\pi} \int E\,dD + \Psi_0(T, \tau), \tag{31.11}$$

где $\Psi_0(T, \tau)$ — постоянная интегрирования. Она, очевидно, имеет смысл свободной энергии диэлектрика при отсутствии в нем электрического поля. Таким образом, мы вновь приходим к выводу, что формула (29.5) выражает не внутреннюю, а *свободную энергию* диэлектрика, точнее, ту ее часть, которая зависит от напряженности электрического поля. Возьмем уравнение состояния в виде $\mathbf{D} = \varepsilon\mathbf{E}$. Величина ε зависит только от T и τ, которые при интегрировании в формуле (31.11) должны

оставаться постоянными. Поэтому интегрирование дает

$$\Psi = \frac{\varepsilon E^2}{8\pi} + \Psi_0 = \frac{D^2}{8\pi\varepsilon} + \Psi_0. \tag{31.12}$$

Найдем теперь выражение для внутренней энергии диэлектрика U. Из формул (31.3) и (31.7) получаем

$$U = \Psi - T\left(\frac{\partial\psi}{\partial T}\right)_{D,\tau}, \tag{31.13}$$

причем мы явно отметили, что при дифференцировании плотность τ должна оставаться постоянной, так как все наше рассмотрение проведено в предположении постоянства τ. Используя выражение (31.2), из формулы (31.13) найдем

$$U = \left(1 + \frac{T}{\varepsilon}\frac{\partial\varepsilon}{\partial T}\right)\frac{D^2}{8\pi\varepsilon} + U_0(T,\tau) = \left(\varepsilon + T\frac{\partial\varepsilon}{\partial T}\right)\frac{E^2}{8\pi} + U_0(T,\tau), \tag{31.14}$$

где функция $U_0(T,\tau)$ имеет смысл *внутренней энергии* диэлектрика при отсутствии в нем электрического поля. *Если диэлектрическая проницаемость ε не зависит от температуры, то электрические составляющие свободной и внутренней энергий диэлектрика равны между собой. При наличии температурной зависимости ε это равенство уже несправедливо.*

4. Если изменение поляризации диэлектрика производится *квазистатически* и *адиабатически*, то оно, вообще говоря, сопровождается изменением *температуры* диэлектрика. Такое изменение температуры называется *электрокалорическим эффектом*. При квазистатическом адиабатическом процессе энтропия S остается постоянной. Рассматривая ее как функцию E и T (τ поддерживается постоянной), можно написать для бесконечно малого процесса

$$\left(\frac{\partial S}{\partial T}\right)_E \Delta T + \left(\frac{\partial S}{\partial E}\right)_T \Delta E = 0.$$

Очевидно,

$$\left(\frac{\partial S}{\partial T}\right)_E = \frac{1}{T}\left(\frac{T\partial S}{\partial T}\right)_E = \frac{1}{T}\left(\frac{\delta Q}{\delta T}\right)_E = \frac{C_E}{T},$$

где C_E — теплоемкость единицы объема диэлектрика при постоянной напряженности электрического поля E. Далее, из формулы (31.8) получаем

$$\left(\frac{\partial S}{\partial E}\right)_T = \frac{1}{4\pi}\left(\frac{\partial D}{\partial T}\right)_E = \frac{E}{4\pi}\frac{\partial\varepsilon}{\partial T}.$$

Таким образом,

$$\Delta T = -\frac{TE}{4\pi C_E}\frac{\partial\varepsilon}{\partial T}\Delta E. \tag{31.15}$$

Величина $\partial\varepsilon/\partial T$ обычно отрицательна. В этом случае диэлектрик, помещенный в диэлектрическое поле, охлаждается при адиабатическом уменьшении напряженности поля. Аналогично ведет себя парамагнетик в магнитном поле при адиабатическом размагничивании. Это явление используется для получения сверхнизких температур (см. § 73). При конечном изменении напряженности поля от E_1 до E_2 температура диэлектрика изменяется на

$$T_2 - T_1 = -\int_{E_1}^{E_2} \frac{TE}{4\pi C_E}\,\frac{\partial\varepsilon}{\partial T}\,dE. \tag{31.16}$$

5. Иногда внутреннюю энергию и прочие термодинамические функции диэлектрика определяют несколько иначе. Исключим из формулы (31.2) D, пользуясь определением (13.3). Получим

$$\delta Q = dU - \frac{1}{4\pi}\,E\,dE - E\,dP.$$

Затем введем новую функцию состояния

$$U' = U - E^2/(8\pi). \tag{31.17}$$

Тогда

$$\delta Q = dU' - E\,dP. \tag{31.18}$$

Функция U' получается вычитанием из U величины $E^2/(8\pi)$, которая может быть истолкована как *плотность электрической энергии в вакууме.* Таким образом, U' имеет смысл той части плотности внутренней энергии диэлектрика, которая связана с его *поляризацией.* Однако в применениях термодинамики такая интерпретация не обязательна. Для термодинамики важно только, что величина U' есть *функция состояния.* Она может быть использована в термодинамических соотношениях вместо внутренней энергии U, и притом с тем же успехом. Величину U' можно было бы назвать *энергией поляризации диэлектрика.* С введением этой величины соотношение (31.6) переходит в

$$dU' = T\,dS + E\,dP. \tag{31.19}$$

Свободная энергия поляризации диэлектрика должна быть определена выражением $F' = U' - TS$. Соответственно изменяются определения (31.4) и (31.5), а также термодинамические соотношения (31.7)–(31.9).

ЗАДАЧА

Вычислить теплоемкости единицы объема диэлектрика C_D и C_E при постоянных D и E соответственно. Как осуществить нагревание при постоянном D и при постоянном E?

Решение.

$$C_D = T\left(\frac{\partial S}{\partial T}\right)_D = -T\left(\frac{\partial^2\Psi}{\partial T^2}\right)_D, \quad C_E = T\left(\frac{\partial S}{\partial T}\right)_E = -T\left(\frac{\partial^2\Phi}{\partial T^2}\right)_E.$$

Подставив сюда

$$\Psi = \Psi_0 + \frac{\varepsilon E^2}{8\pi} \quad \text{и} \quad \Phi = \Psi - \frac{1}{4\pi} ED = \Psi_0 - \frac{\varepsilon E^2}{8\pi}$$

и выполнив дифференцирование, получим

$$C_D = C_V - T\frac{D^2}{8\pi}\frac{d}{dt^2}\left(\frac{1}{\varepsilon}\right) = C_V + \frac{TE^2}{8\pi}\frac{d^2\varepsilon}{dT^2} - \frac{TE^2}{4\pi\varepsilon}\left(\frac{d\varepsilon}{dT}\right)^2, \tag{31.20}$$

$$C_E = C_V + \frac{TE^2}{8\pi}\frac{d^2\varepsilon}{dT^2}, \tag{31.21}$$

где C_V — теплоемкость единицы объема диэлектрика при постоянном объеме в отсутствие электрического поля. Теплоемкость C_D экспериментально можно определить, нагревая диэлектрик в плоском заряженном конденсаторе, отключенном от источника напряжения, а теплоемкость C_E — в конденсаторе, между обкладками которого поддерживается постоянное напряжение.

§ 32. Свободная энергия и силы

1. В электрическом поле на диэлектрики и проводники действуют силы. Их называют *пондеромоторными силами*, т. е. силами, действующими на *весомые тела*. Этот термин был введен в то время, когда в физике, наряду с обычными веществами, признавалось существование многих *невесомых субстанций* (теплород, эфир, электрические и магнитные жидкости и пр.). Теперь он устарел, так как невесомых субстанций не существует. Однако мы сохраняем его за неимением другого. Первопричиной возникновения пондеромоторных сил являются электрические заряды, сообщаемые телам. Однако сообщение зарядов телам осложняется появлением поляризационных зарядов и упругих деформаций в диэлектриках и проводниках. Вычисление пондеромоторных сил с одновременным исследованием механизма их возникновения в общем случае довольно затруднительно. Термодинамика дает общий метод вычисления пондеромоторных сил, отвлекаясь от такого механизма.

2. Поясним этот метод на примере плоского конденсатора, пространство между обкладками которого заполнено каким-то диэлектриком. Зарядим конденсатор, а затем отключим его от источника электричества, поддерживая тем самым заряды на пластинах постоянными. Между пластинами возникнут силы притяжения. Обозначим через $\mathbf{F}$ одну из них, например силу, действующую на положительно заряженную пластину (рис. 84). Пластину, заряженную отрицательно, закрепим, а к положительно заряженной пластине приложим внешнюю силу $\mathbf{F}'$, уравновешивающую силу $\mathbf{F}$. Если нарушить равновесие, бесконечно мало изменив силу $\mathbf{F}'$, то положительная пластина начнет бесконечно

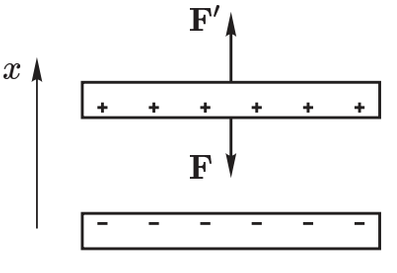


Рис. 84

медленно перемещаться. Кинетическую энергию возникшего движения можно не учитывать, так как процесс можно проводить бесконечно медленно (квазистатически) и достаточно долго. Тогда сила $\mathbf{F}'$ будет отличаться от силы $\mathbf{F}$ бесконечно мало, а потому с точностью до бесконечно малых высшего порядка работы этих сил $\delta A'$ и δA будут равны по величине и противоположны по знаку: $\delta A' = -\delta A$. Процесс может сопровождаться *выделением* или *поглощением теплоты*. Будем эту теплоту отводить, чтобы *температура системы оставалась постоянной*. Тогда работа внешней силы $\mathbf{F}'$ пойдет на приращение свободной энергии системы: $\delta A' = d\Psi$, или

$$\delta A + (d\Psi)_{q,T} = 0. \tag{32.1}$$

Значки q и T указывают, что приращение свободной энергии должно быть вычислено при постоянных q и T. Вычислив по формуле (32.1) работу δA, можно затем найти и искомую силу $\mathbf{F}$.

Формула (32.1) носит общий характер. Она применима к *любым системам*, а не только к плоскому конденсатору. Надо только под δA понимать работу *всех сил*, действующих в системе при произвольных бесконечно малых смещениях входящих в нее тел и диэлектрической среды между ними. Такие смещения называются *возможными* или *виртуальными смещениями* в отличие от *действительных смещений*, которые возникают в системе при том или ином ее движении. В связи с этим отметим, что при получении формулы (32.1) мы допустили некоторую неточность. При смещении пластины конденсатора диэлектрик может входить или выходить из конденсатора. Величина δA должна включать и работу сил при таком смещении диэлектрика.

3. Согласно формуле (28.8) полная свободная энергия Ψ слагается из электрической части $\Psi_{\text{эл}} = W$ и упругой части $\Psi_{\text{упр}}$. На такие же две части можно разделить и виртуальную работу δA, причем

$$\delta A_{\text{эл}} + (dW)_{q,T} = 0, \tag{32.2}$$

$$\delta A_{\text{упр}} + (dW_{\text{упр}})_{q,T} = 0. \tag{32.3}$$

Таким способом пондеромоторные силы, действующие на проводники и диэлектрики, можно разделить на *электрические* и *упругие*.

Электрическую часть виртуальной работы $\delta A_{\text{эл}}$ можно вычислить иначе. Формула (32.2) относится к тому случаю, когда при виртуальных смещениях *заряды тел поддерживаются постоянными*. Допустим теперь, что тела системы соединены с источниками электричества (например, гальваническими батареями), которые при виртуальных смещениях поддерживают постоянными *потенциалы всех тел*. Пондеромоторные силы, а с ними и величина виртуальной работы $\delta A_{\text{эл}}$, конечно, не зависят от того, происходят ли виртуальные смещения при постоянных зарядах, при постоянных потенциалах или как-либо иначе. Введение виртуальных смещений надо рассматривать как *искусственный прием* для вычисления действующих сил. Но виртуальные смещения к вопросу о возникновении этих сил не имеют прямого отношения.

Однако приращение свободной энергии dW при виртуальных смещениях существенно зависит от того, при каких условиях эти смещения производятся. Поясним это сначала на примере конденсатора. При виртуальных смещениях его обкладок или диэлектрической среды между ними меняется емкость конденсатора C. Если конденсатор отключен от источника электричества ($q = \text{const}$), то приращение свободной энергии $W = q^2/(2C)$ будет

$$(dW)_{q,T} = \frac{1}{2}\, q^2\, \delta\left(\frac{1}{C}\right) = -\frac{1}{2}\,\frac{q^2}{C^2}\,\delta C = -\frac{1}{2}\varphi^2\,\delta C,$$

а виртуальная работа, согласно формуле (32.2), $\delta A_{\text{эл}} = (1/2)\varphi^2\,\delta C$. Если же конденсатор подключен к гальванической батарее ($\varphi = \text{const}$), то, представив W в виде $W = (1/2)\,\varphi^2 C$, получим

$$(dW)_{\varphi,T} = \frac{1}{2}\varphi^2\,\delta C.$$

Эта величина отличается от предыдущего выражения знаком. Различие обусловлено тем, что во втором случае гальваническая батарея поставляет конденсатору дополнительный заряд $\delta q = \varphi\,\delta C$, совершая при этом работу $\delta A_{\text{бат}} = \varphi\,\delta q = \varphi^2\,\delta C$. Теперь приращение свободной энергии определяется полной внешней работой, т. е. величиной

$$\delta A_{\text{бат}} - \delta A_{\text{эл}} = \varphi^2\,\delta C - \frac{1}{2}\varphi^2\,\delta C = \frac{1}{2}\varphi^2\,\delta C.$$

Таким образом,

$$(dW)_{q,T} = -(dW)_{\varphi,T}, \tag{32.4}$$

$$\delta A_{\text{эл}} - (dW)_{\varphi,T} = 0. \tag{32.5}$$

4. Применимость полученных формул не ограничивается случаем конденсатора. Формулы (32.4) и (32.5) носят общий характер. Чтобы убедиться в этом, рассмотрим произвольную систему проводников, пространство между которыми заполнено каким угодно диэлектриком. Свободная энергия такой системы определяется выражением (28.4), или

$$W = \frac{1}{2}\sum_{i,j} C_{ij}\varphi_i\varphi_j = \frac{1}{2}\sum_{i,j} V_{ij}q_iq_j.$$

При виртуальных смещениях проводников или диэлектрической среды между ними меняются емкостные и потенциальные коэффициенты. Если смещения происходят при постоянных потенциалах проводников, то

$$(dW)_{\varphi,T} = \frac{1}{2}\sum_{i,j} \varphi_i\varphi_j\,\delta C_{ij}.$$

Если же они происходят при постоянных зарядах, то

$$(dW)_{q,T} = \frac{1}{2}\sum_{i,j} q_iq_j\,\delta V_{ij}.$$

Докажем, что эти выражения удовлетворяют соотношению (32.4). Подставив в формулу (27.5) выражение для φ_j из формулы (27.4), получим

$$q_i = \sum_{j,k} C_{ij} V_{jk} q_k.$$

Отсюда

$$\sum_j C_{ij} V_{jk} = \delta_{ik}, \tag{32.6}$$

где δ_{ik} равно нулю при $i \neq k$ и единице при $i = k$. Варьируя полученное соотношение, найдем

$$\sum_j \delta C_{ij} V_{jk} + \sum_j C_{ij}\, \delta V_{jk} = 0, \tag{32.7}$$

причем здесь, ввиду симметрии потенциальных и емкостных коэффициентов, можно переставлять местами индексы i и j, а также j и k. С использованием соотношений (32.6) и (32.7) можем написать

$$\begin{aligned}(dW)_{q,T} = \frac{1}{2}\sum_{i,j} q_i q_j\, \delta V_{ij} = \frac{1}{2}\sum_{i,j,k,l} C_{ik} C_{jl} \varphi_k \varphi_l\, \delta V_{ij} = \\ = \frac{1}{2}\sum_{j,k,l} C_{jl}\varphi_k\varphi_l \sum_i C_{ki}\, \delta V_{ij} = -\frac{1}{2}\sum_{j,k,l} C_{jl}\varphi_k\varphi_l \sum_i \delta C_{ki} V_{ij} = \\ = -\frac{1}{2}\sum_{i,k,l} \varphi_k\varphi_l\, \delta C_{ki} \sum_j V_{ij} C_{jl} = -\frac{1}{2}\sum_{i,k,l} \varphi_k\varphi_l\, \delta C_{ki}\delta_{il} = \\ = -\frac{1}{2}\sum_{i,k} \varphi_k\varphi_i\, \delta C_{ki} = -(dW)_{\varphi,T}.\end{aligned}$$

Таким образом, соотношение (32.4) доказано. При доказательстве предполагалось, что свободные заряды находятся только на проводниках. Однако доказательство может быть распространено и на тот случай, когда свободные заряды имеются и в диэлектриках. Достаточно мысленно разделить диэлектрики на малые части таким образом, чтобы потенциал каждой из таких частей с достаточной точностью можно было считать одним и тем же во всех ее точках. Для таких малых частей диэлектриков можно ввести емкостные и потенциальные коэффициенты совершенно так же, как это было сделано для проводников. Поэтому доказательство, изложенное выше, может быть распространено без всяких изменений и на случай диэлектриков.

5. Рассмотрим теперь произвольную систему заряженных проводников в вакууме. Заполним все пространство между ними однородной несжимаемой жидкостью с диэлектрической проницаемостью ε, сохраняя заряды проводников неизменными. От этого электрическое смещение поля $\mathbf{D}$ не изменится, а его напряженность $\mathbf{E}$ уменьшится в ε раз (см. § 22, п. 3). Уменьшится в ε раз и электрическая часть свободной энергии W. Так как жидкость несжимаемая, то при виртуальных смещениях упругая часть свободной энергии Ψ изменяться не будет. Поэтому силы взаимодействия между проводниками будут определяться только изменениями величины W. А так как при заполнении пространства диэлектрической жидкостью эта величина уменьшается

в ε раз, то во столько же раз уменьшатся и силы взаимодействия между проводниками.

6. Вернемся в заключение к примеру, с которого мы начали изложение в настоящем параграфе. Доведем до конца вычисление силы притяжения между пластинами плоского конденсатора, предполагая, что все пространство заполнено несжимаемой однородной жидкостью с диэлектрической проницаемостью ε (см. рис. 84). Упругую часть свободной энергии можно не принимать во внимание, так как для несжимаемой жидкости $d\Psi = dW$. Направим ось X от отрицательной пластины к положительной и сместим положительную пластину в направлении оси X на величину δx, сохраняя заряды пластин неизменными. При неизменных зарядах останется постоянным электрическое смещение $\mathbf{D}$ в конденсаторе, а с ним и объемная плотность свободной энергии

$$w = -\frac{D^2}{8\pi\varepsilon}.$$

При смещении объем конденсатора увеличится на $\delta V = S\,\delta x$, где S — площадь пластины. Часть жидкости войдет в конденсатор, а его свободная энергия увеличится на $(d\Psi)_{q,T} = w\,\delta V = Sw\,\delta x$ (влиянием краевых эффектов мы пренебрегаем). Виртуальная работа, совершенная силой $\mathbf{F}$, будет $\delta A = F\,\delta x = Sf\,\delta x$. Подставляя эти значения в формулу (32.1), получим

$$f = -w = -\frac{D^2}{8\pi\varepsilon} = -\frac{\varepsilon E^2}{8\pi}. \tag{32.8}$$

Таким образом, сила f, отнесенная к единице площади, численно равна плотности свободной энергии w. Сила получилась отрицательной: она направлена в отрицательную сторону оси X, т. е. является *силой притяжения.*

7. Выражение (32.8) дает полную силу притяжения между пластинами конденсатора (отнесенную к единице площади). Она слагается из упругой силы $f_{\text{упр}}$ и электрической силы $f_{\text{эл}}$. Чтобы выделить эти составные части, надо несжимаемую жидкость рассматривать как предельный случай сжимаемой. Существенно заметить, что виртуальное смещение δx можно выбрать *каким угодно.* От этого величина полной силы f не изменится. Возьмем такое виртуальное смещение, положительно заряженной пластины δx, чтобы жидкость в конденсатор не поступала. Для этого, конечно, надо ввести добавочные уравновешивающие силы. Однако эти силы не будут совершать работы, поскольку они приложены к неподвижным частям жидкости. При неизменных зарядах электрическое смещение $\mathbf{D}$ по-прежнему меняться не будет. Но плотность жидкости в конденсаторе τ, а с ней и диэлектрическая проницаемость ε изменятся. В случае сжимаемой жидкости надо учитывать изменения и упругой части свободной энергии. Полное изменение свободной энергии будет $d\Psi = d\Psi_{\text{упр}} + dW$. Для изменения упругой свободной энергии имеем

$$d\Psi_{\text{упр}} = -(\mathscr{P} - \mathscr{P}_0)\,\delta V,$$

где $\mathscr{P}$ — гидростатическое давление жидкости в конденсаторе, а $\mathscr{P}_0$ — вне его. Изменение электрической части свободной энергии определяется выражением

$$(dW)_{q,T} = \delta\left(\frac{D^2}{8\pi\varepsilon}\,V\right) = \frac{D^2}{8\pi}\,\delta\left(\frac{V}{\varepsilon}\right) = \frac{D^2}{8\pi\varepsilon}\,\delta V - \frac{D^2}{8\pi\varepsilon^2}\,V\,\delta\varepsilon.$$

Диэлектрическая проницаемость ε является функцией только плотности жидкости τ и температуры T, так что

$$\delta\varepsilon = \left(\frac{\partial\varepsilon}{\partial\tau}\right)_T \delta\tau.$$

Изменение плотности можно найти из условия сохранения массы жидкости в конденсаторе: $\tau V = \text{const}$. Оно дает $V\,\delta\tau + \tau\,\delta V = 0$. Определив отсюда $\delta\tau$ и учтя соотношение $\delta V = S\delta x$, получим

$$(dW)_{q,T} = \left(\varepsilon + \tau\frac{\partial\varepsilon}{\partial\tau}\right)\frac{E^2}{8\pi}\,S\,\delta x.$$

Подстановка полученных выражений в (32.1) приводит к результату

$$f = (\mathscr{P} - \mathscr{P}_0) - \left(\varepsilon + \tau\frac{\partial\varepsilon}{\partial\tau}\right)\frac{E^2}{8\pi}. \tag{32.9}$$

Тем самым выделена упругая часть силы $f_{\text{упр}} = \mathscr{P} - \mathscr{P}_0$. Для несжимаемых (точнее, слабо сжимаемых) жидкостей в состоянии равновесия формула (32.9) должна переходить в (32.8), т. е. должно быть

$$\mathscr{P} - \mathscr{P}_0 = \tau\left(\frac{\partial\varepsilon}{\partial\tau}\right)_T \frac{E^2}{8\pi}. \tag{32.10}$$

Таким образом, дополнительная электрическая сила

$$-\tau\left(\frac{\partial\varepsilon}{\partial\tau}\right)_T \frac{E^2}{8\pi},$$

возникающая в результате зависимости диэлектрической проницаемости от плотности диэлектрика, компенсируется гидростатическим давлением. Более строгое обоснование этого утверждения будет дано в § 33 и 34. Однако такая компенсация имеет место *только в статике.* Если электрическое поле меняется во времени, то компенсации, вообще говоря, нет.

ЗАДАЧИ

1. Получить выражение для силы притяжения между пластинами плоского воздушного конденсатора, рассматривая эту силу как результат взаимодействия электрических зарядов.

Р е ш е н и е. Одна из пластин конденсатора создает электрическое поле $E_1 = 2\pi\sigma$. Это поле действует на вторую пластину с силой $F = S\sigma E_1 = 2\pi\sigma^2 S$. Поверхностная плотность электричества σ связана с напряженностью электрического поля в конденсаторе E соотношением $E = 4\pi\sigma$.

Исключая σ, получим

$$F = \frac{E^2}{8\pi} S. \tag{32.11}$$

2. Между пластинами плоского воздушного конденсатора введена плоскопараллельная пластинка из твердого диэлектрика, так что между ней и пластинами конденсатора остались воздушные зазоры. Как изменится при этом сила притяжения между пластинами конденсатора?

Ответ. Сила притяжения не изменится.

3. Конденсатор переменной емкости состоит из двух неподвижных металлических пластин, расположенных на расстоянии d друг от друга, и подвижной диэлектрической пластины, которая может поворачиваться и входить в зазор между металлическими пластинами (рис. 85). Все пластины имеют форму полукруга радиуса R, причем зазоры между диэлектрической пластиной и пластинами конденсатора пренебрежимо малы по сравнению с d. Пренебрегая краевыми эффектами, найти момент сил M, действующих на диэлектрическую пластину, когда она выведена из положения равновесия. Конденсатор заряжен до разности потенциалов φ, диэлектрическая проницаемость подвижной пластины равна ε.

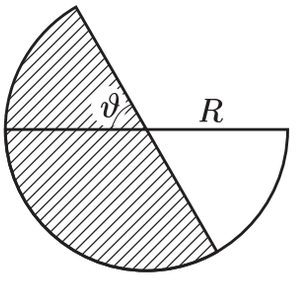


Рис. 85

Ответ. $M = (\varepsilon - 1)R^2\varphi^2/(16\pi d)$. Момент сил M стремится втянуть диэлектрическую пластину внутрь конденсатора.

4. В предыдущей задаче модуль момента M не зависит от угла поворота ϑ диэлектрической пластины. Но в положении равновесия, когда $\vartheta = 0$, момент M должен обращаться в нуль. Объяснить это расхождение.

§ 33. Максвелловские натяжения и давления

1. Найдем пондеромоторные силы, действующие на границе раздела двух диэлектриков с диэлектрическими проницаемостями ε_1 и ε_2. Поверхностная плотность свободных зарядов на границе раздела может быть равна нулю, но может быть и отличной от нуля. Для последующих вычислений это не имеет значения. Рассмотрим сначала случай, когда электрическое поле *однородно и перпендикулярно к границе раздела.* Примером может служить плоский конденсатор, пространство между обкладками которого заполнено двумя однородными диэлектрическими жидкостями, граничащими вдоль плоскости, параллельной пластинам конденсатора (рис. 86). Предположим, что обе жидкости несжимаемы. Тогда при виртуальных смещениях упругая часть свободной энергии изменяться не будет. Сместим изотермически границу раздела вверх на расстояние δx, сохраняя заряды пластин постоянными. При постоянных зарядах останется постоянным и электрическое смещение

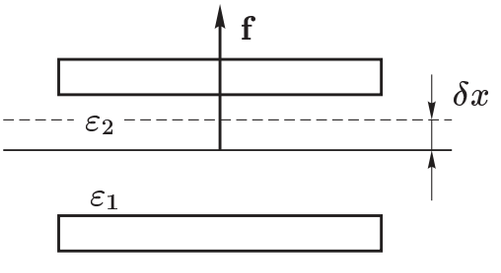


Рис. 86

в диэлектриках $\mathbf{D}_1$ и $\mathbf{D}_2$, а с ним и плотности свободной энергии

$$w_1 = \frac{D_1^2}{8\pi\varepsilon_1} = \frac{\varepsilon_1 E_1^2}{8\pi} \quad \text{и} \quad w_2 = \frac{D_2^2}{8\pi\varepsilon_2} = \frac{\varepsilon_2 E_2^2}{8\pi}. \tag{33.1}$$

При смещении жидкость с диэлектрической проницаемостью ε_1 будет входить в конденсатор, а с диэлектрической проницаемостью ε_2 — выходить. Приращение свободной энергии будет

$$(d\Psi)_{q,T} = (dW)_{q,T} = (w_1 - w_2) S\,\delta x,$$

а работа пондеромоторных сил $\delta A = fS\,\delta x$. Подставляя эти выражения в формулу (32.1), находим силу f, действующую на единицу площади границы раздела:

$$f = w_2 - w_1, \tag{33.2}$$

причем за положительное мы приняли направление вверх.

2. Совершенно аналогично решается вопрос о пондеромоторных силах, когда *электрическое поле параллельно границе раздела диэлектриков.* Рассмотрим снова плоский конденсатор с двумя диэлектрическими жидкостями, заполняющими его, как показано на рис. 87. В рассматриваемом случае виртуальное смещение δx границы раздела удобнее произвести при постоянной разности потенциалов между пластинами конденсатора. Плотности свободной энергии w_1 и w_2 в обоих диэлектриках при этом изменяться не будут. Свободная энергия получит приращение

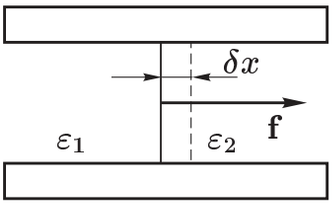


Рис. 87

$$(d\Psi)_{\varphi,T} = (dW)_{\varphi,T} = (w_1 - w_2) S_1\,\delta x,$$

где S_1 — площадь границы раздела между диэлектриками. Виртуальная работа $\delta A = S_1 f\,\delta x$. Подставив эти значения в формулу (32.5), найдем силу f, действующую на единицу площади границы раздела:

$$f = w_1 - w_2, \tag{33.3}$$

причем за положительное мы приняли направление от первого диэлектрика ко второму.

3. Рассматривая формулы (33.2) и (33.3), мы видим, что все происходит так, как если бы вдоль электрических силовых линий существовало *натяжение*, а перпендикулярно к ним — *давление.* Величины натяжения $\mathcal{T}$ и давления $\Pi = -\mathcal{T}$ численно равны плотности электрической свободной энергии w. Формула (33.2) показывает, что действующая сила определяется *разностью натяжений*, а формула (33.3) — *разностью давлений* по разные стороны границы раздела. Такая интерпретация согласуется также с формулой (11.2), определяющей силу, действующую на границе заряженного проводника. Величины $\mathcal{T}$ и Π получили название *максвелловских натяжений* и *давлений.* Фарадей и Максвелл считали эти величины вполне аналогичными упругим натяжениям и давлениям, существующим в диэлектриках и в чистом эфире.

От такой механической интерпретации наука давно отказалась. Но при вычислении пондеромоторных сил можно пользоваться наглядной картиной максвелловских натяжений и давлений, поскольку она приводит к верным результатам.

4. Если электрическое поле $\mathbf{E}$ параллельно границе раздела, то $E_1 = E_2$, так что индексы 1 и 2 можно опустить и написать

$$f = \frac{1}{8\pi}(\varepsilon_1 - \varepsilon_2)E^2. \tag{33.4}$$

Предположим, что $\varepsilon_1 > \varepsilon_2$. Тогда $f > 0$, т. е. сила f направлена вправо (рис. 87) — от диэлектрика с большей к диэлектрику с меньшей диэлектрической проницаемостью. То же будет и в том случае, когда поле $\mathbf{E}$ нормально к границе раздела, если только на этой границе нет свободных электрических зарядов. Действительно, при выполнении этого условия $D_1 = D_2 = D$, и формула (33.2) принимает вид

$$f = \frac{D^2}{8\pi}\left(\frac{1}{\varepsilon_2} - \frac{1}{\varepsilon_1}\right). \tag{33.5}$$

Если $\varepsilon_1 > \varepsilon_2$, то $f > 0$, т. е. сила f направлена вверх — от первого диэлектрика ко второму (см. рис. 86). Вообще, независимо от направления электрического поля, *пондеромоторные силы, действующие на незаряженной границе двух диэлектриков, направлены всегда в сторону диэлектрика с меньшей диэлектрической проницаемостью.* Существованием таких сил объясняется притяжение легких кусочков бумаги наэлектризованной палочкой.

Если две параллельные металлические пластинки частично погрузить в диэлектрическую жидкость, то она немного поднимается под действием капиллярных сил. Если между пластинками создать разность потенциалов в несколько тысяч вольт, то поднятие становится еще больше — жидкость втягивается в конденсатор.

Если диэлектрическая проницаемость тела меньше диэлектрической проницаемости окружающей среды, то тело выталкивается в область более слабого электрического поля. Демонстрацией может служить следующий опыт. Берется стеклянный сосуд, наполненный керосином или дистиллированной водой (рис. 88). Через отверстие A на загнутом конце стеклянной трубочки выдувается воздух под небольшим давлением. Если нет электрического поля, то пузырьки воздуха поднимаются вертикально вверх. Но если в жидкость опустить электрически заряженный шарик, то пузырьки воздуха отталкиваются от него и отклоняются в сторону.

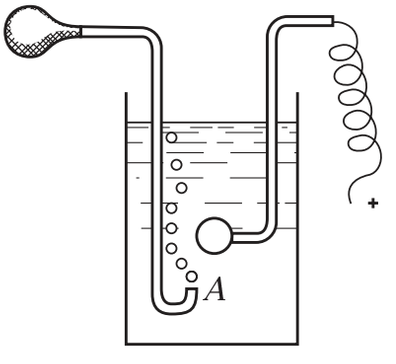


Рис. 88

5. При выводе формул (33.2) и (33.3) мы рассматривали диэлектрики как *абсолютно несжимаемые* жидкости. Таким путем нельзя

отделить упругие силы от электрических. Чтобы сделать это, надо учесть *сжимаемость жидкостей* точно так же, как мы поступали в предыдущем параграфе. Надо рассмотреть такие виртуальные смещения границы раздела диэлектриков, при которых изменялись бы плотности жидкостей. Сделаем это для случая, когда электрическое поле параллельно границе раздела диэлектриков (см. рис. 87). Виртуальное смещение границы раздела произведем так, чтобы количества обоих диэлектриков в конденсаторе оставались без изменения. Приращение упругой части свободной энергии можно не вычислять. Вместо этого, как было показано в предыдущем параграфе, можно к электрическому давлению прибавить, а из электрического натяжения вычесть гидростатическое давление $\mathscr{P}$. Поэтому мы ограничимся вычислением приращений только электрической части свободной энергии. Если V_1 — объем первого, а V_2 — второго диэлектриков, то $W = V_1 w_1 + V_2 w_2$. Виртуальное смещение границы произведем при постоянной разности потенциалов между пластинами конденсатора. Тогда напряженность электрического поля в конденсаторе $\mathbf{E}$ при смещении изменяться не будет, а потому

$$dW_{\varphi,T} = \frac{E^2}{8\pi}\left[\delta(V_1\varepsilon_1) + \delta(V_2\varepsilon_2)\right],$$

или

$$dW_{\varphi,T} = \frac{E^2}{8\pi}\left[(V_1\,\delta\varepsilon_1 + V_2\,\delta\varepsilon_2) + (\varepsilon_1\,\delta V_1 + \varepsilon_2\,\delta V_2)\right].$$

Диэлектрическая проницаемость жидкости зависит только от плотности и температуры последней. При изотермических изменениях

$$\delta\varepsilon = \left(\frac{\partial\varepsilon}{\partial\tau}\right)_T \delta\tau.$$

Так как масса первого диэлектрика остается неизменной, то $V_1\tau_1 = \text{const}$. Отсюда $V_1\,\delta\tau_1 = -\tau_1\,\delta V_1$, и, следовательно,

$$V_1\,\delta\varepsilon_1 = -\tau_1\frac{\partial\varepsilon_1}{\partial\tau_1}\,\delta V_1.$$

Аналогично для второго диэлектрика:

$$V_2\,\delta\varepsilon_2 = -\tau_2\frac{\partial\varepsilon_2}{\partial\tau_2}\,\delta V_2.$$

Далее, очевидно, что $\delta V_1 = -\delta V_2 = S_1\,\delta x$. Учтя все это, найдем $dW_{\varphi,T}$ и полученное выражение подставим в формулу (32.5). Учтем, кроме того, гидростатическое давление. Тогда вместо (33.3) получим формулу

$$f = \mathscr{P}_1 + \frac{E^2}{8\pi}\left(\varepsilon_1 - \tau_1\frac{\partial\varepsilon_1}{\partial\tau_1}\right) - \mathscr{P}_2 - \frac{E^2}{8\pi}\left(\varepsilon_2 - \tau_2\frac{\partial\varepsilon_2}{\partial\tau_2}\right). \qquad (33.6)$$

Если электрическое поле перпендикулярно к границе раздела диэлектриков, то рассуждения совершенно аналогичны. В этом случае

формула (33.6) заменяется на

$$f = -\mathscr{P}_2 + \frac{E^2}{8\pi}\left(\varepsilon_2 + \tau_2\frac{\partial\varepsilon_2}{\partial\tau_2}\right) + \mathscr{P}_1 - \frac{E^2}{8\pi}\left(\varepsilon_1 + \tau_1\frac{\partial\varepsilon_1}{\partial\tau_1}\right). \qquad (33.7)$$

Таким образом, и с учетом зависимости диэлектрической проницаемости от плотности жидкости пондеромоторные силы в диэлектрике сводятся к *натяжению вдоль линий поля* $\mathscr{T}$ и к *давлению* Π *в перпендикулярном направлении.* Меняются только выражения для этих величин. Именно:

$$\mathscr{T} = -\mathscr{P} + \frac{E^2}{8\pi}\left(\varepsilon + \tau\frac{\partial\varepsilon}{\partial\tau}\right), \qquad (33.8)$$

$$\Pi = \mathscr{P} + \frac{E^2}{8\pi}\left(\varepsilon - \tau\frac{\partial\varepsilon}{\partial\tau}\right). \qquad (33.9)$$

Различием этих величин по разные стороны границы раздела диэлектриков и определяются силы, действующие на единицу поверхности указанной границы. Дополнительные силы, возникающие в результате зависимости диэлектрической проницаемости от плотности диэлектрика, называются *электрострикционными силами*, а вызываемое ими изменение гидростатического давления и плотности диэлектрика — *электрострикцией.* Электрострикционные силы впервые были учтены Гельмгольцем (1821–1894). Максвелл их не учитывал, хотя величины $\tau\partial\varepsilon/\partial\tau$, как правило, того же порядка, что и ε. Тем не менее в электростатике для несжимаемых (точнее, слабосжимаемых) жидкостей результаты Максвелла правильны, как это следует из их вывода, приведенного нами выше. Они должны согласовываться с формулами (33.8) и (33.9). Поэтому, *если несжимаемая жидкость находится в равновесии* (*гидростатика*), *то электрострикционные силы должны уравновешиваться силами гидростатического давления.* Иными словами, во всем объеме жидкости должно выполняться соотношение

$$\mathscr{P} - \tau\left(\frac{\partial\varepsilon}{\partial\tau}\right)_T\frac{E^2}{8\pi} = \text{const}. \qquad (33.10)$$

Этот результат согласуется с ранее полученной формулой (32.10). Однако *если электрическое поле быстро меняется во времени*, *то соотношение* (33.10), *вообще говоря*, *несправедливо.*

ЗАДАЧИ

1. Между пластинами плоского воздушного конденсатора введена диэлектрическая пластинка толщины l_2 с диэлектрической проницаемостью ε_2 (рис. 89). Конденсатор частично погружен в жидкость с диэлектрической проницаемостью ε_1 и плотностью τ. Найти высоту поднятия h жидкости в конденсаторе, пренебрегая капиллярными явлениями, если между его обкладками поддерживается разность потенциалов φ. Суммарная толщина столбов жидкости в конденсаторе равна l_1.

Ответ. $h = \dfrac{\varepsilon_1 - 1}{8\pi\tau g}\left(\dfrac{\varphi\varepsilon_2}{\varepsilon_1 l_2 + \varepsilon_2 l_1}\right)^2$.

2. Диэлектрическая пластинка толщины l_2 с диэлектрической проницаемостью ε введена между обкладками плоского воздушного конденсатора (рис. 90). Между поверхностью пластинки и обкладками конденсатора остались воздушные зазоры, суммарная толщина которых равна l_1. Определить силу притяжения F между обкладками, если разность потенциалов между ними равна φ, а площадь пластинки S. Во что переходит выражение для F в предельном случае $l_1 \to 0$?

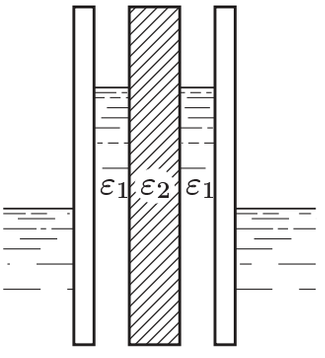


Рис. 89

Ответ. $F = \dfrac{S}{8\pi}\left(\dfrac{\varphi\varepsilon}{l_1\varepsilon + l_2}\right)^2$.

3. Капиллярный вольтметр состоит из капиллярной стеклянной трубочки с металлизированной полупрозрачной внутренней поверхностью, служащей одной из обкладок цилиндрического конденсатора. Второй обкладкой является тонкая металлическая проволока, коаксиальная с внутренней цилиндрической поверхностью трубочки. Определить поднятие мениска воды h в вольтметре при наложении на обкладки напряжения $V = 100$ В, если внутренний диаметр капилляра $D = 0{,}5$ мм, а диаметр проволоки $d = 0{,}05$ мм.

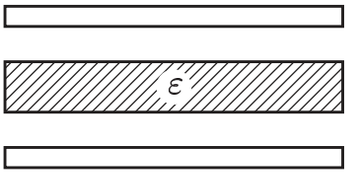


Рис. 90

Ответ. $h = \dfrac{\varphi^2(\varepsilon - 1)}{\pi\tau g(D^2 - d^2)\ln(D/d)} \approx 5$ мм.

4. Капля жидкости заряжена электричеством. Найти зависимость упругости насыщенного пара над поверхностью капли от ее заряда q. Используя полученный результат, объяснить принцип действия камеры Вильсона (р. 1906).

Решение. Искомая зависимость находится в точности так же, как и зависимость упругости насыщенного пара от кривизны поверхности жидкости. Можно дословно повторить все рассуждения, приведенные в § 118 тома II. Дополнительно надо только учесть влияние электрического поля на высоту поднятия жидкости в капилляре. Электрическое поле должно быть перпендикулярно к поверхности мениска жидкости в капилляре. Влияние такого поля эквивалентно уменьшению поверхностного натяжения жидкости σ. Из капиллярного давления $2\sigma/r$ надо вычесть максвелловское натяжение:

$$\frac{D^2}{8\pi}\left(1 - \frac{1}{\varepsilon}\right) = \frac{q^2}{8\pi r^4}\left(1 - \frac{1}{\varepsilon}\right),$$

где ε — диэлектрическая проницаемость капли, r — ее радиус. Для проводящей капли в этой формуле следует положить $\varepsilon = \infty$. То же можно делать для воды ввиду большого значения диэлектрической проницаемости последней ($\varepsilon = 81$). В результате вместо формулы (118.5) тома II получится

$$\ln\frac{P_0}{P} = \frac{\mu v_{\text{ж}}}{RT}\left(P_0 - P - \frac{2\sigma}{r} + \frac{q^2}{8\pi r^4}\right). \qquad (33.11)$$

При $r = 0$ и $r = \infty$ эта формула дает соответственно $P = 0$ и $P = P_0$. В промежутке между этими значениями упругость насыщенного пара P

достигает максимума. Дифференцируя (33.11) по r и полагая $dP/dr = 0$, находим, что это происходит при

$$r = r_0 \equiv \sqrt[3]{q^2/(4\pi\sigma)}. \tag{33.12}$$

Применим полученные результаты к капле воды, полагая q равным элементарному заряду e, находящемуся в центре капли. При 20 °C для воды $\sigma = 73$ дин/см. По формуле (33.12) находим $r_0 = 6{,}3\cdot 10^{-8}$ см. При таких малых размерах капель макроскопические формулы как точные количественные соотношения становятся сомнительными. Тем не менее мы воспользуемся ими, рассчитывая, что грубо качественно результаты получатся правильными. Мы не будем также смущаться тем обстоятельством, что в реальных условиях осаждающиеся ионы не попадают в центр капли, а могут находиться в ней в любом месте. Зависимость упругости насыщенного пара над заряженной каплей от ее радиуса представлена на рис. 91. Та же зависимость для незаряженной капли представляется пунктирной кривой. Заряд капли уменьшает упругость насыщенного пара, причем при $r < r_0$ упругость пара растет с увеличением радиуса капли. Этим и объясняется конденсация пара на ионах (см. т. II, § 119).

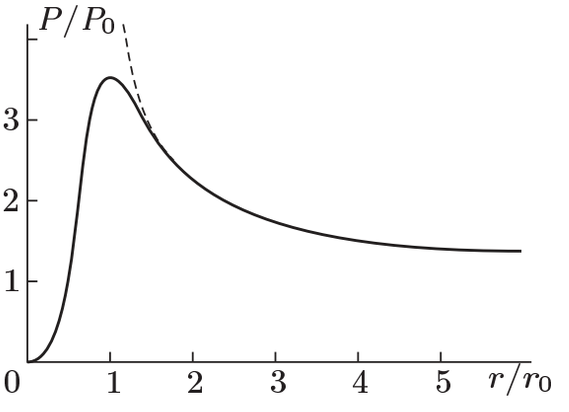


Рис. 91

§ 34. Вычисление пондеромоторных сил в общем виде

1. Без ущерба для общности будем предполагать, что вещество и электричество распределены в пространстве *непрерывно.* Электрическую часть свободной энергии представим в виде

$$W = \int \frac{D^2}{8\pi\varepsilon}\, dV. \tag{34.1}$$

Пусть каждая частица диэлектрика вместе со своим электрическим зарядом претерпела бесконечно малое виртуальное смещение $\mathbf{q} = \mathbf{q}(\mathbf{r})$. Приращение величины W при таком виртуальном смещении будет

$$dW = \int \delta\left(\frac{D^2}{8\pi\varepsilon}\right) dV = \frac{1}{4\pi}\int \frac{\mathbf{D}\,\delta\mathbf{D}}{\varepsilon}\, dV + \frac{1}{8\pi}\int D^2\,\delta\left(\frac{1}{\varepsilon}\right) dV,$$

или

$$dW = \frac{1}{4\pi}\int \mathbf{E}\,\delta\mathbf{D}\, dV - \frac{1}{8\pi}\int E^2\,\delta\varepsilon\, dV.$$

Здесь символ δ применяется для обозначения *локальных* изменений соответствующих величин, т. е. изменений в одном и том же месте пространства, обусловленных виртуальными смещениями вещества и электричества. Интегрирование производится по всему бесконечному пространству. Преобразуем первый интеграл с помощью интегральной теоремы Гаусса–Остроградского.

Интегрируя сначала по конечному объему, ограниченному замкнутой поверхностью S, имеем

$$\int \mathbf{E}\,\delta\mathbf{D}\,dV = -\int \operatorname{grad}\varphi\,\delta\mathbf{D}\,dV = -\int \operatorname{div}(\varphi\,\delta\mathbf{D})\,dV + \int \varphi \operatorname{div}\,\delta\mathbf{D}\,dV =$$

$$= -\oint \varphi\,\delta\mathbf{D}\,d\mathbf{S} + 4\pi\int \varphi\,\delta\rho\,dV.$$

Предполагая, что все заряды находятся в ограниченной области пространства, будем удалять в бесконечность окружающую их замкнутую поверхность S. Тогда в пределе первый интеграл обратится в нуль, и мы получим

$$dW = \int \varphi\,\delta\rho\,dV - \frac{1}{8\pi}\int E^2\,\delta\varepsilon\,dV. \tag{34.2}$$

2. Для упрощения последующих рассуждений предположим, что виртуальное смещение претерпевают только частицы вещества, находящиеся внутри бесконечно тонкого цилиндра, параллельного оси X. Пусть это бесконечно малое смещение происходит параллельно той же оси и является произвольной функцией координаты x: $q = x - x' = q(x)$ (рис. 92). Найдем *локальные* изменения плотности электричества $\delta\rho$ и диэлектрической прони-

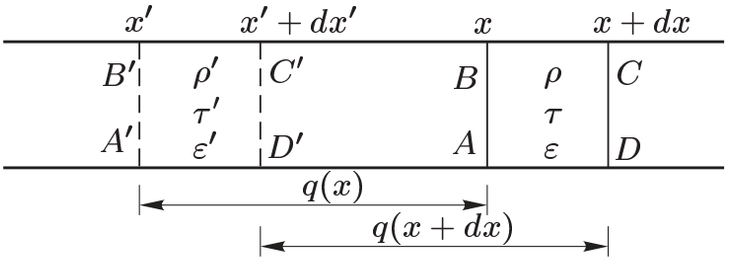


Рис. 92

цаемости $\delta\varepsilon$ в бесконечно малом элементе объема $ABCD$. Через сечение AB внутрь объема $ABCD$ входит заряд $q(x)\rho(x)\,dS$, через сечение CD выходит заряд $q(x+dx)\rho(x+dx)\,dS$, где dS — площадь поперечного сечения. Избыток входящего заряда над выходящим составляет

$$[q(x)\rho(x) - q(x+dx)\rho(x+dx)]\,dS = -\frac{\partial(\rho q)}{\partial x}\,dV,$$

где $dV = dS\,dx$ — величина рассматриваемого объема. Тот же избыток можно представить выражением $\delta\rho\,dV$, а потому

$$\delta\rho = -\frac{\partial(\rho q)}{\partial x}. \tag{34.3}$$

Найдем теперь *локальное* приращение диэлектрической проницаемости $\delta\varepsilon$. Оно обусловлено, во-первых, тем, что в результате смещения в объем $ABCD$ поступает вещество из других областей пространства, где ε имеет другие значения. Во-вторых, тем, что при смещении каждого элемента среды меняется плотность вещества τ, а с ней и диэлектрическая проницаемость. После смещения из элемента объема $A'B'C'D'$ в элемент $ABCD$ плотность вещества $\tau' \equiv \tau(x')$ получает приращение $\Delta\tau$ и становится равной $\tau' + \Delta\tau$. Так как при этом масса смещенного вещества не меняется, то

$$dx'\tau' = dx(\tau' + \Delta\tau),$$

откуда

$$\Delta\tau = \frac{dx' - dx}{dx}\,\tau' = \frac{d(x' - x)}{dx}\,\tau',$$

или с точностью до бесконечно малых высшего порядка

$$\Delta\tau = -\frac{\partial q}{\partial x}\,\tau.$$

Поэтому в объем $ABCD$ вещество вступает, имея диэлектрическую проницаемость

$$\varepsilon' + \frac{\partial\varepsilon}{\partial\tau}\,\Delta\tau = \varepsilon' - \tau\frac{\partial\varepsilon}{\partial\tau}\,\frac{\partial q}{\partial x}.$$

Вычитая отсюда значение ε в объеме $ABCD$ до смещения, получим

$$\delta\varepsilon = \varepsilon' - \varepsilon - \tau\frac{\partial\varepsilon}{\partial\tau}\,\frac{\partial q}{\partial x}.$$

А так как $\varepsilon' - \varepsilon = \varepsilon(x') - \varepsilon(x) = (x' - x)\dfrac{\partial\varepsilon}{\partial x} = -q\dfrac{\partial\varepsilon}{\partial x}$, то

$$\delta\varepsilon = -q\frac{\partial\varepsilon}{\partial x} - \tau\frac{\partial\varepsilon}{\partial\tau}\,\frac{\partial q}{\partial x}. \tag{34.4}$$

3. Поскольку электричество смещается вместе с веществом, электрический заряд каждого смещающегося элемента среды при смещении не изменяется. Поэтому, подставив выражения (34.3) и (34.4) в формулу (34.2), получим приращение величины W при постоянном заряде:

$$(dW)_{q,T} = -dS\int\limits_{-\infty}^{\infty}\varphi\frac{\partial(\rho q)}{\partial x}\,dx + \frac{dS}{8\pi}\int\limits_{-\infty}^{\infty}E^2\left(q\frac{\partial\varepsilon}{\partial x} + \tau\frac{\partial\varepsilon}{\partial\tau}\,\frac{\partial q}{\partial x}\right)dx.$$

Интегрируя по частям и имея в виду, что на пределах интеграла все величины обращаются в нуль, найдем

$$\int\limits_{-\infty}^{\infty}\varphi\frac{\partial(\rho q)}{\partial x}\,dx = -\int\limits_{-\infty}^{\infty}\rho q\frac{\partial\varphi}{\partial x} = \int\limits_{-\infty}^{\infty}q\rho E_x\,dx.$$

Таким же путем преобразуется в (34.2) и второй интеграл. В результате получится

$$(dW)_{q,T} = -dS\int\limits_{-\infty}^{\infty}\left[\rho E_x - \frac{E^2}{8\pi}\,\frac{\partial\varepsilon}{\partial x} + \frac{1}{8\pi}\,\frac{\partial}{\partial x}\left(\tau\frac{\partial\varepsilon}{\partial\tau}\,E^2\right)\right]q\,dx.$$

Если $f^{\text{эл}}$ — объемная плотность электрических пондеромоторных сил, действующих в диэлектрике, то работа их при рассматриваемом виртуальном смещении будет

$$\delta A_{\text{эл}} = \int f_x^{\text{эл}}q\,dV = dS\int\limits_{-\infty}^{\infty}f_x^{\text{эл}}q\,dx.$$

Подставив эти выражения в формулу (32.2) и имея в виду, что функция $q = q(x)$ может быть выбрана какой угодно, найдем $f_x^{\text{эл}}$. К этой силе надо еще

добавить силу гидростатического давления — $\partial\mathscr{P}/dx$. В результате для x-составляющей объемной плотности полной пондеромоторной силы получится

$$f_x = \rho E_x - \frac{E^2}{8\pi}\frac{\partial\varepsilon}{\partial x} + \frac{1}{8\pi}\frac{\partial}{\partial x}\left(\tau\frac{\partial\varepsilon}{\partial\tau}E^2\right) - \frac{\partial\mathscr{P}}{\partial x}.$$

Такие же выражения можно написать для y- и z-составляющих силы $\mathbf{f}$. Объединяя их в одну векторную формулу, получим

$$\mathbf{f} = -\operatorname{grad}\mathscr{P} + \rho E - \frac{E^2}{8\pi}\operatorname{grad}\varepsilon + \frac{1}{8\pi}\operatorname{grad}\left(\tau\frac{\partial\varepsilon}{\partial\tau}E^2\right). \tag{34.5}$$

4. Повторим теперь приведенный вывод, предполагая, что жидкость является «несжимаемой». Тогда при перемещении любого элемента жидкости из одного места пространства в другое величина τ, а с ней ε меняться не будут. Не будет изменяться и упругая часть свободной энергии. В результате в выражении (34.5) выпадут первый и последний члены, и мы получим

$$\mathbf{f} = \rho\mathbf{E} - \frac{E^2}{8\pi}\operatorname{grad}\varepsilon. \tag{34.6}$$

В таком виде выражение для пондеромоторной силы было получено Максвеллом. Добавочный член $\dfrac{1}{8\pi}\operatorname{grad}\left(\tau\dfrac{\partial\varepsilon}{\partial\tau}E^2\right)$ был введен Гельмгольцем. По величине он того же порядка, что и член $\dfrac{E^2}{8\pi}\operatorname{grad}\varepsilon$, учтенный Максвеллом. Однако в случае «несжимаемых» жидкостей применимы оба способа рассуждения, приведенные выше. Поэтому в статических электрических полях должно соблюдаться соотношение

$$\operatorname{grad}\mathscr{P} = \frac{1}{8\pi}\operatorname{grad}\left(\tau\frac{\partial\varepsilon}{\partial\tau}E^2\right), \tag{34.7}$$

или после интегрирования

$$\mathscr{P} = \frac{1}{8\pi}\tau\frac{\partial\varepsilon}{\partial\tau}E^2 + \text{const}. \tag{34.8}$$

Это означает, что *добавочная часть пондеромоторной силы, возникающая в результате зависимости диэлектрической проницаемости от плотности жидкости, компенсируется гидростатическим давлением, появляющимся в жидкости при наложении электрического поля. Но в быстропеременных полях такой компенсации нет. Там соотношение* (34.8) *не соблюдается.* Действительно, допустим, что электрическое поле было включено мгновенно. Электрические возмущения распространяются в диэлектриках со скоростями порядка скорости света, упругие — со скоростью звука. Поэтому в быстропеременных полях гидростатическое давление $\mathscr{P}$ не успеет прийти в равновесие с «электрическим давлением» $\dfrac{1}{8\pi}\tau\dfrac{\partial\varepsilon}{\partial\tau}E^2$. В этом случае формула (34.7) неприменима.

ЗАДАЧА

Показать, что пондеромоторная сила (34.5) сводится к натяжению вдоль силовых линий (33.8) и давлению (33.9) в перпендикулярном направлении.

Р е ш е н и е. Докажем обратное: считая натяжения и давления известными, найдем результирующую силу $\mathbf{f}$, действующую на единицу объема

диэлектрика Убедимся, что она совпадает с (34.5). Вырежем мысленно в диэлектрике бесконечно малый прямоугольный параллелепипед со сторонами dx, dy, dz, параллельными координатным осям (рис. 93). На основание *1* этого параллелепипеда действует сила гидростатического давления $\mathscr{P}(x)\,dy\,dz$, на основание *2* — противоположно направленная сила — $\mathscr{P}(x+dx)\,dy\,dz$. Их результирующая, действующая в положительном направлении оси X, равна

$$[\mathscr{P}(x)-\mathscr{P}(x+dx)]\,dy\,dz = -\frac{\partial\mathscr{P}}{\partial x}\,dx\,dy\,dz.$$

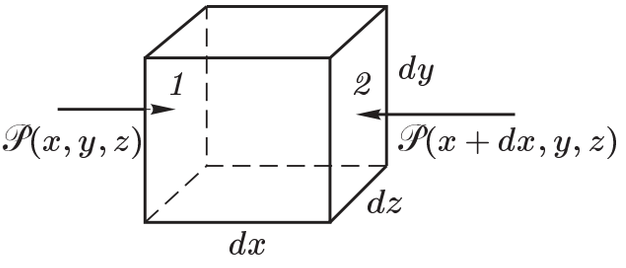


Рис. 93

Аналогично находятся выражения для сил, действующих вдоль двух других координатных осей. Объединяя их и разделив на объем параллелепипеда $dx\,dy\,dz$, получим полную силу гидростатического давления, отнесенную к единице объема диэлектрика:

$$\mathbf{f}_1 = -\left(\frac{\partial\mathscr{P}}{\partial x}\,\mathbf{i}+\frac{\partial\mathscr{P}}{\partial y}\,\mathbf{j}+\frac{\partial\mathscr{P}}{\partial z}\,\mathbf{k}\right) = -\operatorname{grad}\mathscr{P}.$$

К ней надо добавить силу всестороннего изотропного натяжения в электрическом поле:

$$\mathbf{f}_2 = \operatorname{grad}\left(\frac{\tau}{8\pi}\,\frac{d\varepsilon}{d\tau}\,E^2\right),$$

которая вычисляется так же. Остается найти силу $\mathbf{f}_3$, вызываемую натяжением $\varepsilon E^2/(8\pi)$ вдоль поля и равным ему давлением поперек поля. При вычислении этой силы координатную ось X удобно направить вдоль вектора $\mathbf{E}$. Тогда

$$\mathbf{f}_3 = \frac{1}{8\pi}\left[\frac{\partial}{\partial x}\,(\varepsilon E^2)\mathbf{i} - \frac{\partial}{\partial y}\,(\varepsilon E^2)\mathbf{j} - \frac{\partial}{\partial z}\,(\varepsilon E^2)\mathbf{k}\right],$$

или

$$\mathbf{f}_3 = -\frac{1}{8\pi}\,\operatorname{grad}(\varepsilon E^2) + \frac{1}{4\pi}\,\frac{\partial}{\partial x}\,(\varepsilon E^2)\mathbf{i}.$$

Преобразуем это выражение. Прежде всего пишем

$$\operatorname{grad}(\varepsilon E^2) = E^2\operatorname{grad}\varepsilon + 2\varepsilon E\operatorname{grad}E.$$

Так как вектор $\mathbf{E}$ содержит только x-составляющую, то

$$\varepsilon E\operatorname{grad}E = D\operatorname{grad}E_x = D\left(\frac{\partial E_x}{\partial x}\,\mathbf{i}+\frac{\partial E_x}{\partial y}\,\mathbf{j}+\frac{\partial E_x}{\partial z}\,\mathbf{k}\right).$$

Но

$$\frac{\partial E_x}{\partial y} = -\frac{\partial}{\partial y}\left(\frac{\partial\varphi}{\partial x}\right) = -\frac{\partial}{\partial x}\left(\frac{\partial\varphi}{\partial y}\right) = \frac{\partial E_y}{\partial x} = 0.$$

Аналогично, $\partial E_x/\partial z = 0$. Таким образом,

$$\varepsilon E\operatorname{grad}E = D\frac{\partial E}{\partial x}\,\mathbf{i}.$$

Далее,

$$-D\frac{\partial E}{\partial x}\,\mathbf{i}+\frac{\partial}{\partial x}\,(\varepsilon E^2)\mathbf{i}=\mathbf{E}\frac{\partial D}{\partial x}=\mathbf{E}\,\mathrm{div}\,\mathbf{D}=4\pi\rho\mathbf{E}.$$

В результате сложения $\mathbf{f}_1$, $\mathbf{f}_2$ и $\mathbf{f}_3$ получается выражение (34.5).

§ 35. Электронная теория поляризации неполярных диэлектриков

1. В § 12 был качественно рассмотрен электронный механизм возникновения поляризации в диэлектриках. Поляризация диэлектриков во всех случаях возникает в результате *смещений* электронов и атомных ядер, которые они испытывают при внесении диэлектрика в электрическое поле. Остановимся теперь на количественной стороне теории.

Начнем с *неполярных диэлектриков.* Так называются диэлектрики, молекулы которых в отсутствие внешнего электрического поля не имеют дипольных моментов. Неполярными являются все атомы. Согласно полуклассической теории Бора простейший из всех атомов — атом водорода — состоит из ядра и электрона, вращающегося вокруг него по круговой или эллиптической орбите. Электрон смещен относительно ядра, а потому в любой момент времени такая модель атома обладает *дипольным моментом.* Однако из-за быстрого вращения электрона ориентация этого дипольного момента очень быстро меняется, так что в среднем дипольный момент атома равен нулю. В квантовой механике такое усреднение по положениям электрона на боровской орбите становится ненужным. Принцип неопределенности Гейзенберга делает невозможным сохранение представления о движении электрона в атоме по классическим орбитам (см. т. I, § 5). Классическая картина движения заменяется вероятностным описанием его. Можно говорить лишь о *вероятности* нахождения электрона в том или ином месте пространства. Электрон в атоме ведет себя подобно электронному облаку с непрерывным распределением электричества в нем. В стационарных состояниях распределение заряда в таком облаке с течением времени не меняется. Если атом находится в *нормальном состоянии*, т. е. не возбужден, то распределение заряда в облаке сферически симметрично относительно ядра. Поэтому дипольный момент атома точно равен нулю. Это справедливо не только для атома водорода, но и для любого атома или атомного иона в невозбужденном состоянии. Если же атом *возбужден*, то сферическая симметрия распределения заряда в облаке нарушается. Однако ядро по-прежнему остается центром симметрии атома, а в этом случае дипольный момент также должен равняться нулю. Неполярными будут и симметричные молекулы, построенные из нейтральных атомов. Такова, например, молекула водорода H_2.

2. Допустим теперь, что неполярная молекула внесена во внешнее однородное электрическое поле $\mathbf{E}$. Отрицательные заряды молекулы немного сместятся относительно положительных. В результате у молекулы появится *электрический дипольный момент* $\mathbf{p}$, величина которого зависит от напряженности внешнего поля $\mathbf{E}$. Для определения

вида этой зависимости надо принять во внимание, что при поляризации диэлектриков мы имеем дело с внешними электрическими полями, напряженность которых *очень мала* по сравнению с напряженностью внутренних полей, действующих в атомах и молекулах. Например, в нормальном состоянии среднее расстояние электрона от ядра в атоме водорода составляет $r = r_{\text{В}} = 0{,}53 \cdot 10^{-8}$ см (боровский радиус). Электрическое поле, создаваемое ядром на таком расстоянии, будет

$$\frac{e}{r^2} \approx 1{,}7 \cdot 10^7 \text{ СГСЭ-ед.} \approx 5 \cdot 10^9 \text{ В/см}.$$

Чтобы составить представление о таком поле, заметим, что для получения электрической искры между двумя металлическими шариками с радиусами 2,5 см каждый, находящимися на расстоянии 1 см друг от друга, в сухом воздухе требуется разность потенциалов порядка 30 000 В, т. е. напряженность поля $E \sim 3 \cdot 10^4$ В/см. Это поле в сто тысяч раз слабее внутреннего поля в атоме водорода. Понятно, что при таких условиях смещения зарядов внутри молекулы при поляризации ничтожны, и можно считать, что индуцированный дипольный момент $\mathbf{p}$ зависит от напряженности поля $\mathbf{E}$ *линейно.* Если молекула *сферически симметрична*, то можно написать

$$\mathbf{p} = \beta \mathbf{E}. \tag{35.1}$$

Величина β постоянна для данной молекулы и зависит от ее строения. Она называется *поляризуемостью молекулы.*

3. Величину β можно оценить, рассматривая молекулу как проводящий шарик радиуса a. В § 16, а также в § 23 было показано, что во внешнем электрическом поле $\mathbf{E}$ такой шарик приобретает дипольный момент $\mathbf{p} = a^3 \mathbf{E}$. Поэтому для рассматриваемой модели

$$\beta = a^3. \tag{35.2}$$

В квантовой механике классическим аналогом молекулы является *электронное облако*, в центре которого находится положительный точечный заряд. Поляризация такого облака во внешнем однородном поле аналогична поляризации проводящего шарика. Однако облако не имеет резкой границы, и надо решить, что следует понимать под a. Это можно сделать, пользуясь вычислительными методами квантовой механики. Для атома водорода в невозбужденном состоянии квантовая механика приводит к результату

$$\beta = \frac{9}{2} r_{\text{В}}^3, \tag{35.3}$$

где $r_{\text{В}}$ — боровский радиус, числовое значение которого было приведено выше. Таким образом, для атома водорода теория дает $\beta = 0{,}67 \cdot 10^{-24}$ см^3.

4. Если молекула не обладает сферической симметрией, то направления векторов $\mathbf{E}$ и $\mathbf{p}$, вообще говоря, не совпадают. Однако линейная связь между компонентами этих векторов сохраняется. Вместо соотношения (35.1) теперь

следует писать

$$p_i = \sum_j \beta_{ij} E_j \quad (i, j = x, y, z). \tag{35.4}$$

Девять величин β_{ij} образуют так называемый *тензор поляризуемости молекулы.* Он зависит от строения молекулы и от ее ориентации относительно координатных осей.

Тензор поляризуемости молекулы *симметричен*, т. е. $\beta_{ij} = \beta_{ji}$. Это утверждение получается из энергетических соображений. При изменении электрического поля на dE заряды внутри молекулы претерпевают смещения dr_a и над молекулой совершается работа

$$\delta A = \sum_a e_a \mathbf{E}\, d\mathbf{r}_a = \mathbf{E}\, d\mathbf{p}.$$

Эта работа идет на приращение потенциальной энергии деформации молекулы:

$$dU_{\text{д}} = \mathbf{E}\, d\mathbf{p}. \tag{35.5}$$

Интегрируя это выражение, заметим, что величина $U_{\text{д}}$ не может зависеть от способа приведения молекулы в конечное состояние. Будем наращивать поле $\mathbf{E}$ так, чтобы его направление сохранялось неизменным. Пусть $\mathbf{E}$ и $\mathbf{p}$ означают теперь напряженность электрического поля и дипольный момент молекулы в конечном состоянии. Если поле увеличить в λ раз, то, ввиду линейности соотношения между $\mathbf{E}$ и $\mathbf{p}$, во столько же раз увеличится вектор $\mathbf{p}$, причем направление его останется неизменным. Таким образом, промежуточное состояние молекулы может быть охарактеризовано векторами $\mathbf{E}' = \lambda\mathbf{E}$, $\mathbf{p}' = \lambda\mathbf{p}$, а величина $dU_{\text{д}}$ представлена выражением $dU_{\text{д}} = \mathbf{E}\mathbf{p}\lambda\, d\lambda$. Интегрируя это выражение в пределах от $\lambda = 0$ (начальное состояние) до $\lambda = 1$ (конечное состояние), получим

$$U_{\text{д}} = \frac{1}{2}\,\mathbf{p}\mathbf{E}. \tag{35.6}$$

Возьмем отсюда дифференциал $dU_{\text{д}}$ и приравняем его выражению (35.5). Тогда найдем

$$\mathbf{p}\, d\mathbf{E} = \mathbf{E}\, d\mathbf{p}. \tag{35.7}$$

Записанное в координатах, последнее соотношение имеет вид

$$\sum\sum \beta_{ij} E_j\, dE_i = \sum\sum E_i \beta_{ij}\, dE_j,$$

или после перестановки индексов суммирования

$$\sum\sum \beta_{ij} E_j\, dE_i = \sum \beta_{ij} E_j\, dE_i.$$

Ввиду произвольности dE_i отсюда следует $\beta_{ij} = \beta_{ji}$.

5. От рассмотрения отдельных молекул обратимся к исследованию *макроскопических тел*, построенных из множества молекул. Для простоты будем предполагать, что молекулы *сферически симметричны.* Пусть $\mathbf{E}$ — среднее макроскопическое электрическое поле, как оно было определено в § 10. Примем, что это поле совпадает с полем $\mathbf{E}'$, действующим на каждую молекулу диэлектрика. Равенство этих полей может быть только приближенным. Поле, действующее на молекулу, создается внешними зарядами и всеми молекулами, за исключением рассматриваемой. Если соседние молекулы расположены близко одна к другой, то действующее поле $\mathbf{E}'$ неоднородно, т. е. заметно меняется

на протяжении молекулы. В этом случае наше предположение $\mathbf{E}' = \mathbf{E}$ не оправдывается. С таким положением мы сталкиваемся в жидкостях и твердых телах. К этому случаю наши результаты не относятся. Равенство полей $\mathbf{E}'$ и $\mathbf{E}$ может выполняться *в газах*, где средние расстояния между соседними молекулами велики по сравнению с размерами последних. Этот случай мы и имеем в виду. При выполнении условия $\mathbf{E}' = \mathbf{E}$ индуцированный дипольный момент молекулы определяется выражением (35.1), а вектор поляризации среды $\mathbf{P} = n\mathbf{p}$ — выражением

$$\mathbf{P} = \alpha \mathbf{E}, \tag{35.8}$$

где α — поляризуемость единицы объема диэлектрика:

$$\alpha = n\beta \tag{35.9}$$

(n — число молекул в единице объема). Для диэлектрической проницаемости получаем по формуле (15.3)

$$\varepsilon = 1 + 4\pi n\beta. \tag{35.10}$$

Как указывалось выше, поляризуемость молекулы β является *молекулярной константой*, зависящей только от внутреннего строения молекулы. Поэтому величина $\varepsilon - 1$ должна быть пропорциональна плотности газа:

$$\frac{\varepsilon - 1}{\tau} = \text{const}. \tag{35.11}$$

По той же причине диэлектрическая проницаемость газов, состоящих из неполярных молекул, *не зависит от температуры.* Действительно, молекула является *квантовой системой*, внутренняя энергия которой может принимать лишь *дискретные значения.* При обычных температурах молекулы находятся в невозбужденных состояниях, т. е. на самых низших энергетических уровнях. Энергии теплового движения бывает недостаточно, чтобы перевести молекулы в возбужденные состояния. Поэтому с повышением температуры внутреннее строение молекулы, а с ним и диэлектрическая проницаемость газа (при постоянной плотности) остаются *неизменными.* Лишь при высоких температурах заметная доля молекул переходит в *возбужденные состояния*, или *диссоциирует.* Тогда появляется и зависимость величины ε от температуры.

Из независимости ε от температуры следует, что в газах с неполярными молекулами внутренняя энергия U совпадает со свободной энергией, т. е. дается выражением (29.9). Это непосредственно видно из формулы (31.14). Влияние диэлектрика на внутреннюю энергию имеет в этом случае простое физическое объяснение. Плотность внутренней энергии U слагается из двух частей: 1) из плотности энергии поля в вакууме $\mathbf{E}^2/(8\pi)$ и 2) из плотности энергии деформации молекул $n\mathbf{pE}/2 = \mathbf{PE}/2$. Действительно, сумма этих двух частей равна $\mathbf{E}^2/(8\pi) + (\mathbf{PE})/2 = (\mathbf{ED})/(8\pi)$, что совпадает с (29.6).

6. Распространение изложенной теории на случай *плотных сред* (жидкостей и твердых тел) встречает большие трудности. Простейший способ, каким это может быть сделано, состоит в следующем. Пренебрежем размерами молекул и будем считать их «точечными», т. е. такими, линейные размеры которых очень малы по сравнению со средними расстояниями между ними. Тогда можно будет пренебречь и изменениями действующего поля в пределах молекулы, т. е. считать, что поле $\mathbf{E}'$ относится к *центру молекулы*, на которую оно действует. Для нахождения дипольного момента молекулы надо в формуле (35.1) поле $\mathbf{E}$ заменить на $\mathbf{E}'$, т. е. написать

$$\mathbf{p} = \beta \mathbf{E}'. \tag{35.12}$$

Задача сводится к вычислению действующего поля $\mathbf{E}'$. Для этого возьмем в диэлектрике произвольную молекулу A и опишем вокруг нее вспомогательную физически бесконечно малую сферу S с центром O, совпадающим с центром этой молекулы. Интересующее нас поле $\mathbf{E}'$ в точке O возбуждается всеми источниками, за исключением зарядов самой молекулы A. Электрическое поле $\mathbf{E}_{\text{внеш}}$ внутри сферы S, создаваемое зарядами, расположенными вне ее, можно вычислить, предполагая, что эти заряды распределены в пространстве непрерывно. А так как сфера S физически бесконечно мала, то поляризацию диэлектрика в ее окрестности можно считать однородной. Поэтому можно воспользоваться формулой (16.7) и написать

$$\mathbf{E}_{\text{внеш}} = \mathbf{E} + \frac{4\pi}{3}\mathbf{P}.$$

В частности, такое поле будут создавать внешние источники в центре O сферы S. Чтобы получить действующее поле, к вектору $\mathbf{E}_{\text{внеш}}$ надо добавить поле $\mathbf{E}_1$, которое создается в точке O всеми зарядами, находящимися внутри сферы S (за исключением зарядов молекулы A). При выполнении определенных условий можно утверждать, что $\mathbf{E}_1 = 0$. Это применимо для кубических кристаллов, построенных из электрически нейтральных и изотропных точечных молекул. Такие молекулы возбуждают электрические поля, как точечные диполи с индуцированными дипольными моментами $\mathbf{p}$, ориентированными вдоль среднего поля $\mathbf{E}$. Если $\mathbf{r}$ — радиус-вектор, проведенный из центра O к одному из диполей, то поле $\mathbf{E}_1$ в точке O представится суммой

$$\mathbf{E}_1 = \sum \left(\frac{3(pr)\mathbf{r}}{r^5} - \frac{\mathbf{p}}{r^3} \right),$$

где суммирование ведется по всем диполям сферы S, за исключением диполя, находящегося в ее центре. Поскольку сфера S физически бесконечно мала, индуцированные дипольные моменты $\mathbf{p}$ всех молекул внутри нее одинаковы, и их можно вынести из-под знака суммы. Имея это в виду и воспользовавшись соотношением $r^2 = x^2 + y^2 + z^2$, для

x-составляющей вектора $\mathbf{E}_1$ получим

$$E_{1x} = p_x \sum \frac{2x^2 - y^2 - z^2}{r^5} + 3p_y \sum \frac{xy}{r^5} + 3p_z \sum \frac{xz}{r^5}.$$

Последние две суммы обращаются в нуль в силу кубической симметрии кристаллической решетки. В силу той же симметрии

$$\sum \frac{x^2}{r^5} = \sum \frac{y^2}{r^5} = \sum \frac{z^2}{r^5},$$

так что $E_{1x} = 0$. То же справедливо для y- и z-составляющих вектора $\mathbf{E}_1$, а потому $\mathbf{E}_1 = 0$. Таким образом, все действующее поле сводится к $\mathbf{E}_{\text{внеш}}$, т. е.

$$\mathbf{E}' = \mathbf{E} + \frac{4\pi}{3}\mathbf{P}. \tag{35.13}$$

Эта формула была получена голландским физиком Х.А. Лоренцем. Ее можно применять не только для кристаллов кубической системы, но и для плотных газов, построенных из точечных неполярных молекул, если под $\mathbf{E}'$ понимать действующее поле, усредненное по времени. Действительно, молекулы газа распределены в пространстве хаотически, причем положение каждой из них практически не зависит от положения других. Поэтому рассуждения, с помощью которых мы убедились в равенстве нулю вектора $\mathbf{E}_1$, сохраняют силу и в этом случае (после усреднения вектора $\mathbf{E}_1$ по времени). Однако формула (35.13) неприменима для диэлектриков, построенных из полярных молекул. В таких диэлектриках дипольные молекулы испытывают сильное ориентирующее действие со стороны соседних дипольных молекул. Конечно, в отсутствие внешнего электрического поля все направления осей дипольных молекул в пространстве равновероятны. Однако между направлениями дипольных моментов соседних молекул существует определенная корреляция, мало изменяющаяся и при наложении внешнего электрического поля. Это обстоятельство и является причиной неприменимости формулы (35.13) к диэлектрикам с полярными молекулами.

7. Применим теперь формулу (35.13) для расчета поляризации диэлектрика $\mathbf{P} = n\mathbf{p} = n\beta\mathbf{E}'$. Получаем

$$\mathbf{P} = n\beta\left(\mathbf{E} + \frac{4\pi}{3}\mathbf{P}\right),$$

откуда

$$\mathbf{P} = \frac{n\beta}{1 - (4/3)\pi n\beta}\mathbf{E}.$$

Далее,

$$\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P} = \frac{1 + (8/3)\pi n\beta}{1 - (4/3)\pi n\beta}\mathbf{E}.$$

Следовательно,

$$\varepsilon = \frac{1 + (8/3)\pi n\beta}{1 - (4/3)\pi n\beta}. \tag{35.14}$$

Разрешая это уравнение относительно $(4/3)\pi n\beta$, получим

$$\frac{\varepsilon-1}{\varepsilon+2}=\frac{4\pi}{3}\,n\beta. \tag{35.15}$$

Так как величина $n\beta$ пропорциональна плотности диэлектрика τ, то при изменении плотности должно соблюдаться соотношение

$$\frac{1}{\tau}\,\frac{\varepsilon-1}{\varepsilon+2}=\text{const}. \tag{35.16}$$

Это соотношение называется *формулой Мосотти* (1791–1863)–*Клаузиуса* (1822–1888). Приведенный вывод ее принадлежит Х. А. Лоренцу. Для диэлектриков, подчиняющихся формуле Мосотти–Клаузиуса, диэлектрическая проницаемость ε зависит только от плотности, но не зависит от температуры. Следовательно, для таких диэлектриков свободная энергия (точнее, ее электрическая часть) совпадает с внутренней. При малых ε, близких к единице, формула (35.16) переходит в (35.11).

Формула Мосотти–Клаузиуса находится в удовлетворительном согласии с опытом для жидких и газообразных диэлектриков с неполярными молекулами, хотя в случае жидкостей и не выполняется условие точечности молекул, использованное при выводе соотношения (35.16). Так, для газообразного сероуглерода CS_2 при температуре 0 °C и нормальном атмосферном давлении измеренное значение ε равно 1,0029. Пользуясь им, по формуле (35.10) находим $4\pi n\beta=0{,}0029$. Плотность жидкого сероуглерода в 381 раз больше плотности газообразного. Если верно предположение, что поляризуемость молекулы β не меняется при переходе из газообразного состояния в жидкое, то для жидкого сероуглерода величина $4\pi n\beta$ должна быть больше также в 381 раз, т. е. должна составлять 1,11. Этому значению по формуле (35.14) соответствует $\varepsilon=2{,}76$, что достаточно хорошо согласуется с экспериментально найденной величиной $\varepsilon=2{,}64$.

Для диэлектриков с полярными молекулами (вода, спирты, эфиры и пр.) формула Мосотти–Клаузиуса не подтверждается опытом. Например, если бы для воды провести такие же вычисления, какие выше были проведены для сероуглерода, то получилось бы $4\pi n\beta=$ $=13{,}2$. Подставив это значение в формулу (35.14), мы пришли бы к отрицательному значению ε. Между тем для воды опыт дает $\varepsilon=81$.

§ 36. Электронная теория поляризации полярных газообразных диэлектриков

1. При изложении теории поляризуемости диэлектриков, состоящих из полярных молекул, мы ограничимся *случаем газов.* Тогда действующее поле $\mathbf{E}'$ можно отождествить со средним максвелловским полем $\mathbf{E}$. Для типичных полярных молекул (H_2O, HCl, HBr, HI, CO, спирты, эфиры и пр.) дипольные моменты $\mathbf{p}_0$ порядка 10^{-18} СГСЭ-ед. По сравнению с ними индуцированные дипольные моменты на несколько порядков меньше. Действительно, пользуясь оценочной формулой (35.2),

получаем $\beta \sim 10^{-24}$ см^3. В довольно сильном электрическом поле $E \sim 100$ СГСЭ-ед. (30 000 В/см) молекула приобретает дипольный момент всего порядка 10^{-22} СГСЭ-ед. Поэтому индуцированным моментом по сравнению с $\mathbf{p}_0$ можно в первом приближении полностью пренебречь. В этом приближении поляризация диэлектрика обусловлена *только поворотами* осей дипольных молекул в электрическом поле. Введем функцию распределения $f(\mathbf{b})$ для описания распределения осей диполей по направлениям в пространстве. По определению величина $dn = nf(\mathbf{b})\,d\Omega$ дает среднее число диполей в единице объема, оси которых лежат в пределах телесного угла $d\Omega$. Единичный вектор $\mathbf{b}$ указывает направление оси элементарного телесного угла $d\Omega$. В отсутствие электрического поля все направления дипольных осей равновероятны, т. е. функция $f(\mathbf{b})$ постоянна. При наличии электрического поля функция $f(\mathbf{b})$ определяется формулой Больцмана (1844–1906)

$$f = Ce^{-u/kt}, \tag{36.1}$$

где C — постоянная, а $u = -(\mathbf{p}_0\mathbf{E})$ — потенциальная энергия диполя $\mathbf{p}_0$ во внешнем поле $\mathbf{E}$. Поля обычно бывают «слабыми», т. е. удовлетворяют условию

$$\frac{p_0 E}{kT} \ll 1. \tag{36.2}$$

В этих случаях функцию (36.1) можно разложить в степенной ряд и оборвать разложение на первых двух членах. Предполагая, что поле $\mathbf{E}$ направлено параллельно оси Z, получим

$$f(\mathbf{b}) = C\left(1 + \frac{p_0 E}{kT}\right) = C\left(1 + \frac{p_0 E}{kT}\, b_z\right).$$

Постоянная C определяется условием нормировки

$$\int f(\mathbf{b})\,d\Omega = 1, \tag{36.3}$$

в котором предполагается, что интегрирование производится по всем направлениям в пространстве. В разбираемом нами случае это условие принимает вид

$$C\int d\Omega + C\frac{p_0 E}{kT}\int b_z\,d\Omega = 1.$$

Первый интеграл равен 4π, а второй обращается в нуль, поскольку проекция b_z одинаково часто принимает положительные и равные им по модулю отрицательные значения. Следовательно, $C = 1/4\pi$, а потому

$$f(\mathbf{b}) = \frac{1}{4\pi}\left(1 + \frac{p_0 E}{kT}\, b_z\right). \tag{36.4}$$

2. Вектор поляризации среды, очевидно, будет направлен вдоль $\mathbf{E}$, т. е. параллельно оси Z. Его величина равна

$$P = n\int p_{0z} f(\mathbf{b})\,d\Omega = np_0\int b_z f(\mathbf{b})\,d\Omega = \frac{np_0}{4\pi}\int b_z\,d\Omega + \frac{np_0^2 E}{4\pi kT}\int b_z^2\,d\Omega.$$

Первый интеграл в правой части, как мы видели, равен нулю. Для вычисления последнего интеграла используем соотношение $b_x^2 + b_y^2 + b_z^2 = 1$, интегрирование которого дает

$$\int b_x^2\, d\Omega + \int b_y^2\, d\Omega + \int b_z^2\, d\Omega = 4\pi.$$

Все три интеграла здесь равны между собой ввиду симметрии, а потому для каждого из них получаем $4\pi/3$. Таким образом,

$$\mathbf{P} = \frac{np_0^2}{3kT}\,\mathbf{E}, \tag{36.5}$$

$$\alpha = \frac{np_0^2}{3kT}, \tag{36.6}$$

$$\varepsilon = 1 + \frac{4\pi np_0^2}{3kT}. \tag{36.7}$$

Отметим, что полученные результаты полностью, включая и числовой множитель $1/3$, остаются справедливыми и в квантовой механике (пока справедлива статистика Больцмана). В этом существенное отличие формул (36.5)–(36.7) от соответствующих формул теории парамагнетизма (см. § 77).

Отметим еще, что все результаты (36.5)–(36.7) верны *во втором порядке* по параметру $p_0E/(kT)$, а не только в первом, как можно было бы думать на основании их вывода. Действительно, если в разложении функции распределения $f(\mathbf{b})$ сохранить члены, квадратичные относительно указанного параметра, то при вычислении вектора поляризации $\mathbf{P}$ добавятся интегралы первой и третьей степени относительно проекции b_z. Все эти интегралы, в силу симметрии, обращаются в нуль и не влияют на окончательные результаты. Это замечание надо иметь в виду при решении задачи 4.

3. В действительности полярные молекулы в электрическом поле не только поворачиваются, но и *деформируются*, т. е. приобретают индуцированные дипольные моменты. Поляризуемость диэлектрика α в общем случае слагается из *ориентационной части* $np_0^2/(3kT)$ и *деформационной части* $n\beta$. Поэтому

$$\varepsilon = 1 + 4\pi n\beta + 4\pi\frac{np_0^2}{3kT}. \tag{36.8}$$

Деформационная часть поляризуемости *не зависит от температуры*, а ориентационная — *обратно пропорциональна* термодинамической *температуре*. Исследуя температурный ход диэлектрической проницаемости, можно поэтому отделить деформационную часть поляризуемости от ориентационной. Если по оси абсцисс откладывать величину, обратную температуре, а по оси ординат — разность $\varepsilon - 1$, то соответствующий график будет прямолинейным (рис. 94). Экстраполируем его в область высоких температур. Тогда он отсечет на оси ординат отрезок

$OA = 4\pi n\beta$. По длине этого отрезка можно вычислить поляризуемость молекулы β, а по наклону графика — ее постоянный дипольный момент p_0. Именно таким методом чаще всего измеряют дипольные моменты полярных молекул.

Посмотрим, наконец, что конкретно означает условие (36.2). Пусть $p_0 = 10^{-18}$ СГСЭ-ед. При комнатной температуре ($T \approx 300$ К) теория применима при выполнении условия

$$E \ll \frac{kT}{p_0} \approx 4 \cdot 10^4 \text{ СГСЭ-ед.} \approx 10^7 \text{ В/см}.$$

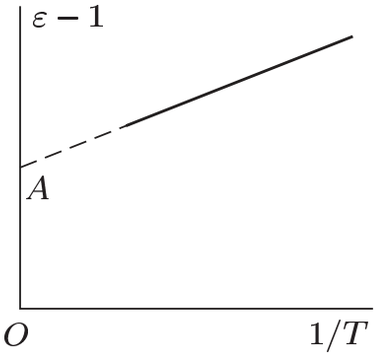


Рис. 94

При $E \gg kT/p_0$ дипольные моменты всех молекул должны были бы ориентироваться вдоль поля $\mathbf{E}$. Наступило бы *состояние насыщения*, при котором поляризация $\mathbf{P}$ уже не изменялась бы при дальнейшем возрастании напряженности электрического поля. Однако такое состояние в газах не достигается, так как значительно раньше наступает *электрический пробой.*

4. Внутреннюю энергию газообразного диэлектрика с полярными молекулами можно вычислить по формуле (31.14). Подставляя в нее значение ε из (36.7), получим

$$U = \frac{E^2}{8\pi}. \tag{36.9}$$

Таким образом, внутренняя энергия U в рассматриваемом случае зависит *только от напряженности электрического поля* $\mathbf{E}$. Вещество само по себе не оказывает непосредственного влияния на величину внутренней энергии. Его влияние сказывается лишь постольку, поскольку оно изменяет напряженность электрического поля $\mathbf{E}$. Этот результат нетрудно понять. Внутренняя энергия слагается из энергии электрического поля $E^2/8\pi$ и энергии вещества в нем. Последняя в свою очередь состоит из кинетической и потенциальной. При возрастании электрического поля потенциальная энергия диполя — $(\mathbf{p}_0\mathbf{E})$ убывает, а кинетическая возрастает на такую же величину. Сумма кинетической и потенциальной энергий диполей остается неизменной, если только не подводится тепло и не совершается дополнительная макроскопическая работа.

ЗАДАЧИ

1. Вычислить плотность пондеромоторной силы $\mathbf{f}$, действующей в однородном незаряженном газообразном диэлектрике.

Решение. Согласно формулам (35.11) и (36.8) для газообразных диэлектриков величина $\varepsilon - 1$ пропорциональна плотности диэлектрика τ. Имея это в виду, по формуле (34.5) находим

$$\mathbf{f} = \frac{\varepsilon - 1}{8\pi} \operatorname{grad} E^2 - \operatorname{grad} \mathscr{P}. \tag{36.10}$$

В электростатическом поле при равновесии сила $\mathbf{f}$ обращается в нуль, т. е. стрикционные силы $\dfrac{\varepsilon-1}{8\pi}\operatorname{grad}E^2$ уравновешиваются силами электростатического давления.

Результат (36.10) легко понять с точки зрения атомистической теории поляризации диэлектриков. В газах молекулы можно считать независимыми. На каждый диполь в электрическом поле действует сила $(\mathbf{p}\nabla)\mathbf{E}$, а на диполи единицы объема — сила $(\mathbf{P}\nabla)\mathbf{E}=\dfrac{\varepsilon-1}{4\pi}(\mathbf{E}\nabla)\mathbf{E}$. Используя потенциальность вектора $\mathbf{E}$, нетрудно показать, что $(\mathbf{E}\nabla)\mathbf{E}=(1/2)\operatorname{grad}E^2$. Таким образом, становится понятным смысл первого слагаемого в формуле (36.10).

2. Ту же задачу решить для диэлектриков, подчиняющихся формуле Мосотти–Клаузиуса (35.16).

Ответ.

$$\mathbf{f}=\frac{(\varepsilon-1)(\varepsilon+2)}{24\pi}\operatorname{grad}E^2-\operatorname{grad}\mathscr{P}. \tag{36.11}$$

3. Вычислить функцию распределения $f(\mathbf{b})$ с точностью до членов второго порядка включительно относительно параметра $(p_0E)/kT$.

Ответ.

$$f=\frac{1}{4\pi}\left(1+\frac{p_0E}{kT}\,b_z+\frac{p_0^2E^2}{2k^2T^2}\,b_z-\frac{p_0^2E^2}{6k^2T^2}\right). \tag{36.12}$$

4. С помощью формулы Больцмана о связи энтропии с вероятностью найти ту часть энтропии газообразного диэлектрика, состоящего из полярных молекул, которая зависит от электрического поля.

Решение. Энтропия единицы объема диэлектрика определяется так:

$$S=-kn\int f\ln f\,d\Omega+\text{const} \tag{36.13}$$

(см. т. II, § 81). Поскольку нас интересует только электрическая часть энтропии, мы можем не принимать во внимание зависимость функции распределения f от кинетической энергии поступательного и вращательного движения молекул и интегрировать только по направлениям осей этих молекул. Если опустить несущественную аддитивную постоянную, то из (36.12) находим во втором порядке

$$\ln f=\frac{p_0E}{kT}\,b_z-\frac{p_0^2E^2}{6k^2T^2}+\frac{p_0^2E^2}{2k^2T^2}\,b_z.$$

Подставляя это значение в формулу (36.13) и отбрасывая все интегралы, обращающиеся в нуль из-за симметрии, получим

$$S=-\frac{kn}{4\pi}\int\left(1+\frac{p_0E}{kT}\,b_z\right)\left(\frac{p_0E}{kT}\,b_z-\frac{p_0^2E^2}{6k^2T^2}\right)d\Omega=$$

$$=-\frac{np_0^2E^2}{4\pi kT^2}\int\left(b_z^2-\frac{1}{6}\right)d\Omega=-\frac{np_0^2}{3kT}\,\frac{E^2}{2T},$$

или на основании формул (36.6) и (36.7)

$$S=-\alpha\frac{E^2}{2T}=-\frac{\varepsilon-1}{8\pi}\,\frac{E^2}{T}. \tag{36.14}$$

Используя формулу $U=\Psi+TS$, получаем плотность внутренней энергии $U=E^2/8\pi$. Как и следовало ожидать, результат совпадает с результатом (36.9), полученным ранее термодинамическим путем.

5. Определить изменение температуры газообразного диэлектрика с полярными молекулами при адиабатическом выключении электрического поля, если объем газа во время процесса выключения поддерживается постоянным (электрокалорический эффект).

Решение. Как показывает формула (36.7), для таких диэлектриков справедливо соотношение

$$\frac{(\varepsilon-1)T}{\tau}=\text{const}, \tag{36.15}$$

откуда

$$\left(\frac{\partial\varepsilon}{\partial T}\right)_\tau=-\frac{\varepsilon-1}{T}.$$

Подставив это значение в формулу (31.16) и выполнив интегрирование по E, получим

$$T_2-T_1=\frac{\varepsilon-1}{8\pi C_E}\,E^2,$$

где E — начальное значение напряженности электрического поля. (При интегрировании изменением ε с температурой мы пренебрегли.) Здесь C_E означает теплоемкость единицы объема диэлектрика при постоянных объеме и напряженности электрического поля. Зависимостью теплоемкости от напряженности поля можно пренебречь. Тогда

$$C_E=\frac{C_V\mathscr{P}}{RT},$$

где C_V — теплоемкость моля газа при постоянном объеме. С помощью соотношения $C_P-C_V=R$ величину C_E можно выразить через адиабатическую постоянную $\gamma=C_P/C_V$. Сделав это и подставив результат в формулу для T_2-T_1, получим

$$T_2-T_1=-\frac{(\varepsilon-1)(\gamma-1)}{8\pi\mathscr{P}}\,E^2T. \tag{36.16}$$

Используя формулу (36.7), а также соотношение $\mathscr{P}=nkT$, преобразуем эту формулу к виду

$$T_2-T_1=-\frac{(\gamma-1)p_0^2}{24\pi k^2T}\,E^2. \tag{36.17}$$

Таким образом, при адиабатическом выключении электрического поля диэлектрик охлаждается. Понижение температуры T_2-T_1 обратно пропорционально абсолютной температуре T. Для примера возьмем $\gamma=1{,}4$, $p_0=10^{-18}$ СГСЭ-ед., $E=100$ СГСЭ-ед., $T=100$ К. Тогда по формуле (36.17) найдем $T_2-T_1\approx3\cdot10^{-5}$ К.

§ 37. Пьезоэлектричество

1. Во многих кристаллах при растяжении и сжатии в определенных направлениях возникает электрическая поляризация. В результате этого на их поверхностях появляются электрические заряды обоих знаков. Это явление, получившее название *прямого пьезоэлектрического эффекта*, было открыто в 1880 г. братьями Пьером и Жаком Кюри. Оно наблюдалось затем на

кристаллах турмалина, цинковой обманки, хлората натрия, винной кислоты, тростникового сахара, сегнетовой соли, титаната бария и многих других веществ. Пьезоэлектрическими свойствами могут обладать только *ионные кристаллы.* Если кристаллические решетки положительных и отрицательных ионов, из которых построены такие кристаллы, под действием внешних сил деформируются по-разному, то в противоположных местах на поверхности кристалла выступают электрические заряды разных знаков. Это и есть пьезоэлектрический эффект. При однородной деформации пьезоэлектрический эффект наблюдается при наличии в кристалле одной или нескольких *полярных осей* (направлений). Под полярной осью (направлением) кристалла понимают всякую прямую, проведенную через кристалл, оба конца которой неравноценны, т. е. невзаимозаменяемы. Иными словами, при повороте кристалла на 180° вокруг любой оси, перпендикулярной к полярной, он не совмещается сам с собою. Вообще, для существования пьезоэлектрическою эффекта при однородной деформации необходимо *отсутствие у кристалла центра симметрии.* Действительно, если бы недеформированный кристалл имел центр симметрии, то последний сохранился бы и при однородной деформации кристалла. С другой стороны, в электрически поляризованном кристалле есть особое направление, а именно направление вектора поляризации. При наличии такового кристалл не может иметь центр симметрии. Получившееся противоречие и доказывает наше утверждение. Из 32 кристаллических классов не имеет центра симметрии 21 класс. У одного из них, однако, сочетание других элементов симметрии делает пьезоэлектрический эффект также невозможным. Таким образом, пьезоэлектрические свойства наблюдаются у 20 кристаллических классов.

2. Рассмотрим пьезоэлектрический эффект на примере кристалла кварца — важнейшего пьезоэлектрического кристалла, нашедшего широкие научно-технические применения благодаря своим превосходным механическим и электрическим свойствам. При обычных температурах и давлениях кварц встречается в так называемой α-модификации. Кристалл α-кварца (рис. 95) относится к тригональной системе и имеет три оси симметрии второго порядка, обозначенные на рис. 95 через X_1, X_2, X_3. Они и являются полярными осями кристалла. Каждая из них соединяет противоположные, но неравнозначные ребра шестигранной призмы. Неравнозначность этих ребер видна из того, что к краям одного из них примыкают маленькие грани, обозначенные на рисунке буквами *а* и *б*, тогда как у краев другого ребра таких граней нет. Четвертая ось Z является осью симметрии третьего порядка. Ее называют *оптической осью*, так как поворот кристалла вокруг этой оси на любой угол не оказывает никакого влияния на распространение света в кристалле.

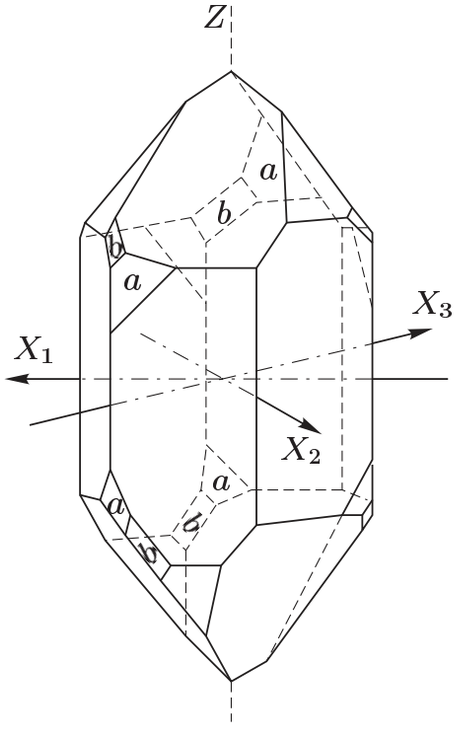


Рис. 95

При механических воздействиях на кристалл кварца на концах полярной оси (точнее, на перпендикулярных к ней гранях) появляются противоположные электрические заряды. Не обязательно, чтобы приложенные внешние силы дей-

ствовали в направлении рассматриваемой полярной оси. Необходимо лишь, чтобы в результате действия внешних сил возникало растяжение или сжатие вдоль этой оси.

При температуре до 200 °С пьезоэлектрические свойства кварца практически не зависят от температуры. С дальнейшим повышением температуры пьезоэлектрический эффект медленно убывает. При 576 °С α-кварц претерпевает фазовое превращение и переходит в β-модификацию. Кристаллы β-кварца относятся к гексагональной системе, а потому пьезоэлектрические явления в них не наблюдаются в согласии с тем, что было сказано выше. При обратном понижении температуры первоначальная структура кварца восстанавливается, причем это восстановление происходит при температуре, несколько более низкой, чем исходная (гистерезис). Ниже всюду речь идет об α-кварце.

3. Возникновение пьезоэлектрического эффекта легко понять с помощью модельного рассмотрения, предложенного Мейсснером. Химическая формула кварца имеет вид SiO_2. Его кристаллическая решетка состоит из положительных ионов кремния и отрицательных ионов кислорода. Каждый ион кремния несет четыре, а каждый ион кислорода — два элементарных заряда. В первом приближении можно представить, что ионы кремния и кислорода расположены в шестигранных ячейках, одна из которых изображена на рис. 96, если смотреть на кристалл вдоль оптической оси (перпендикулярной к плоскости рисунка). Ионы кремния изображены большими шариками *1*, *2*, *3*, ионы кислорода — маленькими. Те и другие ионы расположены по спирали, направление вращения которой определяется тем, какой взят кварц: левый или правый (рис. 95 и 96 относятся к левому кварцу). Ион кремния *3* лежит несколько глубже иона *2*, а ион *2* — глубже иона *1*. Расположение ионов кислорода не требует дополнительных разъяснений. В целом ячейка электрически нейтральна и не имеет дипольного электрического момента.

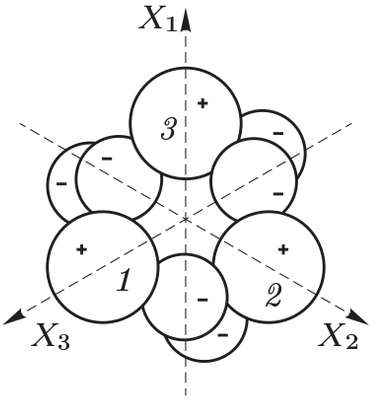


Рис. 96

Для упрощения рассуждений заменим каждую пару соседних ионов кислорода одним отрицательным ионом с удвоенным зарядом. Мы придем к упрощенной модели ячейки, изображенной на рис. 97 *а*. Если подвергнуть такую ячейку сжатию вдоль полярной оси X_1 (рис. 97 *б*), то ион кремния *3* и ион кислорода *4* вклинятся между окружающими их боковыми ионами. В результате на плоскости A пластинки появится отрицательный, а на плоскости B — положительный заряды (*продольный пьезоэлектрический эффект*).

При сжатии в боковом направлении, т. е. перпендикулярно к полярной и оптической осям (рис. 97 *в*), ионы кремния *1* и *2* получают одинаковые, но противоположно направленные смещения внутрь ячейки. Так же ведут себя ионы кислорода *5* и *6*. При этом сохраняется симметрия ячейки относительно плоскости, проходящей посередине между плоскостями C и D, и на этих плоскостях не возникает никаких зарядов. Однако ион кремния *3* и ион кислорода *4* смещаются наружу. Благодаря этому возникает дипольный момент, направленный в положительную сторону полярной оси X_1. На плоскости A появляется положительный, а на плоскости B — отрицательный заряды (*поперечный пьезоэлектрический эффект*). Знаки зарядов в продольном и поперечном эффектах, таким образом, противоположны.

Из рассматриваемой модели видно также, что замена сжатия растяжением приводит к изменению знаков электрических зарядов при пьезо-

электрическом эффекте и что поляризация пропорциональна деформации кристалла (когда деформации малы). А так как между деформацией и силой согласно закону Гука (1635–1703) существует прямая пропорциональность, то поляризация кристалла при пьезоэлектрическом эффекте должна быть

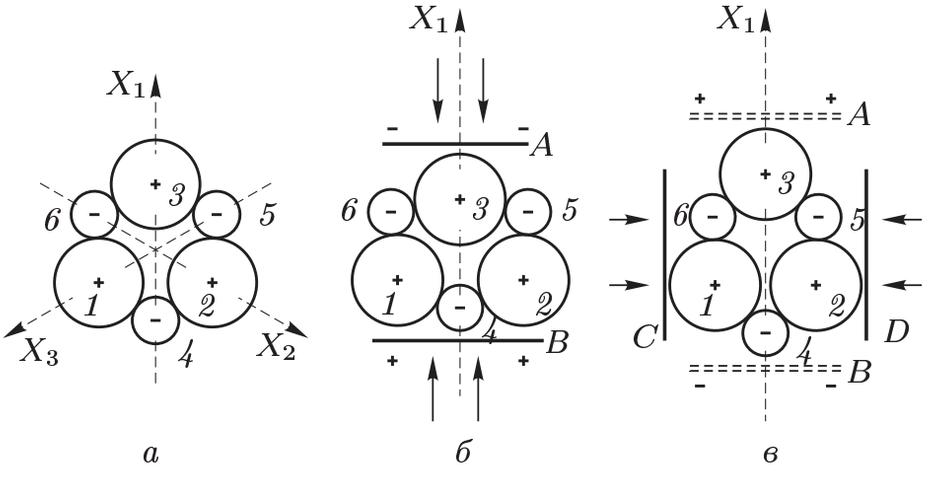


Рис. 97

пропорциональна также приложенной силе. Наконец, из модели видно, что сжатие или растяжение кристалла в направлении оптической оси никакими пьезоэлектрическими эффектами не сопровождается. Все эти заключения подтверждаются опытом.

4. Согласно изложенному для получения максимальных электрических зарядов кристалл кварца надо растягивать или сжимать в направлении одной из полярных осей. В соответствии с этим кварцевые пластинки и стержни, применяемые в пьезоэлектрических опытах и приборах, вырезаются обычно так, чтобы пара плоскостей, образовавшихся при срезе, была перпендикулярна к одной из полярных осей. Такая ось называется также *электрической осью* или *пьезоосью* и обозначается обычно через X. Оптическая ось принимается за ось Z соответствующей правой системы координат. Ось Y такой системы координат называют *механической осью кристалла.* На рис. 98 изображена пластинка, вырезанная указанным образом. Длины ребер пластинки обозначены через l (длина), b (ширина), h (толщина).

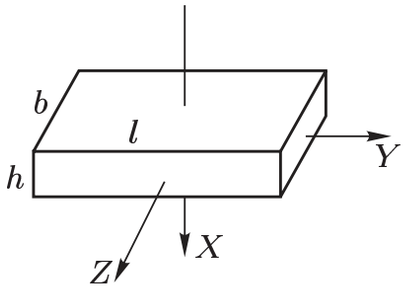


Рис. 98

В соответствии с приведенным выше наглядным объяснением при растяжении или сжатии пластинки в направлении оптической оси Z пьезоэлектрический эффект не возникает. При растяжении вдоль электрической оси X нижняя поверхность пластинки электризуется положительно, а верхняя — отрицательно. То же самое наблюдается при сжатии пластинки в направлении механической оси Y. При замене сжатия растяжением и наоборот знаки зарядов меняются на противоположные. Если нет касательных напряжений, то поляризация кварцевой пластинки при растяжении или сжатии определяется выражением

$$P_x = d_{11}(\tau_x - \tau_y), \tag{37.1}$$

где τ_x и τ_y — механические натяжения, действующие параллельно осям X и Y, а d_{11} — постоянная, называемая *пьезоэлектрическим модулем* (смысл

двойного индекса при d выяснится в п. 9). Для кварца

$$d_{11} = 6{,}99 \cdot 10^{-8} \text{ дин}^{-1/2} \cdot \text{см}.$$

Допустим, например, что $\tau_x = 10^6$ дин/см^2, $\tau_y = 0$. Тогда на нижней поверхности пластинки появится положительный заряд с плотностью $\sigma = P_x = 6{,}99 \cdot 10^{-2}$ СГСЭ-ед. $= 2{,}33 \cdot 10^{-7}$ Кл/м^2. Ему соответствует внутри пластинки электрическое поле $E_x = 4\pi\sigma \approx 0{,}88$ СГСЭ-ед. ≈ 240 В/см. При толщине пластинки $h = 0{,}5$ см она заряжается до разности потенциалов $\varphi \approx 120$ В.

Для того чтобы использовать поляризационные заряды, появляющиеся на противоположных гранях кварцевой пластинки при ее деформации, эти грани снабжают металлическими обкладками. На таких обкладках индуцируются заряды, равные и противоположные по знаку поляризационным, а во внешних проводах, соединяющих обкладки, возникает электрический ток.

5. Значительно сильнее, чем у кварца, пьезоэлектрические свойства выражены у кристаллов сегнетовой соли. Благодаря этому она применяется во многих пьезоэлектрических приборах. Однако сегнетова соль очень хрупка и имеет низкую температуру плавления (+63 °С), что сильно ограничивает возможности ее практического использования. Она удобна для демонстрации прямого пьезоэлектрического эффекта. Пластинка сегнетовой соли слегка зажимается между двумя обкладками из листовой латуни (рис. 99). Обкладки соединены проводами с неоновой лампочкой. Последняя представляет собою стеклянный баллончик, наполненный разреженным неоном. Внутрь баллончика введены два металлических электрода. Когда разность потенциалов между электродами превосходит определенную величину (потенциал зажигания), в лампочке возникает газовый разряд, сопровождающийся свечением неона. Если резко ударять резиновым молотком по пластинке сегнетовой соли, то при каждом ударе появляется кратковременная вспышка неоновой лампочки. Вместо сегнетовой соли в описанной демонстрации можно пользоваться пластинкой из титаната бария.

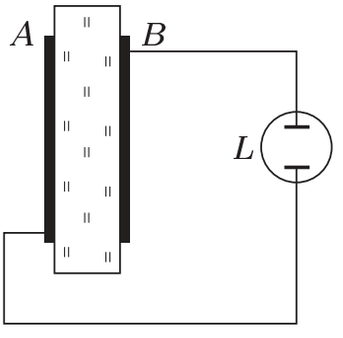


Рис. 99

6. В 1881 г. Липпман (1845–1921), исходя из термодинамических соображений, предсказал *обратный пьезоэлектрический эффект*, который в том же году и был обнаружен братьями Кюри на кристаллах кварца. Обратный пьезоэлектрический эффект состоит в том, что при внесении пьезоэлектрического кристалла в электрическое поле в кристалле возникают механические напряжения, под действием которых кристалл деформируется.

Допустим, что кварцевая пластинка (см. рис. 98) внесена в электрическое поле, направленное параллельно оси X. Пусть она в направлениях X и Y подвержена также действию механических натяжений τ_x и τ_y соответственно. Если $V = hbl$ — объем пластинки, то элементарная работа, которую надо затратить на ее поляризацию при квазистатическом процессе, определяется выражением $\delta A_{\text{пол}} = VE\,dP + VE_x\,dP_x$. Элементарная же механическая работа, совершаемая квазистатически силами натяжения при удлинении ребер h и l, будет $\delta A_{\text{мех}} = bl\tau_x\,dh + hb\tau_y\,dl$. Применим к рассматриваемому процессу термодинамическое соотношение $dU = T\,dS + \delta A$. Разделив его на V и обозначив через s и u значения удельной энтропии и внутренней

энергии, получим

$$du = T\,ds + E_x\,dP_x + \tau_x\frac{dh}{h} + \tau_y\frac{dl}{l},$$

или

$$du = T\,ds + E_x\,dP_x + \tau_x\,d\ln h + \tau_y\,d\ln l.$$

Введя функцию $g = u - Ts - E_xP_x - \tau_x\ln h - \tau_y\ln l$, преобразуем это соотношение к виду

$$dg = -s\,dT - P_x\,dE_x - \ln h\,d\tau_x - \ln l\,d\tau_y.$$

Так как выражение справа есть полный дифференциал функции g, то должно быть

$$\left(\frac{\partial P_x}{\partial\tau_x}\right)_T = \left(\frac{\partial\ln h}{\partial E_x}\right)_T = \frac{1}{h}\,\frac{\partial h}{\partial E_x},$$

$$\left(\frac{\partial P_x}{\partial\tau_y}\right)_T = \left(\frac{\partial\ln l}{\partial E_x}\right)_T = \frac{1}{l}\,\frac{\partial l}{\partial E_x},$$

или с учетом соотношения (37.1)

$$\frac{\partial h}{\partial E_x} = hd_{11}, \quad \frac{\partial l}{\partial E_x} = -ld_{11}. \tag{37.2}$$

Эти формулы и описывают обратный пьезоэлектрический эффект в кварце. В линейном приближении, в котором только и верна излагаемая теория, формулы (37.2) записываются в виде

$$\delta h = d_{11}hE_x = d_{11}\varphi, \tag{37.3}$$

$$\delta l = -d_{11}lE_x = -(l/h)d_{11}\varphi, \tag{37.4}$$

где δh и δl — абсолютные приращения размеров пластинки при наложении электрического поля E_x, а $\varphi = hE_x$ — разность потенциалов между гранью bl и гранью, ей противоположной (рис. 98).

Формула (37.3) выражает *продольный обратный пьезоэлектрический эффект*, а формула (37.4) — *поперечный.* При наложении электрического поля параллельно электрической оси меняется толщина пластинки (продольный эффект) и ее длина (поперечный эффект). Если толщина h увеличивается, то длина l уменьшается, и наоборот, причем относительные изменения этих размеров по абсолютной величине одинаковы, так что объем пластинки остается неизменным. Абсолютное значение δh не зависит от толщины пластинки, а только от приложенной разности потенциалов φ. При $\varphi = 3000$ В $= 10$ СГСЭ-ед. из формулы (37.3) находим $\delta h = 6{,}99\cdot10^{-7}$ см $= 6{,}99\cdot10^{-3}$ мкм. Если $l = 10h$, то поперечный эффект при той же разности потенциалов будет в 10 раз больше. Модуль Юнга (1773–1829) кварца в направлении электрической оси $\mathscr{E} = 7{,}87\cdot10^{11}$ дин/см^2. При толщине пластинки $h = 0{,}5$ см в ней в случае продольного эффекта в приведенном выше примере возникают натяжения или давления $\mathscr{P} = \mathscr{E}\delta h/h \approx 1{,}1\cdot10^6$ дин/см$^2 \approx 1{,}1$ атм.

Термодинамические рассуждения, изложенные выше, проведены в предположении, что температура остается постоянной. Поэтому пьезоэлектрический модуль d_{11} может быть охарактеризован как *изотермический модуль.* Нетрудно видеть, как следует изменить эти рассуждения применительно к адиабатическим процессам. Формулы (37.1), (37.3) и (37.4) остаются верными и для таких процессов. Только изотермический пьезоэлектрический модуль d_{11} надо заменить адиабатическим.

7. Что касается связи между направлениями происходящих изменений в прямом и обратном пьезоэлектрических эффектах, то здесь применим общий принцип Ле Шателье (см. т. II, § 51), как в этом нетрудно убедиться с помощью формул (37.1), (37.3) и (37.4). Например, при растяжении пластинки вдоль оси X (см. рис. 98) или сжатии вдоль оси Y на ее нижней поверхности, как мы видели, возбуждается положительный заряд, а на верхней — отрицательный ($P_x > 0$). Иными словами, в пластинке появляется электрическое поле, направленное вверх ($E_x < 0$). Согласно принципу Ле Шателье появление такого поля можно рассматривать как противодействие системы приложенным растягивающим и сжимающим силам. Это противодействие проявляется в том, что возникают силы, стремящиеся сжать пластинку в направлении оси X и растянуть в направлении оси Y. Если поле E_x усилить, то увеличатся и противодействующие силы. Они появятся и в недеформированной пластинке при внесении ее в электрическое поле. Если электрическое поле направлено вверх ($E_x < 0$), то в направлении оси X пластинка сожмется, а в направлении оси Y — удлинится. Это находится в согласии с формулами (37.3) и (37.4). Так же можно рассуждать и в остальных случаях.

8. Физический механизм обратного пьезоэлектрического эффекта можно разъяснить на той же модели, которая применялась при рассмотрении прямого эффекта. Если, например, на поверхности A и B (см. рис. 97 *б*) нанести электрические заряды указанных знаков, то ион кремния *3* притянется к поверхности A, а ион кислорода *4* — к поверхности B, в результате чего ячейка вытянется в направлении оси X. Ионы кремния *1* и *2* будут отталкиваться от поверхности B, а ионы кислорода — от поверхности A, смещаясь при этом внутрь увеличившегося зазора между ионами *3* и *4*. Это приведет к сжатию ячейки в поперечном направлении (вдоль механической оси Y).

Обратный пьезоэлектрический эффект имеет внешнее сходство с *электрострикцией* (§ 33). Однако между этими двумя явлениями имеется и существенное различие. Электрострикция имеет место *во всех* диэлектриках при помещении их в *неоднородное* электрическое поле. Обратный пьезоэлектрический эффект наблюдается *только в кристаллах*, да и то не во всех. Он существует *и в однородных* электрических полях. Силы электрострикции возникают в результате действия электрического поля на поляризованный диэлектрик, поляризация которого обусловлена тем же полем. Поэтому электрострикционные силы *квадратичны по полю.* Они не меняются при изменении направления электрического поля на противоположное. Напротив, обратный пьезоэлектрический эффект возникает в результате действия *внешнего электрического поля* на уже имеющиеся противоположно заряженные ионные решетки кристалла. Возникающие здесь силы *линейны по полю.* Они меняют свои направления на противоположные при изменении знака электрического поля.

9. В различных кристаллах пьезоэлектрический эффект может возникать не только под действием нормальных сил давления или натяжения, но и под действием *касательных сил.* Внутреннее состояние упругих напряжений кристалла характеризуется симметричным тензором упругих натяжений (см. т. I, § 74):

$$\tau_{xx},\ \tau_{yy},\ \tau_{zz},\ \tau_{yz} = \tau_{zy},\ \tau_{zx} = \tau_{xz},\ \tau_{xy} = \tau_{yx}$$

(первый индекс указывает направление внешней нормали площадки, к которой приложена сила натяжения, а второй — направление координатной оси, на которую проецируется эта сила). Для сокращения записи компоненты тензора натяжений принято нумеровать одним индексом, полагая

$$\tau_1 \equiv \tau_{xx}, \quad \tau_2 \equiv \tau_{yy}, \quad \tau_3 \equiv \tau_{zz},$$

$$\tau_4 \equiv \tau_{yz} = \tau_{zy}, \quad \tau_5 \equiv \tau_{zx} = \tau_{xz}, \quad \tau_6 \equiv \tau_{xy} = \tau_{yx}.$$

Опыт показывает, что в случае малых деформаций между компонентами вектора поляризации $\mathbf{P}$ и компонентами тензора натяжений существует *линейная связь*. Такая зависимость аналогична известному закону Гука и имеет примерно ту же область применимости. Таким образом, в общем случае можно написать

$$\begin{aligned} P_x &= d_{11}\tau_1 + d_{12}\tau_2 + d_{13}\tau_3 + d_{14}\tau_4 + d_{15}\tau_5 + d_{16}\tau_6, \\ P_y &= d_{21}\tau_1 + d_{22}\tau_2 + d_{23}\tau_3 + d_{24}\tau_4 + d_{25}\tau_5 + d_{26}\tau_6, \\ P_z &= d_{31}\tau_1 + d_{32}\tau_2 + d_{33}\tau_3 + d_{34}\tau_4 + d_{35}\tau_5 + d_{36}\tau_6. \end{aligned} \tag{37.5}$$

Отсюда видно, что в общем случае пьезоэлектрические свойства кристалла характеризуются *восемнадцатью постоянными*. Эти постоянные называются *пьезоэлектрическими модулями*. Впрочем, число независимых пьезоэлектрических модулей уменьшается из-за симметрии кристалла. Чем выше симметрия кристалла, тем меньше число независимых пьезоэлектрических модулей, которыми он характеризуется. Так, в случае кварца $d_{12} = -d_{11}$, $d_{25} = -d_{14}$, $d_{26} = -2d_{11}$, а все остальные пьезоэлектрические модули обращаются в нуль. Таким образом, пьезоэлектрические свойства кварца характеризуются только двумя модулями, за которые можно принять d_{11} и d_{14}. Тогда

$$\begin{aligned} P_x &= d_{11}\tau_1 - d_{11}\tau_2 + d_{14}\tau_4, \\ P_y &= -d_{14}\tau_5 - 2d_{11}\tau_6, \\ P_z &= 0. \end{aligned} \tag{37.6}$$

При этом $d_{14} = -2{,}0 \cdot 10^{-8}$ дин$^{-1/2}\cdot$ см. Числовое значение модуля d_{11} было приведено выше.

10. Известны сотни веществ, которые в принципе могли бы быть использованы для практического применения пьезоэлектричества. Однако дополнительные требования (большая величина пьезоэффекта, механическая и электрическая прочность, устойчивость к влаге и пр.) резко ограничивают список практически пригодных кристаллов. Из них на первом месте стоит *кварц*. Он превосходный изолятор, поэтому в нем можно возбуждать сильные поля, порядка 30 000 В/см. Научно-технические применения пьезоэлектрического эффекта (прямого и обратного) весьма многочисленны и разнообразны. Не имея возможности останавливаться на этой стороне вопроса, укажем на *пьезоэлектрический манометр*, широко применяющийся для измерения быстропеременных давлений. В этом приборе кварцевая пластинка, вырезанная определенным образом, помещается внутри исследуемого газа. О давлении газа судят по величине пьезоэлектрических зарядов, появляющихся на пластинке. Укажем далее на разнообразнейшие *пьезоэлектрические преобразователи*: *пьезоэлектрические стабилизаторы* и *фильтры* в радиотехнике, *пьезоэлектрические датчики* в автоматике и телемеханике, *виброметры*, *звукосниматели* в технике звукозаписи, *микрофоны*, *телефоны*, *гидрофоны* в акустике и т. д. Особо важное значение имеют *кварцевые излучатели ультразвука*, предложенные во время первой мировой войны французским физиком Ланжевеном (1872–1946). Смещения, возникающие в кварцевой пластинке при наложении на нее статического электрического поля, ничтожны. Однако их можно увеличить в тысячи, а энергию колебаний — в миллионы раз, если воспользоваться переменным электрическим полем. Для этого следует использовать явление резонанса, т. е. подобрать частоту наложенного электрического поля равной одной из собственных частот механических колебаний кварца. Собственные частоты кварца определяются

соотношением

$$h = \frac{\lambda}{2} n,$$

где λ — длина ультразвуковой волны в кварце, а n — целое число. При $n = 1$ получается основное колебание пластинки, при $n = 2, 3, 4, \ldots$ — соответствующие ему обертоны. При резонансной частоте электрического поля кварцевая пластинка, как показал Ланжевен, является мощным источником ультразвука. Такие источники ультразвука и были предложены Ланжевеном для измерения морских глубин и подводной сигнализации. С этого времени началось бурное развитие практических применений пьезоэлектричества.

§ 38. Пироэлектричество

1. У некоторых пьезоэлектрических кристаллов решетка положительных ионов в состоянии термодинамического равновесия смещена относительно решетки отрицательных ионов таким образом, что кристаллы оказываются *электрически поляризованными* даже в отсутствие электрического поля. Такая поляризация называется *спонтанной*, а кристаллы, у которых она наблюдается, — *пироэлектрическими*. Обычно эффект спонтанной поляризации замаскирован свободными поверхностными зарядами, появляющимися в результате оседания ионов из воздуха и их последующего распространения по поверхности кристалла, обладающего хотя и ничтожной, но все же конечной электропроводностью. Однако при нагревании ионные решетки, из которых построен кристалл, смещаются одна относительно другой, вследствие чего на поверхности кристалла появляются электрические заряды противоположных знаков. Возникновение таких зарядов называется *прямым пироэлектрическим эффектом*. Вещества, в которых наблюдается этот эффект, называются *пироэлектриками* («пир» — огонь).

К наиболее известным пироэлектрикам относится турмалин. Кристалл турмалина, помещенный в горячий пепел, сначала притягивает, а затем отталкивает его. В Индии и на Цейлоне это явление, по-видимому, было известно с незапамятных времен. В Европе о нем узнали впервые около 1703 г., когда голландские купцы привезли с Цейлона турмалин.

Для обнаружения пироэлектрического эффекта Август Кундт (1839–1894) предложил погружать кристалл турмалина в смесь порошков сурика и серы. В результате трения друг о друга при перемешивании сурик электризуется положительно, а сера — отрицательно. Поэтому красный сурик притягивается к тем плоскостям кристалла турмалина, которые при нагревании зарядились отрицательно, а желтая сера — к плоскостям, зарядившимся положительно. Заряженные участки располагаются на противоположных концах кристалла. Заряды обнаруживаются также всюду, где на гранях кристалла имеются трещины, происходящие от местных напряжений. В том, что турмалин поляризован спонтанно и при обычной температуре, можно убедиться, разломив кристалл, — окажется, что плоскости излома всегда заряжены электричеством. Фохт (1850–1919) погружал два куска, на которые только что был разломан кристалл турмалина, в ртутные чашки, соединенные с гальванометром. По количеству электричества, прошедшего через гальванометр, можно было оценить спонтанную поляризацию P кристалла. Фохт нашел, что нижний предел P при 24 °C составляет 33 СГСЭ-ед. Внешнее электрическое поле лишь очень слабо меняет поляризацию пироэлектрического кристалла. Последний уже в отсутствие электрического поля поляризован практически до насыщения.

Длительные колебания температуры комнатного воздуха, даже если они небольшие, могут повести к появлению пыльных фигур вблизи лежащего

кристалла турмалина. Так, если кристалл турмалина положить на лист белой бумаги в том месте комнаты, где воздух застаивается, подвергаясь, однако, небольшим колебаниям температуры, то через несколько месяцев на бумаге у концов кристалла появляются тусклые грязные пятна. Дело в том, что под действием электрических сил частицы пыли устремляются к тем концам кристалла турмалина, где поле наиболее сильное, и оседают на бумагу вблизи этих концов.

При изменении температуры меняется объем кристалла, т. е. возникает его деформация. Такая деформация обычно сопровождается появлением пьезоэлектрических зарядов. Поэтому для наблюдения пироэлектричества в чистом виде надо производить нагревание при постоянном объеме и неизменной форме кристалла. Если при этом нет градиента температуры, то возникающая при таком нагревании поляризация кристалла называется *первичным* (или *истинным*) *пироэлектрическим эффектом.* Электризация же, возникающая при деформации кристалла в результате его нагревания, называется *вторичным пироэлектрическим эффектом.* Обычно вторичный эффект превосходит первичный. В некоторых кристаллах первичный эффект настолько мал, что его не удается даже обнаружить. Наконец, заряды могут возникать при неоднородном нагревании, т. е. при наличии в кристалле градиента температуры. Это явление называют *третичным* (или *ложным*) *пироэлектрическим эффектом.* Дело в том, что неравномерное нагревание кристалла создает градиент температуры, который из-за теплового расширения вызывает появление неоднородных напряжений и деформаций. В результате этого появляются пьезоэлектрические заряды, которые (если опыт выполнен недостаточно тщательно) можно ошибочно принять за первичное или вторичное пироэлектричество.

2. Для возможности пироэлектрического эффекта, т. е. спонтанной поляризации кристалла, отсутствия одного только центра симметрии недостаточно. Действительно, в пироэлектрическом кристалле уже в отсутствие электрического поля должно существовать какое-то одно особенное направление, вдоль которого и устанавливается спонтанная поляризация. Такого особенного направления нет, например, в кристалле кварца. Все три электрических оси кварца X_1, X_2, X_3 (см. рис. 95) совершенно равноправны, и поэтому ни одна из них не выделяет особенное направление. Таким образом, *не всякий пьезоэлектрический кристалл является пироэлектрическим. Однако всякий пироэлектрический кристалл является также и пьезоэлектрическим.* Пироэлектрическими свойствами обладают *десять* из тридцати двух кристаллографических классов.

3. Помимо прямого существует *обратный пироэлектрический эффект.* Он состоит в том, что изменение электрического поля в отсутствие подвода или отвода тепла (адиабатический процесс) сопровождается изменением температуры пироэлектрического кристалла. Необходимость этого эффекта является следствием прямого пироэлектрического эффекта и законов термодинамики. Действительно, напишем термодинамическое соотношение $du = T\,ds + E\,dP$ для единицы объема кристалла, предполагая, что во время процесса объем v сохраняется неизменным. Из этого соотношения следует

$$d(u - Ts - EP) = -s\,dT - P\,dE.$$

Отсюда

$$\left(\frac{\partial s}{\partial E}\right)_T = \left(\frac{\partial P}{\partial T}\right)_E.$$

Между величинами s, E, T при постоянном объеме имеется функциональная связь, а потому

$$\left(\frac{\partial s}{\partial E}\right)_T = -\left(\frac{\partial s}{\partial T}\right)_E \left(\frac{\partial T}{\partial E}\right)_s .$$

Введем далее удельную теплоемкость вещества при постоянных объеме и напряженности электрического поля:

$$c_{v,E} = T\left(\frac{\partial s}{\partial T}\right)_{v,E} .$$

Тогда

$$\left(\frac{\partial T}{\partial E}\right)_{v,s} = -\frac{T}{c_{v,E}}\left(\frac{\partial P}{\partial T}\right)_{v,E} . \tag{38.1}$$

Эта формула и устанавливает связь между прямым и обратным пироэлектрическими эффектами.

§ 39. Сегнетоэлектричество

1. Некоторые диэлектрические кристаллы в определенной области температур, называемой *полярной областью*, являются пироэлектриками, т. е. спонтанно поляризованы уже в отсутствие электрического поля. Однако на границах этой температурной области они испытывают фазовые превращения, переходя в новые кристаллические модификации, в которых спонтанная поляризация не наблюдается. Такие диэлектрики называются *сегнетоэлектриками.* От обычных пироэлектриков сегнетоэлектрики отличаются еще тем, что направление спонтанной поляризации в сегнетоэлектрике может быть изменено на противоположное уже сравнительно слабым электрическим полем, тогда как у обычных пироэлектриков этого не происходит даже в сильных полях. Прямая, параллельная вектору спонтанной поляризации сегнетоэлектрика, называется его *полярной осью.* Существуют сегнетоэлектрики с одной полярной осью (например, *сегнетова соль*) и с несколькими полярными осями (например, *титанат бария*).

Кристаллическая модификация, в которой сегнетоэлектрик спонтанно поляризован, называется *полярной фазой*, а модификация, в которой спонтанной поляризации нет, — *неполярной фазой.* Температура $T_{\text{К}}$, при которой сегнетоэлектрик переходит из полярной фазы в неполярную (или обратно), называется *диэлектрической точкой Кюри* в честь Пьера Кюри, который ввел аналогичное понятие в учение о ферромагнетизме (см. § 74 и 79). Как правило, сегнетоэлектрик имеет только одну точку Кюри, ниже которой он находится в полярной, а выше — в неполярной фазе. Исключение составляют сегнетова соль и изоморфные с ней соединения, а также соли $Ag_2H_3IO_6$ и $Ag_2D_3IO_6$. Они имеют две точки Кюри: *нижнюю* $T_{\text{н}}$ и *верхнюю* $T_{\text{в}}$. Спонтанная поляризация наблюдается в температурной области, ограниченной этими точками, и не наблюдается при других температурах.

Сегнетоэлектрики характеризуются рядом аномальных диэлектрических и других физических свойств (пьезоэлектрических, электрооптических и пр.), связанных между собой термодинамическими соотношениями. Ранее всего эти аномальные свойства были обнаружены на кристаллах сегнетовой соли. От названия этой соли и происходят термины *сегнетоэлектрик* и *сегнетоэлектричество*. Диэлектрические свойства сегнетоэлектриков во многих отношениях аналогичны магнитным свойствам ферромагнетиков (см. § 74 и 79). Поэтому в иностранной литературе их называют также *ферроэлектриками*. Сегнетова соль $NaKC_4H_4O_6 \cdot 4H_2O$ представляет собой двойную натриевокалиевую соль винной кислоты, содержащую четыре молекулы кристаллизационной воды. У нее две точки Кюри: $T_н = 255$ K (−18 °C) и $T_в = 297$ K (+24 °C). Сегнетоэлектрическими свойствами обладают и другие соли винной кислоты (тартраты), например $NaNH_4C_4H_4O_6 \cdot 4H_2O$ (аммониевая сегнетова соль), $NaRbC_4H_4O_6 \cdot 4H_2O$, $NaTlC_4H_4O_6 \cdot 4H_2O$. Аномально большой пьезоэлектрический эффект в сегнетовой соли был обнаружен братьями Кюри еще в 1880 г. Затем Поккельс (1865–1913) в 1894 г. количественно исследовал этот эффект (а также электрооптический эффект, т. е. зависимость показателей преломления от напряженности электрического поля). Аномальные диэлектрические свойства сегнетовой соли были открыты Валашеком в 1921 г., а затем в начале 30-х годов подробно исследованы И. В. Курчатовым (1903–1960) и П. П. Кобеко (1897–1954). В 1944 г. Б. М. Вул (р. 1903) и И. М. Гольдман (р. 1910) в Советском Союзе и независимо от них Вейнер и Соломон в США и Огава в Японии открыли на керамических образцах аномальные диэлектрические свойства титаната бария ($BaTiO_3$) — важнейшего сегнетоэлектрика, обладающего высокой механической прочностью, большой химической устойчивостью и нашедшего благодаря этому широкое научно-техническое применение (температура Кюри $T_К = 393$ K = 120 °C). С этого момента началось быстрое развитие учения о сегнетоэлектричестве и его применения. Оказалось, что сегнетоэлектричество — более распространенное явление, чем предполагалось раньше. В настоящее время известно более сотни сегнетоэлектриков, не считая огромного количества сегнетоэлектрических твердых растворов. Учение о сегнетоэлектрических явлениях представляет обширный, быстро развивающийся раздел физики твердого тела. В общем курсе о сегнетоэлектричестве можно дать лишь весьма краткие сведения.

2. Будем иллюстрировать поведение сегнетоэлектриков на примере сегнетовой соли и титаната бария. В неполярной фазе (при $T < T_н$ и $T > T_в$) кристаллы сегнетовой соли относятся к *ромбической системе* и являются пьезоэлектриками. Форма ромбического кристалла изображена на схематическом рис. 100. Наиболее развиты и типичны грани (001), а также призматические грани (110). Менее развиты грани (210), (120), (010). Грани (100) в большинстве случаев очень малы или совсем отсутствуют. В полярной фазе сегнетова соль принадлежит к

моноклинной системе, причем полярная ось a [100] направлена параллельно исходной оси a ромбического кристалла. (Вообще, полярная фаза сегнетоэлектрика всегда характеризуется меньшей степенью симметрии, чем неполярная.) Аномальные диэлектрические свойства сегнетовой соли выражены максимально, когда поле **E** направлено по оси a. Аномалии не наблюдаются, когда направление **E** параллельно осям b и c.

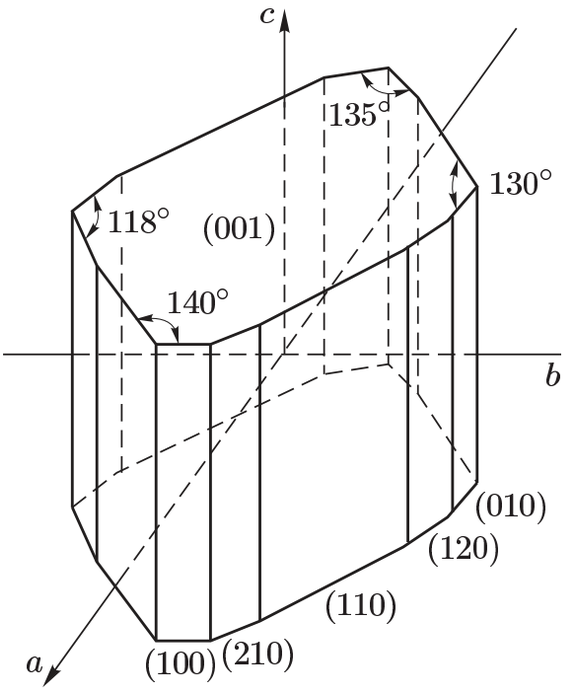


Рис. 100

3. Титанат бария имеет наиболее простую кристаллическую структуру по сравнению со всеми известными сегнетоэлектриками. В неполярной фазе выше 120 °C это есть так называемая *кубическая структура типа перовскита*, приведенная на рис. 101 (такой структурой обладает минерал перовскит $CaTiO_3$, откуда и произошло название). Ввиду наличия центра симметрии титанат бария в неполярной фазе не обладает пьезоэлектрическими свойствами. В полярной области температур между точкой Кюри (120 °C) и температурой 5 °C кристаллы $BaTiO_3$ имеют *тетрагональную симметрию* и становятся пьезоэлектрическими. Фазовый переход при температуре 120 °C сводится к тому, что одно из ребер кубической ячейки удлиняется и становится *полярной тетрагональной осью симметрии*, обозначаемой через c, два других ребра одинаково укорачиваются, переходя в *тетрагональные оси*, обозначаемые через a. Какое из ребер исходной кубической ячейки удлинится и перейдет в полярную ось c — это дело случая. Однако если в результате флуктуации возникнет какое-то случайное удлинение, то оно определит *выделенное направление*, вдоль которого и будет происходить дальнейшее удлинение. Поскольку все три ребра кубической ячейки эквивалентны, каждое из них может перейти в полярную ось. В тетрагональной фазе существует, следовательно, *шесть* возможных направлений спонтанной поляризации — по два взаимно противоположных направления вдоль ребер

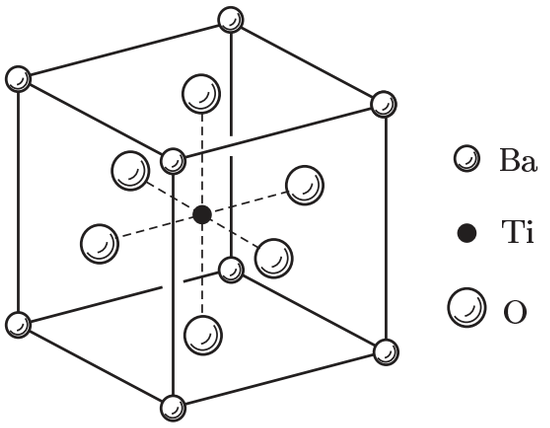


Рис. 101

кубической ячейки. Ниже 5 °C титанат бария испытывает второе *фазовое превращение.* Получается новая сегнетоэлектрическая фаза, устойчивая между 5 и −90 °C и обладающая *орторомбической симметрией.* Элементарная ячейка может быть получена из исходной кубической ячейки, если ее растянуть вдоль диагонали одной из граней куба и сжать вдоль другой диагонали той же грани. Растянутая диагональ служит полярной осью кристалла. Число граней — шесть, число их диагоналей — двенадцать. Однако эти диагонали попарно параллельны. Поэтому в орторомбической фазе существует *двенадцать направлений,* вдоль которых может ориентироваться вектор спонтанной поляризации кристалла. При −90 °C происходит третий фазовый переход. Кристалл становится *ромбоэдрическим* с полярной осью вдоль одной из пространственных диагоналей куба, т. е. диагоналей, соединяющих его противоположные вершины. Так как исходная кубическая ячейка содержит четыре эквивалентных пространственных диагонали и каждой диагонали соответствуют два взаимно противоположных направления спонтанной поляризации, то в ромбоэдрической фазе существует *восемь направлений,* в которых может ориентироваться вектор спонтанной поляризации.

4. Спонтанная поляризация сегнетоэлектрика $P_с$ меняется с температурой, обращаясь в нуль на границах полярной области. Для сегнетовой соли $P_с$ достигает максимума при 5 °C и составляет $P_с = 7{,}5 \cdot 10^2$ СГСЭ-ед. $= 2{,}5 \cdot 10^{-7}$ Кл/см^2. Для титаната бария спонтанная поляризация вдоль полярной оси на два порядка больше, при комнатной температуре она составляет $P_с = 7{,}8 \cdot 10^4$ СГСЭ-ед. $= 2{,}6 \cdot 10^{-5}$ Кл/см^2.

При наложении электрического поля поляризация сегнетоэлектрика меняется. В полярной фазе она складывается из спонтанной поляризации $\mathbf{P}_с$, не зависящей от поля $\mathbf{E}$, и индуцированной поляризации $\mathbf{P}_и$, вызванной этим полем. Связь между $\mathbf{P}$ и $\mathbf{E}$ нелинейна, так что обычное определение поляризуемости α и диэлектрической проницаемости ε к сегнетоэлектрику в полярной фазе неприменимо. Только в слабых полях можно ограничиться линейным приближением, полагая $\mathbf{P} = \mathbf{P}_с + \alpha\mathbf{E}$ или $\mathbf{P}_и = \alpha\mathbf{E}$. Таким образом, обычная связь между поляризацией и полем $\mathbf{P} = \alpha\mathbf{E}$ в случае сегнетоэлектриков в слабых полях относится не к полной, а к *индуцированной поляризации.* Коэффициент α и называется *поляризуемостью* (или *диэлектрической восприимчивостью*) сегнетоэлектрика. Из-за анизотропии кристалла поляризуемость α есть *тензор.* Для простоты мы будем пренебрегать анизотропией. Это можно делать, когда поле $\mathbf{E}$ параллельно одной из трех главных осей тензора α. В этом случае тензорные свойства величины α не проявляются и ее можно считать *скаляром.* В сильных полях поляризуемость α определяется производной $\alpha = \partial P_и / \partial E$ и зависит от напряженности поля. Диэлектрическая проницаемость, как всегда, связана с α соотношением $\varepsilon = 1 + 4\pi\alpha$. Ниже, когда приводятся числовые значения ε и α, всегда предполагается, что эти величины измерены в слабых полях.

5. Одна из характерных особенностей сегнетоэлектриков состоит в том, что значения диэлектрической проницаемости в полярной фазе *аномально велики.* Для сегнетовой соли в максимуме $\varepsilon \approx 10\,000$, для титаната бария $\varepsilon_{\text{макс}} \approx 6000$–$7000$.

В неполярной фазе сегнетоэлектрик ведет себя как обычный *линейный диэлектрик*, в котором поляризация пропорциональна электрическому полю. Однако поляризуемость α и диэлектрическая проницаемость ε меняются с температурой. Вблизи точки Кюри имеет место *закон Кюри–Вейсса* (1865–1940), т. е. соотношение

$$\alpha = \frac{C}{T - T_0}, \tag{39.1}$$

где C и T_0 — постоянные, из которых T_0 называется *температурой Кюри–Вейсса.* Она очень мало отличается от температуры Кюри $T_{\text{К}}$, при которой происходит фазовый переход из полярной фазы в неполярную (или обратно); этим различием обычно пренебрегают. Если точек Кюри две, то в окрестности каждой из них в неполярной фазе выполняется *закон Кюри–Вейсса*: вблизи верхней точки — в прежней форме (39.1), вблизи нижней — в форме

$$\alpha = \frac{C'}{T_0 - T} \tag{39.2}$$

(конечно, с другими постоянными C' и T_0). Мы видим, что и в неполярной фазе в окрестности точки Кюри величины ε и α также аномально велики.

На рис. 102 представлена температурная зависимость диэлектрической проницаемости ε_a монокристалла сегнетовой соли вдоль полярной

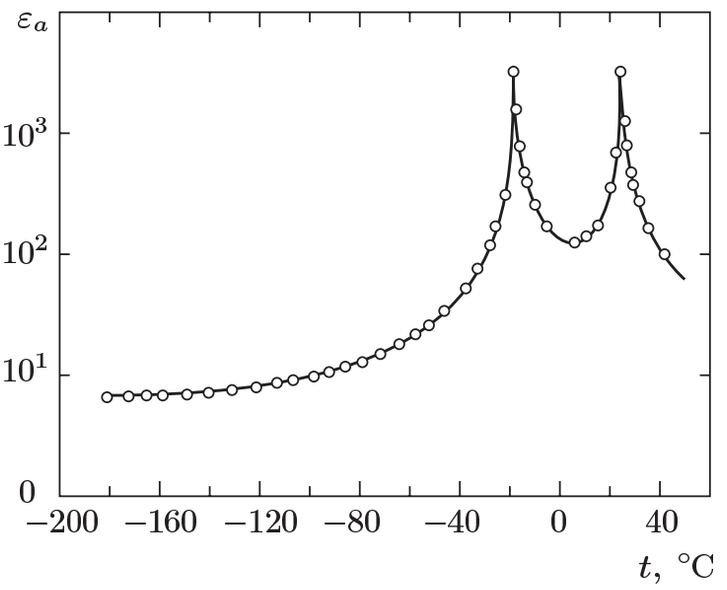


Рис. 102

оси a (по измерениям Хаблютцеля). Аналогичная кривая для титаната бария (по измерениям Мерца) представлена на рис. 103. У титаната

бария в тетрагональной фазе два главных значения диэлектрической проницаемости: ε_c вдоль полярной оси и ε_a вдоль перпендикулярной к ней оси. Они и приведены на графике для однодоменного кристалла

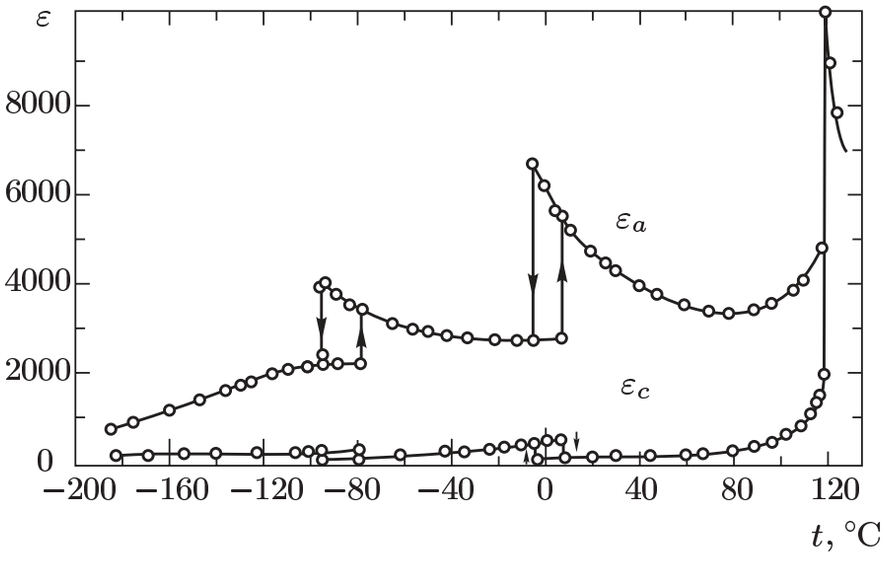


Рис. 103

(см. ниже). В орторомбической и тетрагональной фазах ε измерена в тех же направлениях. В этих фазах кристалл разделяется на *домены* (см. ниже). С разделением на домены связано то обстоятельство, что кривые, снятые при возрастании и убывании электрического поля, в окрестности точек фазовых превращений не совпадают между собой.

6. Спонтанно поляризованный сегнетоэлектрик в полярной области температур должен обладать *большей устойчивостью*, чем тот же сегнетоэлектрик в неполяризованном состоянии. В противном случае в этой области температур должен был бы происходить самопроизвольный переход сегнетоэлектрика в неполяризованную фазу, чего на самом деле не происходит. Поэтому к проблеме сегнетоэлектричества применимы *общие принципы термодинамического равновесия* — условия минимума энергии или какой-либо другой термодинамической функции в зависимости от того, какие термодинамические параметры поддерживаются постоянными. Из этих принципов следует, что уже в отсутствие внешнего электрического поля сегнетоэлектрик, если только размеры его достаточно велики, должен *самопроизвольно* делиться на области с различными направлениями вектора поляризации, называемые *доменами* (точнее, *диэлектрическими доменами*). Действительно, поляризованный кристалл создает электрическое поле не только внутри, но и во внешнем пространстве. Внешнему полю соответствует определенная электрическая энергия. Полная энергия кристалла складывается из трех частей: 1) суммы внутренних энергий всех доменов; 2) энергии внешнего электрического поля; 3) поверхностной энергии на границах раздела доменов. Энергия внешнего электрического поля максимальна, когда кристалл по всему объему поляризован *однородно*. Поэтому

деление на домены может оказаться «энергетически выгодным», т. е. сопровождаться уменьшением полной энергии. Однако процесс деления сопровождается увеличением поверхностной энергии. Он прекратится, когда полная энергия достигнет минимума.

Если сегнетоэлектрик имеет *одну полярную ось*, то возможны только два взаимно противоположных направления поляризации доменов, параллельные этой оси. Такие сегнетоэлектрики имеют *слоистую доменную структуру.* В сегнетоэлектриках с *несколькими полярными осями* доменная структура более разнообразна. Число возможных направлений поляризации доменов в таких кристаллах вдвое больше числа полярных осей.

Экспериментальные доказательства доменной структуры сегнетоэлектриков сначала были косвенными. Прямое доказательство дали наблюдения кристаллов сегнетоэлектрика в *поляризованном свете.* Другой метод основан на *травлении* поверхности сегнетоэлектрика. Например, положительный край домена $BaTiO_3$ при травлении кислотой разрушается сильнее, чем отрицательный.

7. Благодаря доменной структуре дипольный момент кристалла сегнетоэлектрика в отсутствие внешнего электрического поля равен нулю, так как поляризация одних доменов компенсируется противоположно направленной поляризацией других. В этом смысле такой кристалл не поляризован. При наложении электрического поля происходит частичная переориентация доменов, а также рост одних доменов за счет других. Это ведет к появлению в кристалле поляризации $\mathbf{P}$. Зависимость P от напряженности электрического поля E представлена на рис. 104. Сначала рост P происходит вдоль кривой OA. В точке A поляризация всех доменов оказывается ориентированной вдоль поля $\mathbf{E}$. Начиная с этой точки, дальнейшее возрастание P происходит за счет индуцированной поляризации $\mathbf{P}_{\text{и}} = \alpha\mathbf{E}$, и кривая OA переходит в прямолинейный участок AD. Если этот участок продолжить влево, то он отсечет на оси ординат отрезок OC, длина которого равна спонтанной поляризации $P_{\text{с}}$.

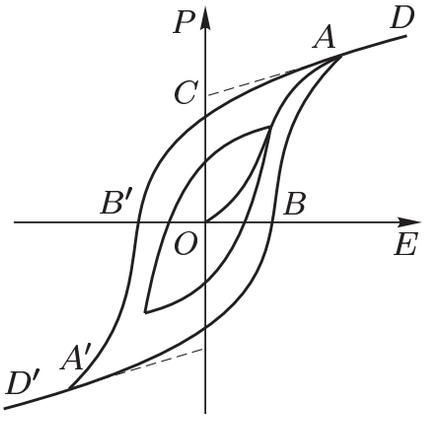


Рис. 104

Будем теперь уменьшать напряженность электрического поля. Оказывается, что изменение поляризации P пойдет не по прежней кривой DAO, а по новой кривой $DAB'A'D'$, расположенной выше. Это явление называется *диэлектрическим гистерезисом* и связано с *доменной структурой диэлектрика*: процесс переориентации и роста доменов в электрическом поле задерживается, напоминая известное явление застоя, обусловленное силами сухого трения (см. т. I, § 17). Таким образом, поляризация $\mathbf{P}$ не определяется однозначно полем $\mathbf{E}$, а зависит также

от *предшествующей истории* сегнетоэлектрика. Если менять электрическое поле в обратном порядке, то зависимость P от E изобразится нижней кривой $D'A'BAD$, симметричной с кривой $D'A'B'AD$ относительно начала координат O. Таким образом, получается замкнутая кривая $AB'A'BA$, называемая *диэлектрической петлей гистерезиса*. Можно получить петли гистерезиса меньших размеров, одна из которых изображена на рис. 104. Совершенно аналогичные петли гистерезиса получатся, если по оси ординат откладывать индукцию $D = E + 4\pi P$. Они практически не отличаются (точнее, отличаются только масштабом) от кривых $P = P(E)$, так как в сегнетоэлектриках $E \ll D$ и слагаемым E можно пренебречь.

8. Петлю гистерезиса легко воспроизвести на осциллографе. С этой целью два конденсатора C_0 и C_x соединяются последовательно и питаются переменным током от одного и того же генератора (рис. 105). Конденсатор C_0 заполнен обычным «линейным» диэлектриком с постоянной диэлектрической проницаемостью ε_0, а конденсатор C_x — сегнетоэлектриком. Ясно, что напряжение на конденсаторе C_x пропорционально полю E в сегнетоэлектрике. Поскольку конденсаторы соединены последовательно, они имеют одинаковые заряды, а следовательно, и одинаковые индукции: $D_0 = D$. А так как величина ε_0 постоянна, то напряжение на конденсаторе C_0 будет пропорционально индукции $D_0 = D$. Если подать на горизонтально отклоняющие пластины осциллографа напряжение с конденсатора C_x, а на вертикально отклоняющие — с конденсатора C_0, то на экране осциллографа будет воспроизведена зависимость $D = D(E)$, т. е. петля гистерезиса.

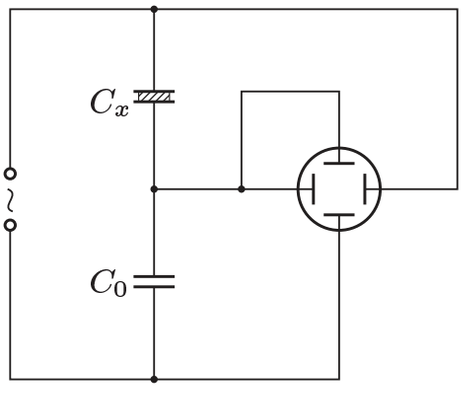


Рис. 105

9. Мы не будем излагать микроскопические теории явления сегнетоэлектричества, которые еще далеки до окончательного завершения. Ограничимся только следующим замечанием. Предположим, что к диэлектрику с индуцированными диполями применимо выражение для действующего поля $\mathbf{E}' = \mathbf{E} + a\mathbf{P}$, где a — постоянная. (Для кубических кристаллов, построенных из точечных изотропных атомов, согласно Лоренцу, $a = 4\pi/3$, см. § 35.) Пусть β — поляризуемость молекулы, а N — число молекул в единице объема. Тогда $N\beta\mathbf{E}' = \mathbf{P} = N\beta\mathbf{E} + Na\beta\mathbf{P}$, или

$$(1 - Na\beta)\mathbf{P} = N\beta\mathbf{E}.$$

Если $Na\beta = 1$, то при $\mathbf{E} = 0$ это уравнение имеет два решения: 1) $P = 0$; 2) $P \neq 0$. Из соответствующих им двух состояний в действительности реализуется то, которое термодинамически более устойчиво. Пусть таковым является второе состояние. Тогда диэлектрик будет

поляризован спонтанно: если сначала $\mathbf{P} = 0$, то всякая флуктуация поляризации $\mathbf{P}$ переведет диэлектрик в более устойчивое состояние, в котором $\mathbf{P} \neq 0$. Величина $Na\beta$ зависит от температуры (хотя бы из-за теплового расширения диэлектрика). Условие $Na\beta = 1$ определяет положение точки Кюри. В окрестности этой точки поляризуемость диэлектрика $\alpha = N\beta/(1 - Na\beta)$ будет аномально велика. Разлагая знаменатель $1 - Na\beta$ в ряд по степеням $T - T_{\mathrm{K}}$ и ограничиваясь при этом линейным членом, придем к соотношению $\alpha = \mathrm{const}/(T - T_{\mathrm{K}})$, т. е. к закону Кюри–Вейсса. На этот пример нельзя смотреть как на «теорию сегнетоэлектричества». Он указывает лишь на *принципиальную возможность* объяснения спонтанной поляризации и связанной с ней аномально большой величины ε в окрестности точки Кюри.

10. Более определенными по сравнению с молекулярными теориями сегнетоэлектричества являются *термодинамические теории.* Точнее их следовало бы назвать *феноменологическими теориями,* поскольку наряду с термодинамикой в их основе лежат некоторые предположения феноменологического характера. Феноменологические теории, разумеется, не разъясняют физическую природу явления сегнетоэлектричества. Они позволяют только установить связь между различными физическими величинами, характеризующими сегнетоэлектрики. Одним из пионеров в создании термодинамической теории сегнетоэлектричества был В.Л. Гинзбург (р. 1916). Ниже приводится краткое изложение теории Гинзбурга.

Термодинамическое соотношение для равновесных процессов в диэлектрике при наличии электрического поля E запишем в виде

$$dU = T\,dS - \mathscr{P}\,dV + E\,dP. \tag{39.3}$$

(Мы пренебрегаем анизотропией диэлектрика и предполагаем, что механических напряжений нет. Вместо $E\,dP$ можно было бы писать $E\,dD/4\pi$, как это подробно было разъяснено в § 31.) Если ввести термодинамический потенциал $\Phi = U - TS + PV - EP$, то это соотношение перейдет в

$$d\Phi = -S\,dT + V\,d\mathscr{P} - P\,dE. \tag{39.4}$$

Независимыми переменными, от которых зависит функция Φ, являются температура T, давление $\mathscr{P}$ и напряженность электрического поля E. Заданием этих величин определяются все прочие внутренние параметры, характеризующие диэлектрик в состоянии термодинамического равновесия. Однако для наших целей необходимо ввести некоторые отклонения от термодинамического равновесия и принять, что поляризация P может меняться *независимо.* Тогда процесс становится *неравновесным* и в соотношении (39.3) вместо знака равенства надо поставить знак $\leqslant 0$. Равенство (39.4) при этом перейдет в неравенство

$$d\Phi \leqslant -S\,dT + V\,d\mathscr{P} - P\,dE.$$

Если величины T, $\mathscr{P}$, E постоянны, то $d\Phi \leqslant 0$. Отсюда следует, что *в состоянии термодинамического равновесия термодинамический потенциал минимален* (*при условии, что* T, $\mathscr{P}$, E *поддерживаются постоянными*). Из условия минимума Φ и найдется равновесное значение поляризации P.

11. После этого термодинамического отступления перейдем к изложению теории Гинзбурга. Основным в теории Гинзбурга является предположение, что *переход из полярной фазы сегнетоэлектрика в неполярную* (*или обратно*) *есть фазовый переход второго рода или близкий к нему переход первого*

рода (см. т. II, § 120). Гинзбург применил к сегнетоэлектрикам основные положения теории фазовых переходов второго рода, развитой Л.Д. Ландау (1908–1968). Допустим сначала, что электрического поля нет ($E = 0$). В полярной фазе $P \neq 0$, в неполярной $P = 0$. Допустим, что в точке Кюри поляризация P меняется непрерывно с температурой и что в окрестности этой точки термодинамический потенциал Φ можно разложить в степенной ряд по P. Это допущение является слабым пунктом теории Гинзбурга, как и аналогичное допущение в общей теории фазовых переходов второго рода Ландау. Но если разложение возможно, то оно должно содержать только четные степени P, так как при изменении направления вектора $\mathbf{P}$ на противоположное в отсутствие электрического поля потенциал Φ не должен меняться. Имея это в виду и оборвав разложение на четных степенях, напишем

$$\Phi = \Phi_0 + AP^2 + \frac{1}{2}\,BP^4, \tag{39.5}$$

где Φ_0, A, B — функции температуры и давления. Равновесная поляризация P найдется из условия минимума Φ. Это приводит к соотношениям

$$\frac{\partial \Phi}{\partial P} = 2(AP + BP^3) = 0, \tag{39.6}$$

$$\frac{\partial^2 \Phi}{\partial P^2} = 2(A + 3BP^2) > 0. \tag{39.7}$$

Уравнение (39.6) имеет два решения:

1-е решение

$$P = 0 \quad (T \geqslant T_{\text{К}}); \tag{39.8}$$

2-е решение

$$P^2 = -A/B \quad (T < T_{\text{К}}). \tag{39.9}$$

Первое решение относится к неполярной фазе ($T \geqslant T_{\text{К}}$), второе — к полярной ($T \leqslant T_{\text{К}}$). В точке Кюри $P = 0$, а потому коэффициент A должен обращаться в нуль: $A(T_{\text{К}}) = 0$. В неполярной фазе ($T \geqslant T_{\text{К}}$) решение $P = 0$ должно удовлетворять условию (39.7), а потому при $T > T_{\text{К}}$ должно быть $A > 0$. В полярной фазе ($T < T_{\text{К}}$) то же решение ($P = 0$) условию минимума не удовлетворяет, и $A < 0$, а $B > 0$, как это следует из условия (39.7), в котором P^2 — существенно положительная величина. Таким образом, в окрестности точки Кюри A является возрастающей функцией температуры: $(\partial A/\partial T)_{T-T_{\text{К}}} > 0$. В близкой окрестности точки Кюри величину A можно аппроксимировать выражением $A = (\partial A/\partial T)_{T_{\text{К}}}(T - T_{\text{К}})$, а зависимостью B от температуры — пренебречь. Тогда (39.9) приводит к соотношению

$$P^2 = \text{const}\left(\frac{\partial A/\partial T}{B}\right)_{T_{\text{К}}}(T_{\text{К}} - T) \quad (T \leqslant T_{\text{К}}). \tag{39.10}$$

12. Энтропия сегнетоэлектрика согласно формуле (39.4) определяется выражением

$$S = -\frac{\partial \Phi}{\partial T} = -\frac{\partial \Phi_0}{\partial T} - P^2\frac{\partial A}{\partial T} - A\frac{\partial P^2}{\partial T} - \frac{P^4}{2}\,\frac{\partial B}{\partial T} - \frac{B}{2}\,\frac{\partial P^4}{\partial T}.$$

В точке Кюри все слагаемые, за исключением $\partial \Phi_0/\partial T$, обращаются в нуль. Значит, при фазовом переходе изменения энтропии не происходит, а следовательно, не выделяется скрытая теплота, как это и должно быть для фазовых переходов второго рода. Зато имеет место скачок теплоемкости.

Действительно, теплоемкость в точке Кюри в неполярной фазе

$$c_{\mathscr{P}}^{\text{непол}} = T\frac{\partial S}{\partial T} = -T\left(\frac{\partial^2 \Phi_0}{\partial T^2}\right)_{\text{K}},$$

а в полярной

$$c_{\mathscr{P}}^{\text{пол}} = -T\frac{\partial^2 \Phi_0}{\partial T^2} - 2T\frac{\partial P^2}{\partial T}\,\frac{\partial A}{\partial T} = -\left(T\frac{\partial^2 \Phi_0}{\partial T^2}\right)_{\text{K}} + 2\left[\frac{T}{B}\left(\frac{\partial A}{\partial T}\right)^2\right]_{\text{K}}.$$

Отсюда

$$\Delta c_{\mathscr{P}} = c_{\mathscr{P}}^{\text{пол}} - c_{\mathscr{P}}^{\text{непол}} = 2\left[\frac{T}{B}\left(\frac{\partial A}{\partial T}\right)_{\mathscr{P}}^2\right]_{\text{K}}. \tag{39.11}$$

Исключая из формул (39.10) и (39.11) величину B, для спонтанной поляризации $P_{\text{сп}}$ получаем

$$P_{\text{сп}}^2 = \frac{\Delta c_{\mathscr{P}}}{2\left[T\left(\dfrac{\partial A}{\partial T}\right)_{\mathscr{P}}\right]_{\text{K}}}(T - T_{\text{K}}). \tag{39.12}$$

Это соотношение допускает экспериментальную проверку, так как все входящие в него величины могут быть независимо измерены на опыте: $\Delta c_{\mathscr{P}}$ определяется калориметрически, $P_{\text{сп}}$ находим по заряду конденсатора с сегнетоэлектриком, $(\partial A/\partial T)_{\mathscr{P}}$ — по значению диэлектрической проницаемости ε.

13. Перейдем теперь к вычислению поляризуемости α (или диэлектрической проницаемости $\varepsilon = 1+4\pi\alpha$ в слабых электрических полях. С этой целью прежде всего введем новую термодинамическую функцию $\Omega = \Phi + EP$. Для нее, как это непосредственно следует из (39.4),

$$d\Omega = -S\,dT + V\,d\mathscr{P} + E\,dP. \tag{39.13}$$

Принимая за независимые переменные T, $\mathscr{P}$ и P, отсюда получаем

$$E = \frac{\partial \Omega}{\partial P} \equiv \left(\frac{\partial \Omega}{\partial P}\right)_{T,\mathscr{P}}. \tag{39.14}$$

Вблизи точки Кюри, как это видно из (39.5),

$$\Omega = \Phi_0 + AP^2 + \frac{1}{2}\,BP^4. \tag{39.15}$$

Так как Φ_0, A, B — функции только температуры и давления, то

$$E = 2(AP + BP^3).$$

Выше точки Кюри ($T > T_{\text{K}}$) членом BP^3 можно пренебречь. Тогда получается линейная связь между P и E с поляризуемостью $\alpha = 1/(2A)$, или после замены A на $(\partial A/\partial T)_{\text{K}}(T - T_{\text{K}})$

$$\alpha^{\text{непол}} = \frac{1}{2(\partial A/\partial T)_{\text{K}}(T - T_{\text{K}})} \qquad (T > T_{\text{K}}). \tag{39.16}$$

Ниже точки Кюри поляризуемость α определяется производной:

$$\alpha = \frac{\partial P}{\partial E} = \frac{1}{2(A + 3BP^2)}.$$

В этом выражении можно пренебречь индуцированной поляризацией, полагая $P = P_c$. Тогда, ввиду формулы (39.10), получится

$$\alpha^{\text{пол}} = \frac{1}{4(\partial A/\partial T)_{\text{К}}(T_{\text{К}} - T)} \qquad (T < T_{\text{К}}). \tag{39.17}$$

Мы видим, что на одинаковых расстояниях от точки Кюри $\alpha^{\text{непол}} = 2\alpha^{\text{пол}}$.

14. В заключение скажем несколько слов об *антисегнетоэлектричестве.* Это явление имеет ту же физическую природу, что и сегнетоэлектричество, и состоит в следующем. В некоторых кристаллах в определенной области температур рядом стоящие ионы одинакового типа самопроизвольно ориентируются не параллельно друг другу, а *антипараллельно.* Такие кристаллы называются *антисегнетоэлектрическими.* Антисегнетоэлектрик можно представить себе в виде вставленных друг в друга подрешеток, причем в одной подрешетке все дипольные моменты ионов параллельно ориентированы в одном, а в другой — в противоположном направлениях, так что дипольный момент всего кристалла равен нулю. Переход от упорядоченного расположения ионов к неупорядоченному происходит в *антисегнетоэлектрической точке Кюри*: упорядоченная ориентация наблюдается обычно ниже, а неупорядоченная — выше этой точки. В самой точке Кюри наблюдается максимум диэлектрической проницаемости, величина которого меньше, чем у многих сегнетоэлектриков. При наложении электрического поля возникает электрическая поляризация $\mathbf{P}$. Поскольку в отсутствие поля $\mathbf{P} = 0$, в слабых полях зависимость $\mathbf{P}$ от $\mathbf{E}$ практически линейна. Однако, если поле достаточно сильное, антисегнетоэлектрик может перейти в *сегнетоэлектрическое состояние.* При таких «вынужденных» фазовых переходах в сильных полях наблюдаются *двойные петли гистерезиса* (рис. 106).

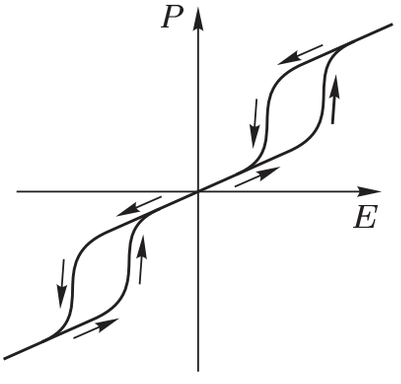


Рис. 106

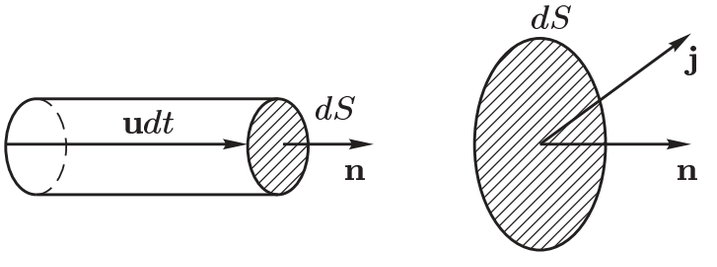
Глава II

ЭЛЕКТРИЧЕСКИЙ ТОК

§ 40. Плотность тока. Закон сохранения электрического заряда

1. Электрический ток есть *упорядоченное движение электрических зарядов.* Эти заряды в учении о токах называются *носителями тока.* В металлах и полупроводниках носителями тока являются *электроны*, в электролитах и ионизационных газах — *положительные* и *отрицательные ионы.*

Рассмотрим сначала простейший случай, когда все носители тока одинаковы (например, электроны в металлах). Выделим мысленно в среде, по которой течет ток, произвольный физически бесконечно малый объем и обозначим через $\mathbf{u}$ средний вектор скорости рассматриваемых носителей в этом объеме. Его называют *средней*, *дрейфовой* или *упорядоченной* скоростью движения носителей тока. Обозначим, далее, через n *концентрацию носителей тока*, т. е. число их в единице объема. Проведем бесконечно малую площадку dS, перпендикулярную к скорости $\mathbf{u}$. Построим на ней бесконечно короткий прямой цилиндр с высотой $u\,dt$, как указано на рис. 107 *а*. Все частицы, заключенные внутри этого цилиндра, за время dt пройдут через площадку dS, перенеся через нее в направлении скорости $\mathbf{u}$ электрический заряд $dq = neu\,dS\,dt$, где e —



а *б*

Рис. 107

заряд отдельной частицы. Таким образом, через единицу площади за единицу времени переносится электрический заряд $j = neu$. Вектор

$$\mathbf{j} = ne\mathbf{u} \tag{40.1}$$

называется *плотностью электрического тока.*

Скаляр j есть заряд, переносимый в единицу времени через единичную площадку, перпендикулярную к току. Направление же вектора $\mathbf{j}$ совпадает с направлением упорядоченного течения положительного электричества.

В случае нескольких типов зарядов, создающих ток, плотность тока определяется выражением

$$\mathbf{j} = \sum_i n_i e_i \mathbf{u}_i, \tag{40.2}$$

где суммирование ведется по всем типам носителей тока (n_i, e_i, $\mathbf{u}_i$ означают концентрацию, заряд и упорядоченную скорость i-го носителя).

Установим произвольно положительное направление нормали к площадке dS и проведем в этом направлении единичный вектор $\mathbf{n}$. Если частицы положительные, то переносимый заряд в направлении нормали $\mathbf{n}$ будет положительным или отрицательным в зависимости от того, движутся ли частицы в направлении вектора $\mathbf{n}$ или в противоположном направлении. Для отрицательных частиц соотношение будет обратным. Вообще, количество переносимого в единицу времени электричества можно записать в виде

$$dq = (\mathbf{j}\,\mathbf{n})\,dS = j_n\,dS. \tag{40.3}$$

Последняя формула остается верной и в том случае, когда площадка dS не перпендикулярна к вектору $\mathbf{j}$ (рис. 107 *б*). Чтобы убедиться в этом, достаточно заметить, что составляющая вектора $\mathbf{j}$, перпендикулярная к вектору $\mathbf{n}$, через площадку dS электричества не переносит.

2. Одним из фундаментальных физических законов является *закон сохранения электрического заряда* (см. § 2). Выразим его математически через макроскопические величины: плотность заряда ρ и плотность электрического тока $\mathbf{j}$. Возьмем в среде произвольную замкнутую поверхность S, ограничивающую объем V (рис. 108). Количество электричества, ежесекундно вытекающее из объема V через поверхность S, представляется интегралом $\oint j_n\,dS$. Ту же величину можно представить в виде $-\partial q/\partial t$, где q — заряд, содержащийся в объеме V. Приравнивая оба выражения, получим

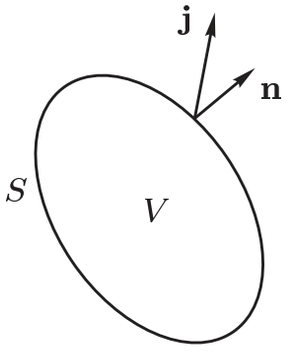


Рис. 108

$$\frac{\partial q}{\partial t} = -\oint\limits_S j_n\,dS. \tag{40.4}$$

(Мы употребляем символ частной производной $\partial/\partial t$, чтобы подчеркнуть, что поверхность S должна оставаться неподвижной.) Представив q в виде $q = \int \rho\,dV$ и преобразовав поверхностный интеграл в объемный $\int \operatorname{div}\mathbf{j}\,dV$, придем к соотношению

$$\frac{\partial}{\partial t}\int \rho\,dV = -\int \operatorname{div}\mathbf{j}\,dV.$$

Это соотношение должно выполняться для произвольного объема V, а потому

$$\frac{\partial \rho}{\partial t} + \operatorname{div}\mathbf{j} = 0. \tag{40.5}$$

Формулы (40.4), (40.5) и выражают *закон сохранения заряда в макроскопической электродинамике.* Последняя формула называется также *уравнением непрерывности* или *уравнением неразрывности.* Эти формулы входят в систему основных уравнений Максвелла, хотя и в неявном виде.

Если токи *стационарны*, т. е. не зависят от времени, то формулы (40.4) и (40.5) переходят в

$$\oint j_n \, dS = 0, \tag{40.6}$$

$$\operatorname{div} \mathbf{j} = 0. \tag{40.7}$$

В настоящей главе рассматриваются в основном стационарные (постоянные) токи.

§ 41. Закон Ома

1. Одним из главных способов возбуждения электрического тока в телах является создание и поддержание в них электрического поля. Как показывает опыт, для многих тел (например, металлов) в широких пределах плотность электрического тока $\mathbf{j}$ пропорциональна напряженности электрического поля $\mathbf{E}$. Это — один из важнейших, хотя и не фундаментальных, законов электродинамики. Он называется *законом Ома* (1787–1854). Математически закон Ома выражается формулой

$$\mathbf{j} = \lambda \mathbf{E}, \tag{41.1}$$

где λ — постоянная для данного материала величина, называемая его *удельной электрической проводимостью.* Она зависит от физического состояния тела (температуры, давления и пр.). Строго говоря, закон Ома справедлив лишь для *физически однородных* тел. Величина, обратная электрической проводимости, называется *удельным сопротивлением материала*:

$$\rho = 1/\lambda. \tag{41.2}$$

В гауссовой (а следовательно, и в электростатической) системе единиц электрическая проводимость λ имеет размерность, обратную времени. Ее единица есть *обратная секунда* (с^{-1}). Удельное сопротивление ρ измеряется в *секундах* (с). Разумеется, совпадение размерностей удельного сопротивления и времени не означает, что эти величины по своей физической природе тождественны. Такое совпадение имеет место только в гауссовой и СГСЭ-системах единиц. В других системах эти величины имеют разные размерности (см. § 44).

2. Если ток стационарный, то объемная плотность электричества в однородном проводнике равна нулю. Действительно, для стационарных токов справедливо уравнение (40.7). Перепишем его в виде $\operatorname{div} \lambda \mathbf{E} = 0$ или $\operatorname{div}\left(\frac{\lambda}{\varepsilon}\mathbf{D}\right) = 0$. Так как среда по предположению однородна, то $\lambda = \text{const}$, $\varepsilon = \text{const}$ и рассматриваемое уравнение

сводится к виду $\operatorname{div}\mathbf{D} = 0$. Отсюда с учетом теоремы Гаусса (13.5) находим $\rho = 0$.

Таким образом, *в случае стационарных токов макроскопические электрические заряды могут находиться только на поверхности или в местах неоднородности проводящей среды.* В этом отношении электрическое поле стационарных токов аналогично электростатическому. Аналогия между этими полями идет еще дальше. Если точки стационарны, то плотность электрических зарядов в каждой точке пространства не меняется во времени, хотя и происходит движение электричества: на место уходящих электрических зарядов непрерывно поступают новые. Такие заряды, как показывает опыт (а также уравнения Максвелла), создают в окружающем пространстве такое же кулоновское электрическое поле, что и неподвижные заряды той же плотности. Отсюда следует, что *электрическое поле стационарных токов есть поле потенциальное.*

Тем не менее электрическое поле стационарных токов существенно отличается от электростатического. Электростатическое поле есть кулоновское поле неподвижных зарядов. Внутри проводников при равновесии зарядов оно равно нулю. Электрическое поле стационарных токов есть также *кулоновское поле*, однако заряды, его возбуждающие, находятся в движении. Поэтому поле стационарных токов существует и внутри проводников. Если бы это было не так, то в проводниках не было бы и электрических токов, как это следует из закона Ома (41.1). Силовые линии электростатического поля всегда нормальны к поверхности проводника. Для электрического поля стационарных токов это не обязательно (см. задачу 1 к этому параграфу).

ЗАДАЧИ

1. Параллельные длинные однородные пластинки AB и CD (рис. 109) сделаны из материала, плохо проводящего электричество (например, из дерева). Боковые края их A и C накоротко соединены хорошим проводником (например, металлом), а между краями B и D поддерживается постоянное напряжение V. Найти напряженность электрического поля и форму электрических силовых линий между пластинками, пренебрегая краевыми эффектами. Расстояние между пластинками равно d, а ширина каждой из них $AB = CD = h$.

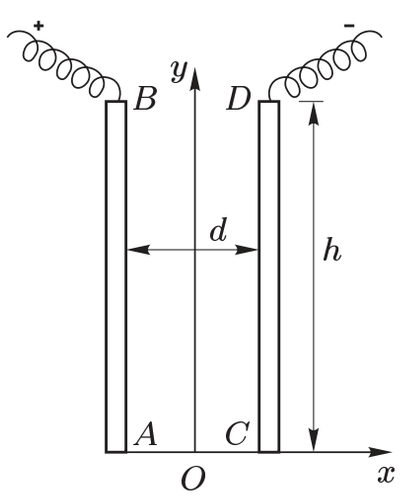


Рис. 109

Р е ш е н и е. Введем прямоугольную систему координат, как указано на рис. 109 (ось z перпендикулярна к плоскости рисунка и параллельна длинным сторонам пластинок). Искомое поле потенциально и удовлетворяет уравнению Лапласа $\partial^2\varphi/\partial x^2 + \partial^2\varphi/\partial y^2 = 0$. На проводнике AC (т. е. при $y = 0$) потенциал должен обращаться в постоянную, которую мы примем равной нулю. Искомое решение будет $\varphi = \alpha xy + \beta y$, где α и β — постоянные. В силу симметрии потенциал φ должен менять знак при замене x на $-x$, а потому $\beta = 0$.

Для напряженности поля получаем

$$E_x = -\frac{\partial\varphi}{\partial x} = -\alpha y, \quad E_y = -\frac{\partial\varphi}{\partial y} = -\alpha x.$$

Постоянная α найдется по разности потенциалов между точками A и B (или между точками C и D). Потенциалы точек B и D равны соответственно $\varphi_B = +V/2$, $\varphi_D = -V/2$. Напряженность поля E_y на поверхности пластины AB (т. е. при $x = -d/2$) будет $E_y = -V/(2h) = \alpha d/2$, откуда $\alpha = -V/(hd)$. Окончательно

$$E_x = \frac{Vy}{hd}, \quad E_y = \frac{Vx}{hd}.$$

Уравнение силовой линии $dx/E_x = dy/E_y$ имеет вид

$$\frac{dx}{y} = \frac{dy}{x},$$

откуда $y^2 - x^2 = K$, т. е. силовыми линиями являются равносторонние гиперболы. При $K > 0$ оси гипербол совпадают с осью y, при $K < 0$ — с осью x. Для выяснения смысла постоянной K обозначим через a расстояние от вершины гиперболы до начала координат. При $K > 0$ координатами вершины гиперболы будут $(0, a)$. Они должны удовлетворять уравнению $a^2 - 0^2 = K$, откуда $K = a^2$. Аналогично, для второго случая $(K < 0)$ $K = -a^2$. Таким образом, получаются два семейства гипербол: $y^2 - x^2 = a^2$ и $x^2 - y^2 = a^2$, асимптотами которых являются биссектрисы соответствующих координатных углов (рис. 110). Гиперболические силовые линии первого семейства легко воспроизводятся экспериментально обычным методом, описанным в § 3. Силовые линии второго семейства экспериментально получить трудно из-за малости составляющей поля E_x.

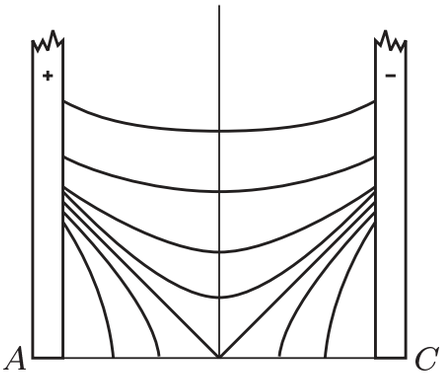


Рис. 110

2. Пространство между пластинами слоистого плоского конденсатора заполнено многослойным диэлектриком, обладающим слабой электрической проводимостью. Диэлектрическая проницаемость и электропроводность изменяются от $\varepsilon_1 = 4$, $\lambda_1 = 10^{-9}$ Ом$^{-1}\cdot$см$^{-1}$ на одной поверхности диэлектрика до $\varepsilon_2 = 3$, $\lambda_2 = 10^{-12}$ Ом$^{-1}\cdot$см$^{-1}$ на другой его поверхности. Конденсатор включен в цепь батареи постоянной электродвижущей силы. Определить величину и знак суммарного свободного заряда q, который возникает в диэлектрике, когда в цепи установится постоянный электрический ток $I = 10^{-7}$ А, текущий через диэлектрик в направлении от стороны *1* к стороне *2*.

Ответ. $q = \dfrac{I}{4\pi}\left(\dfrac{\varepsilon_2}{\lambda_2} - \dfrac{\varepsilon_1}{\lambda_1}\right) = 78$ СГСЭ-ед. $= 2{,}6\cdot 10^{-8}$ Кл.

3. Пространство между пластинами плоского конденсатора заполнено двумя однородными слабо проводящими слоями диэлектрика с толщинами h_1 и h_2. Диэлектрическая проницаемость и электрическая проводимость первого диэлектрика равны соответственно ε_1 и λ_1, второго — ε_2 и λ_2. Найти

плотность поверхностных свободных зарядов σ на границе между диэлектриками, которая установится при наложении на конденсатор постоянного напряжения V.

Ответ. $\sigma = \dfrac{V}{4\pi}\,\dfrac{\varepsilon_2\lambda_1 - \varepsilon_1\lambda_2}{h_1\lambda_2 + h_2\lambda_1}$.

§ 42. Вывод законов Ома и Джоуля–Ленца

1. Будем предполагать в этом параграфе, что электрическое поле $\mathbf{E}$ может меняться во времени. Рассмотрим сначала металлы, хотя наши рассуждения в основном справедливы и в случае других проводящих сред (электролитов, ионизованных газов и пр.). В металлах носителями тока служат *«свободные электроны»*, т. е. электроны, сравнительно слабо связанные с ионами кристаллической решетки, внутри которой они могут свободно перемещаться. Прямое доказательство этого утверждения дают классические опыты Р. Толмена (1881–1948) и Б. Стюарта (1828–1887) (см. § 97). В отсутствие электрического поля или других регулярных сил, действующих на электроны, все направления движения последних равновероятны. В этом отношении движение электронов в металле напоминает тепловое движение молекул газа. Назовем такое движение *беспорядочным*, а соответствующую ему скорость электронов будем обозначать через $\mathbf{v}_б$. Для последующих рассуждений не имеет значения, является ли беспорядочное движение электронов тепловым или нетепловым (см. т. II, § 4, п. 3).

2. При наличии регулярной силы на беспорядочное движение электронов накладывается систематическое — *дрейфовое* — движение. Если поле регулярных сил однородно, то все свободные электроны движутся с одной и той же *дрейфовой скоростью*, обозначаемой ниже через $\mathbf{w}_д$ или $\mathbf{u}$. Полная скорость электрона $\mathbf{v}$ складывается из беспорядочной $\mathbf{v}_б$ и дрейфовой $\mathbf{u}$: $\mathbf{v} = \mathbf{v}_б + \mathbf{u}$. Движение электрона в классической механике описывается уравнением

$$m\frac{d\mathbf{v}}{dt} \equiv m\frac{d}{dt}\,(\mathbf{v}_б + \mathbf{u}) = \mathbf{F} + \mathbf{F}_{ст}, \tag{42.1}$$

где $\mathbf{F}$ — регулярная сила, действующая на электрон со стороны внешнего силового поля, а $\mathbf{F}_{ст}$ — сила, которую он испытывает при столкновениях с ионами или другими электронами. Если уравнение (42.1) усреднить по всем электронам, то производная $d\mathbf{v}_б/dt$ обратится в нуль, а сила $\mathbf{F}_{ст}$ заменится ее средним значением $\langle\mathbf{F}_{ст}\rangle$. Заметим, что при таком усреднении столкновения между электронами можно не принимать во внимание, так как они (столкновения) не влияют на импульс $\sum m\mathbf{v} \equiv \equiv \sum m(\mathbf{v}_б + \mathbf{u})$ всей системы электронов, который только и входит в вычисление среднего значения скорости $\mathbf{v}$. Таким образом, под $\mathbf{F}_{ст}$ следует понимать силы, действующие на электроны при их столкновениях только с ионами кристаллической решетки. При отсутствии дрейфового движения средняя сила $\mathbf{F}_{ст}$ обращается в нуль. При наличии дрейфового движения этого происходить не будет. При малых дрейфовых скоростях величину $\mathbf{F}_{ст}$ можно разложить по степеням u

и ограничиться при этом линейным членом:

$$\mathbf{F}_{\text{ст}} = -m\frac{\mathbf{u}}{\tau_{\text{ин}}}, \tag{42.2}$$

где $\tau_{\text{ин}}$ — постоянная, имеющая размерность времени. В этом приближении уравнение для дрейфового движения электрона принимает вид

$$m\frac{d\mathbf{u}}{dt} + m\frac{\mathbf{u}}{\tau_{\text{ин}}} = \mathbf{F}. \tag{42.3}$$

Сила $\mathbf{F}_{\text{ст}}$, а с ней и время $\tau_{\text{ин}}$ обусловлены инерцией электронов. Поэтому величину $\tau_{\text{ин}}$ можно назвать *инерционным временем* электрона в металле. Конкретное представление о времени $\tau_{\text{ин}}$ дает следующий пример. Предположим, что $\mathbf{F} = 0$ и что в начальный момент времени $t = 0$ электроны совершают дрейфовое движение со скоростью $\mathbf{u} = \mathbf{u}_0$. Тогда из уравнения (42.3) получаем

$$\mathbf{u} = \mathbf{u}_0 e^{-t/\tau_{\text{ин}}}, \tag{42.4}$$

т. е. в отсутствие внешних регулярных сил дрейфовое движение экспоненциально затухает во времени таким образом, что за время $\tau_{\text{ин}}$ скорость u убывает в e раз.

Воспользовавшись соотношением $\mathbf{j} = ne\mathbf{u}$ и введя обозначение

$$\lambda = \frac{ne^2\tau_{\text{ин}}}{m}, \tag{42.5}$$

преобразуем уравнение (42.3) к виду

$$\mathbf{j} + \tau_{\text{ин}}\frac{d\mathbf{j}}{dt} = \lambda\frac{\mathbf{F}}{e}. \tag{42.6}$$

Если регулярная сила $\mathbf{F}$ и коэффициент λ постоянны, то из (42.6) получаем

$$\mathbf{j} = \lambda\frac{\mathbf{F}}{e} + \mathbf{j}_0 e^{-t/\tau_{\text{ин}}}.$$

При $t \gg \tau_{\text{ин}}$

$$\mathbf{j} = \lambda\frac{\mathbf{F}}{e}. \tag{42.7}$$

Все полученные соотношения верны независимо от природы регулярной силы $\mathbf{F}$, возбуждающей электрический ток. Если ток возбуждается электрическим полем $\mathbf{E}$, то $\mathbf{F} = e\mathbf{E}$. Тогда соотношение (42.6) переходит в

$$\mathbf{j} + \tau_{\text{ин}}\frac{d\mathbf{j}}{dt} = \lambda\mathbf{E}, \tag{42.8}$$

а при $\mathbf{E} = \text{const}$ и $\mathbf{j} = \text{const}$ — в

$$\mathbf{j} = \lambda\mathbf{E}. \tag{42.9}$$

Это — *закон Ома* при условии, что концентрация свободных электронов n и инерционное время $\tau_{\text{ин}}$ постоянны, т. е. не зависят от напряженности электрического поля $\mathbf{E}$, причем электрическая проводимость λ

определяется выражением (42.5). Закон Ома (42.9) верен и для переменных полей, как это непосредственно следует из формулы (42.8). Требуется только, чтобы за время $\tau_{\text{ин}}$ ток менялся пренебрежимо мало, т. е. выполнялось условие

$$\left|\tau_{\text{ин}}\frac{d\mathbf{j}}{dt}\right| \ll |\mathbf{j}|. \tag{42.10}$$

3. *Инерционное время электрона в металле можно выразить через среднее время свободного пробега электрона между двумя последовательными столкновениями его с ионами решетки.* Так как масса электрона пренебрежимо мала по сравнению с массой иона, то можно принять, что при каждом столкновении с ионом электрон полностью утрачивает свое упорядоченное движение. От беспорядочного движения можно отвлечься, так как оно на величину тока не влияет. Иначе говоря, можно принять, что при каждом столкновении скорость электрона обращается в нуль. Проведем сначала упрощенный расчет в предположении, что время свободного пробега между двумя последовательными столкновениями τ одно и то же для всех электронов и для всех столкновений. Будем предполагать, что электрон движется в постоянном электрическом поле $\mathbf{E}$. Претерпев столкновение, электрон начинает двигаться с постоянным ускорением $\mathbf{a} = e\mathbf{E}/m$. К следующему столкновению он приходит со скоростью $\mathbf{v} = \mathbf{a}\tau$, так что его средняя скорость между двумя последовательными столкновениями будет $\mathbf{u} = \mathbf{a}\tau/2$, а плотность тока $\mathbf{j} = ne\mathbf{u} = \dfrac{ne^2\tau}{2m}\,\mathbf{E}$. Следовательно,

$$\lambda = \frac{ne^2\tau}{2m}. \tag{42.11}$$

В этих упрощающих предположениях $\tau_{\text{ин}} = \tau/2$.

Приведенный расчет не вполне точен, даже если сохранить идеализацию, что при каждом столкновении электрон полностью утрачивает упорядоченную скорость. Расчет не принимает во внимание, что время свободного пробега τ меняется от столкновения к столкновению. Для учета этого обстоятельства рассмотрим какой-либо один электрон, претерпевший много столкновений. Обозначим через $\tau_1, \tau_2, \ldots, \tau_N$ времена свободного пробега электрона между последовательными столкновениями, а через $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N$ — средние упорядоченные скорости на этих временах. Тогда средняя упорядоченная скорость за все время движения представится выражением

$$\overline{\mathbf{u}} = \frac{\sum \tau_i \mathbf{u}_i}{\sum \tau_i} = \frac{\mathbf{a}}{2}\,\frac{\sum \tau_i^2}{\sum \tau_i}.$$

Разделив числитель и знаменатель на N и перейдя к пределу $N \to \infty$, запишем это выражение в виде $\mathbf{u} = a\overline{\tau^2}/(2\overline{\tau})$, где $\overline{\tau}$ и $\overline{\tau^2}$ — средние

значения времени свободного пробега электрона и его квадрата:

$$\overline{\tau} = \lim_{N\to\infty} \frac{1}{N} \sum \tau_i, \quad \overline{\tau^2} = \lim_{N\to\infty} \frac{1}{N} \sum \tau_i^2.$$

Те же средние значения τ и τ^2 будут характеризовать не только какой-либо определенный электрон, но и всю совокупность электронов. Таким образом, для электрической проводимости λ теперь получается

$$\lambda = \frac{ne^2}{2m} \frac{\overline{\tau^2}}{\tau}.$$

В частном случае, когда все времена $\tau_1, \tau_2, \ldots$ одинаковы, получится прежний результат (42.11).

Выразим теперь отношение $\overline{\tau^2}/\overline{\tau}$ через среднее время свободного пробега электрона $\overline{\tau}$. Возьмем пучок электронов, состоящий из n_0 частиц, находящихся в момент времени $t = 0$ в одинаковых условиях. При дальнейшем движении эти частицы будут сталкиваться с ионами решетки и выбывать из пучка. Пусть $n(t)$ — число частиц, оставшихся в пучке ко времени t. Среднее число частиц $-dn$, выбывших из пучка вследствие столкновений между t и $t + dt$, пропорционально n и может быть представлено выражением $-dn = \alpha n\, dt$, где α — положительная постоянная. Отсюда $n = n_0 e^{-\alpha t}$. Каждая из этих $(-dn)$ частиц двигалась без столкновений в течение времени t. Поэтому

$$\overline{\tau} = -\frac{1}{n_0} \int t\, dn, \quad \overline{\tau^2} = -\frac{1}{n_0} \int t^2\, dn.$$

Интегрируя по частям, преобразуем второй интеграл:

$$\overline{\tau^2} = \int_0^\infty t^2\, de^{-\alpha t} = -2 \int_0^\infty te^{-\alpha t}\, dt = -\frac{2}{n_0} \int t\, dn = 2(\overline{\tau})^2.$$

Таким образом, $\overline{\tau^2}/(2\overline{\tau}) = \overline{\tau}$, и, следовательно, $\tau_{\text{ин}} = \overline{\tau}$,

$$\lambda = \frac{ne^2}{m} \overline{\tau}. \tag{42.12}$$

Тем не менее мы будем пользоваться иногда и упрощенной формулой (42.11). Дело в том, что в кинетической теории газов при изучении явлений переноса мы вводили те же упрощения, что и при выводе формулы (42.11), т. е. предполагали, что длины и времена свободного пробега одинаковы для всех молекул и всех столкновений. Было бы непоследовательно при сопоставлении кинетической теории газов с теорией электрической проводимости вводить различные упрощающие предположения и производить вычисления с различной точностью.

Отметим еще физический смысл постоянной α. Для этого выразим ее через время $\overline{\tau}$:

$$\overline{\tau} = -\frac{1}{n_0} \int t\, dn = \int_0^\infty t\, de^{-\alpha t} = \frac{1}{\alpha}.$$

Таким образом,

$$n = n_0 e^{-t/\overline{\tau}}. \tag{42.13}$$

Эта формула вполне аналогична формуле (88.3) из второго тома нашего курса.

4. Вместо инерционного времени $\tau_{\text{ин}}$ или среднего времени свободного пробега $\overline{\tau}$ можно ввести другие параметры, связанные с этими временами. Часто используют *подвижность частицы* (электрона). Подвижностью частицы называют дрейфовую скорость, приобретаемую частицей либо под действием постоянной силы $\mathbf{F}$, равной единицы, либо под действием постоянного электрического поля $\mathbf{E}$, также равного единице. Первую подвижность будем обозначать большой буквой B, вторую — малой b (см. также т. II, § 64). Таким образом,

$$\mathbf{u} = B\mathbf{F} = b\mathbf{E}. \tag{42.14}$$

Так как $\mathbf{F} = e\mathbf{E}$, то обе подвижности связаны соотношением

$$b = |e|B. \tag{42.15}$$

Через подвижность электрическая проводимость λ выражается формулой

$$\lambda = Be^2 n = b|e|n. \tag{42.16}$$

Понятием подвижности всегда пользуются при рассмотрении электрических токов в электролитах и ионизованных газах. В этих случаях имеется несколько типов носителей тока — положительных и отрицательных ионов. Для получения электропроводности λ выражение (42.16) надо просуммировать по всем типам носителей. Например, в растворах электролитов имеются два типа ионов, и вместо формулы (42.16) следует писать

$$\lambda = B^- e_-^2 n^- + B^+ e_+^2 n^+ = b^-|e_-|n^- + b^+|e_+|n^+, \tag{42.17}$$

где индекс «$-$» относится к отрицательным, а индекс «$+$» к положительным ионам.

Выражение (42.17) можно упростить, использовав то обстоятельство, что для возбуждения электрических полей, встречающихся в действительности, требуется ничтожное разделение положительных и отрицательных зарядов, которым при вычислении можно полностью пренебречь (см. § 10). Иными словами, можно считать, что электролит (или газ) *электрически нейтрален* (точнее, *квазинейтрален*), т. е. для него с большой точностью выполняется соотношение

$$\sum n_i e_i = 0. \tag{42.18}$$

В частности, если заряды положительных и отрицательных ионов одинаковы по абсолютной величине, это соотношение переходит в $n^- = n^+$, т. е. в равенство концентраций обоих ионов. В этом случае, если опустить индексы при n и e, получим

$$\lambda = (B^- + B^+)e^2 n = (b^- + b^+)\,|e|n. \tag{42.19}$$

5. Приведенные рассуждения (в согласии с опытом) приводят к заключению, что при определенных условиях возможны отклонения от закона Ома и даже полное несоблюдение этого закона. Для справедливости закона Ома необходимо, чтобы концентрации носителей тока и инерционные времена $\tau_{\text{ин}}$ (или соответствующие им подвижности) при прохождении тока оставались постоянными. Необходимо также, чтобы в переменных полях инерция носителей тока не играла роли; количественно это условие, как уже отмечалось выше, сводится к требованию, чтобы изменения тока за времена порядка $\tau_{\text{ин}}$ были пренебрежимо малы. В случае периодических переменных электрических полей это означает, что период изменения поля T должен быть очень велик по сравнению с $\tau_{\text{ин}}$.

Закон Ома может нарушаться в *сильных полях*, где могут проявляться нелинейные эффекты. В таких случаях при разложении средней силы $\langle \mathbf{F}_{\text{ст}} \rangle$ по степеням $\mathbf{u}$ линейное приближение (42.2) уже недостаточно. *Сильными* мы называем такие поля, в которых на протяжении среднего свободного пробега носитель тока приобретает скорость, сравнимую со скоростью беспорядочного движения. Количественно условие слабости поля записывается в виде

$$F\tau \ll mv_{\text{б}} \quad \text{или} \quad eE\tau \ll mv_{\text{б}}. \tag{42.20}$$

Посмотрим, соблюдается ли условие (42.20) при прохождении электрического тока в металлах. В качестве примера возьмем медь. Ее проводимость $\lambda = 5{,}3 \cdot 10^{17}$ с$^{-1}$ (при 20 °C), плотность $\rho = 8{,}9$ г/см^3, атомная масса $A = 63$. Если ввести правдоподобное предположение, что на каждый атом меди приходится один свободный электрон, то концентрация свободных электронов будет

$$n = \frac{N}{A} \approx 8{,}5 \cdot 10^{22} \text{ см}^{-3}$$

(N — число Авогадро). Инерционное время электрона вычислим по проводимости λ:

$$\tau_{\text{ин}} = \frac{m\lambda}{ne^2} \approx 2{,}5 \cdot 10^{-14} \text{ с}.$$

(Подвижность электрона $b = e\tau_{\text{ин}}/m = 1{,}3 \cdot 10^4$ СГСЭ-ед. $= 44$ см^2/(с $\cdot$ В). Средняя беспорядочная скорость электронов, если ее оценить по формулам классической кинетической теории газов, будет $v_{\text{б}} \approx \sqrt{3kT/m} \approx 10^7$ см/с. В действительности она примерно на порядок больше, так как для свободных электронов в металлах надо пользоваться не классической статистикой Больцмана, а квантовой статистикой Ферми (1901–1954) (см. т. II, § 82). С учетом этого обстоятельства берем $v_{\text{б}} \sim 10^8$ см/с. Тогда для промежуточных значений между слабыми и сильными полями получаем $E \sim mv_{\text{б}}/(e\tau) \sim$ $\sim 0{,}75 \cdot 10^4$ СГСЭ-ед. $\sim 2 \cdot 10^8$ В/м. Только начиная с таких полей

могли бы проявиться *нелинейные эффекты* при прохождении электрического тока через металл. На самом деле такие поля в металлах невозможны: они мгновенно превратили бы металл в пар. Наибольшая технически допустимая плотность тока в медных проводах принимается равной $j = 10^3$ А/см$^2 = 3 \cdot 10^{12}$ СГСЭ-ед. Ей соответствует напряженность электрического поля $E = j/\lambda \approx 0{,}5 \cdot 10^{-5}$ СГСЭ-ед. $\approx 0{,}15$ В/м, что в 10^9 раз меньше вычисленной выше величины. С этим обстоятельством и связана применимость закона Ома к металлам.

В ионизованных газах закон Ома не соблюдается. При низких давлениях кинетическая энергия, приобретаемая электроном за время свободного пробега, даже в слабых электрических полях становится сравнимой с энергией теплового движения kT. Поэтому уже в таких полях линейное приближение (42.2) становится недействительным и наблюдаются *отступления* от закона Ома. При возрастании напряженности поля энергия электронов становится достаточной, чтобы ионизовать атомы и молекулы газа. Концентрации ионов, а потому и ток в газе сильно возрастают. При дальнейшем увеличении напряженности поля наступает *электрический пробой* газа (искра).

В местах контакта между полупроводниками или между полупроводниками и металлами могут происходить резкие нарушения закона Ома. Эти места являются типичными *«нелинейными»*, или *«неомическими»* проводниками тока, т. е. такими проводниками, в которых связь между $\mathbf{j}$ и $\mathbf{E}$ нелинейна. Их проводимость может быть даже *односторонней*, т. е. ток через них может проходить практически только в одном направлении. Практическое значение нелинейных проводников трудно переоценить. Современная радиотехника и электроника были бы невозможны, если бы все тела, проводящие электричество, подчинялись закону Ома.

6. Над электроном, движущимся со скоростью $\mathbf{v}$ в однородном силовом поле, ежесекундно совершается работа $\mathbf{vF} = (\mathbf{u}+\mathbf{v}_6)\mathbf{F}$. При суммировании по всем электронам члены $\mathbf{v}_6\mathbf{F}$ дают нуль. Остается только регулярная работа, связанная с дрейфовым движением электронов. Эта работа, совершаемая над электронами единицы объема металла, равна $n\mathbf{uF} = \mathbf{jF}/e$. В металлах она идет на приращение *внутренней* (*тепловой*) энергии, поскольку прохождение электрического тока не сопровождается изменениями внутренней структуры металла. Таким образом, *мощность тепла*, выделяемого током в единице объема проводника, дается выражениями

$$Q = \frac{1}{e}(\mathbf{jF}) = \frac{\lambda}{e^2}\mathbf{F}^2 \tag{42.21}$$

или

$$Q = \frac{1}{\lambda}\mathbf{j}^2. \tag{42.22}$$

Последняя формула выражает *закон Джоуля* (1819–1889)*–Ленца* (1804–1865) *в локальной* (*дифференциальной*) *форме*: мощность тепла в единице объема Q пропорциональна квадрату плотности

электрического тока и обратно пропорциональна проводимости среды. В такой форме закон Джоуля–Ленца носит *общий характер*, т. е. *не зависит от природы сил*, возбуждающих электрический ток. Если сила $\mathbf{F}$ чисто электрическая ($\mathbf{F} = e\mathbf{E}$), то

$$Q = (\mathbf{jE}) = \lambda \mathbf{E}^2. \tag{42.23}$$

Из изложенного ясно, что выражение (42.23) носит менее общий характер, чем (42.22).

Закон Джоуля–Ленца, как показывает опыт, справедлив и для электролитов. Отсюда следует, что работа электрического поля в электролитах не тратится на образование ионов. Ионы в растворе образуются в результате *диссоциации молекул* при растворении (электролитическая диссоциация). Приложенное электрическое поле к этому процессу не имеет отношения.

7. Изложенная классическая теория без существенных изменений сохраняется и в квантовой физике. Однако классическая теория носит *кинематический характер*, поскольку она не определяет подвижности и концентрации носителей тока. Для этого требуется *динамическая теория.* Попытки создать такую теорию на классической основе всегда приводили к резким противоречиям с опытом. Только квантовая теория позволяет (или позволяет в принципе) построить динамическую теорию электропроводности, согласующуюся с опытом.

8. Приведем в заключение один результат, полученный классической теорией. Хотя его классический вывод и неверен, но самый результат оказался верным. Речь идет о связи между проводимостью и теплопроводностью металлов. Металлы — хорошие проводники не только электричества, но и тепла. Это связано с тем, что переносчиками электричества и тепла в металлах являются одни и те же частицы — *свободные электроны.* Роль ионов в переносе тепла пренебрежимо мала. Применяя к электронной теплопроводности формулы кинетической теории газов (см. т. II, § 89), для теплопроводности металла можно написать

$$\chi = \frac{1}{3}\, n\overline{v} c_v \overline{l}, \tag{42.24}$$

где $\overline{v}$ — средняя скорость беспорядочного движения электронов, c_v — теплоемкость электронного газа при постоянном объеме, приходящаяся на один электрон, $\overline{l}$ — средняя длина свободного пробега электрона. В том же приближении выведена формула (42.11). Запишем ее в виде

$$\lambda = \frac{ne^2}{2m}\, \frac{\overline{l}}{\overline{v}}. \tag{42.25}$$

Почленным делением (42.24) на (42.25) находим

$$\frac{\chi}{\lambda} \approx \frac{2}{3}\, \frac{m\overline{v}^2}{e^2}\, c_v. \tag{42.26}$$

Эта формула сохраняется и в квантовой теории. Однако квантовая теория применяет к электронам в металлах статистику Ферми–Дирака

(1902–1984), а классическая — статистику Больцмана (что неправильно). В соответствии с этим в классической теории полагают $c_v = (3/2)k$, $m\overline{v}^2 \approx 3kT$ и получают

$$\frac{\chi}{\lambda} = 3\left(\frac{k}{e}\right)^2 T. \tag{42.27}$$

(Мы не учитываем разницы между $(\overline{v})^2$ и $\overline{v^2}$. Учет этой разницы в излагаемой приближенной теории был бы превышением точности расчета.)

Формула (42.27) была получена П. Друде (1863–1906). Друде, как и мы, не учитывал распределения электронов по скоростям. Лоренц, учтя максвелловское распределение тепловых скоростей электронов, получил такую же формулу, но с числовым коэффициентом 2 вместо 3. Важно, однако, не значение числового коэффициента, а то, что отношение χ/λ пропорционально термодинамической температуре T и для всех металлов одно и то же (закон Видемана (1826–1899)–Франца). Этот результат довольно хорошо подтвердился на опыте и долгое время считался доказательством правильности исходных положений классической теории электрической проводимости и теплопроводности металлов, несмотря на то что в вопросе о теплоемкости электронов в металлах эта теория приводила к резкому противоречию с опытом (см. т. II, § 69 и 84). Противоречие было устранено Зоммерфельдом (1868–1951), который применил статистику Ферми–Дирака к проблеме проводимости, теплопроводности и теплоемкости электронного газа в металлах. Он снова получил формулу (42.27) с коэффициентом $\pi^2/3$ вместо 3. Таким образом, классическая теория Друде и квантовая теория Зоммерфельда приводили фактически к одинаковым результатам. Такое совпадение результатов объясняется тем, что классическая теория пользовалась неправильными значениями для $\overline{v^2}$ и c_v. Эти две ошибки случайно компенсировали друг друга, так что произведение $\overline{v^2}c_v$ фактически получилось правильным. В § 99 этот вопрос будет разобран подробно.

ЗАДАЧИ

1. Найти вероятность того, что между моментами времени t_1 и $t > t_1$ свободный электрон в металле не претерпевает столкновений с ионами решетки. Электрическое поле в металле $E = 0$.

Решение. Так как электрического поля нет, то все моменты времени эквивалентны. В силу этого искомая вероятность может зависеть не от времени t и t_1 в отдельности, а только от их разности $t - t_1$. Обозначим эту вероятность через $f(t - t_1)$. Возьмем теперь второй момент времени t_2, более ранний, чем t_1. Вероятность события, что электрон не претерпит столкновений в промежутке времени $t - t_2$, будет $f(t - t_2)$. Но это событие можно рассматривать как сложное. Оно состоит из последовательности двух статистически независимых событий: 1) события, что электрон не претерпевает столкновений в интервале времени (t_2, t_1), и 2) события, что в интервале (t_1, t) он также не испытывает столкновений. По теореме умножения вероятностей вероятность этого сложного события может быть представлена

произведением $f(t-t_2) = f(t_1-t_2)f(t-t_1)$. Решением этого уравнения является

$$f(t) = e^{-\alpha t}, \tag{42.28}$$

где α — постоянная. Очевидно, она положительна, так как с возрастанием аргумента t функция $f(t)$ должна убывать. Таким образом,

$$f(t-t_0) = e^{-\alpha(t-t_0)}. \tag{42.29}$$

Найдем теперь вероятность того, что электрон претерпел столкновение между t и $t+dt$. Очевидно, она равна

$$dw = f(t-t_0) - f(t+dt-t_0) = -\frac{df}{dt}\,dt$$

или

$$dw = \alpha e^{-\alpha(t-t_0)}\,dt. \tag{42.30}$$

Простым интегрированием убеждаемся, что эта вероятность нормирована к единице:

$$\int dw = \int\limits_{t_0}^{\infty} \alpha e^{-\alpha(t-t_0)}\,dt = 1.$$

Так и должно быть, поскольку столкновение электрона на бесконечном интервале времени (t_0, ∞) есть достоверное событие. Постоянную α можно выразить через среднее время свободного пробега электрона $\overline{\tau}$ между двумя последовательными столкновениями его с ионами. Очевидно,

$$\overline{\tau} = \int\limits_{t_0}^{\infty} (t-t_0)\,dw = \alpha \int\limits_{t_0}^{\infty} e^{-\alpha(t-t_0)}(t-t_0)\,dt = \frac{1}{\alpha}.$$

Поэтому

$$f(t-t_0) = e^{-(t-t_0)/\overline{\tau}}. \tag{42.31}$$

2. Разъяснить следующий парадокс. Вероятность того, что последнее столкновение электрона произошло между t_0 и t_0+dt_0, равна

$$dw = f[t-(t_0+dt_0)] - f(t-t_0) = \frac{df}{dt_0}\,dt_0,$$

или

$$dw = \alpha e^{-(t-t_0)}\,dt_0. \tag{42.32}$$

До момента времени t электрон в среднем двигался без столкновений в течение времени

$$\int (t-t_0)\,dw = \int\limits_{-\infty}^{t} (t-t_0)\alpha e^{-(t-t_0)}\,dt_0 = \frac{1}{\alpha},$$

т. е. такое же время, какое он затратит в среднем до следующего столкновения.

Рассмотрим теперь движение электрона между двумя последовательными столкновениями. Пусть момент времени t лежит где-то между моментами столкновений. Пусть от первого столкновения до момента t электрон затрачивает время T_1, а от момента t до второго столкновения — время T_2. Тогда время между столкновениями будет $T = T_1 + T_2$. Усредняя это время, получим $\overline{T} = \overline{T}_1 + \overline{T}_2$. Но согласно доказанному $\overline{T}_1 = \overline{T}_2 = \overline{\tau}$. Таким

образом, среднее время свободного пробега электрона между двумя последовательными столкновениями будет $\overline{T} = 2\overline{\tau}$. На самом деле оно равно $\overline{\tau}$.

Р е ш е н и е. Соотношение $\overline{T} = \overline{T}_1 + \overline{T}_2$ неверно, так как операция усреднения времени T имеет другой смысл, чем операции усреднения времени T_1 и T_2. Возьмем, например, пучок электронов, одновременно претерпевших столкновения. Пусть число частиц в пучке непосредственно после столкновения равно n_0. Для вычисления $\overline{T}$ надо усреднить T по всем n_0 электронам. Но электроны выбывают из пучка из-за столкновений. Пусть к моменту t их останется в пучке n. Для нахождения $\overline{T}_1$ и $\overline{T}_2$ надо усреднять T_1 и T_2 уже по меньшему числу частиц n. Благодаря этому правильное соотношение между средними временами будет $\overline{T} = \overline{T}_1 = \overline{T}_2 = \overline{\tau}$.

3. Применить результаты, полученные в задачах 1 и 2, к вычислению плотности тока в металле, введя предположение, что после каждого столкновения электрон полностью утрачивает упорядоченную скорость. Электрическое поле $\mathbf{E}$ считать слабым, однородным и зависящим от времени.

Р е ш е н и е. В слабых полях формулы (42.31) и (42.32) остаются верными. Если электрон претерпел последнее столкновение в момент t_0, то за время $t - t_0$ его скорость получит приращение

$$\Delta\mathbf{v} = \frac{e}{m}\int\limits_{t_0}^{t} \mathbf{E}(t')\,dt'. \tag{42.33}$$

Усредняя это выражение по всем электронам, мы и найдем упорядоченную (дрейфовую) скорость электрона в электрическом поле:

$$\mathbf{u} = \langle\Delta\mathbf{v}\rangle = \int \Delta\mathbf{v}\,dw = \frac{1}{\overline{\tau}}\int\limits_{-\infty}^{t} \Delta\mathbf{v}e^{-(t-t_0)/\overline{\tau}}\,dt_0.$$

Введем временное обозначение $\Delta\mathbf{v} = \mathbf{k}\xi$, где $\mathbf{k}$ — единичный вектор в направлении электрического поля $\mathbf{E}$. Величину ξ при интегрировании, конечно, надо рассматривать как функцию аргумента t_0. Имея это в виду, интегрированием по частям находим

$$\mathbf{u} = \Delta\mathbf{v}e^{-(t-t_0)/\overline{\tau}}\Big|_{t_0=-\infty}^{t_0=t} - \mathbf{k}\int\limits_{-\infty}^{t} e^{-(t-t_0)/\overline{\tau}}\,d\xi.$$

Множитель $\Delta\mathbf{v}$, как видно из формулы (42.33), обращается в нуль на верхнем пределе $t_0 = t$. На нижнем же пределе $t_0 = -\infty$ обращается в нуль экспоненциальный множитель. Таким образом, в правой части последнего выражения остается только второе слагаемое. Подставляя в него $d\xi = -eE(t_0)\,dt_0/m$ и выполняя интегрирование, получим

$$\mathbf{u} = \frac{e}{m}\int\limits_{-\infty}^{t} \mathbf{E}(t_0)e^{-(t-t_0)/\overline{\tau}}\,dt_0.$$

Плотность тока

$$\mathbf{j} = ne\mathbf{u} = \frac{ne^2}{m}\int\limits_{-\infty}^{t} \mathbf{E}(t_0)e^{-(t-t_0)/\overline{\tau}}\,dt_0.$$

Для исключения входящего сюда интеграла дифференцируем последнее выражение по t и находим

$$\frac{d\mathbf{j}}{dt} = \frac{ne^2}{m}\,\mathbf{E}(t) - \frac{ne^2}{m\overline{\tau}}\int\limits_{-\infty}^{t} \mathbf{E}(t_0)e^{-(t-t_0)/\overline{\tau}}\,dt_0.$$

Следовательно,

$$\mathbf{j} + \overline{\tau}\frac{d\mathbf{j}}{dt} = \frac{ne^2\overline{\tau}}{m}\,\mathbf{E}$$

в согласии с ранее полученными результатами.

§ 43. Сторонние силы. Концентрационный элемент

1. Допустим, что единственными источниками электрического поля $\mathbf{E}$ в проводниках, по которым текут токи, являются электрические заряды, возбуждающие поля по закону Кулона. При прохождении тока непрерывно происходит убыль зарядов, точнее, нейтрализация положительного и отрицательного электричеств. Для того чтобы напряженность поля $\mathbf{E}$, а с ней и плотность электрического тока $\mathbf{j}$ оставались неизменными, необходимы какие-то дополнительные силы или процессы, непрерывно пополняющие электрические заряды.

Плотность электрического тока, как видно из формулы (42.7), определяется полной силой $\mathbf{F}$, действующей на электрон или другой носитель зарядов. Силу $\mathbf{F}$ можно разложить на две части: силу электрическую и силу неэлектрическую, включающую в себя все прочие силы. Эти прочие силы принято называть *сторонними.* В соответствии с этим полагаем $\mathbf{F}/e = \mathbf{E} + \mathbf{E}^{\text{стор}}$, где $\mathbf{E}^{\text{стор}}$ — *напряженность поля сторонних сил*, т. е. сторонняя сила, отнесенная к единице заряда. С учетом сторонних сил закон Ома записывается в виде

$$\mathbf{j} = \lambda(\mathbf{E} + \mathbf{E}^{\text{стор}}). \tag{43.1}$$

2. Приведем пример сторонней силы, не имеющий, правда, практического значения. Если металлический диск равномерно вращается с угловой скоростью ω (рис. 111), то в системе отсчета, связанной с диском, на электрон действует центробежная сила $m\omega^2\mathbf{r}$, где m — масса электрона. Разделив ее на заряд электрона e, найдем напряженность стороннего поля:

$$\mathbf{E}^{\text{стор}} = \frac{m\omega^2}{e}\,\mathbf{r} = \operatorname{grad}\frac{m\omega^2 r^2}{2e} = \frac{4\pi^2 mN^2}{e}\,\mathbf{r}, \tag{43.2}$$

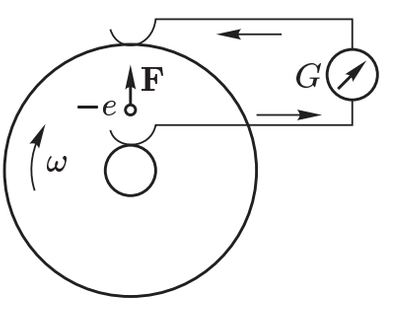


Рис. 111

где N — число оборотов диска в секунду. (Силой Кориолиса (1792–1843) можно пренебречь.) Если к оси и периферии диска подвести скользящие контакты, то через гальванометр потечет электрический ток. Чтобы

составить представление о порядке величины $\mathbf{E}^{\text{стор}}$, подставим в формулу (43.2) $r = 10$ см, $N = 100$ об/с. Получим ничтожную величину $\mathbf{E}^{\text{стор}} \approx 2 \cdot 10^{-9}$ В/см.

3. Второй простой пример, в котором сторонние силы носят несколько формальный характер, дает *концентрационный элемент*. Этот элемент представляет собой два электрода, погруженных в раствор электролита, концентрация которого меняется от точки к точке. Никаких химических реакций между электродами и электролитом не происходит. Электрический ток возникает и поддерживается в результате диффузионного выравнивания концентраций.

Для простоты будем предполагать электроды плоскими (рис. 112). Ось X направим перпендикулярно к поверхностям электродов от отрицательного электрода к положительному. Пусть концентрация электролита зависит только от x. Ради конкретности в качестве электролита возьмем водный раствор соляной кислоты HCl. Молекулы HCl диссоциируют на положительные ионы водорода H^+ и отрицательные ионы хлора Cl^-. Заряд отрицательного иона обозначим через e_-, положительного — через e_+. По абсолютной величине они одинаковы и равны элементарному заряду $|e| = 4{,}80 \times \times 10^{-10}$ СГСЭ-ед. Концентрации отрицательных и положительных ионов $n^-(x)$ и $n^+(x)$ в наших расчетах следует считать одинаковыми, так как электролит во всех случаях является либо электрически нейтральным, либо квазинейтральным (см. § 42). Поэтому можно обозначить эти концентрации одной и той же буквой $n(x)$.

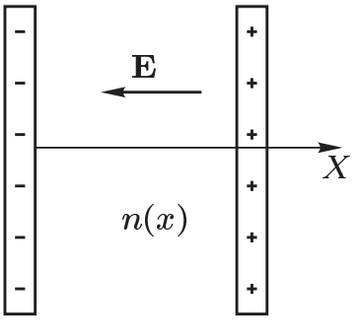


Рис. 112

Из-за наличия градиента концентрации $n(x)$ начнется диффузия обоих ионов в сторону уменьшения концентрации. Скорость диффузии водородных ионов много больше скорости диффузии ионов хлора. Благодаря этому возникнет *разделение зарядов* и появится *электрическое поле*, стремящееся затормозить диффузию ионов водорода и ускорить диффузию ионов хлора. Напряженность электрического поля будет возрастать до тех пор, пока скорости ионов обоих знаков не сделаются одинаковыми. Вместе с диффузией возникнет диффузионный электрический ток. Плотность диффузионного тока равна

$$-D^- \frac{dn^-}{dx}\, e^- - D^+ \frac{dn^+}{dx}\, e^+ = (D^- - D^+) \frac{dn}{dx}\, e,$$

где D^- и D^+ — коэффициенты диффузии ионов водорода и ионов хлора. Этот ток надо прибавить к току $(B^- + B^+)e^2 nE$, вызванному электрическим полем (см. формулу (42.19)). Плотность полного тока будет

$$j = (D^- - D^+) e \frac{dn}{dx} + (B^- + B^+) n e^2 E.$$

Вводя в эту формулу электрическую проводимость $\lambda = (B^- + B^+)ne^2$, получим

$$j = \lambda \left(E + \frac{1}{e} \frac{D^- - D^+}{B^- + B^+} \frac{1}{n} \frac{dn}{dx} \right).$$

Воспользовавшись соотношением Эйнштейна $D = kTB$ (см. т. II, § 91), придадим этой формуле вид

$$j = \lambda \left(E + \frac{kT}{e} \frac{B^- - B^+}{B^- + B^+} \frac{d}{dx} (\ln n) \right), \tag{43.3}$$

где k — постоянная Больцмана, а T — термодинамическая температура. Сравнение формулы (43.3) с (43.1) приводит к соотношению

$$E^{\text{стор}} = \frac{kT}{e} \frac{B^- - B^+}{B^- + B^+} \frac{d}{dx} (\ln n), \tag{43.4}$$

или в векторной форме

$$\mathbf{E}^{\text{стор}} = \frac{kT}{e} \frac{B^- - B^+}{B^- + B^+} \operatorname{grad} \ln n. \tag{43.5}$$

Полученная формула показывает, что от диффузии можно отвлечься и формально рассматривать электролит как однородный, если к напряженности электрического поля $\mathbf{E}$ добавить напряженность поля сторонних сил. Формула (43.4) справедлива и в случае электродов произвольной формы, так как электролит всегда можно мысленно разделить на достаточно малые части, в пределах каждой из которых вектор $\operatorname{grad} \ln n$ может считаться одним и тем же, и применить выражение (43.4).

Концентрационный элемент как источник электрического тока также не применяется на практике. Однако происходящие в нем процессы аналогичны процессам в *гальванических элементах*, где электрические токи поддерживаются за счет *химических реакций* между электродами и электролитами. Для гальванических элементов и прочих источников тока можно также пользоваться формулой (43.1). При этом можно отвлечься от детального рассмотрения физических процессов, возбуждающих и поддерживающих электрический ток в цепи, учитывая эти процессы формально с помощью поля сторонних сил $\mathbf{E}^{\text{стор}}$. Сторонние силы в гальванических элементах действуют на границах между электролитами и электродами. Они действуют также на границе соприкосновения двух разнородных металлов и обусловливают *контактную разность потенциалов* между ними.

4. Как показывают формулы (43.2) и (43.5), в обоих случаях, к которым они относятся, вектор $\mathbf{E}^{\text{стор}}$ выражается через *градиент* какого-то скаляра. Иными словами, в тех областях пространства, в которых действуют сторонние силы, поле этих сил ведет себя как *потенциальное силовое поле.* Это утверждение носит общий характер, если только сторонние силы, как это бывает в большинстве встречающихся случаев, не зависят (или практически не зависят) от силы тока, текущего через

источник. Действительно, сила тока зависит от электрической проводимости и геометрических размеров проводника, который соединяет полюсы источника. Разомкнем цепь, т. е. удалим этот проводник. Тогда, по нашему предположению, поле сторонних сил не изменится. Однако с размыканием цепи прекратятся и все токи, а потому на основании формулы (43.1) должно быть $\lambda(\mathbf{E}+\mathbf{E}^{\text{стор}})=0$. Внутри источника $\lambda\neq$ $\neq 0$, и после сокращения на λ получаем $\mathbf{E}+\mathbf{E}^{\text{стор}}=0$. Поскольку электрическое поле $\mathbf{E}$ потенциально, отсюда следует, что внутри источника $\mathbf{E}^{\text{стор}}$ ведет себя также как потенциальное поле. Вне источника $\lambda=0$, и соотношение $\mathbf{E}+\mathbf{E}^{\text{стор}}=0$ несправедливо. Поэтому во всем пространстве стороннее поле не потенциально. И только благодаря этому оно способно возбуждать и поддерживать постоянные электрические токи (см. следующий параграф).

ЗАДАЧА

Сторонние силы в концентрационном элементе можно рассматривать как силы осмотического давления, действующие в электролите при наличии градиента концентрации. Получить с этой точки зрения формулу (43.4).

Р е ш е н и е. Ввиду квазинейтральности электролита осмотические давления отрицательных и положительных ионов одинаковы. Каждое из них определяется формулой $\mathscr{P}=nkT$. На ионы каждого знака, находящиеся в единице объема, действует сила осмотического давления $-\partial\mathscr{P}/\partial x=-kT\dfrac{\partial n}{\partial x}$, а на один ион — сила $F^{\text{стор}}=-\dfrac{kT}{n}\dfrac{\partial n}{\partial x}$. Направление силы $F^{\text{стор}}$ не зависит от знака заряда иона. Под действием силы $F^{\text{стор}}$ ионы приобретут скорости $u_-=B^-F^{\text{стор}}$, $u_+=B^+F^{\text{стор}}$, и возникнет диффузионный электрический ток с плотностью

$$j_{\text{диф}}=n(e_-u_-+e_+u_+)=ne(B^+-B^-)F^{\text{стор}},$$

или ввиду соотношения (42.19)

$$j_{\text{диф}}=-\lambda\frac{B^--B^+}{(B^-+B^+)e}\,F^{\text{стор}}.$$

Представив это выражение в виде $j_{\text{диф}}=\lambda E^{\text{стор}}$ и воспользовавшись выражением для $F^{\text{стор}}$, найдем напряженность стороннего поля $E^{\text{стор}}$, совпадающую с (43.4).

§ 44. Законы Ома и Джоуля–Ленца в интегральной форме

1. Рассмотрим важнейший случай, когда электрические токи текут вдоль *тонких проводов* (проволок). Направление тока будет совпадать с направлением оси провода. Это автоматически обеспечивается соответствующим распределением электрических зарядов на поверхностях проводников или в местах, где действуют сторонние силы. Площадь поперечного сечения провода S в различных местах его может быть неодинаковой. Для тонких проводов плотность тока j может считаться одной и той же во всех точках поперечного сечения провода. Через

поперечное сечение провода в единицу времени проходит количество электричества

$$I = jS, \tag{44.1}$$

называемое *силой тока* или просто *током*. Если ток постоянен, то из-за сохранения заряда величина I будет одна и та же вдоль всего провода. Для общности будем предполагать, что в проводе действуют сторонние силы, например имеется гальванический элемент. Воспользуемся законом Ома в форме (43.1). Из него получаем

$$E + E^{\text{стор}} = \frac{j}{\lambda} = \frac{I}{\lambda S}.$$

Умножим это соотношение на элемент длины провода dl и проинтегрируем по участку провода от какой-либо точки *1* до другой точки *2*

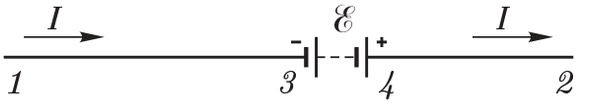


Рис. 113

(рис. 113). Поскольку ток один и тот же во всем проводе, величину I можно вынести из-под знака интеграла. Сделав это, найдем

$$\int\limits_{12} E\,dl + \int E^{\text{стор}}\,dl = I \int \frac{dl}{\lambda S}.$$

Так как электрическое поле стационарных токов потенциально, то первый интеграл выражается через разность потенциалов $\varphi_1 - \varphi_2$. Второй интеграл достаточно распространить на ту часть пути, где $E^{\text{стор}} \neq \neq 0$, т. е. на ту часть, которая проходит внутри источника тока. Этот интеграл не зависит от положения начальной и конечной точек *1* и *2*. Требуется только, чтобы эти точки находились вне источника тока. Ввиду потенциального характера поля $E^{\text{стор}}$ в области, где действуют сторонние силы, интеграл не зависит также от того, как проходит путь интегрирования через источник тока. Значит, этот интеграл есть величина, характеризующая свойства самого тока. Она называется *электродвижущей силой тока*:

$$\mathscr{E} = \int\limits_{12} E^{\text{стор}}\,dl = \int\limits_{34} E^{\text{стор}}\,dl. \tag{44.2}$$

Электродвижущая сила положительна, если путь *12* пересекает источник тока в направлении от отрицательного полюса к положительному. В противном случае она отрицательна [1]. Третий интеграл

$$R = \int \frac{dl}{\lambda S} = \int \rho \frac{dl}{S} \tag{44.3}$$

[1] Термодинамическая теория электродвижущей силы гальванического элемента изложена в т. II, § 49.

есть величина, характеризующая провод, по которому течет электрический ток. Эта величина называется *электрическим сопротивлением* или просто *сопротивлением провода.* Если провод изготовлен из однородного материала и всюду имеет одинаковую толщину, то получается известная формула

$$R = \rho \frac{l}{S}. \tag{44.4}$$

Таким образом,

$$\varphi_1 - \varphi_2 + \mathscr{E} = IR. \tag{44.5}$$

При выводе формулы (44.5) предполагалось, что провод тонкий на протяжении всей своей длины, включая участок, где находится гальванический элемент. Последнее условие, как правило, почти никогда не соблюдается. Тем не менее, как будет показано в конце настоящего параграфа, формула (44.5) справедлива и в этом случае.

2. Формула (44.5) выражает *закон Ома в интегральной форме* в отличие от соотношения (43.1), представляющего тот же закон в *локальной форме.* Эту формулу называют также *законом Ома для участка цепи.* Понятно, что R есть сопротивление *всего участка*, включая сопротивление самого элемента. Если участок не содержит источника тока (или, вообще, на нем не действуют сторонние силы), то формула (44.5) принимает вид

$$\varphi_1 - \varphi_2 = IR. \tag{44.6}$$

Разность потенциалов $\varphi_1 - \varphi_2$ называется в этом случае *электрическим напряжением* или просто *напряжением на концах рассматриваемого провода.* В общем случае электрическое напряжение определяется как интеграл $\int E\,dl$, взятый вдоль длины провода. Такое определение годится и в тех случаях, когда электрическое поле не потенциально.

Если начальная и конечная точки *1* и *2* совпадают, то $\varphi_1 - \varphi_2 = 0$ и формула (44.5) переходит в *закон Ома для всей* (*замкнутой*) *цепи*:

$$\mathscr{E} = IR. \tag{44.7}$$

Здесь R означает уже *полное сопротивление всей цепи.* Если $\varphi_{\text{а}}$ — потенциал анода, а $\varphi_{\text{к}}$ — катода, то $\varphi_{\text{а}} - \varphi_{\text{к}} = R_e I$, где R_e — сопротивление всего внешнего участка цепи. Сравнивая это соотношение с (44.7), получим

$$\frac{\varphi_{\text{а}} - \varphi_{\text{к}}}{\mathscr{E}} = \frac{R_e}{R} = \frac{R_e}{R_e + R_i}, \tag{44.8}$$

где R_i — внутреннее сопротивление самого элемента. Отсюда следует, что всегда, когда по цепи течет ток, разность потенциалов между полюсами элемента $\varphi_{\text{а}} - \varphi_{\text{к}}$ меньше электродвижущей силы $\mathscr{E}$. Только в предельном случае, когда $R_e \to \infty$ (а следовательно, $I \to 0$), получается $\mathscr{E} = \varphi_{\text{а}} - \varphi_{\text{к}}$. Значит, *электродвижущую силу можно определить как разность потенциалов между полюсами разомкнутого источника тока.*

3. Практической единицей тока, употребляемой в электротехнике, является *ампер* (А). Это есть сила такого тока, когда через поперечное сечение провода ежесекундно проходит один кулон электричества. Практическая единица напряжения, или разности потенциалов, есть *вольт* (В). Он определяется как такая разность потенциалов, при прохождении которой над зарядом в один кулон совершается работа в один джоуль. Практическая единица сопротивления есть *ом* (Ом), т. е. сопротивление такого провода, по которому потечет ток в один ампер, если на его концах поддерживать разность потенциалов в один вольт. Соответственно единица удельного сопротивления ρ будет Ом · см, а удельной проводимости — $\text{Ом}^{-1}\cdot\text{см}^{-1}$. Мы видим, что в практической системе единиц ρ уже не имеет размерности времени, как это было в гауссовой системе (см. § 41, п. 1). Очевидно,

$$1\ \text{Ом} = \frac{1\ \text{В}}{1\ \text{А}} = \frac{1/300}{3\cdot 10^9} = \frac{1}{9}\cdot 10^{-11}\ \text{СГСЭ-ед. сопротивления.}$$

4. Рассмотрим n проводов, соединенных параллельно (рис. 114). Будем предполагать, что в проводах действуют электродвижущие силы $\mathscr{E}_k$. Наша схема включает в себя как частный случай и парал-

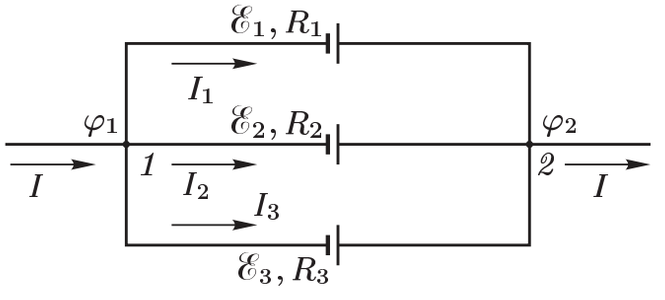


Рис. 114

лельное соединение элементов. Если R_k — сопротивление k-го провода (вместе с внутренним сопротивлением элемента), то ток в нем определяется выражением

$$I_k = \frac{\varphi_1 - \varphi_2}{R_k} + \frac{\mathscr{E}_k}{R_k}.$$

Сложив эти выражения, найдем полный ток, текущий в цепи:

$$I = \frac{\varphi_1 - \varphi_2}{R} + \frac{\mathscr{E}}{R}, \tag{44.9}$$

где введены обозначения

$$\frac{1}{R} = \sum \frac{1}{R_k}, \tag{44.10}$$

$$\mathscr{E} = R\sum \frac{\mathscr{E}_k}{R_k}. \tag{44.11}$$

Формула (44.10) определяет *сопротивление параллельно соединенных проводов.* Величина (44.11) играет роль электродвижущей силы. Для

того чтобы она не зависела от сопротивлений внешних проводов, достаточно понимать под R_k только внутренние сопротивления элементов. Тогда формула (44.11) будет определять *электродвижущую силу батареи параллельно соединенных элементов.* Если все элементы одинаковы, то $\mathscr{E} = \mathscr{E}_k$ т. е. *электродвижущая сила батареи равна электродвижущей силе одного элемента.*

Теперь видно, что для справедливости закона Ома (44.5) участки, содержащие гальванические элементы, не обязательно должны быть тонкими. Достаточно заметить, что область, обтекаемую током, всегда можно мысленно разделить на достаточно тонкие трубки тока, которые можно рассматривать как параллельно соединенные тонкие провода. Электродвижущие силы $\mathscr{E}_k$ в этих проводах одинаковы ввиду потенциального характера поля E в области, где действуют сторонние силы (см. конец предыдущего параграфа). Применив к рассматриваемой системе проводов соотношения (44.9)–(44.11), мы снова придем к закону Ома (44.5). При этом, ввиду одинаковости всех $\mathscr{E}_k$, величина $\mathscr{E}_k$ будет означать просто электродвижущую силу источника тока.

5. *Закон Джоуля–Ленца в интегральной форме* получается из *дифференциальной формы* этого закона интегрированием по объему провода. Представив элемент объема в виде $dV = S\,dl$, получим

$$Q = \int \frac{j^2}{\lambda} S\,dl = I^2 \int \frac{dl}{\lambda S} = I^2 R. \tag{44.12}$$

Это и есть интегральная форма закона Джоуля–Ленца. Формула (44.12) определяет тепло, выделяющееся ежесекундно в рассматриваемом участке провода. Если взять всю замкнутую цепь, то $Q = I\mathscr{E} = I \oint E^{\text{стор}}\,dl$. Отсюда видно, что *тепло производится одними только сторонними силами.* Роль электрического поля сводится к тому, что оно перераспределяет это тепло по различным участкам цепи.

ЗАДАЧИ

1. N одинаковых элементов с электродвижущей силой $\mathscr{E}$ и внутренним сопротивлением R_i разделены на m групп по n элементов в группе. Элементы в каждой группе соединены параллельно одноименными полюсами, а сами группы соединены между собой последовательно. При каком условии ток в цепи (при заданном внешнем сопротивлении R_e) будет максимальным?

Ответ. Ток максимален, если $R_e = mR_i/n$, т. е. когда внешнее сопротивление цепи равно внутреннему сопротивлению батареи. Для соблюдения этого условия необходимо, чтобы было $NR_i \geqslant R_e$. Сила максимального тока $I_{\text{макс}} = m\mathscr{E}/2R_e$.

2. Вычислить электродвижущую силу концентрационного элемента.

Решение. Из формулы (43.5) получаем

$$\mathscr{E} = \frac{kT}{e}\,\frac{B^- - B^+}{B^- + B^+}\,\ln\frac{n_1}{n_2}, \tag{44.13}$$

где n_1 и n_2 — концентрации электролита у электродов элемента. По порядку величины

$$\mathscr{E} = \frac{kT}{e} = \frac{1{,}38 \cdot 10^{-16} \cdot 300}{4{,}8 \cdot 10^{-10}} = 8{,}6 \cdot 10^{-5} \text{ СГСЭ-ед.} = 0{,}029 \text{ В}.$$

§ 45. Правила Кирхгофа

Рассмотрим произвольную разветвленную сеть проводов, в отдельных участках которой включены гальванические элементы или другие источники тока. Электродвижущие силы этих источников постоянны и предполагаются известными. Токи во всех участках цепи и разности потенциалов на них можно рассчитать с помощью закона Ома (44.5) и закона сохранения электрического заряда. Однако более просто задача решается с помощью двух *правил Кирхгофа*. Одно из них выражает *закон сохранения электрического заряда* для линейных проводов, а другое является *следствием закона Ома*. Сформулируем эти правила.

Первое правило Кирхгофа. *В каждой точке разветвления проводов алгебраическая сумма сил токов равна нулю* (рис. 115). *Токи, идущие к точке разветвления, и токи, исходящие из нее, следует считать величинами разных знаков.* Например, применительно к рис. 115 первое правило Кирхгофа запишется так:

$$I_1 + I_2 - I_3 = 0.$$

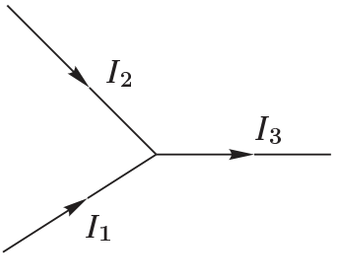


Рис. 115

Если бы это правило не соблюдалось, то в точках разветвления проводов накапливались бы электрические заряды, меняющиеся во времени. Вместе с ними менялось бы во времени и электрическое поле, а потому токи не могли бы оставаться постоянными.

Второе правило Кирхгофа. *Выделим в сети произвольный замкнутый контур, состоящий из проводов. Сумма электродвижущих сил, действующих в таком контуре, равна сумме произведений сил токов в отдельных участках этого контура на их сопротивления.*

Для доказательства достаточно рассмотреть случай, когда контур состоит из трех участков (рис. 116). Применяя к ним закон Ома (44.5), можем написать

$$\varphi_2 - \varphi_3 + \mathscr{E}_1 = I_1 R_1,$$

$$\varphi_3 - \varphi_1 + \mathscr{E}_2 = I_2 R_2,$$

$$\varphi_1 - \varphi_2 + \mathscr{E}_3 = I_3 R_3.$$

Складывая эти равенства, получим

$$\mathscr{E}_1 + \mathscr{E}_2 + \mathscr{E}_3 = I_1 R_1 + I_2 R_2 + I_3 R_3,$$

т. е. второе правило Кирхгофа.

Правила Кирхгофа в каждом конкретном случае позволяют написать *полную систему линейных уравнений*, из которой могут быть найдены все неизвестные токи. В нее совсем не входят неизвестные

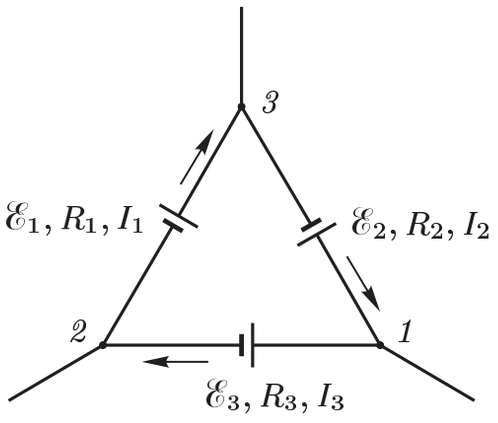


Рис. 116

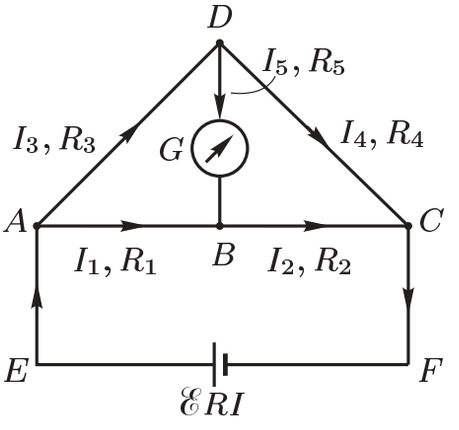


Рис. 117

разности потенциалов. В исключении потенциалов из уравнений для токов и состоит упрощение, вносимое правилами Кирхгофа по сравнению с законом Ома. При применении правил Кирхгофа надо поступать следующим образом:

1) Направления токов во всех участках сети следует обозначить стрелками, *не задумываясь над тем, куда эти стрелки направить. Если вычисление покажет, что ток положителен, то его направление указано правильно. Если же ток отрицателен, то его истинное направление противоположно направлению стрелки.*

2) Выбрав произвольный замкнутый контур, все его участки следует обойти в *одном направлении. Если это направление совпадает с направлением стрелки, то слагаемое RI берется со знаком плюс. Если же эти направления противоположны, то оно берется со знаком минус. Если при обходе контура источник тока проходится от отрицательного полюса к положительному, то его электродвижущую силу следует считать положительной; в противоположном случае ее надо считать отрицательной.*

3) *Все электродвижущие силы и все сопротивления проводов должны входить в систему уравнений.*

Рассмотрим два примера на правила Кирхгофа.

Пример 1. Мостик Уитстона (1802–1875). Схема мостика представлена на рис. 117. Расставим (произвольно) стрелки, указывающие направления токов. Имеется четыре точки разветвления: A, B, C, D. Применение к ним первого правила Кирхгофа приводит к четырем уравнениям:

$$\begin{aligned} &I - I_1 - I_3 = 0, \quad I_1 - I_2 + I_5 = 0, \\ &I_2 + I_4 - I = 0, \quad I_3 - I_4 - I_5 = 0. \end{aligned} \tag{45.1}$$

Из этих уравнений независимы *только три* (при сложении всех уравнений получается тождество $0 = 0$). Для определения шести неизвестных

I, I_1, I_2, I_3, I_4, I_5 требуется *еще три* уравнения. Их дает второе правило Кирхгофа. Применяя его, можно брать разные контуры. Но один из них *должен обязательно содержать источник тока* (с электродвижущей силой $\mathscr{E}$). Можно, например, взять контуры ABD, BDC и $ABCFEA$. Для них второе правило Кирхгофа дает

$$\begin{aligned} I_1R_1 - I_5R_5 - I_3R_3 &= 0, \\ I_2R_2 - I_4R_4 + I_5R_5 &= 0, \\ IR + I_1R_1 + I_2R_2 &= \mathscr{E}. \end{aligned} \tag{45.2}$$

Здесь R — сопротивление участка $CFEA$, включая внутреннее сопротивление источника тока. Использование других контуров не дает новых независимых уравнений. Решая уравнения (45.1) совместно с уравнениями (45.2), можно вычислить все токи. Ограничимся выводом условия, при котором ток в мостике I_5 обращается в нуль. Если $I_5 = 0$, то из уравнений (45.1) следует $I_1 = I_2$, $I_3 = I_4$. После этого из первых двух уравнений (45.2) находим

$$I_1R_1 = I_3R_3, \quad I_1R_2 = I_3R_4.$$

Отсюда почленным делением получаем известное условие

$$\frac{R_1}{R_2} = \frac{R_3}{R_4},$$

на котором основано применение мостика Уитстона для измерения сопротивлений проводов. Ветвь AC (реохорд) изготовляется из длинной однородной проволоки с большим удельным сопротивлением, так что отношение R_1/R_2 можно заменить отношением длин AB/BC.

Пример 2. Сравнение электродвижущих сил элементов методом Поггендорфа (1796–1877). Схема опыта изображена на рис. 118. Предполагается, что $\mathscr{E}_1 > \mathscr{E}_2$. Первое правило Кирхгофа дает

$$I_1 - I_2 - i_1 = 0,$$

а второе

$$\begin{aligned} I_1(R_1 + r_2) + i_1r_1 &= \mathscr{E}_1, \\ i_1r_1 - I_2R_2 &= \mathscr{E}_2. \end{aligned}$$

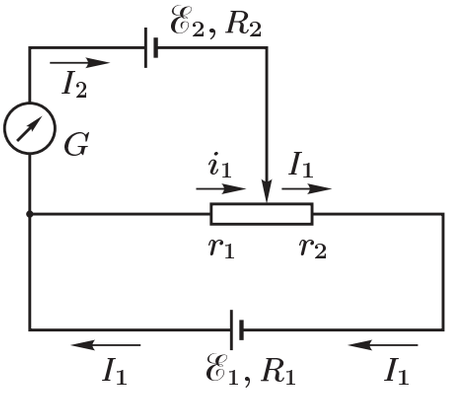


Рис. 118

Под R_2 понимается сумма внутреннего сопротивления элемента *2* и сопротивления гальванометра. Найдем условие, при котором ток через гальванометр не пойдет. Если $I_2 = 0$, то $I_1 = i_1$ и, следовательно,

$$I_1(R_1 + r_1 + r_2) = \mathscr{E}_1, \quad I_1r_1 = \mathscr{E}_2.$$

Отсюда получаем искомое условие

$$\frac{\mathscr{E}_2}{\mathscr{E}_1} = \frac{r_1}{r_1 + r_2 + R_1}. \tag{45.3}$$

В это уравнение входит неизвестное сопротивление R_1. Для его исключения применяется следующий метод. Опыт повторяют, заменив элемент *2* элементом *3* с электродвижущей силой $\mathscr{E}_3 < \mathscr{E}_1$, оставляя все остальные параметры схемы неизменными. Ток через гальванометр не пойдет при новых значениях сопротивлений r_1 и r_2. Обозначая их через r_1' и r_2', можем написать $\mathscr{E}_3/\mathscr{E}_1 = r_1'/(r_1' + r_2' + R_1)$. Так как $r_1 + r_2 = r_1' + r_2'$, то отсюда и из условия (45.3) получаем $\mathscr{E}_3/\mathscr{E}_2 = r_1'/r_1$. Как и в предыдущей схеме, отношение r_1'/r_1 можно свести к отношению длин и после этого вычислить неизвестное отношение электродвижущих сил $\mathscr{E}_3/\mathscr{E}_2$.

§ 46. Стационарные токи в массивных проводниках

1. Пусть в проводящую однородную среду помещены два электрода A и B (рис. 119), электрическая проводимость которых очень велика по сравнению с проводимостью самой среды. В этих условиях изменениями потенциала внутри электродов можно пренебречь, т. е. считать, что все точки каждого электрода находятся при одном и том

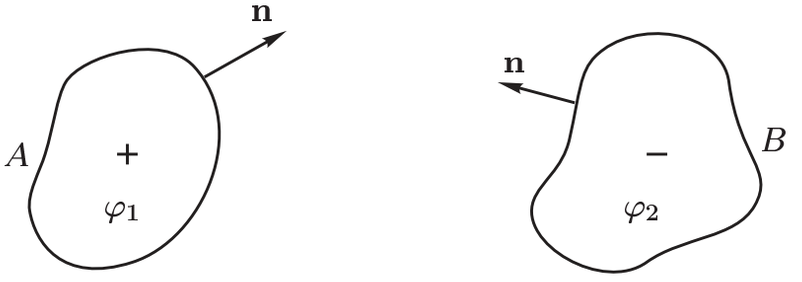


Рис. 119

же потенциале. Будем поддерживать потенциалы электродов φ_1 и φ_2 постоянными. Поставим задачу о нахождении плотности тока, текущего между электродами. Эта задача математически эквивалентна электростатической задаче о поле конденсатора, обкладками которого служат рассматриваемые электроды. Действительно, поскольку токи стационарны, а среда однородна, в ней не могут появиться объемные электрические заряды. Значит, потенциал φ во всем пространстве между электродами должен удовлетворять уравнению Лапласа $\Delta\varphi = 0$ и принимать заданные значения φ_1 и φ_2 на электродах A и B. Задача об электрическом поле в конденсаторе формулируется в точности так же. Обе задачи математически тождественны и имеют единственные решения. Поэтому, если пространство между электродами A и B заполнить однородным диэлектриком и поддерживать электроды при прежних потенциалах φ_1 и φ_2, то электрическое поле во всем пространстве вне электродов A и B останется неизменным. Найдя напряженность электрического поля $\mathbf{E}$, можно вычислить затем и плотность тока по формуле $\mathbf{j} = \lambda\mathbf{E}$.

Вычисление полного тока I между электродами — задача еще более простая. Она сводится к вычислению емкости соответствующего конденсатора. Действительно, ток выражается интегралом

$$I = \oint\!\!\!\!\!\int j_n \, dS = \lambda \oint\!\!\!\!\!\int E_n \, dS,$$

распространенным по поверхности положительно заряженного электрода A. Если задаче о токе сопоставить соответствующую электростатическую задачу, то по теореме Гаусса

$$\oint\!\!\!\!\!\int E_n \, dS = \frac{4\pi}{\varepsilon} \, q = \frac{4\pi}{3} \, C(\varphi_1 - \varphi_2),$$

где ε — диэлектрическая проницаемость межэлектродного пространства, а C — емкость конденсатора с обкладками A и B. Таким образом,

$$I = \frac{4\pi C\lambda}{3} \, (\varphi_1 - \varphi_2).$$

С другой стороны, по закону Ома

$$I = \frac{\varphi_1 - \varphi_2}{R},$$

где R — полное электрическое сопротивление проводящей среды между электродами A и B. Сравнивая обе формулы, получаем

$$R = \frac{\varepsilon}{4\pi C\lambda}, \tag{46.1}$$

после чего находим ток I. Как и следовало ожидать, сопротивление R не зависит от ε, так как емкость конденсатора C сама пропорциональна ε. Не меняя результата, можно было бы взять $\varepsilon = 1$.

Приведем конкретные примеры.

1) Электродами служат концентрические сферы с радиусами a и b. Используя формулу (26.5), получим

$$R = \frac{1}{4\pi\lambda} \left(\frac{1}{a} - \frac{1}{b} \right). \tag{46.2}$$

2) Электродами являются поверхности коаксиальных цилиндров с высотой l и радиусами a и b. Формулы (46.1) и (26.7) дают

$$R = \frac{1}{2\pi\lambda l} \, \ln \frac{b}{a}. \tag{46.3}$$

3) Электродами являются два шарика радиуса a, удаленные друг от друга на большое расстояние. В этом случае

$$R = \frac{1}{2\pi\lambda a}. \tag{46.4}$$

Сопротивление R не зависит от расстояния между шариками. Это объясняет результат, эмпирически найденный телеграфистами, обнаружившими, к своему удивлению, что сопротивление земли между телеграфными станциями не зависит от расстояния между ними.

2. Рассмотрим задачу о сопротивлении заземления в более общей постановке. На рис. 120 схематически изображены две станции *1* и *2*. Связь между ними осуществляется проводом *12*. Другим проводом слу-

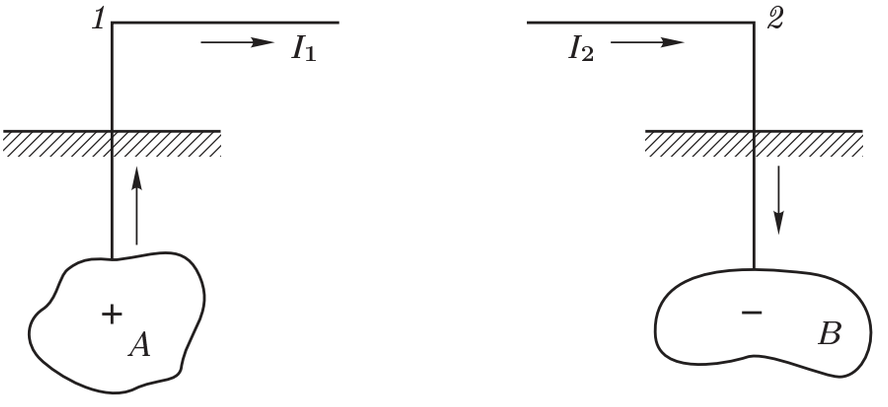


Рис. 120

жит земля. Вблизи каждого из заземленных металлических тел (электродов) A и B почва может считаться однородной средой с удельными проводимостями λ_1 и λ_2. Будем предполагать, что электроды A и B зарыты глубоко, так что влиянием границы между землей и атмосферой на общий ток, текущий между электродами, можно пренебречь. Пусть C_1 и C_2 — емкости электродов A и B, какими они были бы, если бы электроды были уединены и находились в вакууме. Ток I, стекающий с положительного электрода A, зависит только от его потенциала φ_1 и электрической проводимости почвы в окрестности этого электрода. Он практически не зависит от свойств удаленных областей окружающей среды, а потому при вычислении I_1 электрическую проводимость всей среды можно считать постоянной и равной λ_1. Если потенциал бесконечно удаленных точек условиться считать равным нулю, то на основании формулы (46.1) можно написать

$$I_1 = \frac{\varphi_1}{R_1} = 4\pi C_1 \lambda_1 \varphi_1.$$

(Мы приняли $\varepsilon = 1$.) Аналогично, для тока I_2, текущего к отрицательному электроду B, получаем

$$I_2 = -\frac{\varphi_2}{R_2} = -4\pi C_2 \lambda_2 \varphi_2.$$

Станции *1* и *2* соединены между собой проводом *12*, а потому $I_1 = I_2 = I$. Используя это, из предыдущих формул находим

$$\varphi_1 - \varphi_2 = \frac{I}{4\pi}\left(\frac{1}{\lambda_1 C_1} + \frac{1}{\lambda_2 C_2}\right),$$

откуда следует, что сопротивление заземления равно

$$R = \frac{1}{4\pi}\left(\frac{1}{\lambda_1 C_1} + \frac{1}{\lambda_2 C_2}\right). \tag{46.5}$$

В выводе нигде не предполагалось, что весь ток с электрода A, текущий по земле, попадает на электрод B. Такое утверждение неправильно, хотя оно не отразилось бы на формуле (46.5). Было использовано только равенство токов I_1 и I_2, а оно обеспечивается тем, что электроды A и B соединены между собой проводом. Формула (46.5) показывает, что для получения хорошего заземления электроды A и B должны быть больших размеров. Кроме того, окружающая их почва должна обладать хорошей электрической проводимостью.

ЗАДАЧА

Имеется n идеально проводящих тел в вакууме с зарядами $q_1, q_2, \ldots, q_n$ и потенциалами $\varphi_1, \varphi_2, \ldots, \varphi_n$. Какое количество теплоты будет выделяться ежесекундно, если пространство между этими телами заполнить однородной жидкостью с электрической проводимостью λ и диэлектрической проницаемостью ε, а потенциалы тел поддерживать при прежних значениях?

Р е ш е н и е. При заполнении пространства проводящей жидкостью электрическое поле между телами не изменится. Количество же ежесекундно выделяющейся теплоты определится выражением $Q = \sum I_k \varphi_k$. Ток, текущий с поверхности k-го проводника,

$$I_k = \oint j_N \, dS,$$

где N — наружная нормаль к этой поверхности. Заряд на поверхности k-го проводника

$$q_k = \frac{1}{4\pi}\oint D_N \, dS = \frac{\varepsilon}{4\pi}\oint E_N \, dS.$$

В результате находим

$$Q = \frac{4\pi\lambda}{\varepsilon}\sum q_k \varphi_k.$$

§ 47. Электролитическая ванна

1. При конструировании электронных, ионных и многих других приборов надо знать электрические поля между электродами сложной конфигурации. Теоретический расчет таких полей практически невозможен. Экспериментальные измерения полей внутри самих приборов не всегда могут дать достаточную точность ввиду малости размеров отдельных деталей, вблизи которых необходимо измерить распределение поля, а также потому, что многие места внутри конструируемого прибора недоступны для введения зонда. Для экспериментального решения этой задачи применяется *метод электролитической ванны.* Изготовляются увеличенные подобные и подобно расположенные модели электродов, погружаемые затем в однородную слабо проводящую

жидкость (электролит), например в водопроводную воду. Потенциалы электродов модели должны быть пропорциональны потенциалам соответствующих электродов прибора. При этом условии модель воспроизведет в увеличенном масштабе эквипотенциальные поверхности и силовые линии электрического поля заряженных электродов. Поскольку исследуемое пространство теперь заполнено проводящей средой, измерение потенциалов легко осуществить с помощью зонда.

2. При практическом применении метода электролитической ванны возникают экспериментальные трудности. Одна из них заключается в следующем. Размеры ванны должны быть велики по сравнению с размерами исследуемой системы электродов. Сами электроды должны быть погружены в ванну глубоко, чтобы наличие свободной поверхности электролита существенно не искажало исследуемое поле. Но тогда ванна приняла бы слишком большие размеры, затрудняющие экспериментирование с ней. Потребовались бы какие-то приспособления для введения и удержания зонда в нужных точках пространства, а они привели бы к заметным искажениям исследуемого поля. Кроме того, некоторые существенные области изучаемого поля могли бы оказаться недоступными для введения зонда, например все части пространства, полностью окруженные замкнутой металлической оболочкой. Существует изящный способ преодоления подобных трудностей, называемый *методом сечений*. Метод сечений применим для исследования полей, обладающих осевой симметрией. Это — наиболее важный случай, встречающийся на практике. Очевидно, достаточно исследовать распределение электрического потенциала в любой плоскости симметрии, проходящей через ось модели.

Допустим, что исследуемая модель электродов погружена в электролит, заполняющий все бесконечное пространство. Рассечем ее мысленно горизонтальной плоскостью, проходящей через ось симметрии модели. Все силовые линии и линии тока, проходящие через любую точку рассматриваемого сечения, не выходят из его плоскости. Поэтому удаление верхней половины модели и электролита, заполняющего верхнее полупространство, никак не скажется на распределении потенциала и электрическом поле во всем нижнем полупространстве и в самом сечении.

Это может быть доказано и строго математически. Действительно, после удаления всех частей модели и электролита из верхнего полупространства образуется свободная плоская поверхность, ограничивающая нижнее полупространство. В стационарном состоянии токи могут течь вдоль самой границы, но не могут пересекать ее. Так как $\mathbf{j} = \lambda \mathbf{E}$, то отсюда следует, что на границе нормальная составляющая вектора $\mathbf{E}$ равна нулю. Потенциалы всех электродов заданы, в нижнем полупространстве потенциал удовлетворяет уравнению Лапласа $\Delta\varphi = 0$. Этими условиями потенциал и поле $\mathbf{E}$ в нижнем полупространстве определяются однозначно. Но в точности таким же условиям эти величины удовлетворяли и до того, как были удалены

из верхнего полупространства электролит и все верхние части разрезанной модели, так как и в этом случае поле **E** в рассматриваемом сечении не имело нормальной составляющей. Этим наше утверждение доказано.

Теперь ясно, что для воспроизведения изучаемого поля нет необходимости применять цельные электроды. Достаточно взять их половинки, получающиеся разрезанием цельных электродов по плоскости, проходящей через ось симметрии. Это и используется в методе сечений. Изготовленные половинки электродов погружаются в ванну таким образом, чтобы плоскость разреза совместилась со свободной поверхностью электролита. Распределение потенциала достаточно измерить на одной только свободной поверхности, что экспериментально очень удобно.

3. На рис. 121 приведена принципиальная схема электролитической ванны. Электрическое поле образуется между электродами A и B. Зондом S может служить кончик тонкой металлической проволоки. Гальванометр включается между зондом S и подвижной точкой C де-

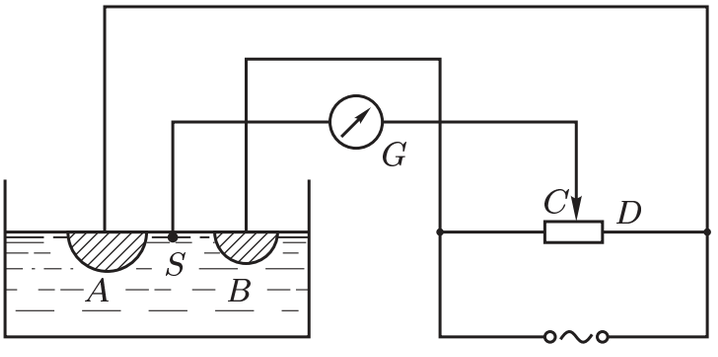


Рис. 121

лителя напряжения. Вместо гальванометра лучше пользоваться электронным осциллографом, соединив точки S и C, например, с его вертикальными пластинами. Перемещая зонд по поверхности электролита, находят такие положения его, при которых ток через гальванометр обращается в нуль. Во всех этих положениях потенциал зонда равен потенциалу точки C делителя напряжения. Все отмеченные точки лежат, таким образом, на линии постоянного потенциала — эквипотенциальной линии. Движение зонда с помощью специальных рычагов передается штифту пантографа, который отмечает положения соответствующих точек на листе бумаги. Соединяя отмеченные точки плавной кривой, находят эквипотенциальную линию в увеличенном масштабе. Затем повторяют то же самое при другом положении движка C, находят другую эквипотенциальную линию и т. д.

Если на электроды подавать постоянное напряжение, то по электролиту будут течь постоянные токи, сопровождающиеся электролизом и выделением на электродах составных частей электролита. Эти токи нарушали бы однородность электролита, приводя к поляризации электродов и к изменению напряжения между ними. Во избежание

этого на электроды подают *переменное напряжение*, например, от клемм городского тока. Условия, при выполнении которых результаты, полученные выше для постоянных полей, могут быть распространены на переменные поля, будут сформулированы ниже (см. задачу к § 144).

§ 48. Процессы установления тока при зарядке и разрядке конденсатора

1. Задачи о разрядке и зарядке конденсатора, строго говоря, выходят за рамки учения о постоянных токах. Приводимые ниже решения их получаются в предположении, что мгновенное значение тока одно и то же во всех поперечных сечениях провода, соединяющего обкладки конденсатора, а мгновенное электрическое поле такое же, как в электростатике при тех же зарядах на обкладках конденсатора. Токи и поля, удовлетворяющие этим условиям, называются *квазистационарными*. В дальнейшем будут точно сформулированы условия, при которых имеет место *квазистационарность* (см. § 123).

Если обкладки заряженного конденсатора соединить проводом, то по проводу потечет ток. Пусть I, Q, φ — мгновенные значения тока, заряда положительной обкладки и разности потенциалов между обкладками. Считая ток в проводе положительным, когда он течет от положительной обкладки к отрицательной, можем написать

$$I = -\frac{dQ}{dt}, \quad RI = \varphi, \quad Q = C\varphi,$$

где C — емкость конденсатора, а R — сопротивление провода. Исключая φ и I, получим

$$\frac{dQ}{dt} + \frac{Q}{RC} = 0. \tag{48.1}$$

После интегрирования этого уравнения придем к соотношению

$$Q = Q_0 e^{-t/\tau}, \tag{48.2}$$

где Q_0 — начальное значение заряда конденсатора, а τ — постоянная:

$$\tau = RC, \tag{48.3}$$

имеющая размерность времени. Она называется *временем релаксации*. Через время τ заряд конденсатора убывает в e раз. Поэтому τ по порядку величины равно времени, в течение которого конденсатор разрядится. Дифференцируя (48.2) по t, находим закон изменения тока во времени:

$$I = \frac{Q_0}{\tau} \exp\left\{-\frac{t}{\tau}\right\} = I_0 \exp\left\{-\frac{t}{\tau}\right\}, \tag{48.4}$$

где $I_0 = Q_0/\tau$ — начальное значение тока, т. е. ток при $t = 0$.

Полезно запомнить, что время релаксации определяется формулой (48.3) *во всех системах единиц.* Так, если сопротивление измерять в омах, а емкость в фарадах, то время (48.3) получится в секундах.

2. Аналогично решается задача о зарядке конденсатора. Пусть в цепь конденсатора включена гальваническая батарея или какой-либо другой источник тока с постоянной электродвижущей силой $\mathscr{E}$. Источник возбуждает ток, заряжающий конденсатор. Электрические заряды на обкладках конденсатора препятствуют прохождению тока и уменьшают его. Уравнение $Q = C\varphi$ остается неизменным. Остальные два уравнения запишутся следующим образом:

$$I = \frac{dQ}{dt}, \quad RI = \mathscr{E} - \varphi,$$

где R означает полное сопротивление провода, соединяющего обкладки конденсатора, включая внутреннее сопротивление батареи. (Теперь ток в проводе считается положительным, когда он течет в направлении к положительной обкладке.) Исключая снова I и φ, придем к уравнению

$$\frac{dQ}{dt} + \frac{Q}{RC} = \frac{\mathscr{E}}{R}. \tag{48.5}$$

Это неоднородное уравнение сведется к однородному, если его записать в виде

$$\frac{d}{dt}(Q - \mathscr{E}C) + \frac{Q - \mathscr{E}C}{RC} = 0.$$

Решая это уравнение, получим

$$Q - \mathscr{E}C = Ae^{-t/\tau}.$$

Значение постоянной интегрирования A найдется из условия, что в начальный момент времени конденсатор не заряжен, т. е. в этот момент $Q = 0$. Это дает $A = -\mathscr{E}C$, и, следовательно,

$$Q = \mathscr{E}C(1 - e^{-t/\tau}). \tag{48.6}$$

При $t \to \infty$ заряд Q стремится к предельному значению $Q_\infty = \mathscr{E}C$. Для тока получаем

$$I = \frac{\mathscr{E}C}{\tau}\, e^{-t/\tau} = \frac{\mathscr{E}}{R}\, e^{-t/\tau}. \tag{48.7}$$

Ток максимален в начальный момент и равен $I_0 = \mathscr{E}/R$. В дальнейшем он убывает по экспоненциальному закону.

3. Если пространство между заряженными электродами заполнено непрерывной средой с электрической проводимостью λ и диэлектрической проницаемостью ε, то можно воспользоваться формулой (46.1). Тогда для времени релаксации получим

$$\tau = RC = \frac{\varepsilon}{4\pi\lambda} = \frac{\varepsilon\rho}{4\pi}. \tag{48.8}$$

Для меди $\rho = 1{,}7 \cdot 10^{-6}$ Ом $\cdot$ см $= 1{,}9 \cdot 10^{-18}$ с. Формальная подстановка в формулу (48.8) дала бы $\tau \approx 10^{-19}$ с, что примерно в сто тысяч раз меньше инерционного времени электронов (см. § 42, п. 5). Это показывает, что формула (48.8) к металлам неприменима: за время 10^{-19} с электрическое состояние не может распространиться более чем на $3 \cdot 10^{10} \cdot 10^{-19} = 3 \cdot 10^{-9}$ см. Значит, условие квазистационарности, использованное при выводе формулы (48.8), в случае металлов не выполняется. Заряженное тело, окруженное металлом, будет терять заряд значительно медленнее (но все же практически мгновенно), чем это следует из формулы (48.8). Практически область применимости формулы (48.8) ограничивается только диэлектриками. Идеальных изоляторов, совершенно не проводящих электричество, не существует. Однако их электрическая проводимость на много порядков меньше, чем у металлов. Так, для эбонита $\rho = 10^{16}$ Ом $\cdot$ см $= 1{,}1 \cdot 10^{4}$ с, что превосходит удельное сопротивление меди в 10^{22} раз. Диэлектрическая проницаемость эбонита $\varepsilon = 2{,}7$, а время релаксации $\tau = 2{,}4 \cdot 10^{3}$ с $=$ $=$ 40 мин.

Глава III

МАГНИТНОЕ ПОЛЕ

§ 49. Силы, действующие в магнитном поле на движущиеся заряды и токи

1. Магнитные явления были ранее всего обнаружены и изучены на естественных и искусственных магнитах. Да и теперь первоначальное знакомство с этими явлениями лучше всего начинать с магнита. Однако понимание процессов, происходящих в магните, требует предварительного изучения более простых и фундаментальных явлений. Вот почему при изложении современного учения о магнетизме мы не можем следовать историческому пути. Мы изберем дедуктивный метод, положив в основу изложения два экспериментальных факта, которые были установлены в XIX веке: 1) *магнитное поле действует на движущиеся заряды*; 2) *движущиеся заряды создают магнитное поле.* Как и в электростатике, сначала изучим магнитное поле в вакууме, а затем перейдем к изучению магнитного поля в веществе.

2. Начнем с простейшего демонстрационного опыта. В осциллографической трубке получим прямолинейный сфокусированный пучок электронов, движущихся в вакууме слева направо. Попадая на флуоресцирующий экран, пучок оставляет след в виде светящегося пятнышка. Поднесем снизу к пучку северный полюс прямолинейного магнита (рис. 122). Пучок сместится вбок по направлению от читателя. Если магнит поднести южным полюсом, смещение произойдет в про-

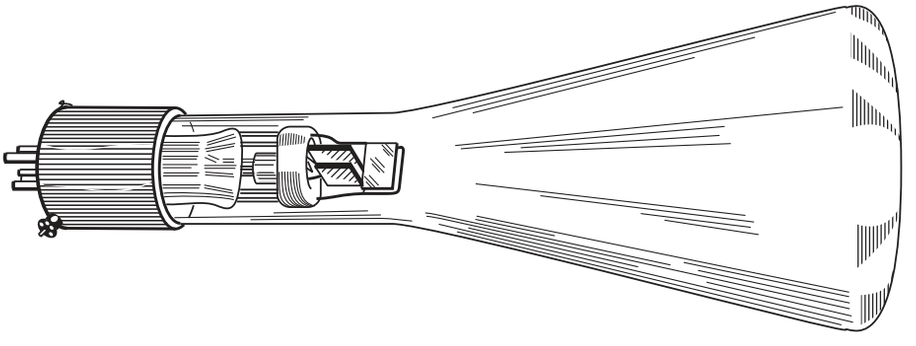


Рис. 122

тивоположную сторону, т. е. к читателю. Если магнит поднести сбоку, то смещение будет происходить вверх или вниз в зависимости от того, каким полюсом и с какой стороны поднесен магнит. Этот и аналогичные опыты показывают, что на движущийся электрон действует сила, перпендикулярная к скорости электрона и к направлению оси магнита,

т. е. к прямой, идущей от одного полюса магнита к другому. Эта сила пропорциональна скорости электрона. Аналогично ведут себя всякие частицы, движущиеся в магнитном поле.

Закон, определяющий силу $\mathbf{F}_m$, действующую на движущийся точечный заряд q в магнитном поле, получен обобщением опытных фактов. Он выражается формулой

$$\mathbf{F}_m = \frac{q}{c}\,[\mathbf{v}\mathbf{B}], \tag{49.1}$$

где вектор $\mathbf{B}$ не зависит от величины заряда q и его движения. Он характеризует только магнитное поле, в котором движется заряд q. Вектор $\mathbf{B}$ называется *напряженностью магнитного поля*. Сила $\mathbf{F}_m$ перпендикулярна как к скорости частицы $\mathbf{v}$, так и к вектору $\mathbf{B}$, а ее величина пропорциональна синусу угла между этими векторами. Когда векторы $\mathbf{B}$ и $\mathbf{v}$ коллинеарны, сила $\mathbf{F}_m$ обращается в нуль. Формула (49.1) справедлива не только для постоянных, но и для переменных магнитных полей, и притом при любых значениях скорости $\mathbf{v}$.

Постоянную c можно выбрать произвольно. Выбором числового значения и размерности этой постоянной определяется *система единиц*. Целесообразно величине c приписать размерность скорости, так как тогда размерности электрического и магнитного полей будут одинаковы. Именно так поступают в *гауссовой системе единиц*. Числовое значение c мы временно оставим неопределенным.

Отметим особо, что *на покоящийся заряд магнитное поле не действует*. В этом существенное отличие магнитного поля от поля электрического. Индикатором электрического поля служит *покоящийся заряд*, индикатором магнитного поля — *движущийся заряд*.

Формула (49.1) указывает принципиальный способ измерения магнитного поля $\mathbf{B}$ по силе, действующей на движущийся заряд. Для этого с помощью неподвижного заряда надо сначала убедиться, что электрического поля нет. Затем надо найти такое направление скорости $\mathbf{v}$, при котором сила $\mathbf{F}_m$ обращается в нуль. Это будет происходить тогда, когда скорость $\mathbf{v}$ параллельна или антипараллельна вектору $\mathbf{B}$. Тем самым с точностью до знака определится направление магнитного поля $\mathbf{B}$. Наконец, надо измерить силу $\mathbf{F}_m$, когда заряд движется перпендикулярно к $\mathbf{B}$ с какой-то скоростью $\mathbf{v}_\perp$. Очевидно,

$$\mathbf{F}_m = \frac{q}{c}\,[\mathbf{v}_\perp\mathbf{B}].$$

Умножая это соотношение векторно на $\mathbf{v}_\perp$ и принимая во внимание, что $(\mathbf{v}_\perp\mathbf{B}) = 0$, получим

$$\mathbf{B} = \frac{c}{q\mathbf{v}_\perp^2}\,[\mathbf{F}_m\mathbf{v}_\perp]. \tag{49.2}$$

Этой формулой вектор $\mathbf{B}$ определяется однозначно и по модулю, и по направлению. Что величина $\mathbf{B}$ есть вектор (точнее, псевдовектор) — это непосредственно следует из формулы (49.2), представляющей эту величину в виде векторного произведения полярных векторов $\mathbf{F}_m$ и $\mathbf{v}_\perp$.

3. В электрическом поле $\mathbf{E}$ на заряд q действует сила $\mathbf{E}_e = q\mathbf{E}$. Если электрическое и магнитное поля действуют независимо, а такое предположение согласуется с опытными фактами, то при совместном действии электрического и магнитного полей возникает сила $\mathbf{F} = \mathbf{F}_e + \mathbf{F}_m$, т. е.

$$\mathbf{F} = q\left(\mathbf{E} + \frac{1}{c}[\mathbf{vB}]\right). \tag{49.3}$$

Она называется *силой Лоренца*.

В нерелятивистском приближении сила $\mathbf{F}$, как и всякая другая сила, не зависит от выбора (инерциальной) системы отсчета. Между тем второе слагаемое в формуле (49.3) меняется при переходе от одной системы отсчета к другой. Поэтому должно меняться и первое слагаемое $q\mathbf{E}$. Таким образом, как мы уже подчеркивали (см. § 2, п. 5), разделение полной силы $\mathbf{F}$ на электрическую и магнитную зависит от выбора системы отсчета. Без указания системы отсчета такое разделение теряет смысл.

4. Опыты по действию магнитного поля на движущиеся заряды проще производить не с отдельными зарядами, а с электрическими токами, когда в движение вовлекается сразу очень много заряженных частиц. Допустим, например, что ток создается движением одинаковых частиц с зарядом e и концентрацией n. Тогда $\mathbf{j} = ne\mathbf{v}$. Число частиц в объеме dV будет $dN = n\,dV$, а сила, действующая в магнитном поле на элемент объема тела dV,

$$d\mathbf{F} = \frac{e}{c}[\mathbf{vB}]\,dN = \frac{ne}{c}[\mathbf{vB}]\,dV,$$

или

$$d\mathbf{F} = \frac{1}{c}[\mathbf{j}\,\mathbf{B}]\,dV. \tag{49.4}$$

Конечно, это выражение справедливо и в более общем случае, когда носителями тока являются разные заряды.

Рассмотрим частный случай, когда ток I течет вдоль бесконечно тонкого провода с площадью поперечного сечения S. Возьмем бесконечно короткий участок провода длины dl и вычислим действующую на него силу $d\mathbf{F}$. Если $dV = S\,dl$ — объем этого участка, то $\mathbf{j}\,dV = j\,S\,d\mathbf{l}$, или

$$\mathbf{j}\,dV = I\,d\mathbf{l}, \tag{49.5}$$

причем направление вектора $d\mathbf{l}$ совпадает с направлением тока. Вектор $\mathbf{j}\,dV$ называется *объемным*, а $I\,d\mathbf{l}$ — *линейным элементом тока*. Из соотношений (49.4) и (49.5) получаем

$$d\mathbf{F} = \frac{I}{c}[d\mathbf{l}\,\mathbf{B}]. \tag{49.6}$$

Формула (49.6), определяющая силу, действующую в магнитном поле на линейный элемент тока, была установлена Ампером и носит название *закона Ампера*. Сила, действующая на провод конечной длины,

найдется из (49.6) интегрированием по всей длине провода:

$$\mathbf{F} = \int \frac{1}{c}\,[d\mathbf{l}\,\mathbf{B}]. \tag{49.7}$$

Силы, действующие на токи в магнитных полях, называются *амперовыми силами.*

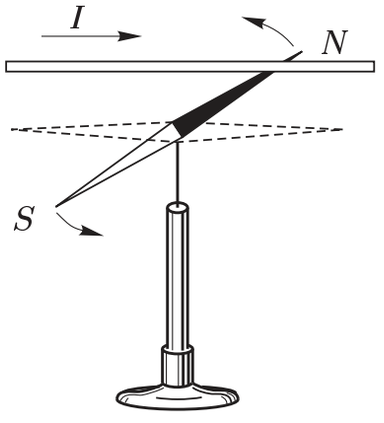


Рис. 123

Итак, *магниты действуют на электрические токи. Токи также действуют на магниты.* Примером может служить классический *опыт Эрстеда* (1777–1851). Эрстед поместил над магнитной стрелкой прямолинейный провод (рис. 123), параллельный стрелке. Стрелка могла свободно вращаться вокруг вертикальной оси. При пропускании по проводу электрического тока стрелка отклонялась в сторону и устанавливалась перпендикулярно к проводу. При изменении направления тока стрелка поворачивалась на 180°. То же самое происходило, когда провод переносился вниз и располагался под стрелкой. Опыт Эрстеда был произведен в 1820 г. На нем впервые была установлена связь между электрическими и магнитными явлениями.

§ 50. Магнитное поле равномерно движущегося заряда. Закон Био и Савара

1. Сформулируем теперь закон, определяющий магнитное поле движущегося точечного заряда q, ограничиваясь при этом равномерными движениями с малыми скоростями. (Точное определение медленности движения будет приведено ниже.) Такой закон был получен обобщением опытных фактов и выражается формулой

$$\mathbf{B} = \frac{q}{c' r^3}\,[\mathbf{vr}],$$

где $\mathbf{r}$ — радиус-вектор, проведенный от заряда q к точке наблюдения, а c' — коэффициент пропорциональности, зависящий от выбора единиц. Электрическое поле неподвижного заряда той же величины q в той же точке наблюдения определяется выражением

$$\mathbf{E} = \frac{q}{r^3}\,\mathbf{r}. \tag{50.1}$$

(Такое же выражение годится и для медленно движущихся зарядов.) С использованием этого выражения предыдущая формула может быть записана в виде $\mathbf{B} = (1/c')[\mathbf{vE}]$. В гауссовой системе единиц величины $\mathbf{B}$ и $\mathbf{E}$ имеют одинаковую размерность. Поэтому постоянная c' должна в этой системе иметь размерность скорости. Для простоты она выбирается равной постоянной c, введенной в предыдущем параграфе.

Условием $c' = c$, как будет видно из дальнейшего изложения, однозначно определяется и числовое значение c. Так определенная скорость c называется *электродинамической постоянной.* Измерения показали, что она совпадает со *скоростью света в вакууме.* Медленность движения, о которой говорилось выше, надо понимать в том смысле, что скорость v должна быть очень мала по сравнению с величиной c ($v \ll c$).

Таким образом,

$$\mathbf{B} = \frac{q}{cr^3}[\mathbf{vr}], \tag{50.2}$$

или

$$\mathbf{B} = \frac{1}{c}[\mathbf{vE}]. \tag{50.3}$$

2. Используем приведенные формулы для вычисления сил взаимодействия двух движущихся точечных зарядов q_1 и q_2. Это взаимодействие складывается из *электрического* (по закону Кулона) и *магнитного.* Ниже речь идет только о магнитном взаимодействии. Пусть $\mathbf{v}_1$ и $\mathbf{v}_2$ означают скорости движущихся зарядов. Магнитное поле, создаваемое зарядом q_1 в точке нахождения заряда q_2, будет

$$\mathbf{B}_1 = \frac{q_1}{cr_{12}^3}[\mathbf{v}_1\mathbf{r}_{12}],$$

где $\mathbf{r}_{12}$ — радиус-вектор, проведенный от первого заряда ко второму. На заряд q_2 это поле действует с силой

$$\mathbf{F}_{12} = \frac{q_2}{c}[\mathbf{v}_2\mathbf{B}_1] = \frac{q_1q_2}{c^2r_{12}^3}[\mathbf{v}_2[\mathbf{v}_1\mathbf{r}_{12}]]. \tag{50.4}$$

Аналогично, заряд q_2 действует на заряд q_1 с силой

$$\mathbf{F}_{21} = \frac{q_2q_1}{c^2r_{21}^3}[\mathbf{v}_1[\mathbf{v}_2\mathbf{r}_{21}]], \tag{50.5}$$

где радиус-вектор $\mathbf{r}_{21}$ проведен от заряда *2* к заряду *1.*

Если скорости $\mathbf{v}_1$ и $\mathbf{v}_2$ параллельны, одинаково направлены и перпендикулярны к вектору $\mathbf{r}_{12}$ (рис. 124), то в случае одноименных зарядов $\mathbf{F}_{12}$ и $\mathbf{F}_{21}$ будут силами притяжения, а в случае разноименных — силами отталкивания. Эти силы определяются выражением

$$F_{12} = F_{21} = F = \frac{q_1q_2}{r_{12}^2}\left(\frac{v_1v_2}{c^2}\right). \tag{50.6}$$

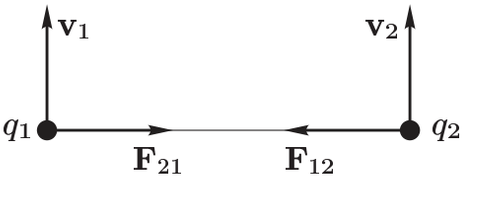


Рис. 124

В частности, когда скорости одинаковы,

$$F = \frac{q_1q_2}{r_{12}^2}\left(\frac{v}{c}\right)^2. \tag{50.7}$$

Если скорости антипараллельны, то при тех же условиях одноименные заряды будут отталкиваться, а разноименные — притягиваться.

В формулы (50.6) и (50.7) магнитное поле не входит. Этими формулами, как уже говорилось выше, электродинамическая постоянная c определяется однозначно, так как все прочие величины, входящие в указанные формулы, определены и могут быть экспериментально измерены. Впрочем, формулы (50.6) и (50.7) указывают лишь на принципиальную возможность измерения постоянной c. Для практических измерений они не пригодны. Практический метод измерения постоянной c будет указан в § 51.

В общем случае силы магнитного взаимодействия $\mathbf{F}_{12}$ и $\mathbf{F}_{21}$ не удовлетворяют принципу равенства действия и противодействия. Особенно резко нарушение этого принципа проявляется тогда, когда скорости $\mathbf{v}_1$ и $\mathbf{v}_2$ взаимно перпендикулярны, причем скорость $\mathbf{v}_2$ направлена вдоль вектора $\mathbf{r}_{12}$ (рис. 125). Тогда $\mathbf{B}_2 \sim [\mathbf{v}_2\mathbf{r}_{21}] = 0$, так что $\mathbf{F}_{21} = 0$. Между тем, как видно из рисунка, $\mathbf{F}_{12} \neq 0$. Но мы уже неоднократно подчеркивали, что для взаимодействий, осуществляющихся посредством полей, соблюдение принципа равенства действия и противодействия не обязательно.

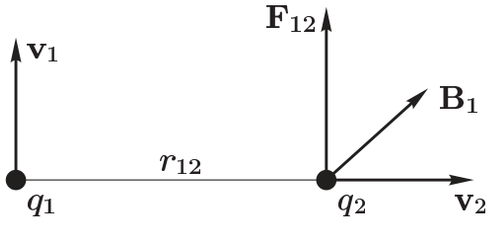


Рис. 125

3. Формула (50.7) показывает, что отношение силы магнитного взаимодействия движущихся зарядов к силе их кулоновского притяжения или отталкивания порядка $(v/c)^2$. Скорости установившегося движения электронов в металлах при прохождении электрического тока не превышают нескольких сантиметров в секунду, а в электролитах они еще меньше. Таким образом, отношение $(v/c)^2$ ничтожно и не превышает примерно 10^{-20}. Почему же электродвигатели приводят в движение именно магнитные (амперовы) силы, по сравнению с которыми силы электростатического взаимодействия не играют никакой роли? Все дело в том, что *в переносе тока участвует громадное количество заряженных частиц*, и это обстоятельство компенсирует малость множителя $(v/c)^2$. При этом существенно, что действие магнитного поля на движущийся заряд q определяется не q и $\mathbf{v}$ в отдельности, а произведением этих величин $q\mathbf{v}$. Когда течет ток, то заряды противоположных знаков движутся в противоположных направлениях, так что произведение $q\mathbf{v}$ имеет для них один и тот же знак. Силы, действующие в магнитном поле на частицы противоположных знаков, *арифметически складываются*, а не вычитаются. Точно так же магнитные поля, возбуждаемые движущимися зарядами, зависят от произведения $q\mathbf{v}$, а потому поля противоположных зарядов также *арифметически складываются*. Совсем иначе ведут себя электрические заряды по отношению к электрическим полям. С одной стороны, в выражения для напряженностей электрических полей и сил, действующих на заряды в таких полях,

скорость $\mathbf{v}$ не входит. Силы, действующие на положительные и отрицательные заряды, направлены противоположно, а потому *арифметически вычитаются.* С другой стороны, даже в электрически заряженном теле *заряды определенного знака в высокой степени скомпенсированы зарядами противоположного знака.* Как бы ни был велик электрический заряд тела, все же он ничтожно мал по сравнению с суммарным зарядом входящих в него частиц одного знака (см. § 10, п. 4). Вот почему магнитные силы намного превосходят электрические силы, действующие на нескомпенсированные заряды тел.

4. Получим теперь закон, определяющий магнитное поле отдельного элемента тока. Как и в электростатике, будем исходить из *принципа суперпозиции* как обобщения опытных фактов. Согласно этому принципу *магнитные поля отдельных движущихся зарядов векторно складываются, причем каждый заряд возбуждает поле, совершенно не зависящее от наличия других зарядов.* С использованием формулы (50.2) принцип суперпозиции приводит к следующему выражению для магнитного поля объемного элемента тока:

$$d\mathbf{B} = \frac{1}{c}\,\frac{[\mathbf{j}\,\mathbf{r}]}{r^3}\,dV. \tag{50.8}$$

Аналогично, для линейного элемента тока:

$$d\mathbf{B} = \frac{I}{c}\,\frac{[d\mathbf{l}\,\mathbf{r}]}{r^3}. \tag{50.9}$$

Эти формулы выражают так называемый *закон Био* (1774–1862) *и Савара* (1791–1841). Полное поле найдется интегрированием выражений (50.8) и (50.9) по всем токам, т. е.

$$\mathbf{B} = \frac{1}{c}\int \frac{[\mathbf{j}\,\mathbf{r}]}{r^3}\,dV, \tag{50.10}$$

или

$$\mathbf{B} = \oint \frac{I}{c}\,\frac{[d\mathbf{l}\,\mathbf{r}]}{r^3}. \tag{50.11}$$

Оба эти выражения применимы лишь *для постоянных токов.* А постоянные токи всегда замкнуты. Все наблюдаемые величины не изменились бы, если бы в правой части формулы (50.9) было добавлено произвольное слагаемое, интеграл от которого по любому замкнутому контуру обращается в нуль. Поэтому в рамках учения о постоянных токах элементарный закон Био и Савара в форме (50.8) или (50.9) *принципиально недоступен опытной проверке*, так как невозможно изолировать отдельные элементы постоянных токов и экспериментировать с ними. *Опытной проверке доступна только интегральная форма закона Био и Савара* (50.10) *или* (50.11). По этой причине в основу учения о магнитном поле постоянных токов мы положили не элементарный закон Био и Савара, как это обычно принято, а закон, определяющий магнитное поле движущегося заряда. *Поле движущегося заряда принципиально всегда может быть измерено на опыте*, хотя практически это и весьма

трудная задача. И действительно, установление того факта, что движущиеся макроскопические заряды создают магнитные поля, было получено с большим трудом. Впервые это удалось сделать в 1877 г. Роуланду (1848–1901) в лаборатории Гельмгольца. Его работа была продолжена многими учеными и завершилась фундаментальными исследованиями А. А. Эйхенвальда (1863–1944), выполненными в Москве в 1901–1904 гг. Из формул (50.4) и (50.5) следует, что параллельные токи должны притягиваться, а антипараллельные — отталкиваться.

§ 51. Расчет магнитных полей с помощью закона Био и Савара. Системы единиц

1. Рассмотрим сначала очень длинный прямолинейный провод, по которому течет постоянный ток I. Подводящие провода должны быть расположены настолько далеко, чтобы их магнитными полями в рассматриваемой области пространства можно было полностью пренебречь. Тогда провод может считаться бесконечно длинным. Магнитное поле элемента тока $I\,d\mathbf{l}$ (рис. 126) дается выражением

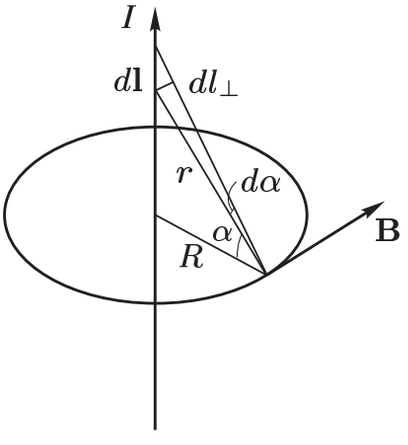


Рис. 126

$$d\mathbf{B} = \frac{I}{cr^3}\,[d\mathbf{l}\,\mathbf{r}] = \frac{I}{cr^3}\,[d\mathbf{l}_\perp \mathbf{r}],$$

где $d\mathbf{l}_\perp$ — составляющая вектора $d\mathbf{l}$, перпендикулярная к $\mathbf{r}$. Магнитные силовые линии будут окружностями, центры которых расположены на оси провода. В скалярной форме:

$$dB = \frac{I}{cr^2}\,dl_\perp = \frac{I}{cr}\,d\alpha,$$

где $d\alpha$ — угол, под которым вектор $d\mathbf{l}$ виден из точки наблюдения. Введя расстояние до провода $R = r\cos\alpha$, получим $dB = I\cos\alpha\,d\alpha/(cR)$. Интегрирование этого выражения от $\alpha = -\pi/2$ до $\alpha = +\pi/2$ дает искомый результат:

$$B = \frac{2I}{cR}. \tag{51.1}$$

Теперь нетрудно рассчитать взаимодействие двух бесконечно длинных параллельных постоянных токов I_1 и I_2. Первый ток в месте нахождения второго создает магнитное поле $B_1 = 2I_1/(cR)$, где R — расстояние между токами. Это поле действует на участок второго тока длины l с силой $F = I_2 l B_1/c$, или

$$F = \frac{2}{Rc^2}\,I_1 I_2 l. \tag{51.2}$$

Измерив F, можно по этой формуле вычислить значение электродинамической постоянной c. Впервые эта постоянная была измерена

несколько иным путем Вильгельмом Вебером и Рудольфом Кольраушем (1809–1858) в 1856 г. Они пришли к поразительному результату, что в пределах ошибок измерений *величина с совпадает со скоростью света в вакууме.* Последующие измерения других ученых не оставили никаких сомнений в том, что электродинамическая постоянная и скорость света в вакууме — это *одна и та же физическая постоянная.* Теоретические исследования Максвелла показали, что этот фундаментальный результат является выражением *электромагнитной природы света.*

2. Знание числового значения c открывает возможность рационального построения системы единиц в учении об электрических и магнитных полях. Если ввести обозначение $q^{(m)} = q/c$, то основные формулы (49.1) и (50.2) перепишутся без множителя c:

$$\mathbf{F} = q^{(m)}[\mathbf{vB}], \tag{51.3}$$

$$\mathbf{B} = \frac{q^{(m)}}{r^3}[\mathbf{vr}]. \tag{51.4}$$

Тем самым вводятся новые единицы заряда (и тока) — в c раз большие соответствующих электростатических единиц и отличающиеся от них размерностью. На них основана так называемая *магнитная система СГС*, обозначаемая кратко СГСМ. Десятая доля СГСМ-единицы заряда называется *кулоном*, а силы тока — *ампером.* Это — *точные определения* кулона и ампера. (Не совсем точные определения, которыми мы пользовались до сих пор, использовали приближенное значение электродинамической постоянной $c \approx 3 \cdot 10^{10}$ см/с.)

Теперь можно дать определение единицы напряженности магнитного поля, которая называется *гауссом.* Допустим, что векторы $\mathbf{v}$ и $\mathbf{B}$ взаимно перпендикулярны и что $q^{(m)} = 1$ СГСМ-ед., $v = 1$ см/с, $B = 1$ Гс. Тогда формула (51.3) дает $F = 1$ дин. Это приводит к следующему определению. *Гаусс есть индукция такого магнитного поля, которое действует на заряд в одну СГСМ-единицу с силой в одну дину, если сам заряд движется перпендикулярно к магнитному полю со скоростью* 1 см/с. Легко переформулировать это определение, используя вместо единицы электрического заряда единицу электрического тока. Для получения конкретного представления о гауссе заметим, что напряженность земного магнитного поля меняется приблизительно от 0,4 Гс (на экваторе) до 0,7 Гс (на полюсе). В геомагнетизме применяется также более мелкая единица — *гамма* (γ). По определению $1\gamma = 10^{-5}$ Гс.

Система СГСЭ применяется только для измерения чисто электрических величин: заряда, напряженности и смещения электрического поля, электрического потенциала, емкости, электродвижущей силы, электрической проводимости, электрического сопротивления и пр. Система СГСМ, напротив, применяется лишь для измерения чисто магнитных величин: напряженности и индукции магнитного поля, магнитного потока, коэффициентов само- и взаимоиндукции, магнитных моментов, вектора намагничивания и пр. Ни одна из этих систем никогда не

используется как единая система для измерения всех электрических и магнитных величин. Гауссова система, которой мы пользуемся, является *комбинированной.* Единицы чисто электрических величин в ней совпадают с единицами СГСЭ, а единицы чисто магнитных величин — с единицами СГСМ.

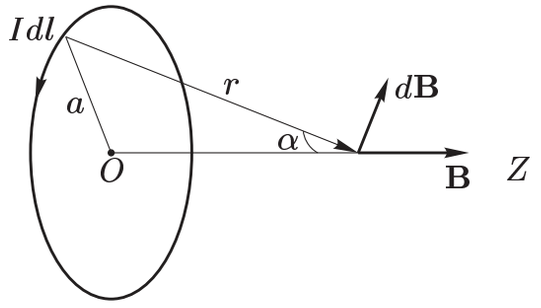


Рис. 127

3. Вычислим напряженность магнитного поля кругового тока на его оси (рис. 127). Элемент тока $I\,d\mathbf{l}$ возбуждает магнитное поле $d\mathbf{B}$, перпендикулярное к радиусу-вектору $\mathbf{r}$. Разложим это поле на две слагающие: осевую $d\mathbf{B}_z$ и радиальную $d\mathbf{B}_r$. При интегрировании по контуру кругового тока радиальные слагающие взаимно уничтожаются. Результирующее поле будет направлено вдоль оси Z, и надо интегрировать только осевую составляющую

$$dB_z = \frac{I\,dl}{cr^2}\sin\alpha.$$

Угол α один и тот же для всех точек кругового тока. Интегрирование сводится к простому умножению на длину контура $2\pi a$. Таким образом,

$$B_z = B = \frac{2\pi a I}{cr^2}\sin\alpha = \frac{2\pi a^2 I}{cr^3}. \tag{51.5}$$

В точках, не лежащих на оси, выражение для поля кругового тока имеет сложный вид.

4. Иногда удобно вводить в рассмотрение *поверхностные токи*, т. е. токи, текущие по тонким поверхностным слоям тел. Проведем на обтекаемой током поверхности линию, перпендикулярную к направлению тока. Ток, приходящийся на единицу длины такой линии, называется *линейной плотностью тока* и рассматривается как вектор $\mathbf{i}$, направленный вдоль тока. За положительное направление обхода вокруг поверхностного, как и всякого другого тока, принимается направление вращения ручки правого буравчика, ориентированного по току, если при таком вращении буравчик ввинчивается в направление тока.

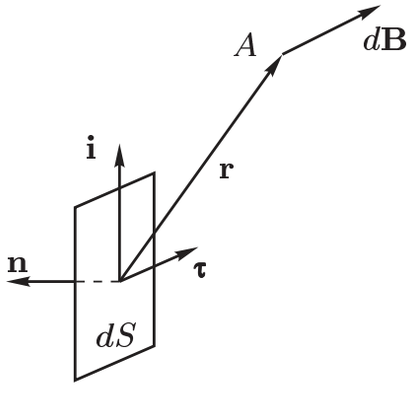


Рис. 128

Пусть по площадке dS (рис. 128) течет постоянный поверхностный ток с линейной плотностью $\mathbf{i}$. Площадку с током можно рассматривать как поверхностный элемент тока $\mathbf{i}\,dS$,

создающий магнитное поле

$$d\mathbf{B} = \frac{[\mathbf{ir}]}{cr^3}\, dS.$$

Вектор $d\mathbf{B}$ перпендикулярен к току $\mathbf{i}$. Его можно разложить на составляющую $d\mathbf{B}_\tau$, касательную к площадке dS, и на составляющую $d\mathbf{B}_n$, перпендикулярную к ней. Найдем составляющую $d\mathbf{B}_\tau$. Единичный вектор $\boldsymbol{\tau}$ расположим в плоскости площадки dS, ориентируя его в положительном направлении обхода перпендикулярно к току $\mathbf{i}$, как указано на рис. 128. Единичный вектор нормали $\mathbf{n}$ можно направить произвольно. Очевидно,

$$dB_\tau = (\boldsymbol{\tau}\, d\mathbf{B}) = i\frac{\boldsymbol{\tau}[\mathbf{i}_1\mathbf{r}]}{cr^3}\, dS = i\frac{[\boldsymbol{\tau}\mathbf{i}_1]\mathbf{r}}{cr^3}\, dS,$$

где $\mathbf{i}_1$ — единичный вектор в направлении тока $\mathbf{i}$. Так как $[\boldsymbol{\tau}\mathbf{i}_1] = -\mathbf{n}$, то

$$dB_\tau = -i\frac{(\mathbf{rn})}{cr^3}\, dS = \frac{i}{c}\, d\Omega, \tag{51.6}$$

где $d\Omega$ — телесный угол, под которым из точки наблюдения A видна внутренняя сторона площадки dS.

5. Применим вспомогательную формулу (51.6) к цилиндрической трубке, по поверхности которой перпендикулярно к ее образующим течет постоянный ток с линейной плотностью i (рис. 129). Такая трубка с током называется *соленоидом*. Найдем магнитное поле соленоида на его оси. Из соображений симметрии ясно, что поле $\mathbf{B}$ направлено вдоль оси соленоида. Поэтому для нахождения $\mathbf{B}$ достаточно просуммировать касательные составляющие dB_τ создаваемые отдельными поверхностными элементами тока. Поскольку величина i постоянна по всей поверхности соленоида, формула (51.6) дает

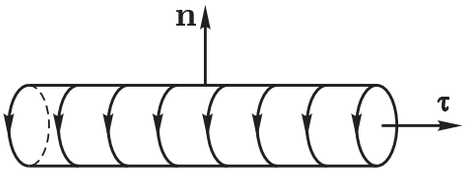


Рис. 129

$$B = \frac{i}{c}\, \Omega, \tag{51.7}$$

где Ω — полный телесный угол, под которым из точки наблюдения видна внутренняя поверхность соленоида. Формула (51.7) верна независимо от того, находится ли точка наблюдения внутри или вне соленоида.

В частном случае, когда соленоид длинный, а точка наблюдения лежит внутри него далеко от его концов, можно пренебречь телесными углами, под которыми видны основания трубки соленоида. Иными словами, соленоид можно считать бесконечно длинным. В таком случае $\Omega = 4\pi$, т. е.

$$B = \frac{4\pi}{c}\, i. \tag{51.8}$$

На концах соленоида напряженность поля будет вдвое меньше. В дальнейшем будет показано, что формула (51.8) справедлива в любой точке внутри соленоида, а не только на его оси.

Проволочную спираль, шаг которой мал по сравнению с радиусом витка, можно приближенно рассматривать как соленоид. Пусть по спирали течет постоянный ток I. Если N — число витков спирали, а l — ее длина, то $i = NI/l$. Поэтому

$$B = \frac{4\pi}{c}\,\frac{NI}{l}. \tag{51.9}$$

Эта формула не учитывает наклон витков спирали. Кроме того, она неприменима в непосредственной близости от проволоки, а также внутри нее.

§ 52. Момент сил, действующих на виток с током в магнитном поле

1. Результирующая сила, действующая на виток с током в постоянном магнитном поле, дается выражением

$$\mathbf{F} = \frac{I}{c}\oint [d\mathbf{l}\,\mathbf{B}],$$

где интегрирование производится по контуру витка. Если магнитное поле однородно, то вектор $\mathbf{B}$ можно вынести из-под знака интеграла. Задача сведется к вычислению векторного интеграла $\oint d\mathbf{l}$, а такой интеграл равен нулю. Значит, в однородном поле равна нулю и сила $\mathbf{F}$. Однако момент этой силы $\mathbf{M}$, вообще говоря, в нуль не обращается. Займемся его вычислением.

2. Рассмотрим сначала плоский виток, плоскость которого параллельна магнитному полю $\mathbf{B}$ (рис. 130). Проведя достаточно часто маг-

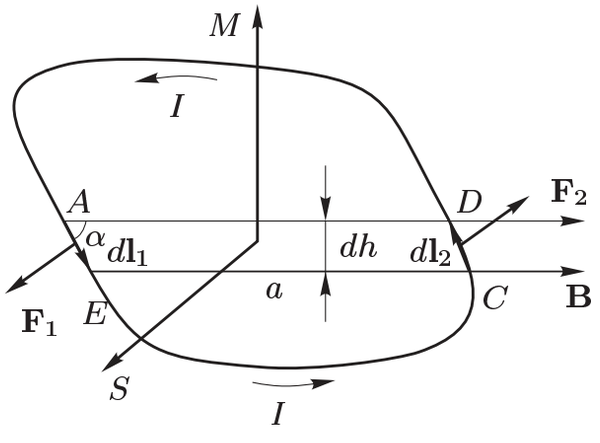


Рис. 130

нитные силовые линии, разобьем виток на пары элементов тока $I\,d\mathbf{l}_1$ и $I\,d\mathbf{l}_2$. Действующие на них амперовы силы $\mathbf{F}_1$ и $\mathbf{F}_2$ перпендикулярны

к плоскости витка и противоположны по направлению. По закону Ампера $F_1 = IB\,dl_1 \sin\alpha/c = IB\,dh/c$, где dh — высота криволинейного четырехугольника $AECD$. Тем же выражением определяется значение силы F_2. Таким образом, $F_1 = F_2$, т. е. $\mathbf{F}_1$ и $\mathbf{F}_2$ образуют пару сил с моментом $dM = IBa\,dh/c = IB\,dS/c$, где a — плечо пары, а dS — площадь четырехугольника $AECD$. Интегрированием получаем $M = IBS/c$, где S — площадь, охватываемая рассматриваемым витком тока. Вращающий момент $\mathbf{M}$ направлен вертикально вверх. Введем вектор площади контура $\mathbf{S}$, образующий с направлением тока правовинтовую систему. (Вектор $\mathbf{S}$ перпендикулярен к плоскости рисунка и направлен к читателю.) Тогда полученный результат можно записать в векторной форме:

$$\mathbf{M} = [\mathfrak{M}\mathbf{B}], \tag{52.1}$$

где введено обозначение

$$\mathfrak{M} = \frac{I}{c}\mathbf{S}. \tag{52.2}$$

Вектор $\mathfrak{M}$ называется *магнитным моментом тока.*

Допустим теперь, что плоскость витка перпендикулярна к магнитному полю. В этом случае амперова сила $d\mathbf{F} = I[d\mathbf{l}\,\mathbf{B}]/c$, действующая на элемент тока $I\,d\mathbf{l}$, будет лежать в плоскости витка и будет равна по модулю $IB\,dl/c$. Такие силы, в зависимости от направления тока, будут только растягивать или сжимать виток. Однако их момент равен нулю. В этом нетрудно убедиться, когда виток имеет форму прямоугольника или треугольника (рис. 131). (Точки означают, что магнитное поле перпендикулярно к плоскости рисунка и направлено

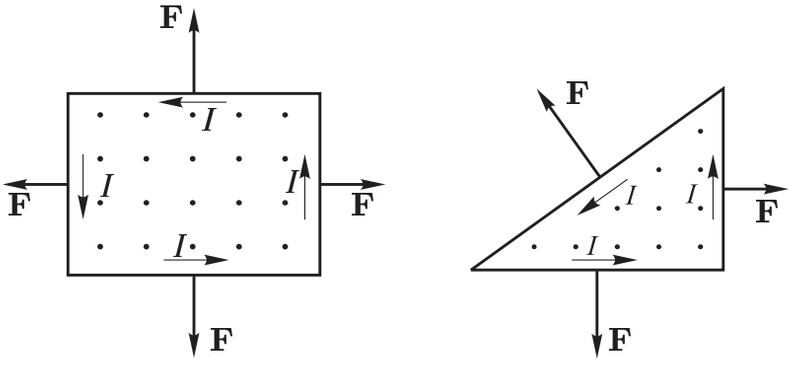


Рис. 131

к читателю.) Но к этим частным случаям сводится и общий случай, когда форма (плоского) витка произвольна. Достаточно провести прямые, разбивающие плоскость витка на бесконечно малые прямоугольники и треугольники, и вообразить, что по этим прямым в противоположных направлениях пропущены равные токи силой I (рис. 132). Добавление таких токов ничего не меняет, так как полный ток, текущий по каждой вспомогательной прямой, равен нулю. Однако теперь полный

момент $\mathbf{M}$ может быть представлен в виде суммы моментов, действующих на элементарные прямоугольники и треугольники. Поскольку эти моменты равны нулю, будет равен нулю и полный момент $\mathbf{M}$.

Рассмотрим, наконец, случай, когда магнитное поле направлено под углом к плоскости контура. Представим вектор $\mathbf{B}$ в виде $\mathbf{B} = \mathbf{B}_{\parallel} + \mathbf{B}_{\perp}$, где $\mathbf{B}_{\parallel}$ — составляющая вектора $\mathbf{B}$, параллельная, а $\mathbf{B}_{\perp}$ — перпендикулярная к плоскости контура. Вторая составляющая не вносит вклада в момент $\mathbf{M}$. Поэтому $\mathbf{M} = [\mathfrak{M}\mathbf{B}_{\parallel}] = [\mathfrak{M}\mathbf{B}]$. Таким образом, и в этом случае справедлива формула (52.1).

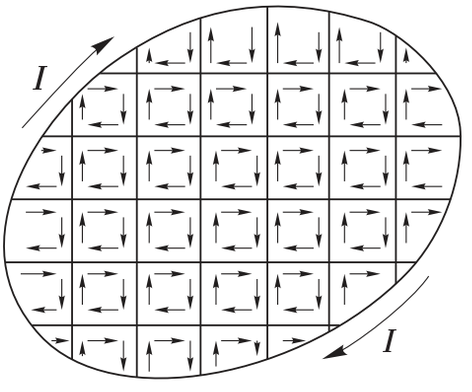


Рис. 132

Формула (52.1) верна и в том случае, когда контур тока не плоский, однако магнитное поле однородно. Чтобы убедиться в этом, натянем на контур с током произвольную поверхность S и разобьем ее вспомогательными линиями на элементарные площадки $d\mathbf{S}$ подобно тому, как это сделано на рис. 132. Пропустив по этим вспомогательным линиям равные и противоположно направленные токи силой I, представим момент $\mathbf{M}$ в виде суммы моментов, действующих на такие элементарные площадки. Но каждая площадка, ввиду ее бесконечной малости, может рассматриваться как плоская и к ней применима формула (52.1). Сложив моменты, действующие на элементарные площадки, снова получим формулу (52.1), причем магнитный момент тока $\mathfrak{M}$ будет определяться прежним выражением (52.2), в котором под вектором $\mathbf{S}$ следует понимать векторный интеграл $\mathbf{S} = \int d\mathbf{S}$, взятый по поверхности S, натянутой на контур тока. Вектор $\mathbf{S}$ не зависит от выбора вспомогательной поверхности S, а только от контура, на который она натянута.

Формула (52.1) справедлива и для соленоида, поскольку последний можно рассматривать как систему кольцевых токов. Магнитный момент соленоида, очевидно, определяется прежней формулой (52.2), если под I понимать полный ток, текущий по боковой поверхности соленоида, а под S — площадь его поперечного сечения. Для проволочной спирали с малым шагом, состоящей из N витков, $\mathfrak{M} = NI\mathbf{S}/c$.

3. Под действием вращающего момента $\mathbf{M}$ виток или катушка будут поворачиваться так, чтобы векторы $\mathbf{M}$ и $\mathbf{B}$ стали параллельными и одинаково направленными. Это — положение устойчивого равновесия. В положении, когда $\mathfrak{M}$ и $\mathbf{B}$ параллельны и направлены противоположно, также имеет место равновесие, но такое равновесие неустойчивое.

Формула (52.1) приближенно применима и для неоднородных магнитных полей. Необходимо только, чтобы размеры витка или катушки были малы. Тогда влиянием неоднородности поля на вращающий момент можно пренебречь. Такие витки и катушки могут быть использованы для практических измерений магнитных полей. Тогда они

называются *пробными.* Пробные витки и катушки играют в магнитных измерениях ту же роль, что и пробные заряды в электрических измерениях. Если пробную катушку поместить в магнитное поле, то ее магнитный момент $\mathfrak{M}$ установится вдоль поля $\mathbf{B}$. Повернем катушку из этого положения на 90°, чтобы вращающий момент $\mathbf{M}$ обратился в максимум. Тогда магнитный момент $\mathfrak{M}$ будет перпендикулярен к вектору $\mathbf{B}$, т. е. $(\mathfrak{M}\mathbf{B}) = 0$. Поэтому, умножая равенство (52.1) векторно на $\mathfrak{M}$, получим

$$\mathbf{B} = \frac{[\mathbf{M}\mathfrak{M}]}{\mathfrak{M}^2}. \tag{52.3}$$

Измерив вращающий момент $\mathbf{M}$, можно по этой формуле найти напряженность магнитного поля как по модулю, так и по направлению. В частности, такие измерения позволяют проверить закон Био и Савара в интегральной форме (50.11).

§ 53. Теорема Гаусса для магнитных полей

Элементарный закон (50.2) определяет магнитное поле равномерно движущегося точечного заряда. Предполагая, что магнитное поле возбуждается только такими зарядами, найдем поток вектора $\mathbf{B}$ (называемый для краткости *магнитным потоком*) через замкнутую поверхность, а также циркуляцию того же вектора по замкнутому контуру.

Начнем с вычисления магнитного потока. Для магнитных полей движущихся зарядов справедлив принцип суперпозиции. Кроме того, поток геометрической суммы нескольких векторов через любую поверхность равен алгебраической сумме потоков этих векторов через ту же поверхность. Поэтому при вычислении магнитного потока достаточно ограничиться частным случаем, когда поле $\mathbf{B}$ создается отдельным точечным зарядом, равномерно движущимся со скоростью $\mathbf{v}$.

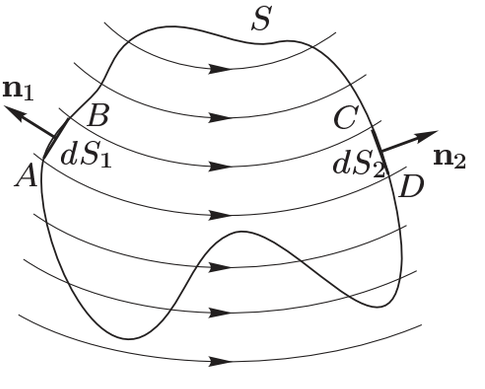


Рис. 133

Докажем, что *магнитный поток через произвольную замкнутую поверхность S равен нулю* (рис. 133). Для простоты будем предполагать, что заряд равномерно движется перпендикулярно к плоскости рисунка. Магнитными силовыми линиями будут коаксиальные окружности, плоскости которых параллельны плоскости рисунка, а центры расположены на прямой, вдоль которой движется заряд. Возьмем бесконечно тонкую кольцевую трубку $ABCD$, образованную магнитными силовыми линиями. Ввиду осевой симметрии магнитный поток через поперечное сечение трубки будет оставаться постоянным на всем ее протяжении.

Трубка пересечет замкнутую поверхность S четное число раз, например два. Магнитные потоки через площадки $d\mathbf{S}_1$ и $d\mathbf{S}_2$, которые трубка вырезает из поверхности S, одинаковы по величине, но противоположны по знаку. Сумма потоков через такие площадки равна нулю. То же справедливо и в том случае, когда трубка пересекает поверхность S произвольное число раз, поскольку это число всегда четное. Но все пространство можно разбить на бесконечно тонкие кольцевые магнитные трубки, и каждая из них не будет вносить никакого вклада в магнитный поток через замкнутую поверхность S. Суммарный магнитный поток через такую поверхность будет равен нулю. Теорема доказана.

Итак,

$$\oint (\mathbf{B}\, d\mathbf{S}) = 0, \tag{53.1}$$

или в дифференциальной форме

$$\operatorname{div} \mathbf{B} = 0. \tag{53.2}$$

В учении об электрическом поле соответствующие уравнения имели вид

$$\oint (\mathbf{E}\, d\mathbf{S}) = 4\pi q, \quad \operatorname{div} \mathbf{E} = 4\pi\rho.$$

Можно было бы предполагать, как это делалось на ранней стадии учения о магнетизме, что источниками магнитного поля являются *магнитные заряды*, взаимодействующие по закону Кулона. Однако такое предположение не согласуется с формулой (53.1). Она показывает, что *магнитных зарядов не существует.* Понятно, что справедливость столь фундаментального результата не может ограничиваться рамками учения о постоянных магнитных полях. Поэтому естественно ожидать, что *уравнение* (53.1) *или эквивалентное ему уравнение* (53.2) *справедливы также для любых магнитных полей.* Все опытные факты подтверждают это заключение. Уравнения (53.1) и (53.2) входят как составные части в *систему уравнений Максвелла.*

Силовые поля, дивергенция которых всюду обращается в нуль, называются *бездивергентными* или *соленоидальными полями.* Следовательно, *магнитное поле есть поле соленоидальное. Его источниками являются не магнитные заряды, а электрические токи.*

§ 54. Дополнительные сведения о телесных углах

1. О телесных углах говорилось в § 5. Для дальнейшего понятие телесного угла необходимо обобщить. Возьмем произвольный замкнутый контур L и натянем на него произвольную замкнутую поверхность S (рис. 134 *а*). На контуре L установим положительное направление обхода, а на поверхности S — положительное направление нормали $\mathbf{n}$. Эти направления должны быть согласованы между собой так, чтобы они находились в правовинтовом соотношении. Для упрощения рассуждений будем пользоваться плоскими, а не пространственными рисунками. Сечение замкнутого контура L плоскостью рисунка изобразим двумя кружками. Кружок с точкой внутри него

означает, что контур пересекает плоскость рисунка в направлении к читателю, а кружок с крестиком — в направлении от читателя. Как и раньше, сторону поверхности S, в которую нормаль $\mathbf{n}$ входит, условимся называть

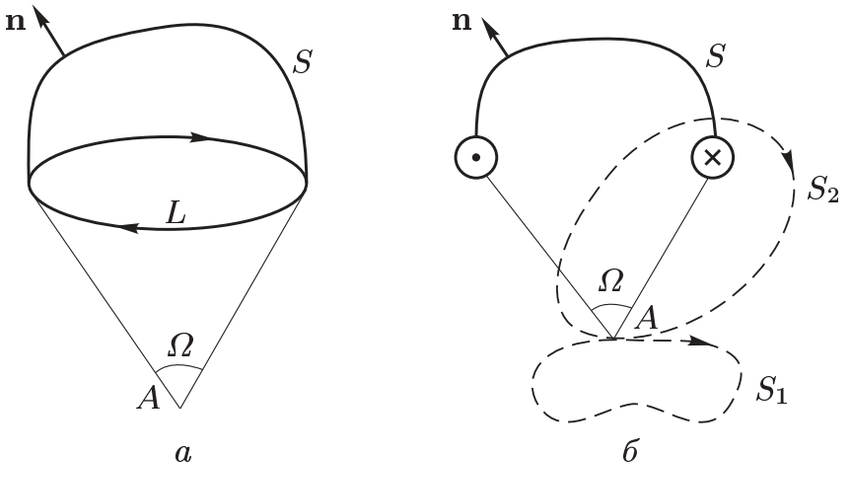


Рис. 134

внутренней, а сторону, из которой она выходит, — *внешней*. Телесный угол Ω мы считали *положительным*, когда из его вершины видна внутренняя сторона поверхности S, и *отрицательным*, когда видна внешняя. Временно сохраним это соглашение и здесь.

2. При перемещении вершины телесного угла Ω его значение меняется, вообще говоря, непрерывно. Исключение составляет случай, когда вершина A пересекает поверхность S. Тогда телесный угол получает скачкообразное приращение, равное $+4\pi$, если точка A переходит с внешней стороны на внутреннюю, и -4π, если переход совершается в противоположном направлении. Для пояснения рассмотрим сначала частный случай, когда контур L и поверхность S плоские. Если вершина A находится бесконечно близко от плоскости S с внешней стороны, то эта плоскость видна из вершины под углом -2π. Если же вершина переходит на внутреннюю сторону, то телесный угол становится равным $+2\pi$, т. е. получает приращение 4π. Допустим теперь, что поверхность ACB, натянутая на контур L, произвольная (рис. 135). Пусть *1* и *2* — две бесконечно близкие точки, лежащие по разные стороны от поверхности ACB. Натянем на тот же контур L бесконечно близкую поверхность $AC'B$ так, чтобы обе точки *1* и *2* оказались по одну сторону от нее. Телесные углы Ω_1' и Ω_2', под которыми видна из точек *1* и *2* эта поверхность, будут отличаться друг от друга бесконечно мало, причем $\Omega_1' = \Omega_1$. Возьмем теперь замкнутую поверхность $ACBC'A$. Так как на ACB нормаль $\mathbf{n}$ проведена наружу, а на $AC'B$ — внутрь указанной замкнутой поверхности, то из точки *2* поверхность $ACBC'A$ будет видна под углом $\Omega_2 - \Omega_2'$. А так как эта поверхность замкнутая, то тот же угол равен 4π, т. е. $\Omega_2 - \Omega_2' = 4\pi$. Но, очевидно, с точностью до бесконечно малых $\Omega_2' = \Omega_1' = \Omega_1$, а потому $\Omega_2 - \Omega_1 = 4\pi$, что и требовалось доказать.

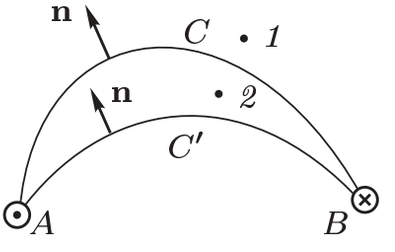
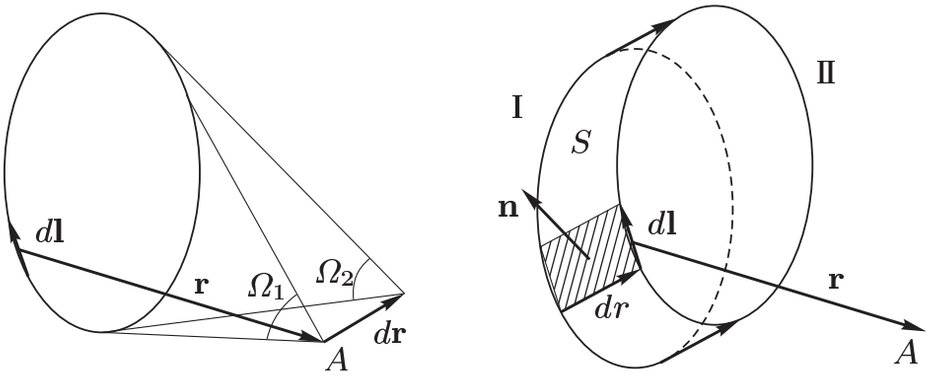


Рис. 135

3. Разрывное поведение телесного угла Ω делает эту величину неудобной при рассмотрении замкнутых контуров, обтекаемых токами. Действительно, точка разрыва величины Ω зависит от того, где проведена вспомогательная

поверхность S, натянутая на контур L. Магнитное же поле тока, текущего по указанному контуру, ни от каких вспомогательных поверхностей не зависит. Однако небольшим обобщением понятия телесного угла нетрудно освободить это понятие от указанного неудобства. Для этого *телесный угол надо определить как многозначную* (точнее, *бесконечнозначную*) *функцию положения его вершины.* Если вершина занимает какое-то произвольное (начальное) положение, то телесный угол можно определить так, как это делалось выше. Когда вершина A описывает произвольный замкнутый путь s_1 (рис. 134 *б*), возвращаясь в исходное положение, не пересекая при этом (или пересекая в противоположных направлениях четное число раз) никакой поверхности, натянутой на контур L, то телесный угол Ω возвращается к своему исходному значению. Если же замкнутый путь s_2 обходит вокруг контура L, то телесный угол получает приращение 4π; при обходе в отрицательном направлении приращение телесного угла будет -4π. Так понимаемый телесный угол Ω определен с точностью до слагаемого вида $4\pi n$, где n — целое число (положительное или отрицательное). При непрерывном перемещении вершины A телесный угол меняется непрерывно. Величина его не зависит ни от какой поверхности, натянутой на контур L. Она определяется только самим контуром L и установленным на нем положительным направлением обхода. Поэтому следует говорить о телесном угле, под которым виден *сам контур* L, а не поверхность, натянутая на него. Именно в таком смысле мы будем понимать телесный угол, когда в следующем параграфе свяжем эту величину с напряженностью магнитного поля.

4. Рассмотрим теперь дифференцирование телесных углов. Пусть контур, на который опирается телесный угол Ω, неподвижно закреплен. Тогда величина телесного угла Ω будет зависеть только от положения его вершины. Найдем производные величины Ω по координатам x, y, z этой вершины. При смещении вершины на dr (рис. 136 *а*) телесный угол получает приращение



а *б*

Рис. 136

$d\Omega = \Omega_2 - \Omega_1$. Такое же приращение он получил бы, если бы вершина была неподвижно закреплена, а контур, как целое, смещен в противоположном направлении на $-dr$ (рис. 136 *б*). При таком смещении точки контура опишут цилиндрическую поверхность S, которая видна из вершины A под телесным углом $\Omega_{\text{бок}}$. Как видно из рисунка, $\Omega_1 = \Omega_2 + \Omega_{\text{бок}}$, т. е. $d\Omega = \Omega_2 - \Omega_1 = -\Omega_{\text{бок}}$. Телесный угол $\Omega_{\text{бок}}$ определяется выражением

$$\Omega_{\text{бок}} = -\int \frac{(d\mathbf{S}\,\mathbf{r})}{r^3},$$

где dS — элемент площади боковой поверхности, причем радиус-вектор $\mathbf{r}$ проведен от площадки $d\mathbf{S}$ к вершине телесного угла, а единичная нормаль n к цилиндрической поверхности направлена наружу, как указано на рисунке. Если $d\mathbf{l}$ — элемент длины контура, то при таком выборе единичной нормали $d\mathbf{S} = [d\mathbf{r}\, d\mathbf{l}]$. Подставляя это выражение в предыдущее и вынося $d\mathbf{r}$ из-под знака интеграла, получим

$$\Omega_{\text{бок}} = -\oint \frac{[d\mathbf{r}\, d\mathbf{l}]\mathbf{r}}{r^3} = -d\mathbf{r} \oint \frac{[d\mathbf{l}\, \mathbf{r}]}{r^3} = -\mathbf{a}\, d\mathbf{r},$$

где введено обозначение

$$\mathbf{a} = \oint \frac{[d\mathbf{l}\, \mathbf{r}]}{r^3}.$$

Интегрирование здесь ведется по контуру, на который опирается телесный угол. Таким образом, $d\Omega = a\, dr = a_x\, dx + a_y\, dy + a_z\, dz$. Отсюда находим

$$\frac{\partial \Omega}{\partial x} = a_x, \quad \frac{\partial \Omega}{\partial y} = a_y, \quad \frac{\partial \Omega}{\partial z} = a_z,$$

или в векторной форме

$$\operatorname{grad} \Omega = \oint \frac{[d\mathbf{l}\, \mathbf{r}]}{r^3}. \tag{54.1}$$

§ 55. Теорема о циркуляции магнитного поля в вакууме

1. Линейный интеграл $\int\limits_{12} \mathbf{B}\, d\mathbf{s}$, называемый *магнитным напряжением* между точками *1* и *2*, зависит не только от положения этих точек, но и от пути интегрирования. Его значение становится определенным только после указания этого пути. Предположим сначала, что магнитное поле $\mathbf{B}$ создается постоянным током I, текущим по бесконечно тонкому замкнутому витку. Напряженность такого поля определяется выражением (50.11), или в силу формулы (54.1)

$$\mathbf{B} = -\operatorname{grad} \varphi_m, \tag{55.1}$$

где

$$\varphi_m = -\frac{I}{c}\, \Omega, \tag{55.2}$$

а Ω — телесный угол, под которым виток с током виден из точки наблюдения. Величина φ_m называется *магнитным потенциалом.* Магнитный потенциал, как и телесный угол Ω, является *многозначной* (*бесконечнозначной*) *функцией* точки наблюдения. Магнитное напряжение выражается через магнитный потенциал соотношением

$$\int\limits_{12} \mathbf{B}\, d\mathbf{s} = -\int\limits_{12} d\varphi_m = \varphi_{m1} - \varphi_{m2}. \tag{55.3}$$

2. Допустим, что точки *1* и *2* совпадают, т. е. путь интегрирования становится замкнутым. Тогда магнитное напряжение переходит в *циркуляцию* $\oint \mathbf{B}\, d\mathbf{s}$ вдоль этого замкнутого пути. Из свойств телесных

углов, установленных в предыдущем параграфе, следует, что циркуляция $\oint \mathbf{B}\, d\mathbf{s}$ будет равна нулю, если путь интегрирования не обходит вокруг тока. Если же он обходит вокруг тока один раз, то

$$\oint \mathbf{B}\, d\mathbf{s} = \frac{4\pi}{c}\, I. \tag{55.4}$$

Ток I считается положительным, если его направление находится в правовинтовом соотношении с направлением пути обхода.

Допустим теперь, что магнитное поле создается несколькими замкнутыми бесконечно тонкими витками с постоянными токами. К таким виткам с токами сводится и общий случай, так как постоянные токи всегда замкнуты и пространство, обтекаемое ими, можно разбить на бесконечно тонкие замкнутые токовые линии, которые и будут играть роль бесконечно тонких витков. Магнитные поля отдельных витков удовлетворяют принципу суперпозиции, а циркуляции этих полей по одному и тому же замкнутому контуру складываются алгебраически. Поэтому и в общем случае получается формула (55.4). Только в ней под I следует понимать сумму токов всех проводников, вокруг которых обходит контур циркуляции. В результате доказано следующее положение: *циркуляция магнитного поля постоянных токов по всякому замкнутому контуру равна сумме токов, пронизывающих контур циркуляции, умноженной на* $4\pi/c$. Это положение называется *теоремой о циркуляции* вектора напряженности магнитного поля.

Если объемная плотность тока $\mathbf{j}$ конечна, то

$$I = \int\limits_S j_n\, dS = \int\limits_S (\mathbf{j}\, d\mathbf{S}),$$

где S — любая поверхность, натянутая на контур, по которому вычисляется циркуляция. В этом случае формула (55.4) перейдет в

$$\oint (\mathbf{B}\, d\mathbf{s}) = \frac{4\pi}{c} \int (\mathbf{j}\, d\mathbf{S}). \tag{55.5}$$

В тех областях пространства, где не текут электрические токи, циркуляция $\oint \mathbf{B}\, d\mathbf{s}$ обращается в нуль по любому замкнутому контуру, т. е. в таких областях магнитное поле *потенциально.* Однако, как видно из формулы (55.5), этого не будет там, где текут электрические токи. Там магнитное поле *не потенциально.*

3. В учении о магнитном поле постоянных токов теорема о циркуляции играет примерно ту же роль, что и теорема Гаусса в электростатике. При наличии определенной симметрии теорема о циркуляции позволяет иногда очень просто рассчитать напряженность магнитного поля. Приведем несколько примеров.

Пример 1. Магнитное поле бесконечного прямолинейного провода с током. Ввиду симметрии магнитные силовые линии имеют форму окружностей с центрами на оси тока (рис. 126). Длина вектора $\mathbf{B}$ одна и та же во всех точках силовой линии. Циркуляция магнитного поля вдоль силовой линии, с одной стороны, равна

$2\pi RB$. С другой стороны, по теореме о циркуляции та же величина равна $4\pi I/c$. Приравнивая оба выражения, получим

$$B = \frac{2I}{cR}. \tag{55.6}$$

Этот результат уже был найден в § 51 непосредственно интегрированием магнитных полей элементов тока.

Рассмотрим теперь бесконечно длинный прямой цилиндрический провод радиуса a, по которому течет ток I с постоянной плотностью. Нетрудно видеть, что наружное поле будет определяться прежней формулой (55.6). Остается найти поле внутри провода. Конечно, и там магнитные силовые линии будут коаксиальными окружностями. Одна из них изображена на рис. 137 штриховой линией. (Предполагается, что ток течет перпендикулярно к плоскости рисунка.) Циркуляция вектора $\mathbf{B}$ по этой линии равна, с одной стороны, $2\pi RB$. С другой стороны, по теореме о циркуляции та же величина равна $4\pi I'/c$, где $I' = IR^2/a^2$ — ток, пронизывающий рассматриваемый контур. Сравнивая эти выражения, находим

$$B = \frac{2I}{ca^2}\,R \qquad (R < a). \tag{55.7}$$

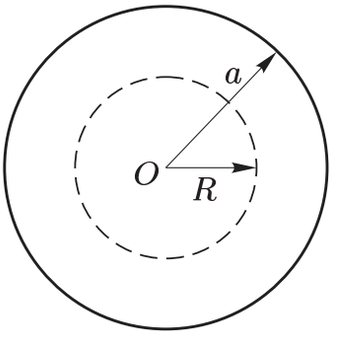


Рис. 137

Если провод полый, а поверхностная плотность тока постоянна, то вне цилиндра по-прежнему верна формула (55.6). Поле внутри цилиндра в этом случае равно нулю.

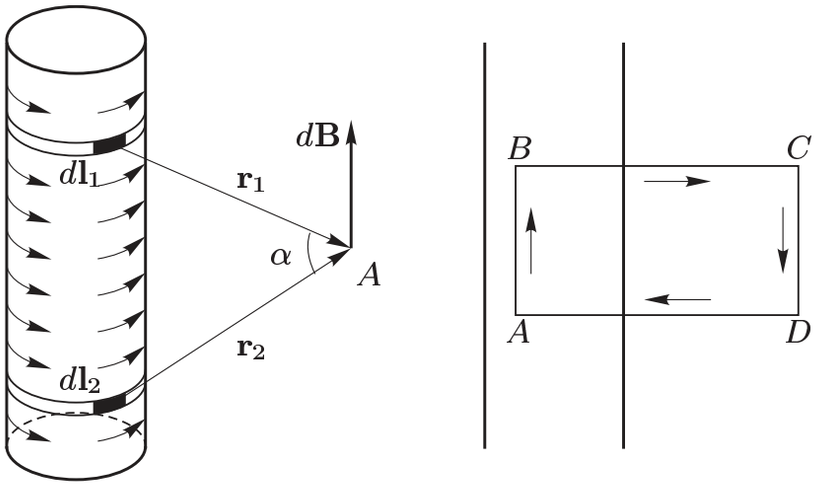
Пример 2. Магнитное поле бесконечно длинного соленоида. Пусть поверхностная плотность тока i, циркулирующего по поверхности соленоида, одна и та же по всей длине соленоида. Покажем, что магнитное поле, если оно отлично от нуля, должно быть направлено параллельно оси соленоида. Возьмем два элемента тока $I\,d\mathbf{l}_1$ и $I\,d\mathbf{l}_2$, симметрично расположенных относительно точки наблюдения A, как это изображено на рис. 138 *а*. По закону Био и Савара результирующее магнитное поле этих двух элементов в точке A дается выражением

$$d\mathbf{B} = \frac{I}{cr_1^3}\,[d\mathbf{l}_1\mathbf{r}_1] + \frac{I}{cr_2^3}\,[d\mathbf{l}_2\mathbf{r}_2] = \frac{I}{cr_1^3}\,[d\mathbf{l}_1(\mathbf{r}_1 + \mathbf{r}_2)].$$

Так как векторы $d\mathbf{l}_1$ и $(\mathbf{r}_1+\mathbf{r}_2)$ перпендикулярны к оси соленоида, то поле $d\mathbf{B}$ параллельно этой оси. Всю поверхность бесконечного соленоида можно разбить на пары элементов, аналогичных $d\mathbf{l}_1$ и $d\mathbf{l}_2$. Магнитное поле каждой такой пары параллельно оси соленоида. Следовательно, то же справедливо и для полного поля соленоида. Доказательство справедливо независимо от того, где лежит точка наблюдения A — внутри или вне соленоида.

Исследуем, как ведет себя магнитное поле бесконечного соленоида при удалении точки наблюдения A в бесконечность. Как будет доказано

в § 57, магнитное поле отдельного витка на больших расстояниях убывает обратно пропорционально кубу расстояния. Рассмотрим конечный участок соленоида, который виден из точки A под углом α (рис. 138 *а*). Разобьем его на большое число N кольцевых токов. Будем удалять



а *б*

Рис. 138

точку A в бесконечность, сохраняя угол α и число N неизменными. При этом сила каждого кольцевого тока будет возрастать пропорционально длине выделенной части соленоида, а значит, и расстоянию до точки наблюдения. Если бы сила тока не изменялась, то из-за увеличения расстояния происходило бы убывание поля — обратно пропорционально кубу расстояния, как указано выше. Из-за увеличения силы тока убывание будет более медленным — обратно пропорционально квадрату расстояния. Но и в этом случае магнитное поле выделенной части соленоида, как и поле всего бесконечного соленоида, на бесконечности обратится в нуль.

Применим теперь теорему о циркуляции к прямоугольному контуру $ABCD$, сторона AB которого проходит внутри соленоида, а сторона CD бесконечно удалена (рис. 138 *б*). Участки BC и AD не вносят никакого вклада в циркуляцию, так как магнитное поле перпендикулярно к ним. По доказанному выше не дает вклада и сторона CD. Вся циркуляция сводится к интегралу по стороне AB и представляется выражением Bl, где l — длина стороны AB. По теореме о циркуляции та же величина равна $4\pi I/c = 4\pi il/c$. Сравнением обоих выражений находим

$$B = \frac{4\pi}{c}\, i. \tag{55.8}$$

Этот результат справедлив не только для круглого соленоида, но и для соленоида с произвольным поперечным сечением. Внутри бесконечного соленоида магнитное поле однородно, а снаружи — равно нулю.

Формула (55.8) с хорошей точностью применима к средней части соленоида конечной длины вдали от его краев. Вблизи краев магнитное поле сильно искажается и становится неоднородным.

Для катушки с проволочной обмоткой

$$B = \frac{4\pi}{c}\,\frac{NI}{l}, \tag{55.9}$$

где N — число витков, l — длина катушки, а I — сила тока в ее обмотке. Эта формула лишь приближенно представляет магнитное поле реального соленоида. Помимо искажения поля вблизи концов соленоида, однородность его нарушается в непосредственной близости каждого витка. Реальный ток лишь приближенно аппроксимируется поверхностным током с постоянной линейной плотностью. Вблизи каждого витка магнитные силовые линии обвиваются вокруг него. Кроме того, из-за наклона витков к оси соленоида имеется слагающая тока, параллельная этой оси. Она также искажает поле.

Пример 3. Магнитное поле тороидальной катушки. Заменим реальную катушку идеальным тором, по поверхности которого циркулирует ток с постоянной линейной плотностью i (рис. 139). Линии тока лежат в меридиональных плоскостях, т. е. в плоскостях, проходящих через ось AA системы. При повороте тора вокруг оси AA на любой угол он совмещается сам с собой. То же произойдет с магнитными силовыми линиями, если их повернуть, оставляя тор неподвижным. Отсюда следует, что силовыми линиями будут окружности с центрами на оси AA. Возьмем внутри тора одну из таких окружностей радиуса R (на рис. 139 она изображена штриховой линией). Циркуляция магнитного поля вдоль этой окружности равна $2\pi RB$. Полный ток, пронизывающий площадь, ограниченную этой окружностью, равен NI, где N — число витков в тороидальной катушке. По теореме о циркуляции $2\pi RB = 4\pi NI/c$, а потому

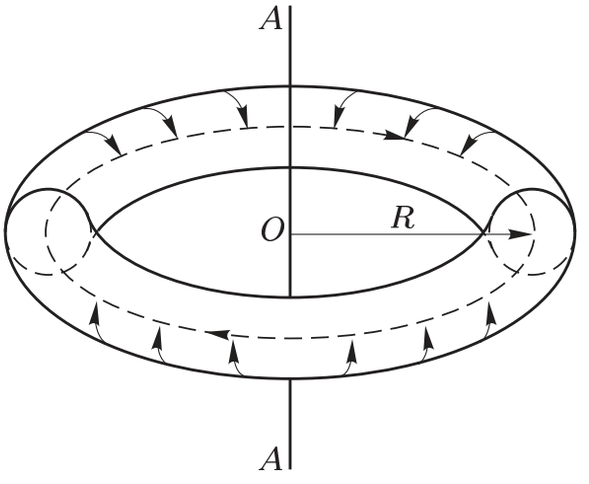


Рис. 139

$$B = \frac{2}{cR}\,NI. \tag{55.10}$$

Таким образом, внутри тора магнитное поле совпадает с полем прямого тока силой NI, текущего вдоль оси AA. Вне тора магнитное поле равно нулю. Устремляя N и R к бесконечности так, чтобы отношение $i = NI/(2\pi R)$ оставалось постоянным, в пределе получим выражение (55.8) для магнитного поля бесконечно длинного соленоида.

§ 56. Дифференциальная форма теоремы о циркуляции

1. Теореме о циркуляции можно придать дифференциальную форму, эквивалентную интегральной форме (55.4) или (55.5). С этой целью применим теорему о циркуляции к бесконечно малому прямоугольному контуру $ABCD$ со сторонами dy и dz, плоскость которого перпендикулярна к оси X (рис. 140). Вклад в циркуляцию, вносимый стороной AB, равен $B_y(x, y, z)\,dy$. Противоположная сторона CD вносит в циркуляцию слагаемое $-B_y(x, y, z + dz)\,dy$. Сумма этих величин равна

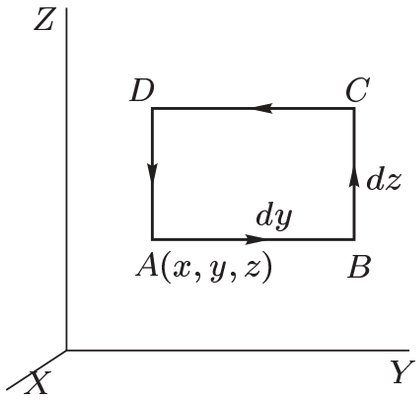


Рис. 140

$$-B_y(x, y, z + dz)\,dy + B_y(x, y, z)\,dy = -\frac{\partial B_y}{\partial z}\,dy\,dz = -\frac{\partial B_y}{\partial z}\,dS,$$

где $dS = dy\,dz$ — площадь прямоугольника $ABCD$. Аналогично, стороны BC и AD вносят в циркуляцию слагаемое $+\dfrac{\partial B_z}{\partial y}\,dS$. Полная циркуляция будет

$$\oint B_s\,ds = \left(\frac{\partial B_z}{\partial y} - \frac{\partial B_y}{\partial z}\right) dS.$$

По теореме о циркуляции та же величина равна $4\pi j_x\,dS/c$, так как $j_x\,dS$ есть полный ток, пронизывающий контур $ABCD$. Приравнивая оба выражения, придем к уравнению

$$\frac{\partial B_z}{\partial y} - \frac{\partial B_y}{\partial z} = \frac{4\pi}{c}\,j_x.$$

Аналогично,

$$\frac{\partial B_x}{\partial z} - \frac{\partial B_z}{\partial x} = \frac{4\pi}{c}\,j_y, \quad \frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y} = \frac{4\pi}{c}\,j_z.$$

Умножив эти уравнения на координатные орты $\mathbf{e}_x$, $\mathbf{e}_y$, $\mathbf{e}_z$ и сложив, получим

$$\operatorname{rot}\mathbf{B} = \frac{4\pi}{c}\,\mathbf{j}. \tag{56.1}$$

Это — одна из важнейших формул в учении об электричестве. Символом $\operatorname{rot}\mathbf{B}$ обозначен вектор

$$\operatorname{rot}\mathbf{B} \equiv \left(\frac{\partial B_z}{\partial y} - \frac{\partial B_y}{\partial z}\right)\mathbf{e}_x + \left(\frac{\partial B_x}{\partial z} - \frac{\partial B_z}{\partial x}\right)\mathbf{e}_y + \left(\frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y}\right)\mathbf{e}_z. \tag{56.2}$$

Дифференциальное выражение (56.2) играет важную роль во многих разделах математики и физики. Оно называется ротором вектора $\mathbf{B}$. Формально $\operatorname{rot}\mathbf{B}$ можно рассматривать как векторное произведение дифференциального оператора набла

$$\nabla = \mathbf{e}_x \frac{\partial}{\partial x} + \mathbf{e}_y \frac{\partial}{\partial y} + \mathbf{e}_z \frac{\partial}{\partial z}$$

на вектор $\mathbf{B}$, т. е.

$$\operatorname{rot}\mathbf{B} \equiv [\nabla \mathbf{B}] \equiv \begin{vmatrix} \mathbf{e}_x & \mathbf{e}_y & \mathbf{e}_z \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ B_x & B_y & B_z \end{vmatrix}. \tag{56.3}$$

2. Векторные поля, ротор которых не равен нулю, называются *вихревыми полями.* Формула (56.1) показывает, что магнитное поле $\mathbf{B}$ является *вихревым* во всех областях пространства, где текут электрические токи, и *безвихревым*, где токов нет. В последнем случае вектор $\mathbf{B}$ может быть представлен в виде градиента магнитного потенциала. Такое представление, однако, невозможно в тех областях, где текут токи. Там понятие магнитного потенциала лишено смысла. Действительно, из соотношений (18.4) и (56.2) нетрудно получить тождество

$$\operatorname{rot}\ \operatorname{grad}\varphi \equiv 0, \tag{56.4}$$

какова бы ни была функция φ. Если бы магнитное поле представлялось выражением $\mathbf{B} = -\operatorname{grad}\varphi_m$, то из этого тождества мы получили бы $\operatorname{rot}\mathbf{B} = 0$. Следовательно, плотность тока $\mathbf{j}$ должна была бы обращаться в нуль, как это следует из сравнения соотношения $\operatorname{rot}\mathbf{B} = 0$ с формулой (56.1). Это и доказывает наше утверждение.

Основные уравнения магнитного поля постоянных токов в вакууме могут быть записаны в виде

$$\operatorname{rot}\mathbf{B} = \frac{4\pi}{c}\mathbf{j}, \quad \operatorname{div}\mathbf{B} = 0. \tag{56.5}$$

Сравним их с основными уравнениями электростатического поля в вакууме:

$$\operatorname{rot}\mathbf{E} = 0, \quad \operatorname{div}\mathbf{E} = 4\pi\rho. \tag{56.6}$$

Из этих уравнений видно, что электростатическое поле всегда потенциально, его источниками являются неподвижные электрические заряды. Магнитное поле, напротив, не потенциально, а соленоидально, его источниками служат электрические токи.

3. Приведем в заключение математическую *теорему Стокса* (1819–1903), широко используемую в математических преобразованиях теории поля. Эта теорема гласит:

$$\oint (\mathbf{A}\, ds) = \int (\operatorname{rot}\mathbf{A}\, dS), \tag{56.7}$$

где $\mathbf{A}$ — произвольный вектор. Интеграл слева берется по произвольному замкнутому контуру s, справа — по произвольной поверхности, натянутой

на этот контур. Имея в виду прием, указанный на рис. 140, достаточно доказать формулу (56.7) для бесконечно малого контура. Бесконечно малый контур можно рассматривать как плоский. Пусть $\mathbf{n}$ — единичная нормаль к плоскости такого контура, а $\mathbf{N}$ — единичная нормаль к самому контуру, лежащая в его плоскости (рис. 141). Введем вектор $\mathbf{C} = [\mathbf{An}]$, получающийся из $\mathbf{A}$ путем поворота вокруг нормали $\mathbf{n}$. Ясно, что вектор $\mathbf{C}$ лежит в плоскости контура. Применим к нему математическую формулу Гаусса–Остроградского

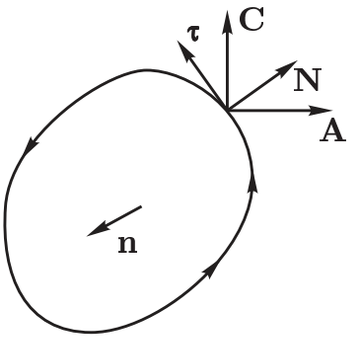


Рис. 141

$$\oint (\mathbf{CN})\, ds = \int \operatorname{div} \mathbf{C}\, dS.$$

Легко показать, что $\operatorname{div} \mathbf{C} = \operatorname{div} [\mathbf{An}] = \mathbf{n} \operatorname{rot} \mathbf{A}$. Кроме того,

$$(\mathbf{CN}) = [\mathbf{An}]\mathbf{N} = \mathbf{A}[\mathbf{nN}] = (\mathbf{A}\boldsymbol{\tau}),$$

где $\boldsymbol{\tau} = [\mathbf{nN}]$ — единичный вектор касательной к контуру s. В результате получим

$$\oint (\mathbf{A}\boldsymbol{\tau})\, ds = \int (\mathbf{n} \operatorname{rot} \mathbf{A})\, dS,$$

что и доказывает теорему. Из доказанного следует:

$$\operatorname{rot}_n \mathbf{A} = \lim_{S \to 0} \frac{1}{S} \oint (\mathbf{A}\, d\mathbf{s}). \tag{56.8}$$

Этим дается *инвариантное определение проекции ротора* $\mathbf{A}$ на направление произвольного вектора $\mathbf{n}$, а следовательно, и *инвариантное определение самого ротора.*

§ 57. Эквивалентность магнитных полей тока и магнитного листка

1. До открытия магнитных действий тока причину магнитных явлений видели в существовании особых субстанций — *северного* и *южного магнетизмов*, взаимодействующих между собой по тому же закону Кулона, которому подчиняются электрические заряды. В одном отношении, однако, магнитные заряды существенно отличались от электрических: в магните невозможно отделить северный полюс от южного. Это вызывало сомнения в реальном существовании магнитных зарядов. Ампер выдвинул гипотезу, подтвержденную всеми последующими исследованиями, согласно которой *магнитных зарядов не существует, а единственными источниками магнитного поля являются токи.* К обычным макроскопическим токам Ампер добавил так называемые *молекулярные токи*, которые, по его представлениям, циркулируют внутри атомов вещества. Как было установлено уже в XX столетии, молекулярные токи Ампера — это просто движущиеся электроны и ядра, из которых построены атомы вещества. Почему же формальное учение о магнетизме, основанное на представлении о магнитных зарядах, приводило к правильным результатам, если в действительности таких зарядов не существует? На этот вопрос дает ответ *теорема Ампера* об эквивалентности магнитных полей тока и магнитного листка.

2. Рассмотрим сначала элементарный виток тока. Магнитное поле вне витка может быть выражено через магнитный потенциал (55.2). Для элементарного витка $\Omega = -(\mathbf{Sr})/r^3$, где $\mathbf{S}$ — вектор элементарной площадки, натянутой на контур тока. Радиус-вектор $\mathbf{r}$ проведен от витка к точке наблюдения, в связи с чем в выражении для Ω и поставлен знак минус. Используя это выражение, получим

$$\varphi_m = \frac{\mathfrak{M}\mathbf{r}}{r^3}, \tag{57.1}$$

где $\mathfrak{M}$ — магнитный момент тока, определяемый выражением (52.2). Формула (57.1) совпадает с соответствующей формулой (19.6) для электрического потенциала точечного диполя. Вместо электрического дипольного момента $\mathbf{p}$ в нее входит магнитный момент тока $\mathfrak{M}$. Таким образом, при нахождении напряженности магнитного поля элементарный виток с током можно заменить элементарным магнитом, состоящим из двух магнитных зарядов противоположных знаков, слегка сдвинутых один относительно другого. Отсюда следует, что магнитное поле элементарного витка (как и поле точечного диполя) меняется обратно пропорционально кубу расстояния от него. Этот результат был использован при выводе формулы (55.8) для напряженности магнитного поля соленоида.

Обобщение полученного результата на конечные витки с током производится обычным приемом разложения конечного витка на элементарные (рис. 132). Заменив каждый элементарный виток эквивалентным ему диполем, получим «магнитный листок», т. е. поверхность S, натянутую на контур тока, одна сторона которой равномерно покрыта северным, а другая — южным магнетизмами. На единицу площади такой поверхности приходится магнитный момент

$$\mathfrak{m} = \frac{I}{c}\,\mathbf{n}, \tag{57.2}$$

называемый *мощностью магнитного листка.* (Единичный вектор нормали $\mathbf{n}$ к поверхности S должен находиться в правовинтовом соотношении с направлением тока.) Магнитное поле такого магнитного листка во внешнем пространстве совпадает с магнитным полем тока I, текущего по контуру листка. В этом и состоит теорема Ампера. В качестве S можно выбрать любую поверхность, опирающуюся на рассматриваемый контур тока.

3. Эквивалентность магнитных полей тока и магнитного листка имеет место *только во внешнем пространстве.* Внутри листка магнитное поле существенно отличается от магнитного поля тока в соответствующих точках пространства. Силовые линии магнитного поля тока тянутся непрерывно, обвиваясь вокруг тока, а магнитный потенциал нигде не претерпевает скачков. Напротив, при прохождении через магнитный листок потенциал меняется скачкообразно на $4\pi\mathfrak{m} = 4\pi I/c$. Это связано с тем, что внутри листка магнитное поле очень сильное и направлено противоположно наружному полю. Магнитный листок подобен конденсатору с близко расположенными обкладками

(рис. 142). На ранней стадии учения о магнетизме изучались магнитные поля вне магнитов. По теореме Ампера они могли быть точно моделированы полями магнитных зарядов, надлежащим образом распределенных внутри магнита. Этим и объясняется, почему расчеты, основанные на представлении о магнитных зарядах, приводили к верным результатам. Однако мы пришли бы к грубым ошибкам, если бы пользовались этими представлениями при расчетах магнитных полей внутри магнитов или в тех областях пространства, по которым текут электрические токи.

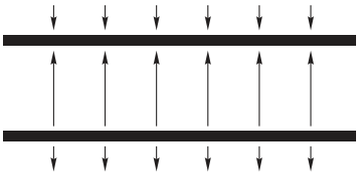


Рис. 142

4. Аналогия между током и магнитным листком относится не только к возбуждаемым ими магнитным полям. Она распространяется также на силы, действующие на токи и листки в магнитных полях. Действительно, ранее уже было показано, что на элементарный виток с током в магнитном поле действует вращающий момент $\mathbf{M} = [\mathfrak{M}\mathbf{B}]$. Остается поэтому доказать, что в неоднородном поле на элементарный виток действует сила

$$\mathbf{F} = (\mathfrak{M}\nabla)\mathbf{B}. \tag{57.3}$$

Для доказательства достаточно ограничиться случаем элементарного витка, имеющего форму бесконечно малого прямоугольника, так как введением добавочных перемычек произвольный контур можно разбить на более мелкие прямоугольные подконтуры (см. рис. 132). Непрямоугольными могут оказаться только подконтуры, примыкающие к границе. Но число перемычек можно взять настолько большим, чтобы площадь, охватываемая такими подконтурами, стала пренебрежимо малой по сравнению с охватываемой площадью всего контура.

Направим оси X и Z параллельно сторонам нашего прямоугольного контура (см. рис. 152). Стороны AB и BC изобразим векторами $\mathbf{a}$ и $\mathbf{b}$. Силы, действующие на стороны AB и CD, равны соответственно $\frac{1}{c}\,I[\mathbf{a}\mathbf{B}_1]$ и $-\frac{1}{c}\,I[\mathbf{a}\mathbf{B}_2]$, где $\mathbf{B}_1$ и $\mathbf{B}_2$ — средние значения соответствующих напряженностей магнитных полей. Так как $\mathbf{B}_1 - \mathbf{B}_2 = -b\,\partial\mathbf{B}/\partial z$, то равнодействующая этих сил равна

$$\mathbf{F}_1 = -\frac{I}{c}\left[\mathbf{a}\frac{\partial\mathbf{B}}{\partial z}\right] b = -\frac{I}{c}\,ab\left[\mathbf{e}_y\frac{\partial\mathbf{B}}{\partial z}\right] = -\frac{I}{c}\,S\left[\mathbf{e}_y\frac{\partial\mathbf{B}}{\partial z}\right],$$

где $S = ab$ — площадь контура. Следовательно, $\mathbf{F}_1 = -\mathfrak{M}[\mathbf{e}_y\,\partial\mathbf{B}/\partial z]$. Аналогично, равнодействующая сил, действующих на противоположные стороны DA и BC, равна $\mathbf{F}_2 = \mathfrak{M}[\mathbf{e}_z\,\partial\mathbf{B}/\partial y]$. Полная сила

$$\mathbf{F} = \mathbf{F}_1 + \mathbf{F}_2 = \mathfrak{M}\left\{\left[\mathbf{e}_z\frac{\partial\mathbf{B}}{\partial y}\right] - \left[\mathbf{e}_y\frac{\partial\mathbf{B}}{\partial z}\right]\right\}.$$

Подставим сюда $\mathbf{B} = B_x\mathbf{e}_x + B_y\mathbf{e}_y + B_z\mathbf{e}_z$. Тогда получим

$$\mathbf{F} = \mathfrak{M}\left(\frac{\partial B_x}{\partial y}\,\mathbf{e}_y + \frac{\partial B_x}{\partial z}\,\mathbf{e}_z - \frac{\partial B_y}{\partial y}\,\mathbf{e}_x - \frac{\partial B_z}{\partial z}\,\mathbf{e}_x\right).$$

В силу уравнения (53.2) $\partial B_y/\partial y + \partial B_z/\partial z = -\partial B_x/\partial x$. Предположим, кроме того, что виток находится в пространстве, где нет токов. Тогда по теореме о циркуляции (56.1) $\partial B_x/\partial y = \partial B_y/\partial x$, $\partial B_x/\partial z = \partial B_z/\partial x$. Следовательно,

$$\mathbf{F} = \mathfrak{M}\left(\frac{\partial B_x}{\partial x}\,\mathbf{e}_x + \frac{\partial B_y}{\partial x}\,\mathbf{e}_y + \frac{\partial B_z}{\partial x}\,\mathbf{e}_z\right) =$$

$$= \mathfrak{M}\frac{\partial}{\partial x}\left(B_x\mathbf{e}_x + B_y\mathbf{e}_y + B_z\mathbf{e}_z\right) = \mathfrak{M}\frac{\partial \mathbf{B}}{\partial x}.$$

Так как магнитный момент $\mathfrak{M}$ направлен по оси X, то это выражение совпадает с выражением (57.3).

ЗАДАЧИ

1. Определить движение заряженной частицы в постоянном однородном магнитном поле $\mathbf{B}$.

Решение. Действующая сила $\mathbf{F} = (e/c)[\mathbf{vB}]$ перпендикулярна к скорости частицы $\mathbf{v}$. Она работы не производит, а потому величина скорости $\mathbf{v}$ остается неизменной. Скорость $\mathbf{v}$ может меняться только по направлению. Допустим сначала, что скорость $\mathbf{v}$ перпендикулярна к магнитному полю $\mathbf{B}$. Тогда частица будет двигаться по окружности. Угловая скорость вращения $\boldsymbol{\omega}$ определится из уравнения

$$-m\omega^2\mathbf{r} = \frac{e}{c}\,[\mathbf{vB}].$$

Подставляя в него $\mathbf{v} = [\boldsymbol{\omega}\mathbf{r}]$, получим

$$-m\omega^2\mathbf{r} = \frac{e}{c}\,[[\boldsymbol{\omega}\mathbf{r}]\mathbf{B}] = \frac{e}{c}\,(\boldsymbol{\omega}\mathbf{B})\mathbf{r},$$

откуда

$$\boldsymbol{\omega} = -\frac{e}{mc}\,\mathbf{B}. \tag{57.4}$$

Величина $|\omega|$ называется *циклотронной частотой*, а радиус окружности

$$r = \frac{v}{|\omega|} = \frac{mcv}{|e|B} \tag{57.5}$$

— *ларморовским радиусом.*

Если скорость $\mathbf{v}$ не перпендикулярна к $\mathbf{B}$, то для нерелятивистских движений ее можно представить в виде $\mathbf{v} = \mathbf{v}_{\|} + \mathbf{v}_{\perp}$, где $\mathbf{v}_{\|}$ — скорость частицы вдоль поля $\mathbf{B}$, а $\mathbf{v}_{\perp}$ — перпендикулярно к нему. Частица будет вращаться по окружности с прежней частотой (57.4). Радиус окружности будет

$$r_{\perp} = \frac{mcv_{\perp}}{|e|B}. \tag{57.6}$$

На это вращение наложится равномерное движение вдоль поля $\mathbf{B}$ со скоростью $\mathbf{v}_{\parallel}$. Таким образом, траекторией частицы будет спираль с осью, направленной вдоль магнитного поля $\mathbf{B}$.

2. Заряженная космическая частица (например, протон) захвачена земным магнитным полем и вращается вокруг Земли в плоскости экватора по окружности радиуса $R = 6700$ км. Напряженность земного магнитного поля на траектории частицы $B = 0{,}4$ Гс. Найти энергию такой частицы.

Ответ. $\mathscr{E} = mc^2 = eBR = 80$ ГэВ.

Указание. Скорость космической частицы настолько мало отличается от скорости света, что ее энергия покоя пренебрежимо мала по сравнению с кинетической энергией. Кинетическую энергию можно вычислять по формуле $\mathscr{E} = mc^2$, где m — релятивистская масса частицы.

3. Серпуховской ускоритель протонов ускоряет эти частицы до энергии $\mathscr{E} = 76$ ГэВ $= 7{,}6 \cdot 10^{10}$ эВ. Если отвлечься от наличия ускоряющих промежутков, то можно считать, что ускоренные протоны движутся по окружности радиуса $R = 236$ м и удерживаются на ней магнитным полем, перпендикулярным к плоскости орбиты. Найти необходимое для этого магнитное поле.

Ответ. $B = \mathscr{E}/(Re) = V/R = 1{,}07 \cdot 10^4$ Гс (V — ускоряющее напряжение, соответствующее энергии $\mathscr{E}$, равное $7{,}6 \cdot 10^{10}$ В $= 2{,}53 \cdot 10^8$ СГСЭ-ед).

Указание. Протоны считать ультрарелятивистскими. Их скорость настолько близка к скорости света, что энергия покоя протона пренебрежимо мала по сравнению с кинетической энергией. При таком условии кинетическую энергию можно вычислить по формуле $\mathscr{E} = mc^2$, где m — релятивистская масса протона.

4. Легкий шарик массы $m = 0{,}5$ г и радиуса $r = 1$ см подвешен на длинной нити и вращается по горизонтальной окружности, радиус которой очень мал по сравнению с длиной нити (конический маятник). Найти изменение угловой скорости вращения шарика $\Delta\omega$ после того, как шарик был заряжен до потенциала $V = 3000$ В и помещен в вертикальное магнитное поле $B = 3000$ Гс. В каком случае угловая скорость увеличится и в каком уменьшится?

Примечание. Окончательную формулу для изменения угловой скорости упростить, использовав соотношение $\omega_0 \equiv qB/(mc) \ll \sqrt{g/l}$, где q — заряд шарика, l — длина нити, g — ускорение свободного падения.

Ответ. $\Delta\omega \equiv \omega - \omega_0 = \pm qB/(2mc) = \pm rVB/(2mc) = 10^{-6}$ с$^{-1}$. Угловая скорость увеличится, когда векторы $\mathbf{B}$ и $\boldsymbol{\omega}_0$ направлены противоположно, и уменьшится, когда их направления совпадают.

5. В центре рамки, имеющей форму равностороннего треугольника со стороной $a = 36$ см, помещена маленькая магнитная стрелка. Плоскость рамки совпадает с плоскостью магнитного меридиана Земли. Какой ток I надо пропустить по рамке, чтобы стрелка повернулась на угол 45°? Горизонтальная составляющая земного магнитного поля $B_0 = 0{,}4$ Гс.

Ответ. $I = (1/18)caB_0 = 2{,}4 \cdot 10^{10}$ СГСЭ-ед. $= 8$ А.

6. Масс-селектор, т. е. прибор, предназначенный для разделения атомных частиц разных масс, состоит из цилиндрического конденсатора с внутренним радиусом $R_1 = 2{,}4$ см и внешним $R_2 = 3$ см (рис. 143). Ионные лучи попадают в селектор через узкую щель S, расположенную посередине между обкладками. Параллельно оси конденсатора (т. е. перпендикулярно к плоскости чертежа) приложено однородное магнитное поле $B = 2000$ Гс.

Какую по модулю и знаку разность потенциалов надо приложить к пластинам конденсатора, чтобы однократно заряженный положительный ион

^{7}Li прошел по средней линии конденсатора, т. е. по окружности радиуса $R = 2{,}7$ см? Найти напряженность электрического поля на этой окружности.

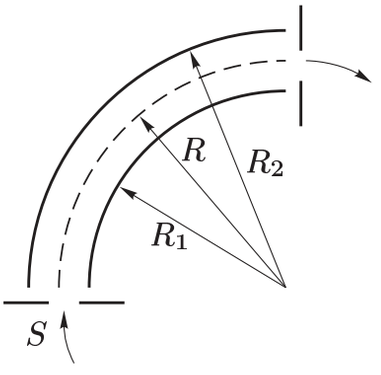


Рис. 143

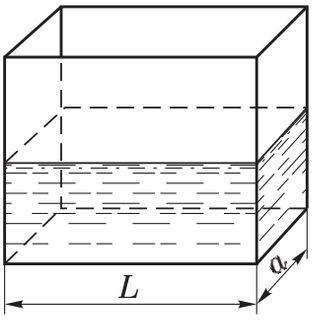


Рис. 144

Энергия иона $\mathscr{E} = 1000$ эВ. Масса атома водорода $1{,}67 \cdot 10^{-24}$ г. На сколько надо изменить эту разность потенциалов, чтобы по той же линии через селектор могли пройти ионы ^{6}Li?

Ответ. Для ^{7}Li $V = ER\ln(R_2/R_1) \approx E(R_2 - R_1) \approx 245$ В, $E = 408$ В/см. Для ^{6}Li $V = 229$ В, $E = 381$ В/см.

7. В прямоугольную кювету (рис. 144), передняя и задняя стенки которой металлические, а боковые диэлектрические, налит электролит с удельной проводимостью $\lambda = 0{,}2$ Ом$^{-1} \cdot$ см$^{-1}$. К металлическим стенкам приложено напряжение $V = 30$ В, и вся кювета помещена в однородное магнитное поле $B = 100$ Гс, направленное вертикально. Размеры кюветы: $L = 20$ см, $a = 2$ см, плотность электролита $\rho = 1$ г/см^3. Определить разность уровней жидкости Δh около правой и левой стенок кюветы.

Ответ. $\Delta h = \lambda VBL/(ac\rho g) = 0{,}6$ см.

8. Для осуществления теплосъема с ядерных реакторов в качестве теплоносителя применяют расплавленные металлы. Перекачка металлов по трубам производится с помощью электромагнитных насосов: участок трубы с расплавленным металлом помещается в поперечное магнитное поле, т. е. поле, перпендикулярное к оси трубы; через тот же участок трубы перпендикулярно к ее оси и к магнитному полю пропускается поперечный электрический ток. При этом возникает сила, приводящая жидкость в движение. Оценить ток, необходимый для перекачки ртути по трубе диаметром $D = 20$ мм и длиной $L = 10$ м со скоростью $V = 1$ л/с. Вязкость ртути $\eta = 1{,}5 \cdot 10^{-2}$ дин $\cdot$ с/см^2, магнитное поле $B = 1000$ Гс.

Решение. Сила, действующая на столб ртути в трубе, $F \approx IDB/c$. Такая сила эквивалентна разности давлений на концах трубы $P_2 - P_1 = 4F/(\pi D^2)$. Подставляя эту разность в формулу Пуазейля (см. т. I, § 97), получим

$$I \approx 32\eta cL/(BD^3) = 1{,}8 \cdot 10^{12} \text{ СГСЭ-ед.} = 600 \text{ А}.$$

9. Найти отношение силы кулоновского расталкивания к силе амперова притяжения двух параллельных пучков, состоящих из электронов, прошедших ускоряющий потенциал $V = 10$ кВ.

Ответ. $F_{\text{кул}}/F_{\text{амп}} = mc^2(2eV) \approx 25$.

10. Две щели S_1 и S_2 шириной $d = 0{,}1$ см каждая (рис. 145), установленные в эвакуированном сосуде, выделяют плоский пучок электронов

с энергией $\mathscr{E} = 400$ эВ. На каком расстоянии x от щели S_2 ширина электронного пучка удвоится из-за кулоновского расталкивания электронов, если электронный ток, приходящийся на единицу длины щели (за щелью S_2), равен $i = 10^{-4}$ А/см? При расчетах щели считать бесконечно длинными.

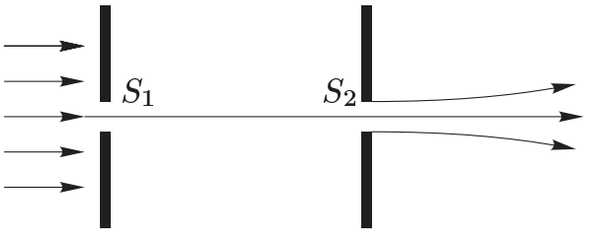


Рис. 145

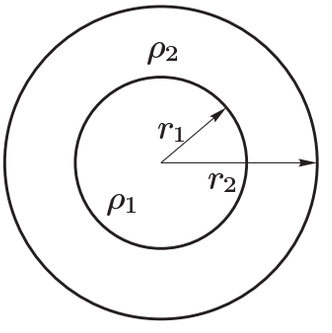


Рис. 146

Решение. Электрическое поле E можно найти, пренебрегая расширением пучка. На границе и вне пучка $E = 2\pi\sigma$, где $\sigma = i/e$ — заряд, приходящийся на единицу площади пучка. Частицы, находящиеся на границе пучка, в поперечном направлении будут двигаться равноускоренно и за время t сместятся на

$$\Delta y = eEt^2/(2m) = eEx^2/(2mv^2).$$

Искомое расстояние x определится из требования $\Delta y = d/2$, которое дает

$$x = \sqrt{mdv^3/(2\pi ie)}, \quad v = \sqrt{2\mathscr{E}/m}.$$

Отсюда $x = 13$ см.

11. Бесконечно длинный цилиндрический провод (рис. 146) состоит из двух коаксиальных цилиндров: центрального сплошного цилиндра радиуса r_1, сделанного из материала с удельным сопротивлением ρ_1 и окружающего его полого цилиндра с внешним радиусом r_2, изготовленного из материала с удельным сопротивлением ρ_2. Внешняя поверхность сплошного и внутренняя поверхность полого цилиндров находятся в электрическом контакте. По проводу параллельно его оси течет постоянный ток I. Найти выражения для напряженности магнитного поля внутри и вне проводов.

Ответ.

$$B = \begin{cases} 2I\rho_2 r/(cA) & (r \leqslant r_1), \\ 2I[\rho_1(r^2 - r_1^2) + \rho_2 r_1^2]/(cAr) & (r_1 \leqslant r \leqslant r_2), \\ 2I/(cr) & (r \geqslant r_2), \end{cases}$$

$A = \rho_1(r_2^2 - r_1^2) + \rho_2 p_1^2$.

12. Вдоль тонкостенной цилиндрической трубки радиуса $a = 25$ мм течет постоянный ток $I = 20$ А. В стенке трубки имеется тонкая щель ширины $d = 1$ мм, параллельная оси трубки. Определить магнитное поле внутри трубки на больших расстояниях r от середины щели ($r \gg d$).

Ответ. Магнитные силовые линии в трубке имеют форму коаксиальных дуг окружностей, центры которых расположены на оси щели. $B = Id/(\pi car)$. На оси трубки (при $r = a$) $B = Id/(\pi ca^2) \approx 0{,}1$ Гс.

13. По двум бесконечно длинным прямолинейным проводникам, сделанным из немагнитного материала и изолированным друг от друга, текут

в противоположных направлениях токи с одной и той же плотностью $j = 1000$ А/см^2. Проводники ограничены участками цилиндрических поверхностей и отделены один от другого полостью $П$. (На рис. 147 поперечные сечения проводников заштрихованы.) Найти величину и направление магнитного поля в полости $П$. Ток в левом проводнике направлен к читателю, а в правом — от читателя. Расстояние между осями цилиндров $AB = D = 5$ см.

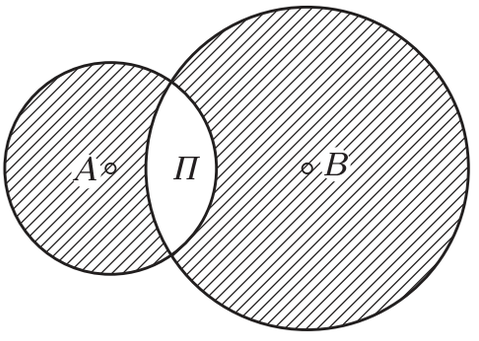


Рис. 147

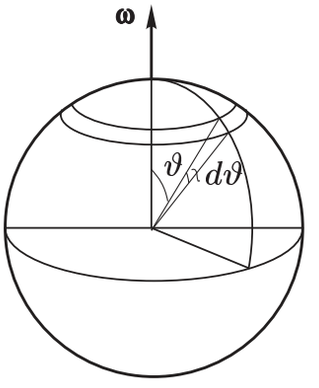


Рис. 148

Ответ. $B = 2\pi jD = 3140$ Гс. Поле $\mathbf{B}$ перпендикулярно к линии AB и направлено вверх.

14. Определить напряженность магнитного поля внутри бесконечной цилиндрической полости, сделанной в бесконечном цилиндрическом проводе, вдоль которого течет постоянный ток плотности j, равномерно распределенный по сечению провода. Расстояние между осями провода и полости равно d.

Ответ. $\mathbf{B} = \dfrac{2\pi}{c}\, j[\mathbf{sd}]$, где $\mathbf{s}$ — единичный вектор, направленный вдоль тока, а вектор $\mathbf{d}$ проведен от оси провода к оси цилиндрической полости.

15. В высокий цилиндрический сосуд радиуса R налит электролит. Внутри сосуда параллельно его оси расположен цилиндрический металлический стержень радиуса r, поверхность которого покрыта изолирующей краской. Расстояние между осями стержня и сосуда равно a. В электролите параллельно оси течет ток силы I, возвращающийся обратно по стержню. Считая плотность тока в электролите постоянной, найти силу, с которой магнитное поле, созданное этими токами, действует на единицу длины стержня. Куда эта сила направлена?

Ответ. $F = \dfrac{2I^2}{c^2(R^2 - r^2)}\, a$. Сила направлена от оси цилиндра к оси стержня.

16. Заряженный шарик радиуса a равномерно вращается вокруг своего диаметра с угловой скоростью $\boldsymbol{\omega}$. Общий заряд шарика равен q. Найти магнитное поле шарика на расстояниях r, больших по сравнению с a, если заряд q равномерно распределен 1) по поверхности шарика; 2) по объему шарика.

Ответ. 1) $\mathbf{B} = \dfrac{qa^2}{3c}\left[\dfrac{3(\boldsymbol{\omega}\mathbf{r})\mathbf{r}}{r^5} - \dfrac{\boldsymbol{\omega}}{r^3}\right]$; 2) $\mathbf{B} = \dfrac{qa^2}{5c}\left[\dfrac{3(\boldsymbol{\omega}\mathbf{r})\mathbf{r}}{r^5} - \dfrac{\boldsymbol{\omega}}{r^3}\right]$. (Вектор $\mathbf{r}$ проведен из центра шарика.)

Решение. Рассмотрим случай 1). Возьмем на поверхности шарика бесконечно узкий поясок, заключенный между углами ϑ и $\vartheta + d\vartheta$ (рис. 148).

Вращаясь с угловой скоростью $\boldsymbol{\omega}$, такой поясок эквивалентен круговому току $dI = \omega q \sin\vartheta\, d\vartheta/(4\pi)$ с магнитным моментом $d\mathfrak{M} = S\, dI/c = \pi a^2 q\omega \sin^3\vartheta\, d\vartheta (4\pi c)^{-1}$. Интегрируя по ϑ, находим магнитный момент всего шара $\mathfrak{M} = qa^2\omega/(3c)$. Отсюда и получается результат, приведенный в ответе. Случай 2) рассматривается аналогично.

17. Для моделирования траектории атомной частицы с зарядом e и импульсом p, движущейся в постоянном магнитном поле, часто пользуются тем обстоятельством, что очень легкий (невесомый) гибкий проводящий шнур с током I, находящийся под постоянным механическим натяжением T, занимает в том же магнитном поле положение, совпадающее с траекторией частицы. (Предполагается, что вне магнитного поля участки шнура прямолинейны и расположены вдоль соответствующих прямолинейных же участков траектории.) Обосновать этот метод. Найти связь между I, e, p, T. Величина магнитного поля B может меняться в пространстве, но его направление должно оставаться неизменным. Шнур и траектория частицы перпендикулярны к магнитному полю.

Р е ш е н и е. Радиус кривизны траектории частицы в магнитном поле $\rho = cp/(eB)$. На единицу длины натянутого шнура действует нормально упругая сила T/ρ_1, которая уравновешивается амперовой силой IB/c. Это дает для радиуса кривизны шнура $\rho_1 = Tc/(IB)$. При выполнении условия $\rho_1 = \rho$, т. е. $Ip = Te$, форма траектории будет совпадать с формой шнура.

§ 58. Магнитное поле в веществе

1. В веществе магнитное поле возбуждается не только электрическими токами, текущими по проводам, но и движениями заряженных частиц внутри самих атомов и молекул. Согласно полуклассической теории Бора, электроны вращаются вокруг атомных ядер по замкнутым орбитам. Кроме того, они совершают вращения вокруг собственных осей подобно вращениям планет вокруг своих осей. С таким внутренним вращением связан определенный момент импульса, называемый *спином электрона*. Спином обладают не только электроны, но и атомные ядра. Орбитальные и спиновые вращения заряженных частиц аналогичны токам и возбуждают магнитные поля. Наглядное представление о движении электронов по классическим орбитам и об их вращениях вокруг собственных осей позднее было заменено более общей и абстрактной картиной движения, которую дает квантовая механика. В этой картине не существует понятия траектории частицы. Термин «орбитальное движение» сохранился, но он носит условный характер. Однако для учения о магнетизме существенна не наглядность движения, а механический и магнитный моменты, связанные с этим движением. Таким образом, по современным представлениям, магнетизм вещества обусловлен тремя причинами: 1) орбитальным движением электронов вокруг атомных ядер, 2) собственным вращением, или спином, электронов, 3) собственным вращением, или спином, атомных ядер. Тяжелые атомные ядра движутся значительно медленнее легких электронов. Поэтому магнитные моменты атомных ядер в тысячи раз меньше орбитальных и спиновых магнитных моментов электронов.

Ядерный магнетизм становится существенным только вблизи абсолютного нуля температур, да и то при условии, что орбитальные и спиновые магнитные моменты электронов скомпенсированы, так что их результирующий момент равен нулю.

Атомы вещества, совершая беспорядочное тепловое движение, в отсутствие внешнего магнитного поля обычно ориентированы хаотически. Возбуждаемые ими магнитные поля в окружающем пространстве компенсируют друг друга. При наложении внешнего магнитного поля атомы полностью или частично ориентируются в направлении этого поля, и тогда компенсация нарушается. В таких случаях говорят, что тело *намагничено.* Тела, способные намагничиваться, называются *магнетиками.* Большинство веществ при внесении в магнитное поле намагничивается слабо. Сильными магнитными свойствами обладают только *ферромагнитные* вещества: железо, никель, кобальт, множество их сплавов, а также элементы редких земель. Постоянные магниты, изготовляемые из стали и различных магнитных сплавов, намагничены и в отсутствие внешнего магнитного поля. Стальной стержень можно намагнитить (правда, очень слабо) в магнитном поле Земли. Для этого стержень ориентируют вдоль магнитного меридиана и постукивают по нему молотком. На конце стержня, обращенном к южному магнитному полюсу Земли (он находится на севере), появляется северный магнитный полюс, а на противоположном конце — южный. Если стержень повернуть к северу другим концом, то при постукивании молотком он перемагнитится и его магнитные полюсы поменяются местами.

2. Магнитное поле, как и поле электрическое, может быть *микроскопическим* и *макроскопическим.* Микроскопическое поле есть истинное поле, возбуждаемое движущимися элементарными зарядами вещества. Оно резко меняется на расстояниях атомного масштаба. Макроскопическое поле получается из микроскопического путем сглаживания, т. е. усреднением по физически бесконечно малым объемам пространства. Напряженность макроскопического поля обозначается буквой $\mathbf{B}$. Вектор $\mathbf{B}$ есть основной вектор, характеризующий макроскопическое поле в веществе. Орбитальные и спиновые вращения электронов и атомных ядер в отношении возбуждаемого ими магнитного поля эквивалентны каким-то токам, циркулирующим в атомах вещества. Они получили общее название *молекулярных токов.* Для вычисления макроскопического поля $\mathbf{B}$ молекулярные токи можно также как-то сгладить, заменив их макроскопическими токами, непрерывно меняющимися в пространстве. Такие макроскопические токи называются *токами намагничивания.* Их плотность в дальнейшем обозначается $\mathbf{j}_m$. Обычные токи, текущие по проводам, связаны с перемещениями в веществе носителей тока — электронов или ионов. Эти токи называются *токами проводимости.* Плотность токов проводимости будем обозначать $\mathbf{j}$. Таким образом, поле $\mathbf{B}$ возбуждается токами проводимости и токами намагничивания. Влияние среды на магнитное поле сводится к действию токов намагничивания. Если известны токи проводимости

и токи намагничивания, то можно как бы забыть о наличии вещества и вычислять напряженность поля $\mathbf{B}$ по формулам для вакуума.

Поскольку вектор $\mathbf{B}$ есть напряженность магнитного поля, возбуждаемого в вакууме какими-то токами, для него справедливо уравнение

$$\oint \mathbf{B}\, d\mathbf{S} = 0, \tag{58.1}$$

или в дифференциальной форме

$$\operatorname{div} \mathbf{B} = 0. \tag{58.2}$$

Каждое из этих уравнений выражает тот факт, что магнитных зарядов не существует.

Для вектора $\mathbf{B}$, очевидно, имеет место теорема о циркуляции. Надо только ток проводимости I дополнить током намагничивания I_m. Тогда получится

$$\oint\limits_L (\mathbf{B}\, d\mathbf{l}) = \frac{4\pi}{c}\,(I + I_m), \tag{58.3}$$

или в дифференциальной форме

$$\operatorname{rot} \mathbf{B} = \frac{4\pi}{c}\,(\mathbf{j} + \mathbf{j}_m). \tag{58.4}$$

В формуле (58.3) I и I_m обозначают полные токи проводимости и намагничивания, пронизывающие замкнутый контур L.

3. Намагниченность среды принято характеризовать не токами намагничивания, как это сделано выше, а *вектором намагничивания* $\mathbf{I}$. Так называют средний магнитный момент единицы объема магнетика, создаваемый молекулярными токами. Через вектор $\mathbf{I}$ можно выразить и плотность токов намагничивания в среде.

Допустим сначала, что магнетик имеет форму намагниченного прямого круглого цилиндра, магнитный момент которого направлен вдоль его оси (рис. 149). Молекулярные токи в намагниченном магнетике текут согласованно, и возбуждаемые ими магнитные поля усиливают друг друга. Если $\mathfrak{M}$ — средний магнитный момент одной молекулы, то, очевидно,

$$\mathbf{I} = n\mathfrak{M}, \tag{58.5}$$

где n — среднее число молекул в единице объема. Полный магнитный момент всего цилиндра равен $V\mathbf{I}$, где $V = SL$ — объем цилиндра (S — площадь его основания, L — высота). Молекулярные токи соседних молекул в местах их соприкосновения текут в противоположных направлениях и макроскопически взаимно компенсируют друг друга. Некомпенсированными остаются только молекулярные токи, выходящие на наружную боковую поверхность цилиндра. Эти токи складываются в макроскопический поверхностный ток I_m, циркулирующий по боковой поверхности цилиндра. Во внешнем пространстве он возбуждает такое же макроскопическое поле, что и молекулярные токи. Этот ток и есть ток намагничивания. Его магнитный момент

равен, с одной стороны, $I_m\mathbf{S}/c$. С другой стороны, тот же магнитный момент равен $V\mathbf{I} = SL\mathbf{I}$. Таким образом, $I_m\mathbf{S}/c = SL\mathbf{I}$. Так как

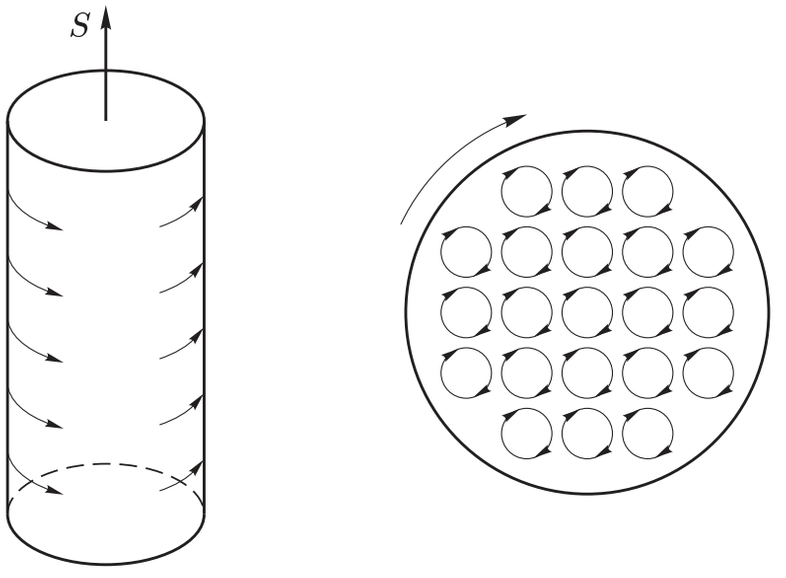


Рис. 149

векторы $\mathbf{S}$ и $\mathbf{I}$ одинаково направлены, то $I_m = cLI$. Следовательно, поверхностный ток намагничивания, приходящийся на единицу длины цилиндра, равен

$$i_m = cI. \tag{58.6}$$

Обобщим теперь результат (58.6) на случай косого цилиндра (рис. 150). Пусть поверхностные токи намагничивания текут в плоскостях, параллельных основаниям цилиндра. Вектор намагничивания $\mathbf{I}$ будет перпендикулярен к этим основаниям. Если α — угол между вектором $\mathbf{I}$ и осью цилиндра, то для магнитного момента последнего можно написать $V\mathbf{I} = SL\mathbf{I}\cos\alpha$. Тот же момент можно представить в виде $I_m\mathbf{S}/c = Li_m\mathbf{S}/c$, где i_m — ток намагничивания, приходящийся на единицу длины образующей боковой поверхности цилиндра. Приравнивая оба выражения, получаем

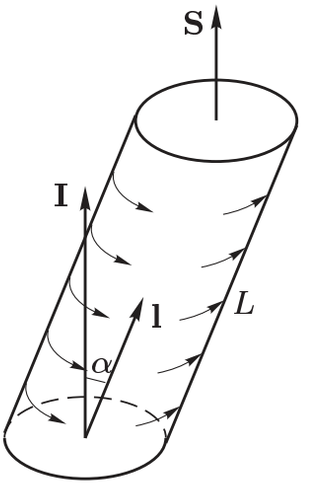


Рис. 150

$$i_m = cI\cos\alpha = c(\mathbf{I}\mathbf{l}) = cI_l. \tag{58.7}$$

Здесь $\mathbf{l}$ — единичный вектор, направленный вдоль оси цилиндра. Таким образом, ток i_m определяется только осевой составляющей вектора намагничивания $\mathbf{I}$. Формула (58.7) и является обобщением формулы (58.6). Ее можно применять и в тех случаях, когда магнетик намагничен неоднородно. Для этого цилиндр следует брать бесконечно малым. В случае неоднородной намагниченности в магнетике возникают не только поверхностные, но и объемные токи намагничивания. Выражение для плотности таких токов будет получено в следующем параграфе.

§ 59. Теорема о циркуляции магнитного поля в веществе

1. Найдем циркуляцию вектора $\mathbf{B}$ по любому замкнутому контуру L. Для этого надо вычислить ток намагничивания I_m, пронизывающий этот контур. Натянем на контур L произвольную поверхность S. На рис. 151 слева изображено сечение этой поверхности и контура L

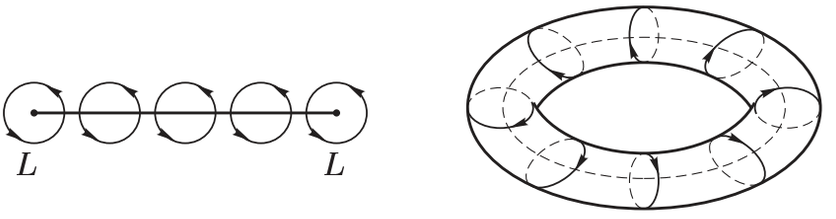


Рис. 151

плоскостью рисунка. Одни молекулярные токи пересекают поверхность S дважды: раз в положительном и раз в отрицательном направлении. Такие токи не вносят никакого вклада в ток намагничивания через поверхность S. Другие молекулярные токи обвиваются вокруг контура L. Каждый из них пересекает поверхность только один раз — противоположно направленный ток в молекуле выходит уже за пределы поверхности S. Такие молекулярные токи и создают макроскопический ток намагничивания I_m, пронизывающий площадь S.

Выразим ток I_m через вектор намагничивания $\mathbf{I}$. Для этого окружим контур L бесконечно узкой трубкой (рис. 151, справа). Согласно формуле (58.7) по поверхности такой трубки циркулирует ток намагничивания с линейной плотностью $i_m = cI_l$. Он только один раз пересекает поверхность S. Ток, приходящийся на элемент длины трубки, равен $i_m\,dl = cI_l\,dl = c(\mathbf{I}\,d\mathbf{l})$. Полный ток намагничивания, пронизывающий поверхность S, найдется интегрированием этого выражения по всему замкнутому контуру L. Это дает

$$I_m = c\oint\limits_L (\mathbf{I}\,d\mathbf{l}). \tag{59.1}$$

Внеся это выражение в формулу (58.3), придадим ей вид

$$\oint (\mathbf{B} - 4\pi\mathbf{I})\,d\mathbf{l} = \frac{4\pi}{3}\,I. \tag{59.2}$$

В дифференциальной форме:

$$\operatorname{rot}\mathbf{B} = \frac{4\pi}{c}\,(\mathbf{j} + c\operatorname{rot}\mathbf{I}). \tag{59.3}$$

Сравнивая это уравнение с уравнением (58.4), находим

$$\mathbf{j}_m = c\operatorname{rot}\mathbf{I}. \tag{59.4}$$

Если намагниченность однородна, т. е. $\mathbf{I} = \text{const}$, то $\mathbf{j}_m = 0$. Если же она неоднородна, то объемная плотность тока намагничивания, вообще говоря, отлична от нуля.

2. Введем теперь вспомогательный вектор

$$\mathbf{H} = \mathbf{B} - 4\pi\mathbf{I}. \tag{59.5}$$

Тогда формулы (59.2) и (59.3) примут вид

$$\oint \mathbf{H}\, d\mathbf{l} = \frac{4\pi}{c}\, I, \tag{59.6}$$

$$\operatorname{rot} \mathbf{H} = \frac{4\pi}{c}\, \mathbf{j}. \tag{59.7}$$

После введения вектора **H** из уравнений (59.6) и (59.7) выпадают токи намагничивания, остаются только токи проводимости. В этом смысл введения вектора **H**. Вектор **H** играет в учении о магнетизме такую же вспомогательную роль, что и вектор **D** в учении о диэлектриках. *Основным вектором является вектор* **B**. Это — *силовой вектор*, и его следовало бы называть напряженностью магнитного поля в веществе. Однако по историческим причинам *напряженностью магнитного поля в веществе называют вектор* **H**, а вектор **B** получил неудачное название *магнитной индукции*. Такая нерациональная терминология сложилась потому, что исторически учение о магнетизме развивалось по аналогии с электростатикой. Источниками магнитного поля считались *магнитные заряды*, а их, как было установлено позднее, в действительности не существует. Мы вынуждены пользоваться этой нерациональной терминологией ввиду того, что она общепринята. Впрочем, в большинстве случаев мы будем избегать употребления терминов «индукция» и «напряженность» магнитного поля, заменяя их соответственно на «вектор **B**» и «вектор **H**». В вакууме векторы **B** и **H** тождественно совпадают.

3. Согласно определению (59.5) векторы **B** и **H** имеют одинаковую размерность. Они должны иметь и общую единицу. Единицей **B** в гауссовой системе является *гаусс*. Та же единица применяется и для измерения **H**. Однако в этом случае ее называют *эрстедом*. Величину **B** измеряют в гауссах, а величину **H** — в эрстедах. Считается ошибкой сказать, что поле **B** равно стольким-то эрстедам, а поле **H** — стольким-то гауссам. Мы не можем с одобрением относиться к такому соглашению, так как *между гауссом и эрстедом нет абсолютно никакой разницы. Это — разные названия одной и той же единицы.* Следовало бы сохранить только одно из этих названий: либо гаусс, либо эрстед.

§ 60. Граничные условия для векторов B и H

1. Из уравнений (58.1) и (59.6) легко получить условия, которым должны удовлетворять векторы **B** и **H** на границе раздела двух магнетиков. Уравнение (58.1) формально не отличается от соответствующего

уравнения для вектора электрического смещения $\mathbf{D}$ при отсутствии электрических зарядов. Отсюда следует, что на границе раздела нормальные слагающие вектора $\mathbf{B}$ должны быть непрерывны:

$$B_{1n} = B_{2n}. \tag{60.1}$$

Перейдем к выводу граничных условий для вектора $\mathbf{H}$. В целях общности будем предполагать, что вдоль границы раздела течет поверхностный ток проводимости с линейной плотностью $\mathbf{i}$. Применим теорему о циркуляции (59.6) к бесконечно малому прямоугольному контуру $ABCD$ (рис. 152), высота которого пренебрежимо мала по сравнению с длиной основания l. Тогда можно пренебречь вкладом в циркуляцию, который вносят боковые стороны AB и CD. В этом приближении циркуляция вектора $\mathbf{H}$ будет $(H_{2t} - H_{1t})l$. По теореме о циркуляции та же величина равна $(4\pi/c)i_N l$, где i_N — слагающая тока $\mathbf{i}$ вдоль нормали к контуру $\mathbf{N} = [\mathbf{nt}]$. Приравнивая оба выражения, получим

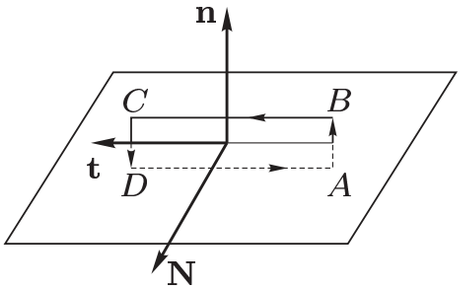


Рис. 152

$$H_{2t} - H_{1t} = \frac{4\pi}{c}\, i_N. \tag{60.2}$$

Придадим этому соотношению векторную форму. Введем единичный вектор касательной к границе раздела $\mathbf{t} = [\mathbf{Nn}]$. Тогда левая часть равенства (60.2) представится в виде

$$(\mathbf{H}_2 - \mathbf{H}_1)\mathbf{t} = (\mathbf{H}_2 - \mathbf{H}_1)\,[\mathbf{Nn}] = ([\mathbf{nH}_2] - [\mathbf{nH}_1])\,\mathbf{N},$$

а правая — в виде $(4\pi)/c(\mathbf{iN})$. Вектор $\mathbf{N}$ в плоскости раздела сред может быть направлен как угодно. Поэтому, приравнивая оба выражения, получим

$$[\mathbf{nH}_2] - [\mathbf{nH}_1] = \frac{4\pi}{c}\,\mathbf{i}. \tag{60.3}$$

Если на границе сред ток проводимости не течет, то

$$H_{1t} = H_{2t}, \tag{60.4}$$

т. е. тангенциальные слагающие вектора $\mathbf{H}$ на границе раздела непрерывны.

2. Граничные условия (60.1) и (60.4) указывают *принципиальный способ* измерения векторов $\mathbf{B}$ и $\mathbf{H}$ в веществе. Непосредственное измерение магнитного поля в веществе с помощью пробного витка или магнитной стрелки не может быть выполнено. Во-первых, потому, что в вещество не всегда можно поместить пробный виток (например, это нельзя сделать, когда тело твердое). Во-вторых, если бы это и можно было сделать, то оставалось бы неясным, как по силе, действующей на виток, и ее моменту найти векторы $\mathbf{B}$ и $\mathbf{H}$. Поэтому для измерения $\mathbf{B}$ и $\mathbf{H}$ в веществе необходимо сделать полость и измерить вектор $\mathbf{B}$

в этой полости. Однако результат будет зависеть от формы полости. Рассмотрим два наиболее важных случая.

Случай 1. В магнетике сделан бесконечно узкий канал, параллельный магнитному полю. Удаление вещества из такого канала лишь бесконечно мало возмущает поле в окружающем магнетике. В силу граничного условия (60.4) векторы $\mathbf{H}$ в канале и в окружающем магнетике должны быть одинаковы. Поместим пробный виток в канал и измерим там величину $\mathbf{B}$. Она будет равна вектору $\mathbf{H}$ в окружающем магнетике.

Случай 2. В магнетике сделана щель, ограниченная двумя бесконечно близкими плоскостями, перпендикулярными к магнитному полю. Удаление вещества из такой бесконечно узкой щели также бесконечно мало возмущает магнитное поле в окружающем магнетике. В силу граничного условия (60.1) векторы $\mathbf{B}$ в щели и в окружающем магнетике одинаковы. Измерив величину $\mathbf{B}$ в щели, мы найдем ее значение и в окружающем магнетике.

§ 61. Магнитная восприимчивость и магнитная проницаемость

1. Железо и все так называемые ферромагнитные вещества (сталь, кобальт, никель, различные магнитные сплавы) обладают не только сильными магнитными свойствами. Они характеризуются весьма сложной зависимостью между векторами $\mathbf{I}$ и $\mathbf{B}$. Эта зависимость — *нелинейная*. Кроме того, для указанных веществ наблюдается *гистерезис*, т. е. зависимость намагничивания от предшествующей истории магнетика. Только для парамагнитных и диамагнитных сред (а такие среды по сравнению с ферромагнитными обладают слабыми магнитными свойствами) зависимость между $\mathbf{I}$ и $\mathbf{B}$ линейная и для изотропных сред может быть записана в виде $\mathbf{I} = \varkappa_1 \mathbf{B}$. Было бы рационально записывать это соотношение именно в такой форме. Однако в силу исторических причин, о которых было сказано в § 59, принято выражать вектор $\mathbf{I}$ не через $\mathbf{B}$, а через $\mathbf{H}$. Вместо соотношения, приведенного выше, пишут

$$\mathbf{I} = \varkappa \mathbf{H}. \tag{61.1}$$

Ошибки в этом нет, так как ввиду соотношения (59.9) из пропорциональности между $\mathbf{I}$ и $\mathbf{B}$ следует и пропорциональность между $\mathbf{I}$ и $\mathbf{H}$. Подставляя выражение (61.1) в соотношение (59.5), получим

$$\mathbf{B} = \mu \mathbf{H}, \tag{61.2}$$

где

$$\mu = 1 + 4\pi\varkappa. \tag{61.3}$$

Величина $\varkappa$ называется *магнитной восприимчивостью*, а μ — *магнитной проницаемостью* вещества. Тела, для которых $\varkappa > 0$ и, следовательно, $\mu > 1$, называются *парамагнитными*, или *парамагнетиками*. К ним относятся, например, кислород, алюминий, платина, хлористое железо ($FeCl_3$) и т. д. Тела, для которых $\varkappa < 0$ и, следовательно, $\mu < 1$,

называются *диамагнитными*, или *диамагнетиками*. Таковы азот, углекислота, вода, серебро, висмут и пр. Парамагнетики намагничиваются вдоль, а диамагнетики — противоположно магнитному полю.

Зависимость вида (61.2) часто используется и для ферромагнетиков, например мягкого железа. Ясно, что она не учитывает явлений гистерезиса. Кроме того, для учета нелинейных эффектов магнитную проницаемость μ надо считать функцией напряженности поля H (или B). В слабых полях B можно разложить по степеням H и ограничиться первым членом этого разложения. В таком приближении величина μ считается постоянной. Более подробно магнитные свойства ферро- и других магнетиков рассматриваются ниже (см. § 76, 77, 79, 80). Пока же при рассмотрении различных явлений мы будем пользоваться соотношениями (61.1) и (61.2), предполагая, что постоянные $\varkappa$ и μ от напряженности магнитного поля не зависят. Разумеется, это строго допустимо только для пара- и диамагнетиков. Для ферромагнетиков результаты, полученные таким путем, должны рассматриваться как оценочные.

2. На границе раздела двух магнетиков магнитные силовые линии должны преломляться. Это следует из граничных условий (60.1) и (60.4), если учесть соотношение (61.2). Допустим, что граница раздела плоская и никакой ток проводимости по ней не течет (рис. 153).

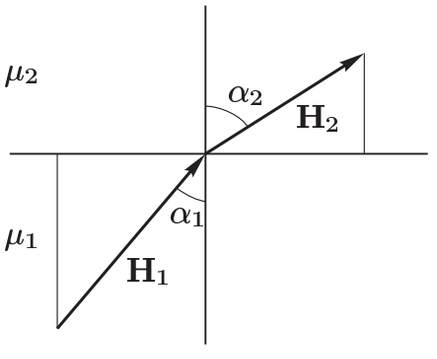


Рис. 153

Непрерывность тангенциальных слагающих вектора $\mathbf{H}$ дает

$$H_1 \sin\alpha_1 = H_2 \sin\alpha_2,$$

а непрерывность нормальных слагающих вектора $\mathbf{H}$ —

$$\mu_1 H_1 \cos\alpha_1 = \mu_2 H_2 \cos\alpha_2.$$

Из этих уравнений находим

$$\frac{\operatorname{tg}\alpha_1}{\operatorname{tg}\alpha_2} = \frac{\mu_1}{\mu_2} \qquad (61.4)$$

— соотношение, аналогичное соотношению (15.9) для электрических силовых линий. При переходе из магнетика с меньшей в магнетик с большей магнитной проницаемостью магнитная силовая линия удаляется от нормали к границе раздела сред. Это приводит к *концентрации магнитных силовых линий* в магнетиках с большей μ.

На рис. 51 показано, как искажаются электрические силовые линии при внесении в однородное электрическое поле полого шара из диэлектрика. Совершенно так же действует полый железный шар, если его внести в магнитное поле. Из-за преломления магнитные силовые линии концентрируются преимущественно в железе. Внутри полости их концентрация мала. Это значит, что магнитное поле в полости сильно ослаблено по сравнению с внешним полем, т. е. оболочка железной полости обладает экранирующим действием по отношению к внешнему магнитному полю. На этом основана *магнитная защита*. Для того

чтобы предохранить какой-либо чувствительный прибор от воздействия внешних магнитных полей, его окружают железной оболочкой.

Но оболочка не защищает внешние тела от воздействия магнитных полей токов или магнитов, помещенных внутри самой оболочки. Например, магнитное поле прямого круглого провода с током не изменится, если окружить его коаксиальной железной трубой. Это непосредственно следует из теоремы о циркуляции. Магнитные силовые линии имеют форму окружностей с центрами на оси провода (см. рис. 126). Циркуляция вектора $\mathbf{H}$ по одной из таких окружностей будет $2\pi RH$. По теореме о циркуляции та же величина равна $4\pi I/c$. Сравнивая оба выражения, получаем $H = 2I/(cR)$ независимо от того, окружен провод железной трубой или не окружен. В этом можно убедиться на демонстрационном опыте. Надо произвести сначала опыт Эрстеда с обычным проводом (см. рис. 134). Затем повторить его с тем же проводом, заключенным в железную трубу. Оказывается, что в обоих случаях отклонение магнитной стрелки одно и то же.

Железная оболочка лишь частично защищает окруженные ею тела от действия внешних магнитных полей. Чем больше магнитная проницаемость μ, тем сильнее это защитное действие. Однако существуют тела, которые в этом отношении являются идеальными. Это — *сверхпроводники.* Оболочка из сверхпроводника, находящегося в сверхпроводящем состоянии, полностью защищает окружаемые ею тела от действия внешнего магнитного поля (см. § 80).

3. Если магнетик ввести в однородное магнитное поле, то он намагнитится. Возникающее при этом магнитное поле внутри магнетика сильно зависит от формы последнего. В общем случае поле внутри тела магнетика неоднородно. Только для тел, имеющих форму эллипсоида, внутреннее магнитное поле однородно. Оно отличается от внешнего поля не только по величине, но, вообще говоря, и по направлению. Предельными случаями эллипсоида являются очень длинный и очень короткий цилиндры. Если ось цилиндра параллельна внешнему магнитному полю, то расчет сильно упрощается. Внутри очень длинного цилиндра вектор $\mathbf{H}$ равен вектору $\mathbf{H}_0$ внешнего магнитного поля. Поэтому

$$\mathbf{I} = \varkappa\mathbf{H} = \frac{\mu-1}{4\pi}\,\mathbf{H}_0 = \frac{\mu-1}{4\pi}\,\mathbf{B}_0,$$

так как во внешнем пространстве $\mathbf{H}_0 = \mathbf{B}_0$. В случае очень короткого цилиндра

$$\mathbf{I} = \varkappa\mathbf{H} = \frac{\varkappa}{\mu}\,\mathbf{B} = \frac{\mu-1}{4\pi\mu}\,\mathbf{B}_0.$$

При одном и том же внешнем поле $\mathbf{B}_0$ намагниченность магнетиков с большой проницаемостью μ во втором случае много меньше, чем в первом. Для цилиндров промежуточных размеров намагниченность $\mathbf{I}$ будет неоднородна. Однако ее значение будет лежать между найденными двумя предельными значениями. Для шара, как легко рассчитать,

намагниченность однородна и определяется формулой

$$\mathbf{I} = \frac{3\mu - 1}{4\pi\mu + 2}\,\mathbf{B}_0.$$

Зависимость намагничивания от формы тела можно демонстрировать с помощью следующего эффектного опыта. Берется пучок тонких железных стержней, перевязанных нитками. Пучок помещается на столе в вертикальном положении. Верхний конец пучка немного входит внутрь вертикальной катушки, расположенной над пучком. По обмотке катушки пропускается постоянный ток. Ток подбирается таким, чтобы сила, втягивающая пучок в катушку, была несколько меньше веса пучка. Пучок подобен толстому стержню и поэтому намагничивается относительно слабо. Если пережечь нитки, то стержни с силой втягиваются в катушку и в дальнейшем удерживаются в ней. Дело в том, что после пережигания ниток каждый стержень ведет себя почти независимо от других стержней и намагничивается значительно сильнее.

Обычно для объяснения описанных явлений вводят понятия *размагничивающего поля* и *размагничивающего фактора*. Пусть $\mathbf{H}_0$ — внешнее однородное магнитное поле, в которое внесено какое-либо тело из магнетика, имеющее эллипсоидальную форму. В этом случае, как было указано выше, поле $\mathbf{H}$ внутри магнетика будет однородным. Его можно представить в виде $\mathbf{H} = \mathbf{H}_0 + \mathbf{H}_\text{р}$. Поле $\mathbf{H}_\text{р}$ и называется *размагничивающим полем*, так как в ферро- и парамагнетиках оно направлено противоположно внешнему полю $\mathbf{H}_0$ и ослабляет последнее. Размагничивающее поле можно представить в виде $\mathbf{H}_\text{р} = -N\,\mathbf{I}$, где коэффициент N зависит только от формы тела. Этот коэффициент и называется *размагничивающим фактором*. Нам представляется, что введение понятий размагничивающего поля и размагничивающего фактора является ненужным и только запутывает истолкование явлений. Реальная задача состоит в том, чтобы выяснить влияние формы тела на вектор намагничивания $\mathbf{I}$, если это тело поместить в заданное однородное магнитное поле $\mathbf{B}_0$.

4. Вектор $\mathbf{B}$ в магнетике был определен как напряженность результирующего магнитного поля, возбуждаемого токами проводимости и токами намагничивания. В некоторых случаях и вектор $\mathbf{H}$ допускает аналогичную интерпретацию. Рассмотрим, например, бесконечно длинный соленоид, по поверхности которого циркулирует ток проводимости с постоянной линейной плотностью. Соленоид заполнен однородным магнетиком. По оси соленоида в магнетике сделан бесконечно узкий цилиндрический канал, из которого вещество удалено. Поперечное сечение соленоида приведено на рис. 154. Как показано в предыдущем параграфе, вектор $\mathbf{H}$ в магнетике совпадает с вектором $\mathbf{H}$ в канале. Токи намагничивания сводятся к поверхностным токам, текущим по наружной поверхности цилиндра и по внутренней поверхности канала. Направления этих токов противоположны. Возбуждаемые ими магнитные поля в канале имеют противоположные направления и взаимно уничтожают друг друга. Полное поле в канале есть результат действия

одних только токов проводимости. Следовательно, в рассматриваемом случае вектор $\mathbf{H}$ можно интерпретировать как напряженность магнитного поля, возбуждаемого только токами проводимости.

Такая интерпретация всегда справедлива, если намагничивание однородно во всем пространстве. Действительно, взяв дивергенцию от обеих частей уравнения (59.5), получим

$$\operatorname{div}\mathbf{H} = -4\pi \operatorname{div}\mathbf{I}. \qquad (61.5)$$

Если $\mathbf{I} = \text{const}$, то получится система уравнений

$$\operatorname{rot}\mathbf{H} = \frac{4\pi}{c}\mathbf{j}, \quad \operatorname{div}\mathbf{H} = 0,$$

тождественная с соответствующими уравнениями для вакуума. Будут тождественны и решения этих уравнений в вакууме и веществе, если только токи проводимости $\mathbf{j}$ одинаковы. Отсюда следует, что если во всем пространстве $\operatorname{div}\mathbf{I} = 0$, то вектор $\mathbf{H}$ в магнетике может быть истолкован как напряженность магнитного поля, возбуждаемого только токами проводимости. Однако если $\operatorname{div}\mathbf{I}$ не везде равна нулю, то такая интерпретация становится неправильной. В этом случае вектор $\mathbf{H}$ не допускает простого физического толкования и должен определяться формально с помощью соотношения (59.5). Если, например, среда неоднородна, то

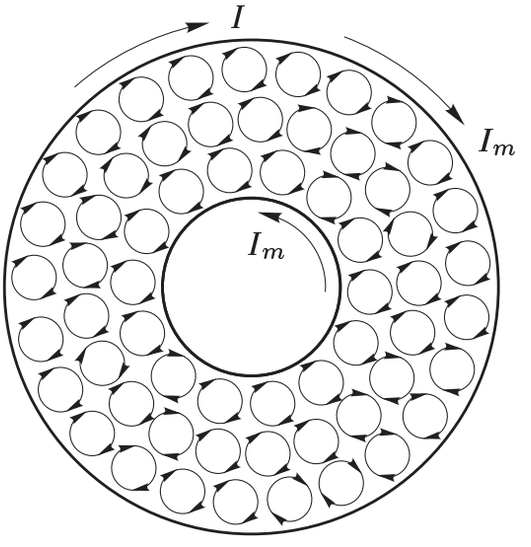


Рис. 154

$$\operatorname{div}\mathbf{B} = \operatorname{div}(\mu\mathbf{H}) = \mu \operatorname{div}\mathbf{H} + \mathbf{H}\operatorname{grad}\mu = 0,$$

или $\operatorname{div}\mathbf{H} = -\mathbf{H}\operatorname{grad}\mu/\mu$, т. е., вообще говоря, $\operatorname{div}\mathbf{H} \neq 0$. Поэтому в рассматриваемом случае источниками поля $\mathbf{H}$ не могут быть одни только токи проводимости.

5. На ток, текущий по проводнику, помещенному в намагничивающейся среде (магнетике), действуют как токи проводимости, так и токи намагничивания. Закон Ампера в магнетике надо поэтому писать в прежнем виде:

$$\mathbf{F} = \frac{I}{c}[d\mathbf{s} \cdot \mathbf{B}]. \qquad (61.6)$$

Под $\mathbf{B}$ следует понимать магнитное поле, создаваемое всеми токами (проводимости и намагничивания), за исключением самого элемента тока $I\,d\mathbf{s}$.

В однородной среде при одних и тех же токах проводимости вектор $\mathbf{B}$ пропорционален магнитной проницаемости μ. Действительно, в этом случае уравнения (58.2) и (59.7) можно записать в виде

$$\operatorname{div}\mathbf{B} = 0, \quad \operatorname{rot}\mathbf{B} = \frac{4\pi\mu}{c}\mathbf{j}.$$

Если токи $\mathbf{j}$ заданы, то $\operatorname{rot}\mathbf{B}$, а с ним и сам вектор $\mathbf{B}$ пропорциональны μ. Отсюда следует, что при заполнении пространства между проводами однородным магнетиком сила взаимодействия токов возрастает в μ раз.

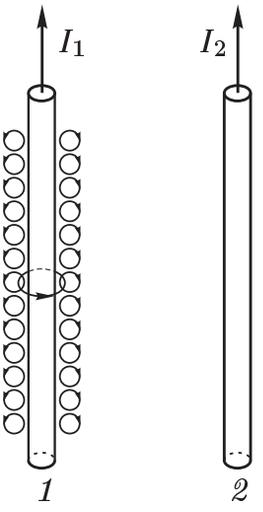


Рис. 155

Поясним это важное заключение на примере взаимодействия двух прямолинейных параллельных проводов, по которым текут токи I_1 и I_2 (рис. 155). Сначала провода находятся в вакууме. Затем все пространство заполняется однородным магнетиком. Вокруг провода *1* возникают молекулярные токи, усиливающие ток I_1. Вследствие этого магнитное поле $\mathbf{B}_1$ первого провода усиливается в μ раз. Во столько же раз возрастает и сила, с которой поле $\mathbf{B}_1$ действует на ток I_2, текущий по второму проводу. Вокруг второго провода также возникают токи намагничивания. Однако они не вносят никакого вклада в силу, действующую на ток I_2. Это объясняется тем, что они текут по поверхности второго провода параллельно его оси. Магнитное поле таких токов равно нулю во всем объеме, занятом вторым проводом.

§ 62. Работа при перемещении витка с током в постоянном магнитном поле

1. Рассмотрим сначала частный случай. Пусть параллельные провода AB и CD (рис. 156) помещены в однородное постоянное магнитное поле, перпендикулярное к плоскости рисунка и направленное к читателю. Слева находится источник тока, не показанный на рисунке. По проводам может свободно перемещаться проводящий мостик KL, замыкающий ток I, текущий по проводам левее мостика. Если l — длина мостика, то на него магнитное поле действует с силой $F = (I/c)lB$. При перемещении мостика на dx эта сила совершает работу

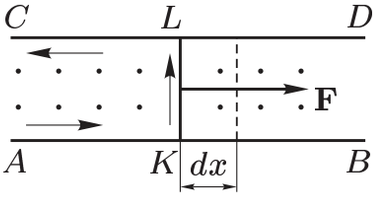


Рис. 156

$$\delta A = \frac{I}{c}\, lB\, dx = \frac{I}{c}\, d(BS),$$

где S — площадь прямоугольника $AKLC$. Величина BS есть магнитный поток через тот же прямоугольник. Обозначив его через Φ, получим для элементарной работы

$$\delta A = \frac{I}{c}\, d\Phi, \tag{62.1}$$

а для конечной работы

$$A_{12} = \frac{I}{c}\,(\Phi_2 - \Phi_1). \tag{62.2}$$

Таким образом, работа, совершаемая магнитным полем над током, равна приращению магнитного потока, умноженному на I/c. При выводе предполагалось, что ток I при перемещении мостика KL поддерживался постоянным.

2. Результат справедлив и в том случае, когда магнитное поле направлено произвольно. Чтобы убедиться в этом, разложим вектор $\mathbf{B}$ на три составляющие: $\mathbf{B} = \mathbf{B}_n + \mathbf{B}_l + \mathbf{B}_x$. Составляющая $\mathbf{B}_l$ вдоль мостика параллельна току в нем, а потому не оказывает на мостик силового действия. Составляющая $\mathbf{B}_x$ вдоль перемещения дает силу, перпендикулярную к перемещению, и работы не производит. Работа производится лишь составляющей $\mathbf{B}_n$, перпендикулярной к плоскости рисунка, в которой перемещается мостик KL. Эта работа представляется выражениями (62.1) и (62.2).

3. Докажем, наконец, что формулы (62.1) и (62.2) справедливы для любого витка с током при произвольном перемещении его в постоянном неоднородном магнитном поле. Виток может не только перемещаться как целое, но и произвольно деформироваться. Для доказательства достаточно мысленно разбить виток на бесконечно малые элементы тока и рассмотреть бесконечно малые перемещения их. При бесконечно малом перемещении элемента тока магнитное поле, в котором он перемещается, может считаться однородным. К такому перемещению применимо выражение (62.1) для элементарной работы. Сложением таких элементарных работ для всех элементов тока, на которые разбит виток, снова получается выражение (62.1), в котором $d\Phi$ означает приращение магнитного потока через весь виток. После этого переход от формулы (62.1) к формуле (62.2) совершается простым интегрированием. Подчеркнем еще раз, что при перемещении витка сила тока в нем *должна поддерживаться постоянной.* Это достигается путем надлежащего увеличения электродвижущей силы источника.

§ 63. Способ Гаусса измерения магнитных полей

Постоянные магниты являются магнетиками, вектор намагничивания $\mathbf{I}$ которых практически не изменяется при внесении магнита во внешнее магнитное поле (если последнее не слишком сильное). На этом основан *метод Гаусса* измерения напряженности магнитного поля. Пусть магнит имеет форму прямого стержня, намагниченного параллельно его оси. Обозначим через $\mathfrak{M}$ его магнитный момент. В однородном магнитном поле $\mathbf{B}$ на магнит действует вращающий момент $[\mathfrak{M}\mathbf{B}]$. Если магнит может свободно вращаться вокруг своего центра масс, то под действием этого вращающего момента вектор $\mathfrak{M}$ установится вдоль $\mathbf{B}$. Выведем немного магнит из положения равновесия. Возникнут малые колебания с периодом

$$T = 2\pi\sqrt{\frac{\Theta}{\mathfrak{M}\mathbf{B}}}, \tag{63.1}$$

где Θ — момент инерции магнита.

Закрепим теперь магнит перпендикулярно к магнитному полю $\mathbf{B}$ и поместим вдали от магнита на продолжении его оси маленькую магнитную стрелку. Считая магнит точечным диполем, для магнитного поля $\mathbf{B}_1$ магнита в месте нахождения стрелки можно написать $B_1 = 2\mathfrak{M}/r^3$, где r — расстояние между центрами стрелки и магнита. Это поле направлено вдоль оси магнита, т. е. перпендикулярно к измеряемому полю $\mathbf{B}$. Под действием полей $\mathbf{B}$ и $\mathbf{B}_1$ стрелка установится под углом α к полю $\mathbf{B}$, который определяется соотношением

$$\operatorname{tg}\alpha = \frac{B_1}{B} = \frac{2\mathfrak{M}}{Br^3}. \tag{63.2}$$

Измерив время T и угол α, можно по формулам (63.1) и (63.2) вычислить как напряженность поля $\mathbf{B}$, так и магнитный момент магнита $\mathfrak{M}$.

Можно поступить и иначе. Магнит по-прежнему закрепляют перпендикулярно к полю $\mathbf{B}$, но магнитную стрелку помещают на линии, перпендикулярной к оси магнита и проходящей через его центр. Тогда поле магнита $\mathbf{B}_2$ в месте нахождения стрелки определится по формуле $B_2 = \mathfrak{M}/r^3$ и будет направлено противоположно вектору $\mathfrak{M}$, т. е. по-прежнему перпендикулярно к вектору $\mathbf{B}$. Поэтому угол α между $\mathbf{B}$ и осью стрелки в положении равновесия определится соотношением

$$\operatorname{tg}\alpha = \frac{B_2}{B} = \frac{\mathfrak{M}}{Br^3}, \tag{63.3}$$

которым можно воспользоваться вместо соотношения (63.2).

Со времен Гаусса описанный способ широко применяется для измерения магнитного поля Земли. При этом магнит и магнитная стрелка располагаются в горизонтальной плоскости и могут вращаться вокруг вертикальной оси. В таком случае методом Гаусса определяется не полное магнитное поле Земли $\mathbf{B}$, а только его *горизонтальная составляющая.*

§ 64. Электромагнитная индукция

1. Открытие *электромагнитной индукции* Фарадеем в 1831 г. было одним из наиболее фундаментальных открытий в электродинамике. Для демонстрации этого явления возьмем неподвижный магнит и проволочную катушку, концы которой соединим с гальванометром. Если катушку приближать к одному из полюсов магнита, то во время движения стрелка гальванометра отклоняется — в катушке возбуждается электрический ток. При движении катушки в обратном направлении направление тока меняется на противоположное. То же самое происходит, если повернуть магнит на 180°, не меняя направления движения катушки. Магнит можно заменить другой катушкой с током или электромагнитом. Вообще, при движении катушки в постоянном магнитном поле в ней (за исключением некоторых специальных случаев, которые выяснятся ниже) возбуждается электрический ток, прекращающийся, когда катушка останавливается. Этот ток называется *индукционным током*, а самое явление — *электромагнитной индукцией.* В частности,

когда катушка равномерно вращается в постоянном магнитном поле, индукционный ток периодически меняет свою силу и направление.

2. Возбуждение электрического тока при движении проводника в магнитном поле объясняется действием силы Лоренца, возникающей при движении проводника. Рассмотрим сначала простейший случай, когда два параллельных провода AB и CD помещены в постоянное однородное магнитное поле, перпендикулярное к плоскости рисунка и направленное к читателю (рис. 157). Слева провода AB и CD замкнуты, справа — разомкнуты. Вдоль проводов может свободно скользить проводящий мостик BC. Когда мостик движется вправо со скоростью $\mathbf{v}$, вместе с ним движутся электроны и положительные ионы. На каждый движущийся заряд e в магнитном поле действует сила Лоренца $\mathbf{F} = (e/c)[\mathbf{vB}]$. На положительный ион она действует вниз, на отрицательный электрон — вверх. В результате электроны начнут перемещаться по мостику вверх, т. е. по нему потечет электрический ток, направленный вниз. Это и есть индукционный ток. Перераспределившиеся заряды создадут электрическое поле, которое возбудит токи и в остальных участках контура $ABCD$. На рис. 157 эти токи изображены сплошными стрелками.

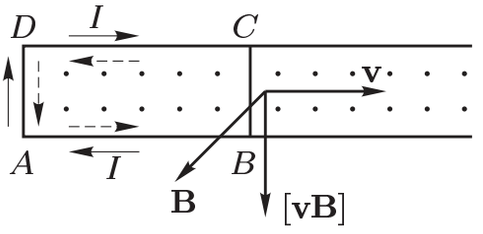


Рис. 157

Сила Лоренца $\mathbf{F}$ в описанном опыте играет роль сторонней силы, возбуждающей электрический ток. Соответствующая напряженность стороннего поля равна $\mathbf{E}^{\text{стор}} = \dfrac{\mathbf{F}}{e} = \dfrac{1}{c}[\mathbf{vB}]$. Электродвижущая сила, создаваемая этим полем, называется *электродвижущей силой индукции* и обозначается $\mathscr{E}^{\text{инд}}$. В рассматриваемом случае $\mathscr{E}^{\text{инд}} = -(v/c)Bl$, где l — длина мостика. Знак минус поставлен потому, что стороннее поле $(1/c)[\mathbf{vB}]$ направлено против положительного обхода контура, определяемого вектором $\mathbf{B}$ по правилу правого винта. На рис. 157 это направление показано штриховыми стрелками. Величина lv есть приращение площади контура $ABCD$ в единицу времени, или скорость приращения этой площади. Поэтому величина vBl равна $d\Phi/dt$, т. е. скорости приращения магнитного потока, пронизывающего площадь контура $ABCD$. Таким образом,

$$\mathscr{E}^{\text{инд}} = -\frac{1}{c}\frac{d\Phi}{dt}. \tag{64.1}$$

Результат (64.1) справедлив и в том случае, когда однородное магнитное поле $\mathbf{B}$ направлено под любым углом к плоскости контура $ABCD$. Действительно, представим вектор $\mathbf{B}$ в виде $\mathbf{B}_t + \mathbf{B}_n$, где $\mathbf{B}_t$ — тангенциальная, а $\mathbf{B}_n$ — нормальная к плоскости контура слагающие этого вектора. Вектор $\mathbf{B}_t$ вносит в стороннее поле слагаемое $(1/c)[\mathbf{vB}_t]$, перпендикулярное к мостику. Оно вызывает лишь перераспределение электрических зарядов поперек мостика, но тока не

дает. Ток вызывается только нормальной слагающей $\mathbf{B}_n$, а потому $\mathscr{E}^{\text{инд}}$ определяется прежней формулой (64.1).

Теперь не составляет труда распространить формулу (64.1) на случай любого замкнутого провода, движущегося произвольным образом в постоянном неоднородном магнитном поле. Для этого надо мысленно разбить провод на бесконечно малые участки и рассмотреть движение каждого из них. При бесконечно малом перемещении каждого из таких участков магнитное поле, в котором он движется, можно считать однородным. Поэтому электродвижущая сила, действующая между концами участка, может быть представлена выражением (64.1). Путем суммирования таких выражений получится формула того же вида, в которой, однако, под $\mathscr{E}^{\text{инд}}$ следует понимать полную электродвижущую силу, действующую в замкнутом проводе, а под $d\Phi/dt$ — скорость изменения магнитного потока через любую поверхность, натянутую на контур провода.

Формула (64.1) выражает *основной закон электромагнитной индукции*. Она показывает, что при движении замкнутого провода в магнитном поле в нем возбуждается электродвижущая сила, пропорциональная скорости приращения магнитного потока, пронизывающего контур провода.

3. К формуле (64.1) можно прийти также с помощью закона сохранения энергии, как это впервые сделал Гельмгольц.

Рассмотрим, следуя Гельмгольцу, замкнутый виток провода, в который включен гальванический элемент с электродвижущей силой $\mathscr{E}$. Виток движется в постоянном магнитном поле (вообще говоря, неоднородном). За время dt амперовы силы совершают над витком работу $(I/c)\,d\Phi$. Кроме того, в витке выделяется джоулево тепло $RI^2\,dt$. Сумма этих величин должна равняться работе гальванического элемента $\mathscr{E}I\,dt$, т. е.

$$\frac{1}{c}\,I\,d\Phi + I^2R\,dt = \mathscr{E}I\,dt. \tag{64.2}$$

Отсюда

$$I = \frac{\mathscr{E} - (1/c)\,d\Phi/dt}{R}. \tag{64.3}$$

Таким образом, в движущемся витке ток определяется не только электродвижущей силой гальванического элемента. К ней добавляется слагаемое $-(1/c)\,d\Phi/dt$. Это слагаемое и есть электродвижущая сила индукции.

Заметим, что уравнению сохранения энергии (64.2) можно также удовлетворить, положив $I = 0$. Какое из двух решений выбрать: решение $I = 0$ или решение (64.3) — на это закон сохранения энергии не дает никаких указаний. Следовательно, без привлечения дополнительных соображений он не позволяет предсказать явление электромагнитной индукции. Нужно как-то исключить решение $I = 0$. С этой целью, как это сделал Гельмгольц, в виток и включен гальванический элемент с электродвижущей силой $\mathscr{E}$. То обстоятельство, что добавочная электродвижущая сила $-(1/c)\,d\Phi/dt$, появляющаяся при движении

проводника, не зависит от $\mathscr{E}$, делает правдоподобным заключение, что и при отсутствии гальванического элемента в движущемся витке должна возникнуть такая же электродвижущая сила. Можно обойтись и без введения гальванического элемента, если предположить, что при движении проводника должен возникать индукционный ток. Тогда закон сохранения энергии позволяет определить силу этого тока, а следовательно, и электродвижущую силу индукции. В этом истинный смысл и содержание рассуждения Гельмгольца.

4. Индукционные токи могут возникать и в неподвижных проводниках. Действительно, возьмем замкнутый провод и постоянный магнит. При движении провода возникает индукционный ток. Что произойдет, если, оставляя провод неподвижным, двигать магнит? Покой и движение — понятия относительные. Явление индукционного тока должно зависеть только от *относительного движения провода и магнита.* Отсюда следует, что при движении магнита будет возбуждаться такой же индукционный ток, что и при соответствующем движении провода. Опыт подтверждает это заключение. Возьмем прежнюю катушку, соединенную с гальванометром, и будем приближать к ней магнит. Стрелка гальванометра отклонится — в катушке возбудился электрический ток. При удалении магнита стрелка отклоняется в противоположную сторону, т. е. индукционный ток меняет направление. То же самое происходит, если магнит повернуть к катушке другим полюсом, не меняя направления его движения. Если магнит вращать, то индукционный ток в катушке будет периодически менять свое направление. Когда магнит останавливается, индукционный ток в катушке прекращается. Вместо магнита можно взять электромагнит или другую катушку, по которой течет ток, возбуждающий магнитное поле. При их движении в неподвижной катушке возбуждается электрический ток.

В описанных опытах с движением магнита менялся магнитный поток, пронизывающий неподвижную катушку. Но такое же изменение магнитного потока можно получить и без движения магнита. Достаточно поместить катушку в переменное магнитное поле. Последнее можно подобрать так, чтобы в месте нахождения катушки оно в точности совпадало с магнитным полем движущегося магнита. От такой замены объективные физические условия, в которых находится катушка, не изменятся. Поэтому естественно ожидать, что не изменится и индукционный ток, возбуждаемый в катушке. Опыт подтверждает и это заключение. Возьмем две неподвижные катушки, одна из которых помещена внутри другой. Если через одну из катушек пропускать переменный ток, то в другой катушке появляется индукционный электрический ток. Таким образом, для возбуждения индукционного тока существенно *изменение магнитного потока* через контур проводника, а не способ, каким это изменение достигается.

Вот другая демонстрация, подтверждающая это заключение. На подковообразный магнит надевается проволочная катушка, соединенная с гальванометром (рис. 158). Если полюсы магнита замкнуть же-

лезным якорем, то изменится магнитный поток через катушку. В ней возникает индукционный ток, и стрелка гальванометра отклоняется. При удалении якоря магнитный поток меняется в обратном направлении. В результате меняется направление индукционного тока, а с ним отклонение стрелки гальванометра.

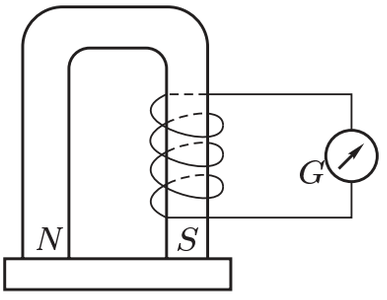


Рис. 158

Таким образом, *всякий раз, когда меняется магнитный поток, пронизывающий контур неподвижного или движущегося замкнутого провода, в проводе возникает индукционный ток, причем во всех случаях электродвижущая сила индукции определяется формулой* (64.1).

§ 65. Правило Ленца

1. Формула (64.1) определяет не только величину, но и *направление* индукционного тока. Действительно, возьмем в магнитном поле замкнутый проволочный виток, положительное направление обхода которого составляет с направлением поля правовинтовую систему (на рис. 159 магнитное поле направлено к читателю). Допустим, что магнитный поток Φ возрастает. Тогда, согласно формуле (64.1), величина $\mathscr{E}^{\text{инд}}$ будет отрицательна, а потому индукционный ток в витке потечет в отрицательном направлении. Такой ток, ослабляя внешнее магнитное поле, будет препятствовать возрастанию магнитного потока. Пусть теперь магнитный поток Φ убывает. Тогда величина $\mathscr{E}^{\text{инд}}$ станет положительной, а индукционный ток в витке потечет в положительном направлении и будет препятствовать убыванию магнитного поля и магнитного потока. Таким образом, *индукционный ток всегда имеет такое направление, что он ослабляет действие причины, возбуждающей этот ток.* Это правило впервые было сформулировано Ленцем (1804–1865) и носит его имя. Ле Шателье (1850–1936), а затем Браун (1850–1918) обобщили правило Ленца и распространили его на все физические явления (см. т. II, § 51).

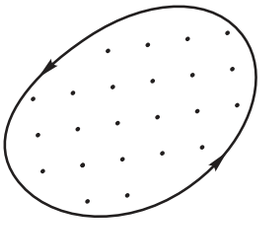


Рис. 159

2. Подвесим над проволочной катушкой прямой магнит NS, уравновешенный грузом на неподвижном блоке (рис. 160). В цепь катушки включим гальванометр. Быстро вдвинем магнит в катушку — появится индукционный ток, и гальванометр даст отброс. Заметив направление отклонения стрелки гальванометра, переключим катушку с гальванометром на батарею. (Во избежание порчи гальванометра его надо шунтировать, так как батарея даст значительно более сильный ток, чем ток индукции.) Пусть переключение выполнено так, что ток от батареи отклоняет стрелку гальванометра в ту же сторону, что и ток индукции, т. е. оба тока направлены одинаково. Опыт показывает, что

в этом случае ток от батареи выталкивает магнит из катушки, как и должно быть согласно правилу Ленца. Если же ток от батареи направлен противоположно току индукции, то, в согласии с тем же правилом, магнит втягивается в катушку. Аналогичные явления наблюдаются при выдвигании магнита.

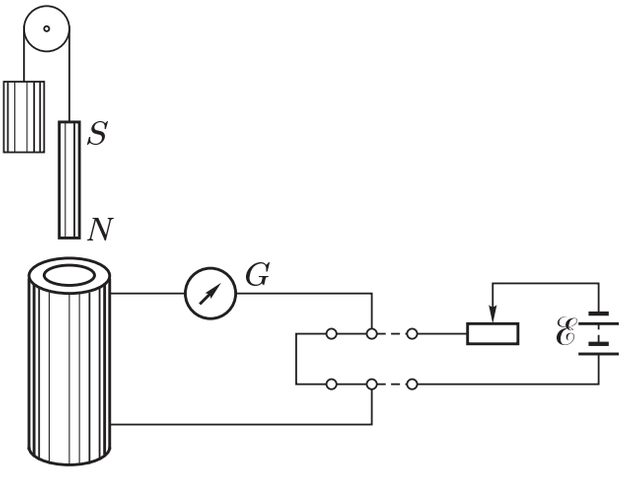


Рис. 160

Возьмем горизонтальную катушку с сердечником, состоящим из многих железных прутьев. Через центр катушки проходит гладко отполированный и закрепленный горизонтально железный стержень, длина которого (~ 1 м) примерно в три раза больше длины катушки. На стержень наденем легкое алюминиевое кольцо, диаметр которого немного больше диаметра стержня, так что оно может свободно перемещаться по стержню. Если замкнуть ток в катушке, то в кольце возникнет индукционный ток противоположного направления, и кольцо оттолкнется от катушки. При надлежаще подобранной силе тока кольцо отлетит почти до самого конца стержня. При размыкании цепи индукционный ток будет противоположного направления и вернет кольцо почти в исходное положение. Вместо кольца можно взять короткую катушку, бифилярно подвешенную на длинных нитях.

Опыт выглядит значительно эффектнее, если воспользоваться переменным током от городской сети, как это сделал Элиу Томсон (1853–1937). Катушка с сердечником ставится теперь вертикально. На железный сердечник надевается широкое толстое алюминиевое кольцо. (Алюминий берется потому, что он легкий и имеет малое удельное сопротивление.) В катушку посылается переменный ток от городской сети, возбуждающий в кольце индукционный ток противоположного направления. Эти токи отталкиваются. Сила отталкивания не постоянна, а периодически колеблется с частотой, вдвое превосходящей частоту колебаний городского тока. Действительно, пусть в некоторый момент оба тока достигли максимума и текут в противоположных направлениях. Тогда будет максимальна и сила отталкивания. Через половину

периода городской и индукционный токи изменят направления на противоположные, и сила отталкивания достигнет прежнего максимального значения. Если замкнуть ток в катушке, то сила отталкивания подбросит алюминиевое кольцо до потолка большой аудитории. Если кольцо удерживать щипцами в неподвижном положении, то оно сильно нагревается: сопротивление кольца ничтожно, а потому индукционные токи в нем очень велики. Опыт с нагреванием можно произвести, взяв вместо алюминиевого тонкое медное кольцо. Из-за большой плотности меди и большого сопротивления кольца оно не подбрасывается вверх, а только сильно разогревается.

3. В массивных проводниках, движущихся в магнитных полях или помещенных в переменные магнитные поля, возбуждаются вихревые индукционные токи, называемые *токами Фуко*. По физической природе они ничем не отличаются от индукционных токов, возбуждаемых в линейных проводах. С помощью токов Фуко очень эффектно демонстрируется правило Ленца. Возьмем маятник, изготовленный из толстой листовой меди и имеющий форму усеченного сектора (рис. 161). Маятник подвешен на стержне и может свободно колебаться вокруг го-

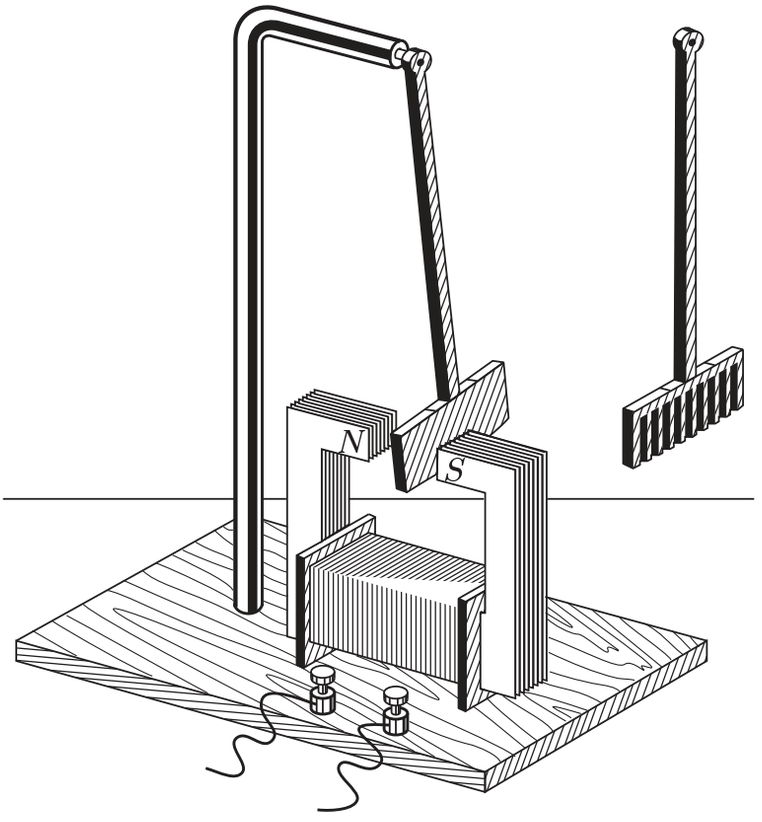


Рис. 161

ризонтальной оси между полюсами сильного электромагнита, создающего магнитное поле порядка 5000 Гс. Пока не включено магнитное поле, маятник колеблется почти без затухания. Замыкая ток в обмотке электромагнита, мы создаем магнитное поле. Тогда при колебаниях

маятника возбуждаются индукционные токи Фуко, согласно правилу Ленца тормозящие движение маятника; колебания последнего моментально прекращаются. Если сплошной сектор маятника заменить гребенкой с длинными зубцами (рис. 161, справа), то возбуждение токов Фуко будет сильно затруднено. Маятник будет колебаться в магнитном поле почти без затухания. Этот опыт объясняет, почему сердечники электромагнитов и рамы трансформаторов делают не из сплошного куска железа, а из многих листов, наложенных друг на друга. В результате токи Фуко возбуждаются слабо и сильно уменьшается вредное влияние джоулева тепла, выделяемого ими.

Возьмем медный или алюминиевый диск диаметром 4–5 см и толщиной 4–6 мм. Заставим его падать в узком зазоре между полюсами электромагнита. Пока магнитное поле не включено, диск движется вниз быстро, как при свободном падении. Включим магнитное поле порядка 5000 Гс. Тогда падение становится очень медленным и напоминает движение в сильно вязкой среде.

Если над магнитной стрелкой поместить горизонтальный медный диск и привести его во вращение вокруг вертикальной оси, то магнитная стрелка придет во вращение в том же направлении. Дело в том, что при относительном движении стрелки и медного диска в диске возбуждаются токи Фуко, которые, согласно правилу Ленца, замедляют это относительное движение. В результате стрелка начинает вращаться вместе с диском.

§ 66. Максвелловская трактовка явления электромагнитной индукции

1. Когда проводник движется в постоянном магнитном поле, индукционный ток вызывается магнитной составляющей силы Лоренца $(e/c)[\mathbf{vB}]$. Какая же сила возбуждает индукционный ток в неподвижном проводнике, находящемся в переменном магнитном поле? Ответ был дан Максвеллом. Согласно Максвеллу, *всякое переменное магнитное поле возбуждает в окружающем пространстве электрическое поле.* Последнее и является причиной возникновения индукционного тока в проводнике. Максвеллу принадлежит следующая углубленная формулировка закона электромагнитной индукции.

Всякое изменение магнитного поля во времени возбуждает в окружающем пространстве электрическое поле. Циркуляция вектора напряженности $\mathbf{E}$ *этого поля по любому неподвижному замкнутому контуру* s *определяется выражением*

$$\oint_s (\mathbf{E}\, d\mathbf{s}) = -\frac{1}{c}\frac{\partial \Phi}{\partial t}, \tag{66.1}$$

где Φ *— магнитный поток, пронизывающий контур* s. Мы использовали для обозначения скорости изменения магнитного потока знак частной, а не полной производной. Этим мы хотим подчеркнуть, что контур s должен быть неподвижным.

Между максвелловым и фарадеевым пониманием явления электромагнитной индукции имеется существенное различие. Согласно Фарадею, электромагнитная индукция состоит *в возбуждении электрического тока.* Для ее наблюдения необходимо *наличие замкнутого проводника.* Максвелл, напротив, видит сущность электромагнитной индукции прежде всего *в возбуждении электрического поля*, а не тока. Электромагнитная индукция может наблюдаться и тогда, когда в пространстве вообще *нет никаких проводников.* Появление индукционного тока в замкнутом проводнике при внесении последнего в переменное магнитное поле есть лишь *одно из проявлений* электрического поля $\mathbf{E}$, возникшего в результате изменения поля магнитного. Но поле $\mathbf{E}$ может производить и другие действия, например поляризовать диэлектрик, вызвать пробой конденсатора, ускорять и тормозить заряженные частицы и т. п. Оно может вызвать электрический ток и в незамкнутом проводнике, как показывает, например, следующий опыт.

Возьмем две катушки, расположенные близко одна от другой приблизительно так, чтобы ось одной катушки была продолжением оси другой. Концы первой катушки присоединим к звуковому генератору, т. е. прибору, который может возбуждать переменные токи с частотами, лежащими в звуковом диапазоне. Концы второй катушки соединим с горизонтальными пластинами электронного осциллографа. Когда в первой катушке течет переменный ток, луч осциллографа отклоняется, хотя цепь второй катушки разомкнута. Луч бегает вверх и вниз, и на экране видна светлая вертикальная полоска, переходящая в синусоиду после включения горизонтальной развертки. Это доказывает, что между горизонтальными пластинами осциллографа появилось переменное электрическое поле. Пластины оказались заряженными, причем их заряды периодически меняются во времени, а во второй катушке текут переменные индукционные токи, несмотря на то, что цепь разомкнута.

Максвеллова формулировка закона индукции более общая, чем формулировка Фарадея. Она принадлежит к числу наиболее важных обобщений электродинамики. Математически закон индукции в понимании Максвелла выражается формулой (66.1), где s — произвольный математический замкнутый контур, который может быть проведен и в диэлектрике, а не обязательно в проводнике, как было у Фарадея. Магнитный поток Φ определяется интегралом

$$\Phi = \oint\limits_s \mathbf{B}\, d\mathbf{S}, \tag{66.2}$$

взятым по произвольной поверхности S, натянутой на контур s. Поэтому формулу (66.1) можно представить в виде

$$\oint\limits_s (\mathbf{E}\, d\mathbf{s}) = -\frac{1}{c}\,\frac{\partial}{\partial t}\int\limits_S \mathbf{B}\, d\mathbf{S} = -\frac{1}{c}\int\limits_S \frac{\partial \mathbf{B}}{\partial t}\, d\mathbf{S}. \tag{66.3}$$

Это уравнение имеет ту же математическую структуру, что и уравнение (55.5). Роль вектора $4\pi\mathbf{j}$ играет вектор $-\partial\mathbf{B}/\partial t$. Поэтому оно может быть преобразовано в дифференциальную форму совершенно

так же, как это было сделано с уравнением (55.5). В результате такого преобразования получится

$$\operatorname{rot}\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}. \tag{66.4}$$

Это — *дифференциальная форма закона электромагнитной индукции.* Уравнение (66.3) или эквивалентное ему уравнение (66.4) — одно из основных соотношений теории электромагнитного поля. Оно входит в систему уравнений Максвелла.

2. В электростатике источниками электрического поля являются неподвижные электрические заряды. Для такого поля интеграл $\oint \mathbf{E}\, d\mathbf{s}$ обращается в нуль по любому замкнутому контуру. По этой причине одно только электростатическое поле не может обеспечить непрерывное течение электричества вдоль замкнутых проводов. Напротив, электрическое поле, возбуждаемое магнитным полем, меняющимся во времени, — не потенциальное, а *вихревое.* Ротор такого поля и его циркуляция, вообще говоря, отличны от нуля. Благодаря этому вихревое поле без каких бы то ни было добавочных сил может вызвать непрерывное течение электричества по замкнутым проводам. Это течение и наблюдается в виде индукционных токов.

3. В общем случае, когда проводник движется в переменном магнитном поле, индукционный ток возбуждается как электрической силой $e\mathbf{E}$, так и магнитной силой $(e/c)[\mathbf{vB}]$. Объединяя обе силы, можно сказать, что *во всех случаях индукционный ток вызывается полной силой Лоренца*

$$\mathbf{F} = e\left(\mathbf{E} + \frac{1}{c}[\mathbf{vB}]\right). \tag{66.5}$$

Какая часть индукционного тока вызывается электрической, а какая магнитной составляющей силы Лоренца — это зависит *от выбора системы отсчета.* Дело в том, что деление электромагнитного поля на электрическое и магнитное определяется системой отсчета, в которой рассматриваются явления. Возьмем две системы отсчета: «неподвижную» систему S и «движущуюся» S'. Пусть V — скорость системы S' относительно системы S. При переходе от одной системы отсчета к другой векторы $\mathbf{E}$ и $\mathbf{B}$ определенным образом преобразуются. Законы этого преобразования устанавливаются в теории относительности и будут рассмотрены в т. IV нашего курса. Но и здесь, если только скорость V мала по сравнению со скоростью света c, можно составить об этих законах предварительное представление. Приводимые ниже формулы совпадают с точными релятивистскими формулами в первом порядке относительно малого отношения V/c. Они содержат ошибки порядка $(V/c)^2$.

Пусть $\mathbf{v}$ — скорость частицы относительно системы S, а $\mathbf{v}'$ — относительно системы S'. Эти скорости связаны соотношением $\mathbf{v} = \mathbf{v}' + \mathbf{V}$. В системе S сила, действующая на частицу с зарядом e, определяется

выражением (66.5), а в системе S' — выражением

$$\mathbf{F} = e\left(\mathbf{E}' + \frac{1}{c}[\mathbf{v}'\mathbf{B}']\right),$$

где $\mathbf{E}'$ и $\mathbf{B}'$ — электрическое и магнитное поля в системе S'. В нерелятивистской механике сила — инвариант: $\mathbf{F}' = \mathbf{F}$. Поэтому, заменяя в формуле (66.5) скорость $\mathbf{v}$ на $\mathbf{v}' + \mathbf{V}$, получим

$$\mathbf{F}' = e\left(\mathbf{E} + \frac{1}{c}[\mathbf{V}\mathbf{B}]\right) + \frac{e}{c}[\mathbf{v}'\mathbf{B}].$$

Чтобы найти напряженность поля $\mathbf{E}'$, надо взять частицу, покоящуюся в системе S', т. е. в предыдущем выражении положить $\mathbf{v}' = 0$. Это дает

$$\mathbf{F}' = e\left(\mathbf{E} + \frac{1}{c}[\mathbf{V}\mathbf{B}]\right).$$

С другой стороны, для той же силы можно написать $\mathbf{F}' = e\mathbf{E}$. Сравнивая оба выражения, находим

$$\mathbf{E}' = \mathbf{E} + \frac{1}{c}[\mathbf{V}\mathbf{B}]. \tag{66.6}$$

Это и есть нерелятивистский закон преобразования электрического поля. Аналогичный закон для магнитного поля не может быть выведен без использования теории относительности, а потому мы приведем его здесь без доказательства:

$$\mathbf{B}' = \mathbf{B} - \frac{1}{c}[\mathbf{V}\mathbf{E}]. \tag{66.7}$$

Возникает вопрос: можно ли перейти к такой системе отсчета, в которой электромагнитное поле сделалось бы чисто электрическим или чисто магнитным? Для того чтобы поле стало чисто электрическим, должно быть $\mathbf{B}' = 0$, или, ввиду формулы (66.7), $\mathbf{B} - (1/c)[\mathbf{V}\mathbf{E}] = 0$. Записанное в виде $\mathbf{B} - (1/c)[\mathbf{V}_\perp\mathbf{E}] = 0$, это уравнение определяет поперечную (относительно вектора $\mathbf{E}$) составляющую скорости $\mathbf{V}$. Продольная составляющая скорости $\mathbf{V}$ остается неопределенной. Векторным умножением последнего уравнения на $\mathbf{E}$ получаем

$$\mathbf{V}_\perp = \frac{c}{E^2}[\mathbf{E}\mathbf{B}].$$

Отсюда видно, что система отсчета, обладающая требуемыми свойствами, существует не всегда. Может случиться, что последнее уравнение даст $V_\perp > c$, а это невозможно. Правда, для такого заключения в нерелятивистской теории нет оснований, так как формулы (66.6) и (66.7) верны только при $V \ll c$. Однако само заключение верно, как в этом можно убедиться с помощью точных релятивистских формул преобразования полей. Точно так же не всегда возможно перейти к такой системе отсчета, в которой электромагнитное поле становится чисто магнитным. Отсюда следует, что в общем случае индукционный ток вызывается как электрическим, так и магнитным полями. Не всегда

возможно, перейдя к другой системе отсчета, рассматривать его как проявление только одного из этих полей.

§ 67. Флюксметр и пояс Роговского

На явлении электромагнитной индукции основан простой и удобный метод измерения напряженности магнитных полей. Соединим с баллистическим гальванометром концы небольшого витка проволоки. Расположим плоскость витка перпендикулярно к магнитному полю. Пусть его пронизывает магнитный поток Φ. Если быстро убрать виток из поля или повернуть его вокруг диаметра на 90°, то магнитный поток обратится в нуль. Того же можно достигнуть, выключив ток, возбуждающий магнитное поле. При изменении магнитного потока через виток течет кратковременный ток

$$I = -\frac{1}{Rc}\frac{d\Phi}{dt},$$

где R — сумма сопротивлений витка, баллистического гальванометра и подводящих проводов. За все время изменения магнитного потока от Φ до 0 через гальванометр пройдет количество электричества

$$q = -\frac{1}{Rc}\int_0^t \frac{d\Phi}{dt}\,dt = \frac{\Phi}{Rc}. \tag{67.1}$$

Отклонение баллистического гальванометра пропорционально заряду q, а потому он позволяет измерить этот заряд. После этого по формуле (67.1) можно вычислить магнитный поток Φ, а затем и индукцию B. Для увеличения чувствительности вместо одного витка лучше взять маленькую плоскую катушечку, состоящую из многих витков. Если n — общее число витков, а S — площадь одного витка, то $\Phi = nSB$. Такая катушечка, служащая для измерения магнитного потока Φ, а с ним и индукции B, называется *флюксметром*. Прибор можно проградуировать, чтобы он прямо указывал значение потока Φ или индукцию B.

Явление электромагнитной индукции можно также использовать для измерения *магнитного напряжения*, т. е. линейного интеграла $\int_{12} \mathbf{B}\,d\mathbf{s}$. Вообще говоря, такой интеграл зависит не только от положения начальной и конечной точек *1* и *2*, но и от кривой *12*, соединяющей эти точки. Однако при определенных условиях можно выделить такую совокупность кривых, соединяющих точки *1* и *2*, что вдоль всех этих кривых интеграл $\int \mathbf{B}\,d\mathbf{s}$ будет иметь одно и то же значение. Так будет всегда, когда от одной кривой можно перейти к другой непрерывной деформацией, нигде не пересекая электрических токов. Возьмем проволочную спираль, навитую на гибкий ремень, концы которой соединены с баллистическим гальванометром (рис. 162 *а*). Расположим ось спирали по линии, вдоль которой требуется измерить магнитное напряжение между точками *1* и *2*. Магнитный поток Φ через спираль может быть

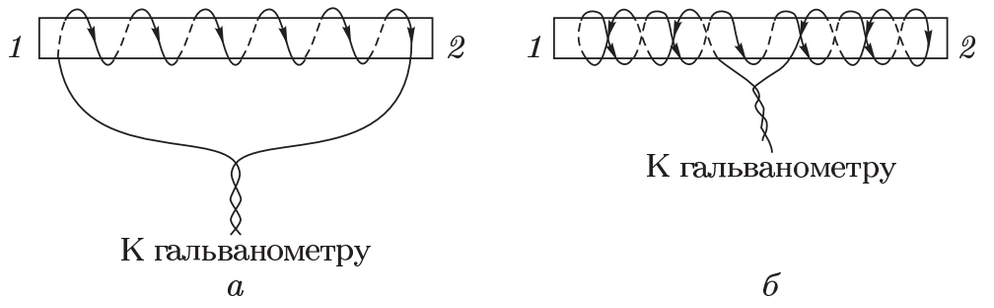


Рис. 162

измерен по отбросу баллистического гальванометра совершенно так же, как это делалось в случае флюксметра. Поток может быть представлен в виде

$$\Phi = \int Sn\mathbf{B}\,d\mathbf{s}.$$

Здесь S — площадь одного витка, а n — число витков, приходящееся на единицу длины спирали. Если величины n и S не меняются вдоль спирали, то

$$\Phi = Sn\int(\mathbf{B}\,d\mathbf{s}).$$

Отсюда

$$\int(\mathbf{B}\,d\mathbf{s}) = \frac{\Phi}{Sn} = \frac{cR}{nS}\,q. \tag{67.2}$$

В приведенных рассуждениях допущены некоторые неточности. Магнитное поле спирали слагается из магнитных полей круговых токов, которые мы учли. Но есть составляющая тока, параллельная оси спирали. Ее магнитное поле не учтено. Кроме того, есть магнитное поле подводящих проводов. Чтобы влияние этих полей исключить, спираль навивают на ремень в два слоя, идущие навстречу друг другу (рис. 162 *б*). Концы проводов выводят в одном месте, например в середине ремня. Подводящие провода скручивают. В таком виде прибор называется *поясом Роговского*. Для измерения магнитного напряжения пояс Роговского располагают между нужными точками вдоль заданной кривой. Затем выключают ток, создающий магнитное поле. Баллистический гальванометр даст отброс, пропорциональный искомому магнитному напряжению.

ЗАДАЧИ

1. Механическая аналогия явления электромагнитной индукции. Труба, свернутая в кольцо, заполнена жидкостью и находится в поле тяжести Земли. При повороте кольца вокруг его диаметра жидкость приходит в движение вдоль оси трубы. Это движение аналогично индукционному току, возникающему при движении проводника в магнитном поле. Роль силы Лоренца играет *сила инерции Кориолиса*, вызванная вращением Земли вокруг своей оси. Опыт может служить для доказательства вращения Земли и измерения угловой скорости этого вращения. На основе аналогии описанного явления с электромагнитной индукцией дать количественную

теорию его, предполагая, что жидкость несжимаема и не обладает вязкостью, а площадь поперечного сечения трубы всюду одинакова.

Р е ш е н и е. Будем рассматривать движение жидкости относительно системы отсчета, связанной с вращающейся Землей. Это относительное движение происходит под действием гравитационных сил, градиентов давления жидкости, центробежной силы инерции и силы инерции Кориолиса. Все эти силы, за исключением силы Кориолиса, потенциальны. Поэтому относительное движение жидкости можно представить в виде

$$\mathbf{a} = 2[\mathbf{v}\boldsymbol{\Omega}] + \mathbf{a}',$$

где $\boldsymbol{\Omega}$ — угловая скорость осевого вращения Земли, $\mathbf{v} \equiv \mathbf{v}_{\text{отн}}$ — скорость жидкости относительно Земли, $\mathbf{a}'$ — ускорение, вызванное потенциальными силами. Потенциальный вектор $\mathbf{a}'$ можно исключить, взяв циркуляцию вектора $\mathbf{a}$ по замкнутому контуру. В качестве такового выберем геометрическую ось кольцевой трубы. Тогда

$$\oint \mathbf{a}\, ds = \oint_s \mathbf{f}_{\text{K}}\, ds,$$

где $\mathbf{f}_{\text{K}} = 2[\mathbf{v}\boldsymbol{\Omega}]$ — кориолисова сила, действующая на единицу массы жидкости. Так как жидкость несжимаема, а площадь поперечного сечения трубы всюду одна и та же, то осевая составляющая ускорения a_s будет также всюду одинакова. Вынося ее из-под знака интеграла, получим $a_s = \mathscr{E}/s$, где s — длина трубы, а $\mathscr{E}$ означает интеграл $\oint \mathbf{f}_{\text{K}}\, ds$. Эта формула аналогична закону Ома. Роль силы тока, сопротивления и электродвижущей силы играют величины a_s, s и $\mathscr{E}$. Аналогом магнитного поля $\mathbf{B}$ служит удвоенная угловая скорость вращения Земли $2\boldsymbol{\Omega}$. Поэтому на основании закона электромагнитной индукции можно написать $\mathscr{E} = -d\Phi/dt$, где Φ — поток вектора $2\boldsymbol{\Omega}$, пронизывающий контур s. Таким образом,

$$a_s \equiv \frac{dV}{dt} = -\frac{1}{s}\,\frac{d\Phi}{dt},$$

откуда

$$V = -\frac{1}{s}\,\Delta\Phi. \tag{67.3}$$

Через V обозначена осевая составляющая скорости жидкости.

Допустим, что кольцо имеет форму круга радиуса R и в начальный момент расположено горизонтально. Пусть S — вектор площади кольца в этом положении. Повернем кольцо на 180° вокруг горизонтальной оси. Тогда

$$\Phi_{\text{нач}} = 2(\boldsymbol{\Omega}\mathbf{S}), \quad \Phi_{\text{конеч}} = -2(\boldsymbol{\Omega}\mathbf{S}), \quad \Delta\Phi = -4(\boldsymbol{\Omega}\mathbf{S}) = -4\Omega S\sin\vartheta.$$

Здесь ϑ — географическая широта места, где производится опыт. Подставляя в формулу (67.3) $S = \pi R^2$, $s = 2\pi R$ и предполагая, что в начальный момент жидкость была неподвижна, получим

$$V = 2\Omega R\sin\vartheta.$$

Пусть в конечном положении кольцо находится в покое. Тогда жидкость будет двигаться в нем со скоростью $V = 2\Omega R\sin\vartheta$. На полюсе $\vartheta = 90^\circ$, $V = 2\Omega R$. Этот результат легко получить также, относя все движения к «неподвижной» системе отсчета. На экваторе $V = 0$.

2. Металлический шар радиуса a из немагнитного материала движется равномерно в постоянном и однородном магнитном поле $\mathbf{B}$ со скоростью $\mathbf{v}$,

направленной под углом к магнитному полю. Найти напряженность электрического поля внутри и вне шара в «неподвижной» системе отсчета, относительно которой шар движется со скоростью $\mathbf{v}$. Найти также объемную и поверхностную плотность индуцированных зарядов. Магнитным полем движущихся индуцированных зарядов пренебречь.

Решение. Так как в стационарном состоянии ток внутри шара должен отсутствовать, то электрическое поле $\mathbf{E}^{(i)}$ должно компенсироваться силой $(1/c)[\mathbf{vB}]$. Это дает для электрического поля внутри шара

$$\mathbf{E}^{(i)} = -\frac{1}{c}[\mathbf{vB}].$$

Объемных зарядов внутри шара не будет, так как $\operatorname{div}\mathbf{E}^{(i)} = 0$. Касательная составляющая поля $\mathbf{E}^{(i)}$, а следовательно, и внешнего поля $\mathbf{E}^{(e)}$ на поверхности шара будет $E_\vartheta = -E^{(i)}\sin\vartheta$, где ϑ — угол между направлением вектора $\mathbf{E}^{(i)}$ и радиусом $\mathbf{r}$, проведенным из центра шара. Наружное поле $\mathbf{E}^{(e)}$ ищем как поле диполя $\mathbf{p}$, помещенного в центре шара:

$$\mathbf{E}^{(e)} = \frac{3(\mathbf{pr})}{r^5}\,\mathbf{r} - \frac{\mathbf{p}}{r^3}.$$

Вектор $\mathbf{p}$ легко найти по значению касательной составляющей E_ϑ. Таким путем получаем

$$\mathbf{E}^{(e)} = \frac{3a^3}{cr^5}([\mathbf{vB}]\mathbf{r})\mathbf{r} - \frac{a^3}{cr^3}[\mathbf{vB}].$$

Поверхностная плотность зарядов σ определится по скачку нормальных составляющих электрического поля. Она равна

$$\sigma = \frac{3}{4\pi c}([\mathbf{vB}]\mathbf{n}),$$

где $\mathbf{n}$ — единичный вектор внешней нормали к поверхности шара.

3. Как с помощью флюксметра можно измерить магнитное поле внутри длинного соленоида, закрытого с обоих торцов крышками из немагнитного материала, не вводя флюксметр внутрь соленоида? Обмотка соленоида доходит до самых концов цилиндра, на который она намотана.

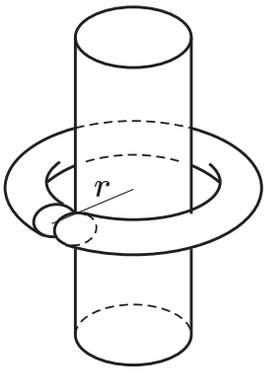


Рис. 163

Ответ. Надо с помощью флюксметра, размеры которого малы по сравнению с радиусом цилиндра, измерить магнитное поле в центре одного из оснований цилиндра. Поле внутри соленоида будет в два раза больше.

4. На цилиндрический железный сердечник, через который проходит однородный магнитный поток $\Phi = \Phi_0\cos\omega t$, надет тор из диэлектрика с диэлектрической проницаемостью ε (рис. 163). В торе имеется бесконечно узкий воздушный зазор, сделанный двумя бесконечно близкими разрезами вдоль меридиональных плоскостей. Найти напряженность электрического поля E в зазоре в зависимости от расстояния r до оси цилиндра.

Ответ. $E = \dfrac{\varepsilon\omega}{2\pi cr}\,\Phi_0\sin\omega t.$

§ 68. Индуктивность проводов. Явления при замыкании и размыкании тока

1. Рассмотрим тонкий замкнутый провод, по которому течет постоянный ток I (рис. 164). Пусть $\mathbf{B}$ — магнитное поле этого тока. Внутри провода параллельно его оси проведем произвольный замкнутый математический контур s и установим на нем положительное направление. Пусть Φ — магнитный поток, посылаемый вектором $\mathbf{B}$ через контур s. Если в пространстве нет ферромагнитных тел, то величины $\mathbf{B}$ и Φ будут пропорциональны току, и можно написать

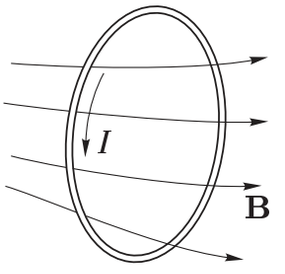


Рис. 164

$$\Phi = LI^{(m)} = \frac{1}{c}\,LI. \qquad (68.1)$$

Здесь I — сила тока в гауссовой системе единиц, а $I^{(m)}$ — сила того же тока в системе СГСМ. Коэффициент L не зависит от силы тока. Он определяется только размерами и конфигурацией самого провода и называется *индуктивностью* этого провода. Его называют также *самоиндукцией*, или *коэффициентом самоиндукции* провода. Это определение содержит известный элемент неопределенности, поскольку точно не фиксировано, как внутри провода проведен вспомогательный математический контур s. Но для тонкого провода эта неопределенность пренебрежимо мала и не имеет никакого значения. Кроме того, от нее можно полностью освободиться, что будет сделано в § 69. Для краткости величину Φ называют просто *магнитным потоком* через рассматриваемый замкнутый провод. Однако точный смысл этого понятия раскрывается введением вспомогательного математического контура s, как сделано выше.

При вычислении индуктивности тонкий провод нельзя заменить проводом бесконечно тонким — геометрической линией. Действительно, в этом случае магнитное поле вблизи провода было бы пропорционально $1/r$, где r — расстояние до провода. Для магнитного потока и индуктивности мы получили бы бесконечные значения. Чем тоньше провод, тем при прочих равных условиях больше его индуктивность.

2. Для примера вычислим индуктивность соленоида, пренебрегая при этом краевыми эффектами. Пусть l — длина соленоида, N — общее число витков, S — площадь одного витка. Индукция магнитного поля внутри соленоида

$$B = \frac{4\pi}{c}\,\frac{IN\mu}{l}.$$

Магнитный поток через один виток равен BS, а через N витков — BSN, т. е.

$$\Phi = \frac{4\pi}{c}\,\frac{\mu N^2 S}{l}\,I.$$

Сравнивая эту формулу с формулой (68.1), получим

$$L = \frac{4\pi\mu N^2 S}{l}. \tag{68.2}$$

3. За единицу магнитного потока в гауссовой системе единиц и в системе СГСМ принимают *максвелл*. Максвелл есть магнитный поток, создаваемый магнитным полем в один гаусс через перпендикулярную к нему площадку в один квадратный сантиметр. Как следует из закона Био и Савара

$$d\mathbf{B} = \frac{I}{cr^3}\,[d\mathbf{l}\,\mathbf{r}],$$

магнитный поток имеет размерность величины $I\mathrm{l}/c$. Учитывая это, из формулы (68.1) находим, что в гауссовой системе и СГСМ коэффициент самоиндукции имеет размерность длины. Его единица в этих системах называется *сантиметром*. Сантиметр есть индуктивность такого витка, в котором ток силой в одну СГСМ-единицу создает магнитный поток в один максвелл. Формула (68.2) дает индуктивность соленоида в сантиметрах.

В практических единицах (вольт, ампер, ом и т. д.) закон электромагнитной индукции и формулу (68.1) записывают в виде

$$\mathscr{E}'^{\text{инд}} = \frac{d\Phi'}{dt}, \tag{68.3}$$

$$\Phi' = L'I'. \tag{68.4}$$

Над всеми буквами здесь поставлены штрихи, которые означают, что величины, обозначаемые этими буквами, измеряются в практических единицах. Практической единицей магнитного потока является *вебер*. Эта единица определяется условием, чтобы при скорости изменения магнитного потока в 1 Вб/с в контуре возбуждалась электродвижущая сила в один вольт. Можно также сказать, что вебер есть *вольт-секунда*. Найдем соотношение между вебером и максвеллом. В гауссовой системе

$$\mathscr{E}^{\text{инд}} = -\frac{1}{c}\,\frac{d\Phi}{dt}.$$

Так как 1 В $= \dfrac{1}{300}$ СГСЭ-ед. напряжения (приближенно), то

$$\mathscr{E}'^{\text{инд}}\ (\text{вольты}) = 10^{-8}\,\frac{d\Phi\ (\text{максвеллы})}{dt}\ (\text{точно}).$$

Сопоставляя эту формулу с формулой (68.3), видим, что вебер в 10^8 раз больше максвелла:

$$1\ \text{Вб} = 10^8\ \text{Мкс}.$$

Практической единицей индуктивности является *генри*. Это есть индуктивность такого провода, в котором при силе тока в один ампер

возбуждается магнитный поток в один вебер:

$$1\ \Г = \frac{1\ \text{Вб}}{1\ \text{А}} = \frac{10^8\ \text{Мкс}}{\frac{1}{10}\ \text{СГСМ-ед. тока}} = 10^9\ \text{см}.$$

4. Рассмотрим явления при замыкании и размыкании постоянного тока, обусловленные индуктивностью цепи. Пусть цепь состоит из источника постоянной ЭДС $\mathscr{E}$, катушки самоиндукции и омического сопротивления (рис. 165). Полную индуктивность цепи обозначим через L, а полное сопротивление — через R. При замыкании ключа K ток не сразу достигает предельного значения $\mathscr{E}/R$, определяемого законом Ома, а нарастает постепенно. При этом возрастает также магнитный поток, пронизывающий контур цепи. Возникает электродвижущая сила индукции и соответствующий ей индукционный ток. Этот ток называется *экстратоком замыкания*. Согласно правилу Ленца направление экстратока замыкания противоположно направлению основного тока.

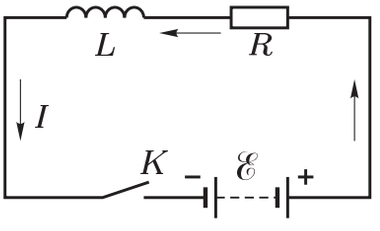


Рис. 165

Сила переменного тока не обязательно должна быть одной и той же на всех участках провода, так как в отдельных местах возможно накопление электрических зарядов. Однако мы рассмотрим здесь только такие переменные токи, которые меняются во времени сравнительно медленно. Тогда мгновенные значения токов во всех участках неразветвленной цепи с высокой степенью точности одикаковы, а магнитные поля внутри проводов могут вычисляться по закону Био и Савара, как если бы токи были постоянными. Такие токи называются *квазистационарными*. Для них справедливы формулы (68.3) и (68.4). Сила тока определяется выражением

$$I = \frac{\mathscr{E} + \mathscr{E}^{\text{инд}}}{R}.$$

В практических единицах

$$I' = \frac{\mathscr{E}' - d\Phi/dt}{R'}. \tag{68.5}$$

Это — дифференциальное уравнение для квазистационарных токов. Его можно записать в виде

$$\frac{d}{dt}(L'I') + R'I' = \mathscr{E}'. \tag{68.6}$$

Если за время изменения тока провода не деформируются, то индуктивность L' постоянна и может быть вынесена из-под знака производной:

$$L'\frac{dI'}{dt} + R'I' = \mathscr{E}'. \tag{68.7}$$

При постоянном значении $\mathscr{E}'$ общее решение этого уравнения имеет вид

$$I' = C \exp\left(-\frac{R'}{L'}\, t\right) + \frac{\mathscr{E}'}{R'}.$$

Постоянная интегрирования C должна определяться из начального условия: в момент замыкания, т. е. при $t = 0$, ток равен нулю. Используя это условие, без труда находим

$$I = \frac{\mathscr{E}}{R}\,(1 - e^{-t/\tau}), \tag{68.8}$$

где τ — постоянная, имеющая размерность времени:

$$\tau = \frac{L'}{R'}. \tag{68.9}$$

Она называется *временем установления тока.* В формуле (68.8) всюду опущены штрихи, так как эта формула применима в любой системе единиц. Меняется только выражение для времени установления тока. В гауссовой системе единиц

$$\tau = \frac{L}{c^2 R}. \tag{68.10}$$

Полный ток I состоит из двух слагаемых, из которых второе, т. е. $-(\mathscr{E}/R)e^{-t/\tau}$, определяет силу экстратока замыкания. При $t \to \infty$ экстраток стремится к нулю, а полный ток I — к своему предельному значению $\mathscr{E}/R$. Таким образом, окончательное значение тока устанавливается постепенно. Быстрота установления определяется временем τ: по истечении времени τ сила экстратока убывает в e раз.

5. Исследуем теперь процесс размыкания тока. Схема опыта изображена на рис. 166. Ключ K сначала замкнут. Направления токов показаны сплошными стрелками. Общий ток распределяется между параллельно включенными самоиндукцией L и омическим сопротивлением R. Если внутреннее сопротивление батареи пренебрежимо мало, то ток в катушке самоиндукции будет равен $I_0 = \mathscr{E}/r$. После размыкания ключа K замкнутым останется только контур $ABCD$. Первоначальный ток, существовавший в катушке самоиндукции, обладал определенным запасом магнитной энергии, которая исчезает не сразу. Магнитное поле начнет убывать. Это возбудит электродвижущую силу и индукционный ток в контуре $ABCD$. Такой ток называется *экстратоком размыкания.* На рис. 166 его направление показано пунктирными стрелками. В катушке

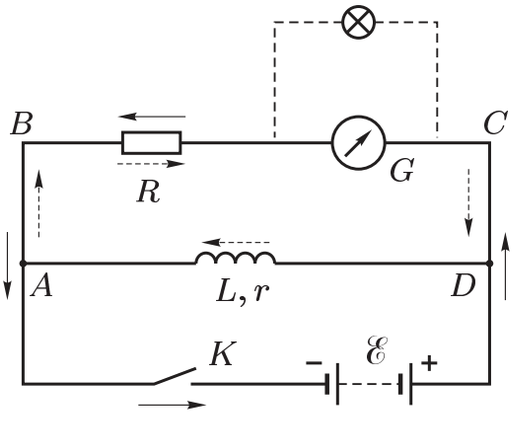


Рис. 166

самоиндукции экстраток течет в том же направлении, что и первоначальный ток, в остальных участках контура $ABCD$ — в противоположном направлении. Если R' — общее сопротивление контура $ABCD$, то сила тока определится из дифференциального уравнения

$$L'\frac{dI'}{dt} + R'I' = 0$$

и начального условия: $I' = I'_0$ при $t = 0$. Это дает

$$I' = I'_0 e^{-t/\tau}, \tag{68.11}$$

где τ определяется прежним выражением (68.9). Электродвижущая сила индукции равна

$$\mathscr{E}'_{\text{инд}} = -L'\frac{dI'}{dt} = \frac{L'I'_0}{\tau}\,e^{-t/\tau} = \frac{R'}{r'}\,\mathscr{E}'e^{-t/\tau}. \tag{68.12}$$

Если $R' \gg r'$, то эта величина может значительно превзойти ЭДС батареи. В этом причина электрического пробоя, наблюдающегося иногда при выключении тока в цепях, содержащих большие индуктивности.

Для демонстрации явления можно взять катушку длиной 50–60 см и диаметром 8–10 см с сердечником из железных прутьев и обмоткой из нескольких слоев проволоки диаметром около 1 мм. Параллельно катушке присоединена лампочка, как указано на рис. 166 штриховой линией. Лампочка рассчитана на напряжение, несколько превышающее ЭДС батареи. (При ЭДС батареи в 4 В можно, например, взять лампочку на 12 В.) При замкнутой цепи лампочка горит тускло. При размыкании ключа K она ярко вспыхивает и даже может перегореть, так как ЭДС индукции превосходит в несколько раз ЭДС батареи.

6. Рассмотрим теперь два витка (или две катушки), по которым текут постоянные токи I_1 и I_2. Установим произвольно на этих витках положительные направления обхода. Если в окружающем пространстве нет ферромагнетиков, то магнитные потоки через витки Φ_1 и Φ_2 пропорциональны токам и могут быть представлены в виде

$$\begin{aligned}\Phi_1 &= \frac{1}{c}\,L_{11}I_1 + \frac{1}{c}\,L_{12}I_2,\\ \Phi_2 &= \frac{1}{c}\,L_{21}I_1 + \frac{1}{c}\,L_{22}I_2.\end{aligned} \tag{68.13}$$

Коэффициенты L_{11}, L_{12}, L_{21}, L_{22} не зависят от токов, а определяются лишь формой, размерами и взаимным расположением витков. Они называются *коэффициентами индуктивности*. Если $I_2 = 0$, то $\Phi_1 = L_{11}I_1/c$; если $I_1 = 0$, то $\Phi_2 = L_{22}I_2/c$. Поэтому L_{11} есть индуктивность первого, а L_{22} — второго витков. Оставшиеся два коэффициента L_{12} и L_{21} называются *взаимными индуктивностями* или *коэффициентами взаимной индукции*. Они, разумеется, измеряются теми же единицами, что и коэффициенты самоиндукции. В практической системе и системе СГСМ множитель c в формулах (68.13) опускают.

Распространение этих понятий на случай системы произвольного числа проволочных витков совершенно тривиально и не нуждается в пояснении. В следующем параграфе будет доказано соотношение $L_{ik} = L_{ki}$, называемое *теоремой взаимности*.

§ 69. Магнитная энергия токов

1. Электрический ток обладает запасом энергии, называемой *магнитной*. При ее вычислении можно полностью отвлечься от сопротивления проводов, по которым текут токи, полагая это сопротивление равным нулю. Это не может отразиться на общности результата, так как *магнитная энергия может зависеть только от величины и распределения токов, а также от магнитных свойств среды, заполняющей пространство.* Считая же провода идеально проводящими, мы упростим рассуждение, так как в расчетах не надо будет учитывать потери энергии на джоулево тепло.

Рассмотрим сначала одиночный неподвижный замкнутый виток проволоки. Пусть в начальный момент сила тока в нем равна нулю. Будем каким-либо способом создавать и наращивать ток в витке I. Тогда будет нарастать и магнитный поток через виток Φ. Возникнет электродвижущая сила индукции. Элементарная работа, которую должен совершить внешний источник против электродвижущей силы индукции, будет

$$\delta A^{\text{внеш}} = -\mathscr{E}^{\text{инд}} I\, dt,$$

или ввиду соотношения (64.1)

$$\delta A^{\text{внеш}} = \frac{1}{c}\, I\, d\Phi. \tag{69.1}$$

Полученное соотношение носит общий характер. Оно справедливо и для ферромагнитных материалов, так как при его выводе относительно магнитных свойств среды не вводилось никаких предположений. Однако если среда не обладает гистерезисом, в частности является пара- или диамагнитной, то работа $\delta A^{\text{внеш}}$ пойдет только на увеличение магнитной энергии W_m, так что

$$dW_m = \frac{I}{c}\, d\Phi. \tag{69.2}$$

В этом параграфе мы будем предполагать, что ферромагнетики отсутствуют. Тогда $\Phi = LI/c$, причем для неподвижного провода самоиндукция L остается постоянной. Используя это и интегрируя, получим

$$W_m = \frac{L}{2}\left(\frac{I}{c}\right)^2 = \frac{1}{2c}\, I\Phi = \frac{\Phi^2}{2L}. \tag{69.3}$$

(В практической системе и системе СГСМ вместо I/c следует писать просто I.) Для справедливости формулы (69.3) несущественно, что во время нарастания тока виток оставался неподвижным, так как энергия зависит только *от состояния системы*, но не от способа, каким было достигнуто это состояние. Например, переход в конечное состояние

можно осуществлять следующим образом. Пока в проводе нет тока, деформируем его, чтобы он принял окончательную конфигурацию. Это не требует затраты работы. Затем, сохраняя провод неподвижным, будем наращивать ток до конечного значения I. На это потребуется работа $LI^2/(2c^2)$. Она и будет равна искомому приращению магнитной энергии. Приведенное рассуждение показывает, что в формуле (69.3) под L следует понимать самоиндукцию витка в конечном состоянии.

Формула (69.3) освобождает понятие самоиндукции от той неопределенности, на которую было указано в § 68. Действительно, ток I и магнитная энергия W_m — величины, определяемые совершенно однозначно. Поэтому по формуле (69.3) можно вычислить также совершенно однозначно и коэффициент самоиндукции L. Более того, эта формула может служить для определения L и в тех случаях, когда провод толстый.

2. Обобщим теперь формулу (69.3) на случай произвольного числа витков. Предполагая опять все витки неподвижными, будем увеличивать токи в них. Тогда для элементарной работы против электродвижущей силы индукции можно по аналогии с формулой (69.1) написать

$$\delta A^{\text{внеш}} = \frac{1}{c}\sum_i I_i\, d\Phi_i, \tag{69.4}$$

где суммирование ведется по всем виткам. Магнитная энергия в конечном состоянии представится интегралом

$$W_m = \frac{1}{c}\int \sum I'_i\, d\Phi'_i,$$

где штрихованными буквами I'_i и Φ'_i обозначены переменные (текущие) значения соответствующих величин. Символы I_i и Φ_i сохранены для обозначения токов и магнитных потоков в конечном состоянии.

Для вычисления интеграла заметим, что его величина не зависит от «пути интегрирования», т. е. от характера изменения силы токов в проводах. Можно, например, возбуждать токи последовательно: сначала создать ток только в первом витке, доведя его значение до величины I_1, затем, не меняя I_1, начать возбуждать ток во втором витке и т. д. Но можно возбуждать токи сразу во всех витках и притом независимо друг от друга. Результат вычисления магнитной энергии во всех случаях будет один и тот же. Чтобы упростить расчет, будем наращивать все токи одновременно и притом так, чтобы они оставались пропорциональными друг другу. Таким образом, в любой момент будет соблюдаться соотношение $I'_i = \lambda I_i$, где λ — переменная величина, не зависящая от i. В начальном состоянии $\lambda = 0$, в конечном $\lambda = 1$. Так как при отсутствии ферромагнитных материалов магнитные потоки связаны с токами линейно, то для них справедливы такие же соотношения, т. е. $\Phi'_i = \lambda\Phi_i$, а потому $d\Phi'_i = \Phi_i\, d\lambda$. Таким образом,

$$W_m = \frac{1}{c}\sum I_i\Phi_i \int_0^1 \lambda\, d\lambda,$$

или после интегрирования

$$W_m = \frac{1}{2c}\sum I_i \Phi_i = \frac{1}{2c^2}\sum\sum L_{ik} I_i I_k. \tag{69.5}$$

3. При расчете предполагалось, что в процессе намагничивания магнитная проницаемость μ оставалась постоянной. Только при этом условии связь между токами и магнитными потоками будет линейной. Можно показать (см. § 73), что если μ зависит от температуры, то при намагничивании температура магнетика, а с ней и величина μ будут меняться. В этом случае расчет неприменим. Однако он становится применимым, если при намагничивании температуру поддерживать постоянной. Тогда величина W_m будет иметь смысл работы, совершаемой над системой при изотермическом и квазистатическом нарастании тока в проводах. Такая работа в термодинамике называется *свободной энергией.* Таким образом, формулы (69.3) и (69.5) в общем случае определяют не внутреннюю, а *свободную магнитную энергию системы.* Здесь все обстоит так же, как и в аналогичном вопросе электростатики (см. § 28, п. 4).

4. Используя формулы (69.4) и (69.5), докажем теорему взаимности для коэффициентов взаимной индукции, о которой говорилось в предыдущем параграфе. Согласно этой теореме матрица коэффициентов L_{ik} симметрична, т. е.

$$L_{ik} = L_{ki}. \tag{69.6}$$

Достаточно доказать это соотношение для какой-либо пары индексов, например $i = 1$, $k = 2$. Так как коэффициенты L_{ik} не зависят от токов, то с целью упрощения вычислений можно предположить, что токи текут только по виткам *1* и *2*, а в остальных витках токи равны нулю. Если бесконечно мало изменить токи I_1 и I_2, то на это потребуется затратить работу

$$\delta A^{\text{внеш}} = \frac{1}{c}\,(I_1\, d\Phi_1 + I_2\, d\Phi_2).$$

Она пойдет на приращение магнитной энергии токов $dW_m = \delta A^{\text{внеш}}$. Но эта энергия в рассматриваемом случае дается выражением

$$W_m = \frac{1}{2c}\,(I_1\Phi_1 + I_2\Phi_2),$$

а ее приращение — выражением

$$dW_m = \frac{1}{2c}\,(I_1\, d\Phi_1 + I_2\, d\Phi_2) + \frac{1}{2c}\,(\Phi_1\, dI_1 + \Phi_2\, dI_2).$$

Сравнивая оба выражения для dW_m, получим

$$I_1\, d\Phi_1 + I_2\, d\Phi_2 = \Phi_1\, dI_1 + \Phi_2\, dI_2. \tag{69.7}$$

Это соотношение справедливо при любых значениях dI_1 и dI_2. Поэтому для сокращения последующих вычислений можно взять $dI_2 = 0$, т. е. $I_2 = \text{const}$. Магнитные потоки определяются выражениями

$$\Phi_1 = \frac{1}{c}\,(L_{11}I_1 + L_{12}I_2), \quad \Phi_2 = \frac{1}{c}\,(L_{21}I_1 + L_{22}I_2).$$

Из них дифференцированием при постоянном I_2 находим

$$d\Phi_1 = \frac{1}{c}\, L_{11}\, dI_1, \quad d\Phi_2 = \frac{1}{c}\, L_{21}\, dI_1.$$

Подставляя эти выражения в соотношение (69.7), получим

$$L_{11} I_1\, dI_1 + L_{21} I_2\, dI_1 = (L_{11} I_1 + L_{12} I_2)\, dI_1.$$

Отсюда следует $L_{12} = L_{21}$, и теорема взаимности доказана. При доказательстве сопротивления проводов не учитывались. Но это не имеет значения, так как коэффициенты взаимной индукции L_{ik} зависят только от формы и расположения проводов, а также от распределения плотности электрического тока по их сечениям.

5. В заключение этого параграфа разберем следующий парадокс. Сила $\mathbf{f} = (e/c)[\mathbf{vB}]$, действующая в магнитном поле на движущийся заряд, перпендикулярна к его скорости $\mathbf{v}$, а потому работы не производит. Почему же при движении витка с током (например, якоря электромотора) в магнитном поле механическая работа над проводником, несомненно, производится? Дело в том, что это не есть полная работа магнитного поля над движущимися зарядами, являющимися носителями тока в проводнике. Рассмотрим тождество

$$\frac{1}{c}\,[\mathbf{j}\,\mathbf{B}]\mathbf{v} = -\frac{1}{c}\,[\mathbf{vB}]\,\mathbf{j},$$

где $\mathbf{v}$ — скорость проводника, а $\mathbf{j}$ — плотность электрического тока в нем. Слева стоит работа амперовой силы $\mathbf{f}_{\text{амп}} = (1/c)[\mathbf{j}\,\mathbf{B}]$ над единицей объема проводника в единицу времени, справа — взятая со знаком минус такая же работа электрического поля индукции $\mathbf{E}_{\text{инд}}$, возникающего при движении проводника. Полная работа $A_{\text{полн}} = A_{\text{амп}} + A_{\text{инд}}$ равна нулю. Допустим теперь, что в цепь включена батарея или какой-либо другой источник тока. Тогда добавится работа батареи; $A_{\text{полн}} = A_{\text{амп}} + A_{\text{инд}} + A_{\text{бат}}$. Пусть электродвижущая сила батареи подобрана так, что она в каждый момент времени компенсирует электродвижущую силу индукции, поддерживая ток в цепи постоянным. Тогда $A_{\text{инд}} + A_{\text{бат}} = 0$, и если потери на джоулево тепло пренебрежимо малы, то $A_{\text{полн}} = A_{\text{амп}}$. Таким образом, работа $A_{\text{амп}}$ производится за счет энергии батареи.

§ 70. Локализация магнитной энергии в пространстве

1. Формулы (69.2) выражают магнитную энергию *через токи и магнитные потоки.* В таком виде величина (69.3) может рассматриваться как *потенциальная энергия токов*, взаимодействующих по закону Ампера. Это соответствует представлению о непосредственном действии на расстоянии. Но выражение для магнитной энергии можно преобразовать в другую форму, которая соответствует иному представлению о месте нахождения энергии. Покажем это на примере длинного соленоида, по поверхности которого циркулирует ток с линейной

плотностью $i = I/l$ (l — длина соленоида). Мы не будем пользоваться выражениями (69.3) для энергии токов, так как они справедливы лишь при условии, что векторы $\mathbf{B}$ и $\mathbf{H}$ связаны соотношением $\mathbf{B} = \mu\mathbf{H}$, а воспользуемся общей формулой (69.2), справедливость которой предполагает только, что между $\mathbf{B}$ и $\mathbf{H}$ существует какая-то однозначная, но не обязательно линейная функциональная связь (нет гистерезиса). Пренебрегая краевыми эффектами, можно написать для поля H внутри соленоида $H = 4\pi i/c = 4\pi I/(cl)$, откуда $I = clH/(4\pi)$. Пусть S — площадь поперечного сечения соленоида. Тогда $\Phi = BS$, и, следовательно,

$$dW_m = \frac{I}{c}\,d\Phi = \frac{1}{4\pi}\,lSH\,dB = \frac{V}{4\pi}\,(\mathbf{H}\,d\mathbf{B})$$

($V = Sl$ — объем соленоида). Если w_m — магнитная энергия, приходящаяся на единицу объема соленоида, то для ее дифференциала можно написать

$$\delta A^{\text{внеш}} = dw_m = \frac{1}{4\pi}\,(\mathbf{H}\,d\mathbf{B}). \tag{70.1}$$

И в общем случае постоянных электрических токов, произвольным образом текущих в пространстве, можно доказать, что выражение для магнитной энергии может быть преобразовано к виду

$$W_m = \int w_m\,dV, \tag{70.2}$$

где w_m определяется прежней формулой (70.1). Это — чисто математический вопрос, совершенно аналогичный соответствующему вопросу в электростатике. Опуская здесь математические преобразования, остановимся только на физическом смысле формулы (70.2). Ее можно истолковать в том смысле, что *магнитная энергия локализована в пространстве с объемной плотностью* w_m. Это соответствует представлениям теории поля. В рамках учения о постоянных токах и постоянных магнитных полях нельзя указать ни одного опыта, который бы решал вопрос в пользу одного из двух представлений о локализации магнитной энергии: представления теории непосредственного действия на расстоянии и представления теории поля. Здесь дело обстоит совершенно так же, как в электростатике. Лишь явления в быстропеременных полях, например распространение электромагнитных волн, позволяют сделать соответствующий выбор. Они согласуются только с представлением теории поля о локализации магнитной энергии в пространстве. В случае быстропеременных полей формулы (69.3) просто лишены смысла.

В случае пара- и диамагнитных сред $\mathbf{B} = \mu\mathbf{H}$ и выражение (70.1) можно проинтегрировать. Таким путем получим

$$w_m = \frac{1}{8\pi}\,\mu H^2 = \frac{1}{8\pi}\,\mathbf{HB} = \frac{B^2}{8\pi\mu}. \tag{70.3}$$

2. Приведем теперь математическое доказательство формулы (70.2) Как будет видно из доказательства, можно ограничиться магнитным полем одного витка. Обобщение на случай многих витков чисто формальное и не встречает никаких затруднений. Считая виток неподвижным и полагая в формуле (69.2) $d\Phi = \int\limits_S d\mathbf{B}\, d\mathbf{S}$, получим

$$dW_m = \frac{I}{c} \int\limits_S d\mathbf{B}\, d\mathbf{S},$$

где интегрирование ведется по произвольной поверхности, натянутой на контур тока l. Дальнейшие преобразования используют понятие векторного потенциала. Можно доказать (см. задачи 1 и 2 к этому параграфу), что всякий вектор, дивергенция которого равна нулю, может быть представлен в виде ротора другого вектора. Так как $\operatorname{div} \mathbf{B} = 0$, то на этом основании можно написать

$$\mathbf{B} = \operatorname{rot} \mathbf{A}. \tag{70.4}$$

Вектор $\mathbf{A}$ и называется *векторным потенциалом* магнитного поля. Из формулы (70.4) следует: $d\mathbf{B} = \operatorname{rot} d\mathbf{A}$. Используя это соотношение и применяя теорему Стокса, находим

$$dW_m = \frac{I}{c} \int\limits_S d\mathbf{S} \operatorname{rot}\, d\mathbf{A} = \frac{I}{c} \oint\limits_l (\, d\mathbf{l}\, d\mathbf{A}).$$

Вместо линейного введем объемный элемент тока $\mathbf{j}\, dV = I\, d\mathbf{l}$ и воспользуемся теоремой о циркуляции $\operatorname{rot} \mathbf{H} = 4\pi \mathbf{j}/c$. Тогда

$$dW_m = \frac{1}{4\pi} \int\limits_V (d\mathbf{A} \operatorname{rot} \mathbf{H})\, dV,$$

где интеграл распространен по всему пространству, по которому течет ток. Применяя известное тождество векторного анализа (см. задачу 3 к этому параграфу), преобразуем подынтегральное выражение к виду

$$d\mathbf{A} \operatorname{rot} \mathbf{H} = \operatorname{div} [\mathbf{H}\, d\mathbf{A}] + \mathbf{H} \operatorname{rot}\, d\mathbf{A} = \operatorname{div} [\mathbf{H}\, d\mathbf{A}] + \mathbf{H}\, d\mathbf{B}.$$

Интеграл от дивергенции можно преобразовать в поверхностный, взяв в качестве поверхности интегрирования бесконечно удаленную поверхность. Если все токи текут в конечной области пространства, то магнитное поле будет убывать на бесконечности достаточно быстро и рассматриваемый интеграл обратится в нуль.

В результате получится

$$dW_m = \frac{1}{4\pi} \int\limits_V (\mathbf{H}\, d\mathbf{B})\, dV.$$

Отсюда после интегрирования по $d\mathbf{B}$ получим

$$W_m = \int w_m\, dV, \quad \text{где} \quad w_m = \frac{1}{4\pi} \int \mathbf{H}\, d\mathbf{B}.$$

ЗАДАЧИ

1. Доказать, что для произвольного вектора $\mathbf{A}$ справедливо соотношение $\operatorname{div} \operatorname{rot} \mathbf{A} = 0$.

2. Доказать, что если $\operatorname{div} \mathbf{B} = 0$, то вектор $\mathbf{B}$ может быть представлен в виде $\mathbf{B} = \operatorname{rot} \mathbf{A}$.

Решение. Содержание теоремы сводится к утверждению, что уравнение $\operatorname{rot}\mathbf{A} = \mathbf{B}$, где $\mathbf{B}$ — заданный, а $\mathbf{A}$ — неизвестный вектор, имеет решение. Покажем, например, что существует решение, в котором $A_z = 0$. В этом случае рассматриваемое уравнение сводится к системе трех скалярных уравнений:

$$\frac{\partial A_y}{\partial z} = -B_x, \quad \frac{\partial A_x}{\partial z} = B_y, \quad \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} = B_z.$$

Первым двум уравнениям можно удовлетворить, полагая

$$A_y = -\int_{z_0}^{z} B_x\,dz + f(x,y), \quad A_x = \int_{z_0}^{z} B_y\,dz,$$

где $f(x,y)$ — произвольная функция. Подставляя это решение в третье уравнение и учитывая, что $\partial B_x/\partial x + \partial B_y/\partial y = -(\partial B_z/\partial z)$, получим

$$\int_{z_0}^{z} \frac{\partial B_z}{\partial z}\,dz + \frac{\partial f}{\partial x} = B_z,$$

откуда

$$\frac{\partial f}{\partial x} = B_z(x,y,z_0)$$

и, далее,

$$f(x,y) = \int B_z(x,y,z_0)\,dx.$$

Таким образом, найдено одно из решений уравнения $\operatorname{rot}\mathbf{A} = \mathbf{B}$, и теорема доказана. Заметим, что это уравнение при условии $\operatorname{div}\mathbf{B} = 0$ имеет бесчисленное множество решений, т. е. векторный потенциал определен не однозначно.

3. Доказать тождество

$$\operatorname{div}[\mathbf{AB}] = \mathbf{B}\operatorname{rot}\mathbf{A} - \mathbf{A}\operatorname{rot}\mathbf{B}. \tag{70.5}$$

Решение. В справедливости этого тождества нетрудно убедиться непосредственной проверкой, если записать его в прямоугольных координатах. Однако выкладки получатся более простыми и естественными, если применить символический метод. На основании определения векторного произведения и дивергенции

$$\operatorname{div}[\mathbf{AB}] = \operatorname{div}\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ A_x & A_y & A_z \\ B_x & B_y & B_z \end{vmatrix} = \begin{vmatrix} \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ A_x & A_y & A_z \\ B_x & B_y & B_z \end{vmatrix}.$$

Переставив первую строку полученного определителя со второй, а затем с третьей, соблюдая при этом известное правило знаков, получим два новых определителя:

$$-\begin{vmatrix} A_x & A_y & A_z \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ B_x & B_y & B_z \end{vmatrix} \quad \text{и} \quad +\begin{vmatrix} B_x & B_y & B_z \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ A_x & A_y & A_z \end{vmatrix}.$$

Оба эти определителя, однако, не равны исходному. Это связано с тем, что к рассматриваемому определителю обычное правило перестановки строк неприменимо, так как первая строка его состоит не из чисел, а из операторов.

Пользуясь правилом дифференцирования произведения, нетрудно, однако, заметить, что сумма определителей, полученных в результате перестановки, равна исходному определителю, т. е.

$$\operatorname{div}[\mathbf{AB}] = \begin{vmatrix} B_x & B_y & B_z \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ A_x & A_y & A_z \end{vmatrix} - \begin{vmatrix} A_x & A_y & A_z \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ B_x & B_y & B_z \end{vmatrix}.$$

Очевидно, это соотношение можно переписать так:

$$\operatorname{div}[\mathbf{AB}] = \mathbf{B}\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ A_x & A_y & A_z \end{vmatrix} - \mathbf{A}\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ B_x & B_y & B_z \end{vmatrix},$$

или

$$\operatorname{div}[\mathbf{AB}] = \mathbf{B}\operatorname{rot}\mathbf{A} - \mathbf{A}\operatorname{rot}\mathbf{B}.$$

§ 71. Теорема о сохранении магнитного потока

Допустим, что виток с током находится в произвольном магнитном поле — постоянном или переменном. Пусть он движется и деформируется произвольным образом. При этом в витке возбуждается индукционный ток

$$I = \frac{\mathscr{E}^{\text{инд}}}{R} = -\frac{1}{cR}\frac{d\Phi}{dt}.$$

Если омическое сопротивление R равно нулю, то должно быть $\mathscr{E}^{\text{инд}} = 0$, так как в противном случае в проводнике возникли бы бесконечно большие токи, что физически невозможно. Значит, должно быть $d\Phi/dt = 0$, а потому $\Phi = \text{const}$. Таким образом, *при движении идеально проводящего замкнутого провода в магнитном поле остается постоянным магнитный поток, пронизывающий контур провода.* Это положение называется *теоремой о сохранении магнитного потока.* Такое сохранение обусловлено индукционными токами, которые, согласно правилу Ленца, препятствуют всякому изменению магнитного потока через контур провода. Магнитный поток, обусловленный внешним магнитным полем, не остается постоянным. Магнитный поток, создаваемый индукционными токами, также меняется во времени. Однако сумма этих двух потоков остается постоянной.

Представим себе теперь идеально проводящую жидкость, движущуюся в магнитном поле. Выделим в ней произвольный жидкий замкнутый контур, т. е. контур, движущийся вместе с частицами самой жидкости. Такой контур может играть роль идеально проводящего провода, и к нему применима теорема о сохранении магнитного потока. Из нее следует, что *при любых движениях идеально проводящей жидкости магнитный поток, пронизывающий всякий замкнутый жидкий контур, не меняется во времени.* Идеально проводящая жидкость может свободно течь вдоль магнитных силовых трубок. Но всякое движение ее поперек магнитного поля увлекает и эти силовые трубки.

Явление происходит так, как если бы магнитные силовые линии были *вморожены в вещество* и двигались вместе с ним. Такое представление о вмороженности магнитных силовых линий широко применяется в магнитной гидродинамике при рассмотрении движений жидкостей, обладающих высокой электрической проводимостью. Оно применяется также в астрофизике и физике горячей плазмы, поскольку последняя также обладает высокой электрической проводимостью (см. § 121).

ЗАДАЧИ

1. Сверхсильные магнитные поля можно получать взрывным сжатием отрезка проводящей цилиндрической трубы, внутри которой создано начальное магнитное поле B_0. Определить магнитное поле B в трубе в момент максимального сжатия, если $B_0 = 5 \cdot 10^4$ Гс, начальный внутренний радиус трубы $R = 5$ см, радиус в момент максимального сжатия $r = 0{,}5$ см. Оболочку, окружающую магнитное поле, считать идеально проводящей. Определить также давление $\mathscr{P}$, необходимое для получения такого сжатия.

Ответ. $B = B_0(R/r)^2 = 5 \cdot 10^6$ Гс, $\mathscr{P} = B^2/(8\pi) \approx 10^{12}$ дин/см^2 $= 10^6$ атм.

2. По длинному идеально проводящему соленоиду длины l_0, отключенному от источника напряжения, течет постоянный ток I_0. Как будет меняться ток I в соленоиде при его растяжениях и сжатиях?

Ответ. $I = I_0 l/l_0$, где l — мгновенная длина соленоида.

§ 72. Энергия и силы

1. Наиболее общим методом расчета сил взаимодействия проводов с токами, а также натяжений и давлений, возникающих в среде при наличии магнитного поля, является *энергетический метод*. В этом методе используется выражение для свободной энергии магнитного поля. Как и сама свободная энергия, указанные силы зависят от *величины* и *конфигурации токов*, но не зависят при прочих равных условиях от удельного сопротивления проводов. Поэтому можно упростить вычисления, отвлекаясь от сопротивления проводов, и не принимать во внимание потери энергии на джоулево тепло. Под действием внутренних сил рассматриваемая система тел, вообще говоря, не будет находиться в равновесии и придет в движение. Для предотвращения этого приложим внешние силы, уравновешивающие внутренние силы. Если бесконечно мало нарушить равновесие, то начнется бесконечно медленный (квазистатический) процесс, сопровождающийся перемещениями и деформациями проводов и окружающей среды. Кинетическая энергия, как и во всяком квазистатическом процессе, при этом возникать не будет. Будем поддерживать во время процесса температуру постоянной. Тогда работа внешних сил $\delta A^{\text{внеш}}$ пойдет на приращение свободной энергии системы. Среду между проводами будем предполагать изотропной — жидкой или газообразной. В статических и медленно меняющихся магнитных полях, как это было выяснено в электростатике (см. § 33, п. 5), упругая часть свободной энергии компенсируется членами, содержащими производные магнитной проницаемости

по плотности среды. Поэтому при расчете сил можно отвлечься от наличия упругой части свободной энергии, если при этом одновременно не учитывать зависимость магнитной проницаемости μ от плотности среды. Так мы и поступим. Опуская упругую часть свободной энергии, пишем $\delta A^{\text{внеш}} = \delta W_m$. А так как для квазистатического процесса $\delta A^{\text{внеш}} = -\delta A$, где δA — элементарная работа внутренних сил, то

$$\delta A = -[\delta W_m]_{\Phi=\text{const}}. \tag{72.1}$$

При доказательстве предполагалось, что провода идеально проводящие, а потому магнитные потоки, пронизывающие их, *остаются постоянными.* Это явно отмечено в формуле (72.1). Однако сама формула (72.1) остается справедливой и для проводов с *конечным омическим сопротивлением.* Дело в том, что силы взаимодействия и магнитное поле в среде явно зависят только от сил токов и их распределения по проводам, но *не зависят от сопротивления проводов.* Поэтому если при проведении квазистатического процесса с реальными проводами каким-либо способом поддерживать магнитные потоки Φ неизменными, то при прочих равных условиях работа δA останется той же самой, что и в случае идеально проводящих проводов. Различие заключается только в том, что в случае идеально проводящих проводов магнитные потоки сохраняются автоматически, а в случае проводов с конечным омическим сопротивлением требуются специальные меры, чтобы обеспечить такое сохранение. Но для вычисления сил взаимодействия это обстоятельство не имеет никакого значения.

2. Формула (72.1) является *основной* при расчете сил в магнитном поле энергетическим методом. Однако в ряде случаев более удобна другая формула, в которой варьирование магнитной энергии W_m производится при сохранении постоянными *сил токов* в проводах. Выведем эту формулу.

Для поддержания постоянства токов во всех витках введем внешние электродвижущие силы $\mathscr{E}_i^{\text{внеш}}$, которые бы в каждый момент времени уравновешивали электродвижущие силы индукции, возникающие во время квазистатического процесса. Для этого должно быть $\mathscr{E}_i^{\text{внеш}} = -\mathscr{E}_i^{\text{инд}} = (1/c)\, d\Phi_i/dt$. Работа этих внешних электродвижущих сил

$$\delta A^{\text{внеш}} = \frac{1}{c} \sum \frac{d\Phi_i}{dt}\, I_i\, dt = \frac{1}{c} \sum I_i\, d\Phi_i$$

пойдет на работу системы δA и на приращение магнитной энергии:

$$\delta A + \delta W_m = \frac{1}{c} \sum I_i\, d\Phi_i.$$

(Мы по-прежнему проводим рассуждение в предположении идеальной проводимости проводов.) Но если токи поддерживаются постоянными, то для вариации магнитной энергии можно написать

$$\delta W_m = \delta \left(\frac{1}{2c} \sum I_i \Phi_i \right) = \frac{1}{2c} \sum I_i\, \delta\Phi_i.$$

Введя это выражение в предыдущее соотношение, получим

$$\delta A = [\,\delta W_m]_{I=\text{const}}, \tag{72.2}$$

где вариация магнитной энергии производится уже при постоянных токах. Это и есть *другая основная формула*, на которой основан энергетический метод расчета сил.

Приведем примеры на применение формул (72.1) и (72.2).

3. Магнитное взаимодействие замкнутых постоянных токов в однородной среде. Магнитная энергия двух витков с токами определяется выражением

$$W_m = \frac{1}{2}\,L_{11}I_1^2 + L_{12}I_1I_2 + \frac{1}{2}\,L_{22}I_2^2.$$

Будем пользоваться формулой (72.2). Если произвольным образом, но без деформаций сместить витки *1* и *2*, то ввиду однородности среды коэффициенты самоиндукции L_{11} и L_{22} меняться не будут. Если при этом поддерживать токи в витках постоянными, то единственным переменным слагаемым в выражении для W_m будет $L_{12}I_1I_2$, так что

$$[\delta W_m]_{I=\text{const}} = I_1I_2\,\delta L_{12}. \tag{72.3}$$

Оставляя виток *2* неподвижным, сместим виток *1* как целое на отрезок $\delta\mathbf{r}_1$. Элементарная работа, совершаемая системой, при этом будет $\delta A = \mathbf{F}_1\delta\mathbf{r}_1$, где $\mathbf{F}_1$ — результирующая амперовых сил, действующих на виток *1*. Согласно формуле (72.2) $\mathbf{F}_1\,\delta\mathbf{r}_1 = I_1I_2\,\delta L_{12}$. Сместим теперь виток *2* на отрезок $\delta\mathbf{r}_2 = -\delta\mathbf{r}_1$, сохраняя неподвижным виток *1*. Изменение коэффициента взаимной индукции будет тем же самым, так как этот коэффициент зависит только от взаимного расположения витков. Поэтому $\mathbf{F}_2\,\delta\mathbf{r}_2 = I_1I_2\,\delta L_{12}$. Таким образом, $\mathbf{F}_1\,\delta\mathbf{r}_1 = \mathbf{F}_2\,\delta\mathbf{r}_2$, или $\mathbf{F}_1\,\delta\mathbf{r}_1 = -\mathbf{F}_2\,\delta\mathbf{r}_1$. Отсюда, ввиду произвольности смещения $\delta\mathbf{r}_1$, следует, что $\mathbf{F}_1 = -\mathbf{F}_2$. Поворачивая один виток относительно другого, таким же путем докажем, что $\mathbf{M}_1 = -\mathbf{M}_2$, где $\mathbf{M}_1$ и $\mathbf{M}_2$ — моменты амперовых сил, действующих на витки. Таким образом, магнитное взаимодействие замкнутых постоянных токов удовлетворяет принципу равенства действия и противодействия.

4. Так как коэффициент взаимной индукции L_{12} пропорционален магнитной проницаемости μ промежуточной среды, то из приведенного рассуждения следует также, что и *силы взаимодействия между проводами в однородной среде пропорциональны ее магнитной проницаемости.* При одних и тех же токах сила взаимодействия между проводами в вакууме возрастает в μ раз, если все пространство заполнить однородной средой с магнитной проницаемостью μ.

5. Силы, действующие на границе раздела двух магнетиков. Допустим сначала, что магнитное поле перпендикулярно к границе раздела магнетиков. Это можно реализовать, взяв достаточно длинный соленоид, одна половина которого заполнена магнетиком

с магнитной проницаемостью μ_1, а другая — с магнитной проницаемостью μ_2. Магнетики граничат между собой вдоль плоскости, перпендикулярной к оси соленоида (рис. 167). Токи, циркулирующие по боковой поверхности соленоида, можно подобрать так, чтобы поле $\mathbf{B}$ внутри соленоида вдали от его концов было однородно. Пространство вне соленоида должно быть заполнено соответствующими магнетиками, чтобы последние могли свободно входить и выходить из соленоида. В выражении для магнитной энергии можно отвлечься от краевых эффектов, так как при расчете сил существенна не сама энергия, а ее вариации, возникающие при смещении границы раздела. Эти же вариации, если только соленоид достаточно длинный, не зависят от неоднородности поля вблизи его краев. Для вычисления силы $\mathbf{F}$, действующей на границу раздела магнетиков, сместим эту границу вправо на величину δx. При этом система совершит работу $F\,\delta x$. Будем производить это смещение с сохранением магнитного потока, пронизывающего соленоид, или, что то же самое, с сохранением индукции $\mathbf{B}$. Тогда согласно формуле (72.1) $F\,\delta x = -[dW_m]_{B=\text{const}}$. При смещении границы на δx вещество первого магнетика будет входить в соленоид, его объем внутри соленоида увеличится на $S\,\delta x$, где S — площадь поперечного сечения соленоида. Вследствие этого магнитная энергия системы увеличится на $Sw_1\,\delta x$. Такой же объем второго магнетика выйдет из соленоида и унесет с собой энергию $Sw_2\,\delta x$. Энергия магнитного поля вне соленоида и вблизи его концов не изменится. Таким образом, увеличение магнитной энергии системы будет $\delta W_m = S(w_1 - w_2)\,\delta x$, где w_1 и w_2 — плотности магнитной энергии по разные стороны границы раздела. Приравнивая эту величину выражению $-F\,\delta x$, получим

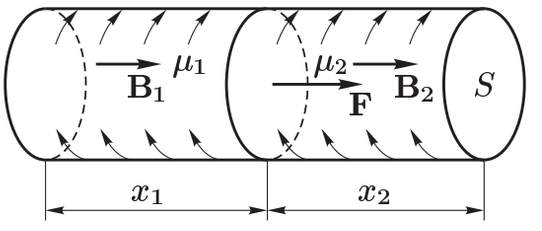


Рис. 167

$$F = S(w_2 - w_1). \tag{72.4}$$

За положительное мы приняли направление вправо, т. е. от первого магнетика ко второму. Поэтому полученное выражение для F может быть истолковано как *разность натяжений*, действующих на границу раздела со стороны обоих магнетиков. Натяжение, приходящееся на единицу площади, равно плотности магнитной энергии. Так как $B = \text{const}$, то

$$F = S\frac{B^2}{8\pi}\left(\frac{1}{\mu_2} - \frac{1}{\mu_1}\right) = \frac{S}{8\pi}\,(\mu_2 H_2^2 - \mu_1 H_1^2). \tag{72.5}$$

Сила F положительна при $\mu_1 > \mu_2$ и отрицательна при $\mu_1 < \mu_2$. В обоих случаях эта сила направлена от магнетика с большей к магнетику с меньшей магнитной проницаемостью.

Разберем теперь второй случай, когда магнитное поле параллельно границе раздела магнетиков. Здесь удобнее взять соленоид прямоугольной формы, по боковой поверхности которого циркулирует ток,

перпендикулярный к его оси (рис. 168). Пространство внутри соленоида заполнено двумя магнетиками, граничащими друг с другом вдоль плоскости, параллельной одной из боковых граней соленоида. Рассуждения будут такими же, как и в предыдущем случае. Однако теперь удобнее воспользоваться формулой (72.2), т. е. варьировать энергию W_m при постоянном значении поля $\mathbf{H}$. Так как формулы (72.1) и (72.2) отличаются знаками, то вместо выражения (72.4) мы придем к выражению

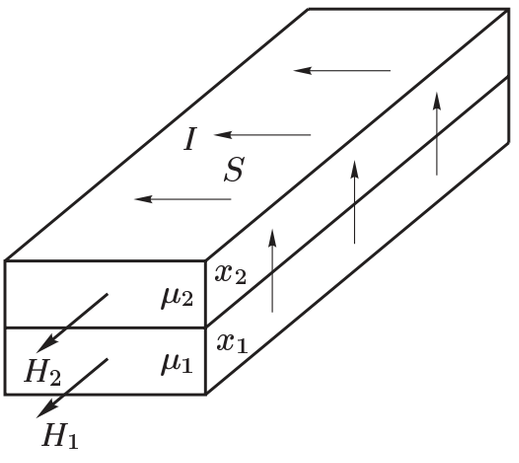


Рис. 168

$$F = S(w_1 - w_2), \qquad (72.6)$$

отличающемуся от (72.4) знаком. Оно может быть истолковано как *разность давлений*, действующих на границу раздела со стороны обоих магнетиков. Величина давления равна плотности магнитной энергии в среде. Так как поле $\mathbf{H}$ тангенциально к границе раздела, то в обоих магнетиках его величина одинакова. Поэтому можно написать

$$F = S\frac{H^2}{8\pi}(\mu_1 - \mu_2) = \frac{S}{8\pi}\left(\frac{B_1^2}{\mu_1} - \frac{B_2^2}{\mu_2}\right).$$

Сила F положительна, когда $\mu_1 > \mu_2$, и отрицательна, когда $\mu_1 < \mu_2$. Как и в предыдущем случае, она всегда направлена от магнетика с большей к магнетику с меньшей магнитной проницаемостью.

6. Когда магнитное поле и плотность среды неоднородны и являются непрерывными функциями координат, расчет сил, действующих на среду в магнитном поле, производится аналогично тому, как это было сделано в электростатике (см. § 34). Мы не будем производить этот расчет, а ограничимся приведением окончательного результата. *Механические силы, действующие в магнитном поле, сводятся к натяжению T вдоль поля и к давлению $\mathscr{P}$ в перпендикулярном направлении. Натяжение и давление, отнесенные к единице площади, на которую они действуют, численно одинаковы и равны плотности магнитной энергии в среде*:

$$T = \mathscr{P} = \frac{\mu H^2}{8\pi} = \frac{HB}{8\pi} = \frac{B^2}{8\pi\mu}. \qquad (72.7)$$

7. Поместим одно из колен U-образной трубки с раствором хлористого железа между полюсами электромагнита (рис. 169). При включении тока в обмотке электромагнита жидкость в этом колене поднимается: раствор хлористого железа, как парамагнетик, втягивается в область более сильного магнитного поля. Диамагнетик, наоборот, выталкивается из магнитного поля. Сильным диамагнетизмом обладает висмут. Кусочек висмута, внесенный в пространство между полюсами

электромагнита, выталкивается из него, если включить ток в обмотке электромагнита. Так же ведет себя пламя свечи (углекислый газ диамагнитен). Разумеется, все эти явления наблюдаются только в неоднородных полях. В однородном поле результирующая сила, действующая на внесенное в него тело, равна нулю.

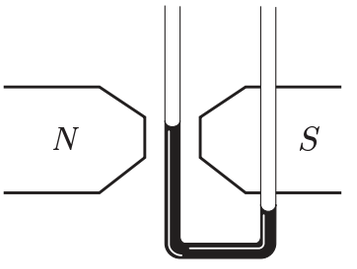


Рис. 169

Продолговатые тела, имеющие, например, форму палочек, подвешенные на нити, устанавливаются вдоль магнитного поля, если они парамагнитны, и поперек поля, если они диамагнитны. Например, палочка висмута устанавливается поперек магнитного поля. Явление наблюдается и в том случае, когда магнитное поле, в которое вносится магнетик, однородно. Это явление объясняется магнитными натяжениями и давлениями, действующими на концах магнетика. Характер явления зависит только от того, что больше — магнитная проницаемость магнетика или окружающей среды. Ампула с раствором хлористого железа в воздухе устанавливается вдоль магнитного поля. Но если ту же ампулу поместить в более сильный раствор хлористого железа, то она установится перпендикулярно к полю.

§ 73. Термодинамика магнетиков

1. Термодинамика магнетиков аналогична термодинамике диэлектриков, изложенной в § 31 нашего курса. Полученные там результаты могут быть перенесены и в термодинамику магнетиков. Надо только выражение для элементарной внешней работы $(1/4\pi)E\,dD$ заменить на $\delta A^{\text{внеш}} = (1/4\pi)H\,dB$. Таким путем в тех же предположениях, которые были введены в § 31, получаем основные уравнения термодинамики магнетиков:

$$\delta Q = dU - \frac{1}{4\pi}\,H\,dB, \tag{73.1}$$

$$dU = T\,dS + \frac{1}{4\pi}\,H\,dB, \tag{73.2}$$

$$d\Psi = -S\,dT + \frac{1}{4\pi}\,H\,dB, \tag{73.3}$$

$$d\Phi = -S\,dT - \frac{1}{4\pi}\,B\,dH, \tag{73.4}$$

$$dI = T\,dS - \frac{1}{4\pi}\,B\,dH. \tag{73.5}$$

Роль уравнения состояния играет соотношение $B = f(H, T, \tau)$. Используя его, получаем для свободной энергии магнетика

$$\Psi = \frac{1}{4\pi}\int H\,dB + \Psi_0(T, \tau), \tag{73.6}$$

где Ψ_0 — значение свободной энергии при отсутствии магнитного поля. (При интегрировании температура T и плотность магнетика τ должны оставаться постоянными.) В частности, при справедливости соотношения $B = \mu H$

$$\Psi = \frac{\mu H^2}{8\pi} + \Psi_0 = \frac{B^2}{8\pi\mu} + \Psi_0. \tag{73.7}$$

Внутренняя энергия магнетика определяется выражением

$$U = \left(\mu + T\frac{\partial\mu}{\partial T}\right)\frac{H^2}{8\pi} + U_0(T,\tau), \tag{73.8}$$

или

$$U = \left(1 + \frac{T}{\mu}\,\frac{\partial\mu}{\partial T}\right)\frac{B^2}{8\pi\mu} + U_0(T,\tau), \tag{73.9}$$

где производная $\partial\mu/\partial T$ берется при постоянной плотности τ.

2. Если квазистатически и адиабатически изменять намагниченность I, то температура магнетика, вообще говоря, будет меняться (*магнитокалорический эффект*). Изменение температуры можно рассчитать из условия постоянства энтропии S. Рассматривая последнюю как функцию T и B (τ поддерживается постоянной), имеем

$$dS = \left(\frac{\partial S}{\partial T}\right)_B dT + \left(\frac{\partial S}{\partial B}\right)_T dB = 0.$$

Введем теплоемкость единицы объема магнетика при постоянной индукции: $c_B = T(\partial S/\partial T)_B$ и воспользуемся соотношением $(\partial S/\partial T)_T = (1/4\pi)(\partial H/\partial T)_B$, которое вытекает из (73.3). Тогда

$$dT = \frac{TB}{4\pi c_B}\,\frac{d}{dT}\left(\frac{1}{\mu}\right)dB = -\frac{TB}{4\pi\mu^2 c_B}\,\frac{d\mu}{dT}\,dB,$$

или

$$dT = -\frac{TB}{\mu^2 c_B}\,\frac{d\varkappa}{dT}\,dB. \tag{73.10}$$

Если рассматривать S как функцию T и H, то таким же путем можно получить формулу

$$dT = -\frac{TH}{c_H}\,\frac{d\varkappa}{dT}\,dH, \tag{73.11}$$

где c_H — теплоемкость единицы объема магнетика при постоянной напряженности магнитного поля H (см. § 31).

3. Применим формулу (73.11) к парамагнетикам и воспользуемся законом Кюри, согласно которому магнитная восприимчивость парамагнетика обратно пропорциональна абсолютной температуре (см. § 77): $\varkappa = \text{const}/T$. Отсюда находим $d\varkappa/dT = -\varkappa/T$, и, следовательно,

$$dT = \frac{\varkappa}{c_H}\,H\,dH.$$

Из этой формулы видно, что при обратимом адиабатическом размагничивании парамагнетик охлаждается. Пренебрежем зависимостью теплоемкости c_H от магнитного поля. Тогда, если начальное магнитное поле равно H, а конечное — нулю, то для изменения температуры из последней формулы получаем

$$\Delta T = -\frac{\varkappa H^2}{2c_H}. \tag{73.12}$$

В качестве примера оценим эффект для парамагнитного газа, к которому применима классическая теория теплоемкостей. Так как мы пренебрегаем зависимостью c_H от H, то под c_H следует понимать теплоемкость при постоянном объеме. Пусть f — число степеней свободы молекулы идеального газа. Тогда по классической теории молярная теплоемкость газа при постоянном объеме будет $C_V = fR/2$, где R — газовая постоянная. Разделив на молярный объем $V = RT/\mathscr{P}$, получим $c_H = f\mathscr{P}/(2T)$, и, следовательно,

$$\Delta T = -\frac{\varkappa H^2}{f\mathscr{P}}\, T. \tag{73.13}$$

Для кислорода ($f = 5$) при нормальных условиях ($T = 293$ К, $\mathscr{P} = 10^6$ дин/см^2) $\varkappa = 0{,}16 \cdot 10^{-6}$. Полагая в формуле (73.13) $H = 3 \cdot 10^4$ Гс, получаем при этих условиях $\Delta T = -0{,}007$ К, т. е. понижение температуры — ничтожное.

Однако при приближении к абсолютному нулю температур теплоемкость очень резко стремится к нулю, и понижение температуры ΔT может стать значительным. (Вблизи самого абсолютного нуля закон Кюри неприменим.) Поэтому П. Дебай (1884–1966) и независимо от него Уильям Джиок (р. 1895) предложили применять обратимое адиабатическое размагничивание для приближения к абсолютному нулю. Этот метод стал основным методом получения сверхнизких температур. Обычно в качестве парамагнетика применяют парамагнитные соли типа квасцов, в которые входят ионы переходных элементов группы железа. Парамагнитная соль, помещенная в сильное магнитное поле, предварительно охлаждается до гелиевых температур, а затем магнитное поле снимается. Таким образом, де Гааз (1878–1960) и Вирсма достигли температуры $3 \cdot 10^{-3}$ К. Еще большее охлаждение можно получить, если вместо электронных брать «ядерные» парамагнетики, т. е. такие вещества, парамагнетизм которых обусловлен ориентацией магнитных моментов атомных ядер. Этим методом Симон с сотрудниками в 1956 г. достигли температуры 10^{-5} К.

§ 74. Ферромагнетизм

1. По своим магнитным свойствам все вещества можно разделить на *слабомагнитные* и *сильно магнитные*. К слабомагнитным веществам относятся *парамагнетики* и *диамагнетики*, к сильномагнитным — *ферромагнетики*, *антиферромагнетики* и *ферримагнетики*. Пара- и диамагнетиками называются вещества, которые в отсутствие магнитного

поля всегда не намагничены и которые характеризуются однозначной зависимостью между вектором намагничивания $\mathbf{I}$ и напряженностью (статического) магнитного поля $\mathbf{H}$. В частности, в слабых магнитных полях эта зависимость линейна: $\mathbf{I} = \varkappa\mathbf{H}$, причем для парамагнетиков $\varkappa > 0$, а для диамагнетиков $\varkappa < 0$. Магнитные свойства таких веществ с феноменологической точки зрения были рассмотрены в предшествующих параграфах. Сильный магнетизм наблюдается только у веществ в твердом состоянии, и притом далеко не у всех: необходимо (но недостаточно), чтобы в состав кристаллической решетки вещества входили атомы с недостроенными внутренними оболочками. Определение антиферромагнетизма и ферримагнетизма будет дано в § 79, а сейчас мы обратимся к феноменологической характеристике ферромагнетиков и ферромагнетизма.

2. Ферромагнетиками называются твердые тела, которые могут обладать *спонтанной намагниченностью*, т. е. намагничены уже в отсутствие магнитного поля. В этом отношении они аналогичны сегнетоэлектрикам. Типичными представителями ферромагнетиков являются переходные металлы: железо, кобальт, никель — и многие их сплавы. Ферромагнетизмом обладают некоторые элементы группы редких земель при низких температурах (гадолиний, тербий, диспрозий, гольмий, эрбий, тулий).

3. Характерной особенностью ферромагнетиков является сложная нелинейная зависимость между $\mathbf{I}$ и $\mathbf{H}$ или между $\mathbf{B}$ и $\mathbf{H}$, впервые систематически исследованная для железа А.Г. Столетовым (1839–1896). Характер этой зависимости представлен на рисунках 170 и 171. Предполагается, что в исходном состоянии тело не намагничено. По

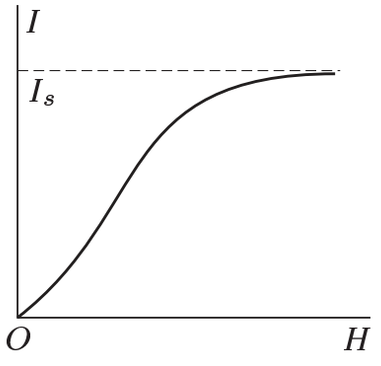


Рис. 170

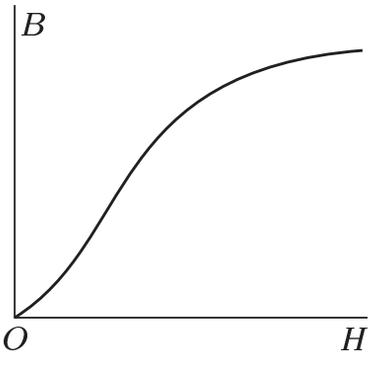


Рис. 171

мере возрастания $\mathbf{H}$ намагниченность $\mathbf{I}$ сначала быстро растет, а затем становится практически постоянной: $\mathbf{I} = \mathbf{I}_s$ (насыщение), т. е. кривая $I = I(H)$ переходит в горизонтальную прямую. Магнитная индукция $\mathbf{B} = \mathbf{H} + 4\pi\mathbf{I}$ также растет с возрастанием поля $\mathbf{H}$, а в состоянии насыщения $\mathbf{B} = \mathbf{H} + 4\pi\mathbf{I}_s = \mathbf{H} + \text{const}$, т. е. кривая $B = B(H)$ переходит в прямую, наклоненную к $\mathbf{H}$ и $\mathbf{B}$ под углом 45° (если H и B откладывать на осях координат в одинаковых масштабах). Ввиду нелинейной связи между $\mathbf{I}$ и $\mathbf{H}$, а также между $\mathbf{B}$ и $\mathbf{H}$ для ферромагнетиков нельзя ввести магнитную восприимчивость $\varkappa$ и магнитную проницаемость μ

как определенные постоянные величины. Правда, по-прежнему можно написать $\mathbf{I} = \varkappa\mathbf{H}$, $\mathbf{B} = \mu\mathbf{H}$, но тогда $\varkappa$ и μ надо рассматривать не как постоянные, а как функции напряженности поля H. Эти функции сначала возрастают с H, затем проходят через максимум, и, наконец, в сильных полях, когда достигнуто состояние насыщения, μ стремится к единице (рис. 172), а $\varkappa$ — к нулю. Значения μ в максимуме у большинства ферромагнетиков при обычных температурах составляют многие сотни и тысячи единиц, а у некоторых специально приготовленных сплавов достигают миллиона. На больших значениях μ основаны технические применения ферромагнетиков, например, в электромагнитах. Следует, однако, заметить, что благодаря насыщению намагничивания использование ферромагнитных сердечников для получения сильных магнитных полей (превышающих 10^4 Гс) не имеет смысла. Здесь сердечники не только бесполезны, но и вредны, так как вносят дополнительные потери энергии. Сильные и сверхсильные магнитные поля получают только с помощью катушек с током без ферромагнитных сердечников.

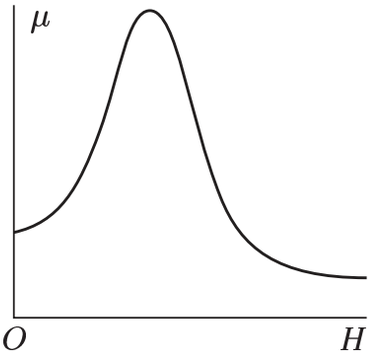


Рис. 172

Монокристаллические образцы в отношении магнитных свойств анизотропны. В каждом монокристалле существует одно или несколько направлений, вдоль которых магнитная восприимчивость особенно велика (направление легкого намагничивания). Существуют также направления трудного намагничивания, вдоль которых магнитная восприимчивость μ минимальна. Однако если тело состоит из мелких поликристалликов, беспорядочно ориентированных в пространстве, то оно в магнитном отношении макроскопически изотропно. К таким телам и относятся кривые на рис. 170 и 171.

4. Вторая характерная особенность ферромагнетиков состоит в том, что для них зависимость $\mathbf{B}$ от $\mathbf{H}$ или $\mathbf{I}$ от $\mathbf{H}$ не однозначна, а определяется предшествующей историей нагмагничивания ферромагнитного образца. Это явление называется *магнитным гистерезисом*. Возьмем ферромагнитный образец в ненамагниченном состоянии и будем намагничивать его, увеличивая магнитное поле от нуля до какого-то значения H_1. Зависимость $B = B(H)$ изобразится кривой OA_1 (рис. 173). Затем будем уменьшать H от $+H_1$ до $-H_1$. Кривая намагничивания, как показывает опыт, не пойдет по прежнему пути A_1O, а пройдет выше, по пути $A_1C_1D_1$. Если дальше изменять поле от $-H_1$ до $+H_1$, то кривая намагничивания пройдет ниже и вернется в прежнюю точку A_1. Получится замкнутая кривая $A_1C_1D_1F_1A_1$, называемая *петлей гистерезиса*. Если первоначальную кривую OA_1 довести до точки A, где намагничивание уже достигло насыщения, то описанным путем получится *наибольшая*, или *предельная*, *петля гистерезиса* $ACDFA$. Из нее видно, что при $H = 0$ индукция B не обращается в нуль, а изображается отрезком OC. Ему соответствует остаточное намагничивание

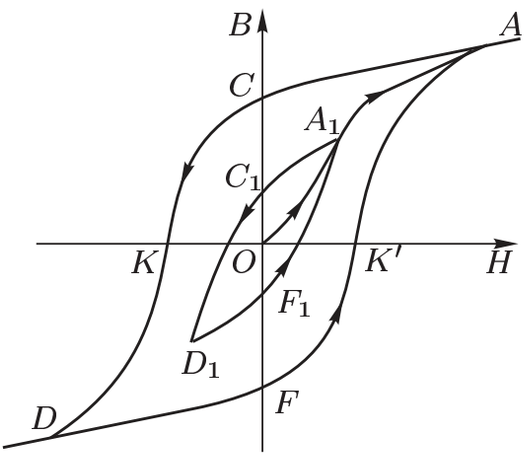


Рис. 173

$I = B/(4\pi) = OC/(4\pi)$. С наличием такого остаточного намагничивания связано существование *постоянных магнитов*. Для того чтобы размагнитить образец, надо довести кривую намагничивания до точки K или K'. Этим точкам соответствует магнитное поле $H_K = |OK|$. Оно называется *задерживающей* или *коэрцитивной силой* ферромагнетика. Значения остаточного намагничивания и коэрцитивной силы для разных ферромагнетиков меняются в широких пределах. Для мягкого железа петля гистерезиса узкая (коэрцитивная сила мала), для стали и всех материалов, идущих на изготовление постоянных магнитов, — широкая (коэрцитивная сила велика). Совершенно такой же характер имеет петля гистерезиса, когда по вертикальной оси откладывается не индукция B, а намагничивание I.

С описанными особенностями кривых намагничивания связан удобный практический прием размагничивания ферромагнетика (например, намагниченных часов). Намагниченный ферромагнетик помещают в катушку, по которой пропускают переменный ток. Плавно уменьшая амплитуду тока, подвергают ферромагнетик циклическим перемагничиваниям, в которых описываются все суживающиеся и суживающиеся петли гистерезиса, пока не будет достигнута точка O, где намагничивание равно нулю.

В демонстрационных опытах петлю гистерезиса легко получить на экране осциллографа. Для этого собирается схема, изображенная на рис. 174, питаемая переменным током. Катушка A, в которой находится исследуемый образец из длинной проволоки или пучок проволок, помещается в переменное магнитное поле, намагничивающее образец. Напряженность магнитного поля H пропорциональна току I в обмотке электромагнита. Поэтому напряжение, подаваемое с сопротивления R на горизонтально отклоняющие пластины осциллографа, пропорционально H. С другой стороны, напряжение, возникающее на концах катушки A, пропорционально производной dB/dt. Это напряжение подается на вход интегрирующей ячейки CR (действие такой ячейки разъяснено в § 122). На конденсаторе последней возникает напряжение, пропорциональное $\int \frac{dB}{dt}\,dt$, т. е. B. Оно подается на вертикально отклоняющие пластины осциллографа. При включении переменного тока на экране осциллографа наблюдается петля гистерезиса. Для исключения влияния краев образца лучше брать замкнутый образец в виде тора и помещать его в тороидальную катушку.

5. Благодаря гистерезису намагничивание и перемагничивание ферромагнетиков сопровождается выделением тепла, называемого *теплотой гистерезиса*. Для бесконечно малого квазистатического

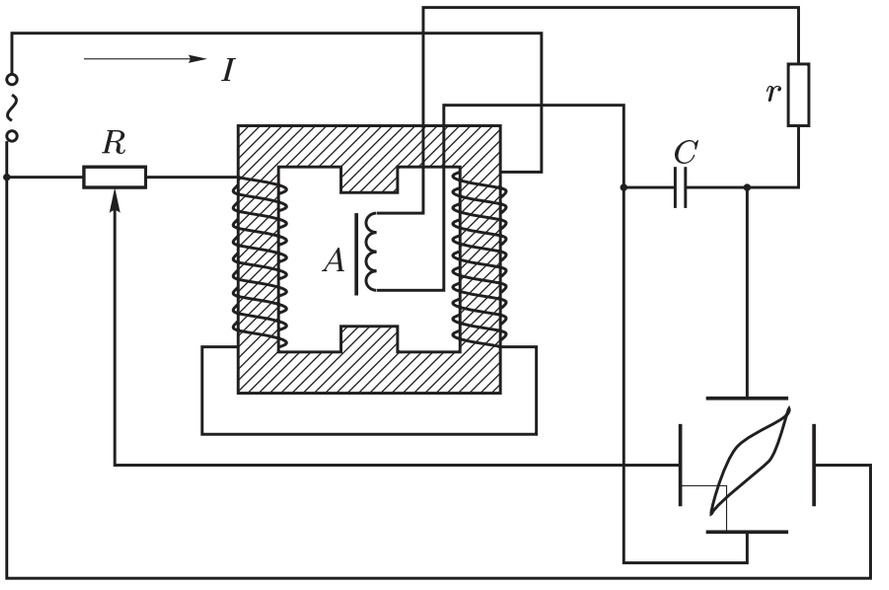


Рис. 174

процесса теплота гистерезиса δQ выражается формулой (73.1). Проинтегрируем ее по одному циклу намагничивания, т. е. вдоль замкнутой петли гистерезиса. Тогда интеграл от dU обратится в нуль, так как в круговом процессе ферромагнетик возвращается в исходное состояние, а потому внутренняя энергия U принимает первоначальное значение. В результате получится

$$Q = \frac{1}{4\pi}\int H\,dB. \tag{74.1}$$

Отсюда видно, что теплота гистерезиса, выделяющаяся в одном цикле намагничивания, с точностью до множителя $1/(4\pi)$ численно равна площади петли гистерезиса.

Подвесим небольшой стальной цилиндр внутри проволочной катушки, питаемой городским переменным током. Цилиндр, испытывая 50 циклов перемагничивания в секунду, сильно нагреется уже через одну-две минуты. Чтобы доказать, что это нагревание обусловлено гистерезисом, а не токами Фуко, подвесим внутри той же катушки медный цилиндр тех же размеров, что и стальной. Электрическая проводимость меди больше, чем у железа, а потому токи Фуко в ней будут сильнее. Между тем при значительном нагревании стального цилиндра медный остается холодным.

6. Третья характерная особенность ферромагнетиков, которую мы здесь отметим, состоит в том, что для всякого ферромагнетика существует определенная температура $T = T_{\text{К}}$, называемая *температурой*

или *точкой Кюри*, при переходе через которую вещество ферромагнетика претерпевает фазовый переход (второго рода). Вещество является ферромагнетиком только ниже точки Кюри. Выше точки Кюри оно становится парамагнетиком, причем магнитная восприимчивость в окрестности точки Кюри подчиняется закону Кюри–Вейсса

$$\varkappa = \frac{C}{T - T_{\text{К}}}, \tag{74.2}$$

где C — постоянная, зависящая от рода вещества.

Для никеля температура Кюри равна 360 °С (633 К). Возьмем никелевую фольгу, свернутую в трубочку. Подвесим ее вблизи сильного постоянного магнита. Фольга притянется к магниту. Поднесем к фольге зажженную газовую горелку. Когда фольга нагреется выше точки Кюри, она перестанет притягиваться, отойдет от магнита и выйдет при этом из пламени горелки. Через короткое время фольга остынет и снова притянется к магниту, затем отойдет от него и т. д. Такие маятникообразные колебания никелевой фольги будут продолжаться все время, пока горит горелка.

Дальнейшие особенности ферромагнетизма будут рассмотрены в связи с его теорией (см. § 79).

§ 75. Магнитные свойства атомов

1. Переходя к объяснению магнитных свойств материальных сред с атомистической точки зрения, заметим прежде всего, что в последовательно классической теории магнетизм должен отсутствовать. Бор в 1911 г. и независимо от него Ван-Лёвен в 1920 г., пользуясь методами классической статистики, строго доказали следующую теорему. *В состоянии термодинамического равновесия система электрически заряженных частиц* (*электронов*, *атомных ядер и пр.*), *помещенная в постоянное магнитное поле*, *не могла бы обладать магнитным моментом*, *если бы она строго подчинялась законам классической физики.* Такая система может быть намагничена только в неравновесном состоянии. Если она перейдет в равновесное состояние, то намагничивание исчезнет. Причина этого, грубо говоря, заключается в том, что постоянное магнитное поле, действуя на заряженную частицу с силой, перпендикулярной к скорости, не может изменить кинетическую энергию частицы. Для объяснения магнетизма вещества требуется привлечение квантовых представлений.

Между тем парамагнетизм и диамагнетизм были объяснены, и притом довольно успешно, Ланжевеном (1872–1946) в 1905 г. без использования квантовых представлений. Причина этого заключается в том, что в классических теориях намагничивания молчаливо вводились представления сугубо квантового характера. Именно, предполагалось, что из электрически заряженных частиц можно построить устойчивые образования — атомы и молекулы. От последовательно классической теории надо требовать объяснения не только намагничивания, но и существования самих атомов, что удалось сделать только квантовой

механике. Поскольку последняя в нашем курсе еще не излагалась, при объяснении намагничивания мы будем пользоваться *полуклассическими представлениями.* Несмотря на свою непоследовательность и недостаточность, полуклассическая теория позволяет в основном уяснить природу намагничивания.

2. Начнем с краткого рассмотрения магнитных свойств атомов. Более подробно этот вопрос будет разобран в т. V нашего курса — в атомной физике. В простейшей боровской модели атома водорода электрон вращается вокруг ядра по окружности. Заряд электрона будем обозначать через $-e$. Вращающийся по окружности электрон в среднем возбуждает магнитное поле как ток $I = -e/T$, где $T = 2\pi r/v$ — период обращения электрона. Поэтому вращающемуся электрону присущ не только орбитальный момент импульса (или механический момент) $L = mrv$, но и магнитный момент $\mathfrak{M} = IS/c = -erv/(2c)$. Отношение этих величин называется *гиромагнитным отношением* и для нашей модели атома равно

$$\Gamma = \frac{\mathfrak{M}}{L} = -\frac{e}{2mc}. \tag{75.1}$$

Тот же результат справедлив для движений электрона по эллиптическим орбитам. Он верен и для многоэлектронных атомов, поскольку для всех электронов отношение e/m одно и то же.

Согласно теории Бора момент импульса атома *квантуется*, т. е. может принимать не непрерывный, а только *дискретный* ряд значений. Допустимыми являются значения $L = n\hbar$, где n — целое число, которое может принимать значения 1, 2, 3, ..., а $\hbar = h/(2\pi) = 1{,}05 \cdot 10^{-27}$ эрг · с — постоянная Планка (1858–1947), деленная на 2π. (Эта величина также называется постоянной Планка и более удобна в теоретических вопросах.) Вместе с механическим моментом магнитный момент также квантуется в соответствии с формулой

$$\mathfrak{M} = -\frac{e\hbar}{2mc}\, n. \tag{75.2}$$

Таким образом, наименьшее значение магнитного момента атома равно

$$\mathfrak{M}_{\text{B}} = \frac{e\hbar}{2mc} = 9{,}28 \cdot 10^{-21} \text{ эрг/Гс}.$$

Эта величина играет роль атома магнитного момента и называется *магнетоном Бора.*

3. Квантовая механика оставила представление о движении электронов по классическим орбитам и уточнила правила квантования теории Бора. Вместо движения самих электронов квантовая механика ввела представление о движении некоторой величины, имеющей смысл *плотности вероятности* нахождения электрона в пространстве. Классическим, однако отнюдь не адекватным аналогом такого представления, может служить облако, в котором масса и соответствующий ей электрический заряд распределены в пространстве непрерывно с определенной плотностью. Существует дискретный ряд так называемых *стационарных состояний*, в которых эти величины не меняются во

времени. К таким стационарным состояниям и относятся квантованные значения физических величин. Классическая формула (75.1) для гиромагнитного отношения справедлива и в квантовой механике, как это непосредственно очевидно, если воспользоваться классическим аналогом квантовомеханического представления, о котором только что говорилось.

В теории Бора электрон, обращающийся по орбите, становится эквивалентным току только после усреднения его положения вдоль орбиты. В квантовой механике, напротив, орбит нет и электрон в атоме, если его уподобить классической модели, вполне аналогичен току, непрерывно распределенному вокруг ядра атома.

В боровской модели невозможны состояния, в которых орбитальные механический и магнитный моменты атомов равны нулю, так как в этом случае электрон должен был бы совершать радиальное движение, в котором он непременно столкнулся бы с ядром. Напротив, в квантовой механике возможны состояния со *сферически симметричным* распределением вероятности нахождения электрона вокруг ядра. В таких состояниях орбитальные механический и магнитный моменты электрона в атоме строго равны нулю.

Наконец в квантовой механике формула $L = n\hbar$ определяет не полный момент количества движения электрона в атоме, а только *проекцию* этого вектора на избранное направление — направление магнитного поля, в которое помещен атом. Остальные две проекции не имеют определенных значений, что, конечно, невозможно представить в рамках классических моделей.

4. Помимо орбитального электрон обладает еще *собственным*, или *спиновым*, *моментом* количества движения (короче, *спином*). В стационарных состояниях проекция спина на избранное направление может принимать только два значения: $+\hbar/2$ и $-\hbar/2$. Спину соответствует магнитный момент, проекция которого на избранное направление равна магнетону Бора. Таким образом, со спином электрона связано гиромагнитное отношение $\Gamma = -e/(mc)$, которое вдвое больше орбитального. Механический и магнитный моменты всякого атома, в том числе и многоэлектронного, векторно складываются из орбитальных и спиновых моментов. Могут существовать состояния, в которых механические и магнитные моменты скомпенсированы, т. е. полный момент атома равен нулю.

§ 76. Объяснение диамагнетизма

1. Диамагнетизм наблюдается у таких веществ, атомы которых в отсутствие магнитного поля не обладают магнитными моментами. Если нет магнитного поля, то на электрон в атоме действуют силы только со стороны атомного ядра и прочих электронов. В постоянном магнитном поле $\mathbf{B}$ к этим силам добавится сила $-(e/c)[\mathbf{vB}]$, где $\mathbf{v}$ — скорость электрона. Определим, как изменится движение атома в стационарном состоянии при наличии этой силы. С этой целью рассмотрим движение относительно системы отсчета, равномерно вращающейся

вокруг направления магнитного поля с угловой скоростью $\boldsymbol{\Omega}$. Величину $\boldsymbol{\Omega}$ определим несколько ниже. Сейчас же будем предполагать, что она мала по сравнению с угловой скоростью $\boldsymbol{\omega}$ собственного вращения электрона вокруг атомного ядра. При этом условии можно пренебречь всеми членами порядка $(\Omega/\omega)^2$, т. е. членами, квадратичными по Ω. Во вращающейся системе отсчета к прежним силам добавятся две силы инерции: сила Кориолиса $2m[\mathbf{v}_{\text{отн}}\boldsymbol{\Omega}]$ и центробежная сила. Центробежной силой мы пренебрежем, как величиной, пропорциональной Ω^2, а в выражении кориолисовой силы относительную скорость электрона $\mathbf{v}_{\text{отн}}$ заменим абсолютной скоростью $\mathbf{v} = \mathbf{v}_{\text{отн}} + [\boldsymbol{\Omega}\mathbf{r}]$. Такая замена также допустима, так как она меняет силу Кориолиса на величину, пропорциональную Ω^2. В этом приближении кориолисова сила представится в виде $2m[\mathbf{v}\boldsymbol{\Omega}]$. Подберем теперь величину $\boldsymbol{\Omega}$ так, чтобы выполнялось условие $2m[\mathbf{v}\boldsymbol{\Omega}] - (e/c)[\mathbf{v}\mathbf{B}] = 0$, т. е. положим

$$\boldsymbol{\Omega} = \frac{e}{2mc}\,\mathbf{B}. \tag{76.1}$$

Тогда во вращающейся системе отсчета к прежним силам, действующим на электрон, не добавится никаких новых сил. Поэтому во вращающейся системе отсчета атом придет в то же стационарное состояние, в котором он находился в неподвижной системе в отсутствие магнитного поля. Если движение по-прежнему отнести к неподвижной системе отсчета, то получается следующий результат. *При наличии внешнего постоянного магнитного поля внутреннее движение электронов атома не изменяется, но атом в целом получает дополнительное вращение с угловой скоростью* (76.1). Этот результат называется *теоремой Лармора* (1857–1942), а величина $\boldsymbol{\Omega}$ — *ларморовской частотой.*

Остается проверить, выполняется ли условие $|\Omega| \ll |\omega|$, использованное при доказательстве. Очевидно, это условие можно записать так:

$$B \ll \left|\frac{2mc\omega}{e}\right|. \tag{76.2}$$

Подставляя сюда численные значения $|e| = 4{,}8 \cdot 10^{-10}$ СГСЭ-ед., $m = 9{,}11 \cdot 10^{-28}$ г, $\omega \sim 10^{15}$ с$^{-1}$, получим $B \ll 10^8$ Гс. Рекордные магнитные поля, полученные до настоящего времени, не превосходят 10^7 Гс. Поэтому условие (76.2) хорошо выполняется.

2. Как видно из формулы (76.1), угловая скорость ларморовского вращения электронов совпадает по направлению с вектором $\mathbf{B}$. Так как заряд электрона отрицательный, то магнитный момент, связанный с этим вращением, направлен против поля $\mathbf{B}$. В результате создается намагничивание среды $\mathbf{I}$, направленное также против поля $\mathbf{B}$. Это и есть диамагнетизм.

Рассчитаем теперь магнитную восприимчивость вещества. Электрон, вращающийся по окружности радиуса r с ларморовской частотой $\boldsymbol{\Omega}$, обладает моментом импульса $mr^2\boldsymbol{\Omega}$ и, следовательно, магнитным моментом $-er^2\boldsymbol{\Omega}/2c = -e^2r^2\mathbf{H}/(4mc^2)$. В последнем выражении вектор $\mathbf{B}$ заменен на $\mathbf{H}$, так как в диамагнетиках различие между этими векторами пренебрежимо мало. Если ось Z перпендикулярна

к плоскости круговой орбиты электрона, то $r^2 = x^2 + y^2$. Чтобы найти магнитный момент атома, надо просуммировать магнитные моменты всех его электронов. Электроны в атомах диамагнетика распределены сферически симметрично. В этом случае $\overline{x^2} = \overline{y^2} = \overline{z^2} = (1/3)\overline{R^2}$, где R — расстояние электрона до ядра. Если атом содержит Z электронов, то его средний магнитный момент в магнитном поле будет

$$\mathfrak{M} = -\frac{Ze^2}{4mc^2}(\overline{x^2} + \overline{y^2})H = -\frac{Ze^2}{6mc^2}\overline{R^2}H,$$

а вектор намагничивания среды

$$\mathbf{I} = -\frac{NZe^2}{6mc^2}\overline{R^2}\,\mathbf{H}, \tag{76.3}$$

где N — число атомов в единице объема. Затем находим

$$\varkappa = -\frac{NZe^2}{6mc^2}\overline{R^2}, \tag{76.4}$$

$$\mu = 1 - \frac{4\pi NZe^2}{6mc^2}\overline{R^2}. \tag{76.5}$$

Энергия теплового движения слишком мала, чтобы изменить внутреннее (квантованное) состояние атома. Поэтому для диамагнетиков величины $\varkappa$ и μ не должны зависеть от температуры. Этот вывод теории находится в согласии с опытом.

Чтобы подтвердить, что теория находится на верном пути, оценим размеры атома, пользуясь значениями восприимчивости $\varkappa$. Величина $\overline{R^2}$ должна вычисляться с помощью квантовой механики.

Однако и без вычислений ясно, что квадратный корень из нее по порядку величины определяет размеры атома. Благородные газы, ввиду сферической симметрии электронных оболочек их атомов, диамагнитны. Диамагнитные свойства удобно характеризовать *магнитной восприимчивостью на моль вещества* $\varkappa_A$. Она связана с $\varkappa$ соотношением $\varkappa_A = V\varkappa$, где V — объем одного моля. Величину $\varkappa_A$ можно вычислять по формуле (76.4), если под N понимать число Авогадро (1776–1856). В случае гелия опыт дает $\varkappa_A = -2{,}2 \cdot 10^{-6}$; в случае аргона $\varkappa_A = -2{,}5 \cdot 10^{-4}$. Для гелия $Z = 2$, для аргона $Z = 18$. Подставляя эти данные в формулу (76.4), получим: He, $\sqrt{\overline{R^2}} = 0{,}63 \cdot 10^{-8}$ см; Ar, $\sqrt{\overline{R^2}} = 0{,}67 \cdot 10^{-8}$ см. Эти результаты удовлетворительно согласуются с размерами атомов, найденными другими способами.

3. Необходимо еще выяснить, какие силы сообщают атому ларморовское вращение. Этого не могут сделать магнитные силы, так как они перпендикулярны к скорости электрона и работы не производят. А с ларморовским вращением связана дополнительная кинетическая энергия атома. *Магнитные силы могут только поддерживать, но не создавать ларморовское вращение. Последнее возникает во время включения магнитного поля.* Переменное магнитное поле возбуждает вихревое электрическое поле. Оно-то и сообщает атому ларморовское

вращение. Для пояснения допустим, что электрон вращается по окружности радиуса r, плоскость которой перпендикулярна к (однородному) магнитному полю $\mathbf{B}$. Пусть магнитное поле включается *адиабатически*, т. е. настолько медленно, что за время одного оборота электрона по окружности поле почти остается постоянным. Ввиду симметрии вихревое электрическое поле $\mathbf{E}$ будет направлено по касательной к окружности. На основании закона электромагнитной индукции $2\pi rE = -(1/c)\,d\Phi/dt$, где Φ — магнитный поток, пронизывающий площадь, ограниченную той же окружностью. Отсюда и найдется поле $\mathbf{E}$. Момент сил, действующих на электрон, $M = -reE = (e/2\pi c)\,d\Phi/dt$. На основании уравнения моментов

$$mr^2\frac{d\Omega}{dt} = M = \frac{e}{2\pi c}\,\frac{d\Phi}{dt}.$$

В начальный момент $B = \Omega = 0$. Поэтому, интегрируя предыдущее уравнение, получим

$$mr^2\Omega = \frac{e}{2\pi c}\,\Phi,$$

откуда $\Omega = eB/(2mc)$. Если изменение магнитного поля прекратится, то прекратится и дальнейшее изменение угловой скорости вращения атома. Последний будет продолжать вращаться с постоянной угловой скоростью, определяемой формулой Лармора (76.1).

Из изложенного видно, что *ларморовское вращение есть одно из проявлений электромагнитной индукции.* То обстоятельство, что электромагнитная индукция должна приводить именно к диамагнетизму, а не к парамагнетизму, проще всего понять, руководствуясь *принципом Ленца.* Действительно, в соответствии с этим принципом магнитное поле $\mathbf{B}_{\text{инд}}$, возбуждаемое ларморовским вращением электронов, должно иметь такое направление, чтобы препятствовать всяким изменениям внешнего приложенного поля $\mathbf{B}$. Поэтому поле $\mathbf{B}_{\text{инд}}$, а с ним и вектор намагничивания среды $\mathbf{I}$ должны иметь направление, противоположное направлению внешнего поля $\mathbf{B}$. Явление электромагнитной индукции имеет место во всех средах. Поэтому и обусловленный им диамагнетизм есть *универсальное явление*, которое должно проявляться во всех средах. Однако в тех случаях, когда атомы обладают собственными магнитными моментами, диамагнитный эффект перекрывается значительно более сильным парамагнитным эффектом.

§ 77. Объяснение парамагнетизма

1. Парамагнетизм наблюдается у тех веществ, атомы которых обладают магнитными моментами уже в отсутствие внешнего магнитного поля. Пока нет магнитного поля, атомы совершают беспорядочное тепловое движение, а их магнитные моменты ориентированы в пространстве также беспорядочно. В этом случае тело не намагничено. В магнитном поле магнитные моменты атомов ориентируются преимущественно в направлении поля. Появляется намагничивание и обусловленный им парамагнетизм.

Излагаемая ниже теория парамагнетизма относится к парамагнитным газам, взаимодействие между атомами которых слабое. Качественно результаты этой теории применимы также к твердым и жидким парамагнетикам, электронные оболочки атомов или ионов которых могут более или менее свободно вращаться вокруг атомных ядер. Это имеет место, например, тогда, когда электронные оболочки обладают сферической симметрией. Таковы электронные оболочки атомов благородных газов или ионов с таким же числом электронов, как у благородных газов.

2. Поместим изолированный атом в постоянное магнитное поле $\mathbf{B}$. Отвлечемся от наличия спинов, предполагая, что все спины электронной оболочки скомпенсированы. Пусть $\mathbf{v}$ — скорость какого-либо электрона атома до внесения в магнитное поле. Тогда, согласно теореме Лармора, в магнитном поле скорость того же электрона будет $\mathbf{v}+[\boldsymbol{\Omega}\mathbf{r}]$, а его кинетическая энергия $(1/2)m(\mathbf{v}+[\boldsymbol{\Omega}\mathbf{r}])^2$, где $\boldsymbol{\Omega}$ — ларморовская частота. Если пренебречь квадратами величины $\boldsymbol{\Omega}$, то для приращения кинетической энергии электрона можно написать $m(\mathbf{v}[\boldsymbol{\Omega}\mathbf{r}])$ или $m([\mathbf{r}\mathbf{v}]\boldsymbol{\Omega})$. Просуммировав по всем электронам оболочки, находим приращение $\mathscr{E}$ кинетической энергии атома в магнитном поле: $\mathscr{E}=(\mathbf{L}\,\boldsymbol{\Omega})$, где $\mathbf{L}$ — момент импульса электронной оболочки. С помощью формул (75.1) и (76.1) это выражение приводится к виду

$$\mathscr{E}=-(\mathfrak{M}\mathbf{B}). \tag{77.1}$$

Такое же изменение энергии (но уже потенциальной) мы получили бы для магнитного диполя с магнитным моментом $\mathfrak{M}$ при внесении его в магнитное поле. Этого и следовало ожидать, так как, согласно теореме Ампера, во внешнем магнитном поле элементарный виток с током (например, вращающийся по орбите электрон) испытывает те же силы, что и магнитный диполь. Поэтому формула (77.1) для приращения энергии справедлива и в том случае, когда магнитный момент $\mathfrak{M}$ обусловлен спинами, а не орбитальным движением электронов. Вообще, формула (77.1) верна независимо от происхождения магнитного момента атома $\mathfrak{M}$, как это можно подтвердить независимым расчетом (см. п. 8).

3. На основании изложенного ясно, что классическая теория парамагнетизма газов по существу не должна отличаться от соответствующей теории Дебая поляризации газообразных диэлектриков с полярными молекулами, которая была изложена в § 36. Теория парамагнетизма была создана Ланжевеном раньше теории диэлектриков Дебая. Последняя просто копирует теорию Ланжевена. Как видно из формулы (77.1), энергия атома больше, если его магнитный момент ориентирован против магнитного поля, и меньше, если он ориентирован по полю. Поэтому в состоянии статистического равновесия больше магнитных моментов будет ориентировано по полю, чем в противоположном направлении, т. е. будет наблюдаться парамагнетизм. Если выполнено условие

$$\frac{\mathfrak{M}B}{kT}\ll 1, \tag{77.2}$$

то, по аналогии с формулой (36.5), в классической теории можно написать

$$\mathbf{I} = \frac{n\mathfrak{M}^2}{3kT}\,\mathbf{B}. \tag{77.3}$$

Отсюда, используя соотношения $\mathbf{I} = \varkappa\mathbf{H}$, $\mathbf{B} = (1 + 4\pi\varkappa)\mathbf{H}$, получим

$$\frac{\varkappa}{1 + 4\pi\varkappa} = \frac{n\mathfrak{M}^2}{3kT}. \tag{77.4}$$

Этой формулой и определяется магнитная восприимчивость парамагнетика. Впрочем, магнитная восприимчивость неферромагнитных тел настолько мала, что в предыдущей формуле, как это обычно делается, величиной $4\pi\varkappa$ можно пренебречь по сравнению с единицей [1]. Тогда

$$\varkappa = \frac{n\mathfrak{M}^2}{3kT}. \tag{77.5}$$

4. Обобщим теперь формулы (77.3) и (77.5) на случай сильных магнитных полей или низких температур, когда условие (77.2) не соблюдается. Это было сделано также Ланжевеном. Однако в классической теории Ланжевена не учитывается квантование магнитных моментов атомов, что существенно в области низких температур. Для учета квантования надо считать магнитный момент атома $\mathfrak{M}$ кратным магнетону Бора $\mathfrak{M}_\mathrm{B}$ и принять во внимание все ориентации его, допускаемые правилами квантования. Примем, что магнитный момент атома $\mathfrak{M}$ спиновый и равен одному магнетону Бора. Такое предположение не затрагивает ничего существенного, обусловленного квантованием, но сильно упрощает расчет, так как в этом случае в магнитном поле возможны только две ориентации атома: параллельная и антипараллельная. В параллельной ориентации проекция магнитного момента на направление магнитного поля равна $+\mathfrak{M}_\mathrm{B}$, а антипараллельной $-\mathfrak{M}_\mathrm{B}$. (В дальнейшем значок B будем опускать.) Этим ориентациям соответствуют энергии $-\mathfrak{M}B$ и $+\mathfrak{M}B$. Согласно распределению Больцмана числа атомов (в единице объема) с параллельной и антипараллельной ориентациями будут равны соответственно

$$n_1 = Ce^x, \quad n_2 = Ce^{-x},$$

где введено обозначение

$$x = \frac{\mathfrak{M}B}{kT}. \tag{77.6}$$

Нормировочная постоянная C определится из условия $n_1 + n_2 = n$, где n — полное число атомов в единице объема. Это дает $C(e^x + e^{-x}) = n$.

[1] Было бы непоследовательно учитывать различие между $\mathbf{B}$ и $\mathbf{H}$ и пренебрегать, как мы всюду делали в этом параграфе, различием между средним макроскопическим полем $\mathbf{B}$ и полем $\mathbf{B}'$, действующим на атом парамагнетика, между квадратом среднего поля $\mathbf{B}$ и средним квадратом микроскопического поля и т. п. Все эти различия одного и того же порядка.

Для намагничивания получаем $I = (n_1 - n_2)\mathfrak{M}$, откуда

$$I = n\mathfrak{M}\frac{e^x - e^{-x}}{e^x + e^{-x}} = n\mathfrak{M}\,\mathrm{th}\,x. \tag{77.7}$$

5. Прежде чем исследовать полученную формулу, заметим, что в случае, когда магнитный момент атома больше одного магнетона Бора, расчет проводится по той же схеме. Увеличивается только число возможных ориентаций и значения проекций магнитных моментов атомов на направление магнитного поля. Во всех случаях результат записывается в виде

$$I = n\mathfrak{M}\,L(x), \tag{77.8}$$

где $L(x)$ называется *функцией Ланжевена.* В разобранном нами случае функция Ланжевена равна $\mathrm{th}\,x$ и обозначается через $L_{1/2}(x)$, поскольку механический момент атома (в единицах $\hbar$) равен $1/2$. В других случаях вид функции $L(x)$ изменяется, но ее общий характер сохраняется. В классическом пределе, когда квантования нет, т. е. допустимы все ориентации магнитных моментов, функцию Ланжевена обозначают через $L_\infty(x)$. Она равна

$$L_\infty(x) = \mathrm{cth}\,x - \frac{1}{x} \tag{77.9}$$

(см. задачу к этому параграфу). Этот результат был получен самим Ланжевеном.

6. Графики функций Ланжевена $L_{1/2}(x)$ и $L_\infty(x)$ приведены на рис. 175. При малых x

$$L_{1/2}(x) = x - \frac{1}{3}x^3 + \ldots, \quad L_\infty(x) = \frac{1}{3}\,x + \ldots$$

В соответствии с этим в слабых полях ($x \ll 1$) $\mathbf{I}$ зависит от $\mathbf{B}$ линейно. По формуле (77.7) получается

$$\mathbf{I} = \frac{n\mathfrak{M}^2}{kT}\,\mathbf{B}, \tag{77.10}$$

а по формуле (77.9) — классический результат (77.3), т. е. величина втрое меньшая. Магнитные моменты $\mathfrak{M}$ определяются *внутренним строением атома.* Тепловое движение недостаточно интенсивно, чтобы изменить их. Поэтому, как видно из формулы (77.10) или (77.3), *магнитная восприимчивость парамагнитных газов должна меняться обратно пропорционально абсолютной температуре* T. Такая закономерность была обнаружена экспериментально П. Кюри еще до разработки соответствующей теории и носит название *закона Кюри.* Закон Кюри хорошо согласуется с опытом для газообразных парамагнетиков, а также для ряда твердых парамагнетиков (например, солей редких земель). С другой стороны, для многих жидких и твердых парамагнетиков изложенная элементарная теория недостаточна, поскольку она допускает свободное вращение электронных оболочек атомов вокруг направления магнитного поля. Впрочем, и в тех случаях, когда это

допущение справедливо, в сильных полях ($x \gg 1$) и при очень низких температурах (порядка нескольких Кельвинов) должны наблюдаться отклонения не только от закона Кюри, но и от самой пропорциональности между величинами I и B.

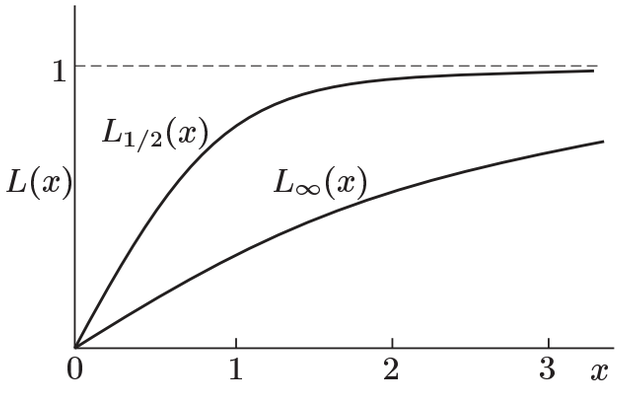


Рис. 175

7. В сильных полях ($x \gg 1$) обе функции $L_{1/2}(x)$ и $L_\infty(x)$ асимптотически стремятся к единице. Этому соответствует *насыщение намагничивания* $I = n\mathfrak{M}$, когда все магнитные моменты устанавливаются параллельно магнитному полю. Однако классическая функция $L_\infty(x)$, как и следовало ожидать, не удовлетворяет тепловой теореме Нернста (1864–1941) (см. т. II, § 84), тогда как функция $L_{1/2}(x)$ согласуется с ней. В самом деле, в отсутствие квантования справедлива классическая теорема о равномерном распределении кинетической энергии по степеням свободы. Согласно этой теореме вращательные степени свободы атома должны быть возбуждены — на каждую из них придется кинетическая энергия $(1/2)kT$ и теплоемкость $(1/2)k$, не зависящая от температуры. При абсолютном нуле получилось бы конечное значение теплоемкости, что противоречит теореме Нернста. Другое противоречие с теоремой Нернста заключается в следующем. Из формул (73.3) и (73.4) следует

$$\left(\frac{\partial S}{\partial B}\right)_T = -\frac{1}{4\pi}\left(\frac{\partial H}{\partial T}\right)_B, \quad \left(\frac{\partial S}{\partial H}\right)_T = \frac{1}{4\pi}\left(\frac{\partial B}{\partial T}\right)_H.$$

Согласно теореме Нернста все процессы при абсолютном нуле температур не сопровождаются изменениями энтропии. Поэтому левые, а с ними и правые части полученных соотношений при $T = 0$ должны обращаться в нуль:

$$\left(\frac{\partial H}{\partial T}\right)_B = \left(\frac{\partial B}{\partial T}\right)_H = 0.$$

Если воспользоваться формулой $\mathbf{B} = \mathbf{H} + 4\pi\mathbf{I}$, то полученные результаты можно записать в виде

$$\left(\frac{\partial I}{\partial T}\right)_B = \left(\frac{\partial I}{\partial T}\right)_H = 0. \tag{77.11}$$

Но классическая формула (77.9) не согласуется с этими соотношениями. Действительно, при $x \gg 1$ справедлива асимптотика $L_\infty(x) = 1 - 1/x$ и, следовательно,

$$I = n\mathfrak{M}\left(1 - \frac{1}{x}\right) = n\mathfrak{M} - \frac{nkT}{B}.$$

Из нее следует $(\partial I/\partial T)_B = nk/B$, т. е. при абсолютном нуле производная $(\partial I/\partial T)_B$ не обращается в нуль. Но если взять формулу (77.7), то можно воспользоваться асимптотикой $L_{1/2}(x) = 1 - 2e^{-2x}$, и, следовательно, $I = n\mathfrak{M}(1-2e^{-2x})$. Отсюда видно, что при $x \to \infty$ I стремится к постоянному значению $n\mathfrak{M}$ (насыщение), а производная $\partial I/\partial x$ — к нулю. Значит, квантовая формула (77.7) согласуется с теоремой Нернста.

8. Чтобы завершить излагаемую теорию, необходимо еще объяснить, *как возникает намагничивание парамагнетика.* Как показано в предыдущем параграфе, в магнитном поле атом в целом вращается с ларморовской частотой, т. е. совершает регулярную прецессию с той же частотой вокруг направления магнитного поля. При такой прецессии угол между магнитным моментом $\mathfrak{M}$ и полем $\mathbf{B}$ остается неизменным. Остается неизменной и проекция вектора $\mathfrak{M}$ на направление магнитного поля. Ясно, что прецессия сама по себе не может привести к намагничиванию парамагнетика. *Намагничивание возникает в результате взаимодействий атомов между собой.* Схематизируя, будем рассматривать эти взаимодействия как столкновения, а атом — как волчок с механическим моментом $\mathbf{L}$ и магнитным моментом $\mathfrak{M} = \Gamma\mathbf{L}$, где Γ — гиромагнитное отношение. Ради конкретности на рис. 176 векторы $\mathbf{L}$ и $\mathfrak{M}$ будем считать направленными в одну сторону, хотя все наши результаты останутся справедливыми и в том случае, когда эти направления противоположны. В магнитном поле на атом действует момент сил $\mathbf{M} = [\mathfrak{M}\mathbf{B}]$, под действием которого и происходит прецессия. Угловая скорость прецессии $\mathbf{\Omega}$ найдется из уравнения моментов $\dot{\mathbf{L}} = \mathbf{M}$. Подставив в него $\dot{\mathbf{L}} = [\mathbf{\Omega}\mathbf{L}]$, получим $[\mathbf{\Omega}\mathbf{L}] = \Gamma[\mathbf{L}\mathbf{B}]$, откуда

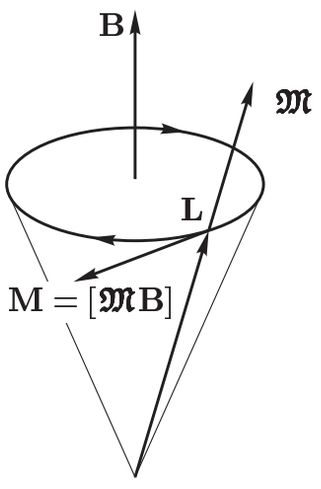


Рис. 176

$$\mathbf{\Omega} = -\Gamma\mathbf{B}. \qquad (77.12)$$

Если момент $\mathbf{L}$ создается орбитальным движением электронов, то $\Gamma = -e/(2mc)$ и, следовательно, $\mathbf{\Omega} = e\mathbf{B}/(2mc)$, т. е. прецессия происходит с ларморовской частотой. Это показывает, что наша классическая модель приводит к правильному результату. Если же момент $\mathbf{L}$ спиновый, то $\Gamma = -e/(mc)$ и угловая скорость прецессии будет вдвое больше. В остальных случаях явление становится более сложным, но его разбор будет приведен в т. V, ч. 1 нашего курса — в атомной физике.

Учтем теперь столкновения. Если атом получает толчок в направлении прецессионного вращения, то соответствующий ему момент сил вызовет прецессию вокруг оси, перпендикулярной к магнитному полю. С помощью уравнения моментов нетрудно убедиться, что в результате такого толчка угол между векторами $\mathfrak{M}$ и $\mathbf{B}$ увеличится. Если же толчок произведен в направлении, противоположном прецессионному вращению, то этот угол уменьшится. Толчки первого типа размагничивают, а второго — намагничивают парамагнетик. Однако эффект намагничивания будет преобладать над эффектом размагничивания, так как толчки против прецессионного вращения в среднем сильнее толчков противоположного направления, подобно тому как сила сопротивления, испытываемая человеком, будет больше, когда он бежит против ветра, а не по ветру. Мы видим, что *магнитное поле только поддерживает, а не создает намагничивание парамагнетика. Намагничивание создается и устанавливается в результате столкновений атомов между собой.*

9. Некоторые металлы, например щелочные, обладают парамагнетизмом, но не подчиняются закону Кюри. Магнитная восприимчивость таких металлов в широких пределах не зависит от температуры. Объяснение этого факта было дано Паули (1900–1958) в 1927 г. Он предположил, что парамагнетизм в этих случаях обусловлен не магнитными моментами ионов кристаллической решетки, а *спиновыми магнитными моментами электронов проводимости.* Рассматривая эти электроны как газ, подчиняющийся статистике Ферми–Дирака (см. т. II, § 82), Паули рассчитал его магнитную восприимчивость. Этот расчет будет приведен в § 99. Здесь же заметим, что независимость магнитной восприимчивости электронного газа в металлах от температуры является *следствием теоремы Нернста.* Действительно, при обычных температурах электронный газ в металлах находится в состоянии *вырождения.* Это означает, что для него такие температуры следует считать *близкими к абсолютному нулю* и можно воспользоваться теоремой Нернста. Из формулы (73.4) следует

$$\frac{1}{4\pi}\left(\frac{\partial B}{\partial T}\right)_H = \left(\frac{\partial S}{\partial H}\right)_T.$$

Правая часть этого соотношения обращается в нуль в силу теоремы Нернста, а левая равна $(H/4\pi)\,d\mu/dT$. Здесь символ H в производной $\partial\mu/\partial T$ опущен, а сама производная написана в виде $d\mu/dT$, так как μ от H не зависит. Таким образом, $d\mu/dt = 4\pi\, d\varkappa/dT = 0$.

ЗАДАЧА

Получить выражение (77.9) для классической функции Ланжевена.

Р е ш е н и е. Число атомов в единице объема, магнитные моменты которых направлены в пределах телесного угла $d\Omega = 2\pi\sin\vartheta\, d\vartheta$, определяется формулой Больцмана

$$dn = C\exp\left(\frac{\mathfrak{M}B}{kT}\cos\vartheta\right)\sin\vartheta\, d\vartheta = Ce^{x\cos\vartheta}\sin\vartheta\, d\vartheta,$$

где ϑ — угол между направлениями магнитного момента $\mathfrak{M}$ и магнитного поля $\mathbf{B}$, а постоянная нормировки C определяется условием $\int dn = n$. В намагничивание среды эти атомы вносят величину $dI = \mathfrak{M}\, dn \cos\vartheta$. Подставив сюда выражение для dn и интегрируя, получим

$$I = n\mathfrak{M}\,\frac{\int\limits_0^\pi e^{x\cos\vartheta}\cos\vartheta\sin\vartheta\,d\vartheta}{\int\limits_0^\pi e^{x\cos\vartheta}\sin\vartheta\,d\vartheta} = n\mathfrak{M}\left(\operatorname{cth} x - \frac{1}{x}\right),$$

откуда и следует формула (77.9).

§ 78. Гиромагнитные явления

1. Подвесим парамагнитное или ферромагнитное тело на нити, вокруг которой оно может вращаться. При намагничивании в магнитном поле атомы тела и их магнитные моменты поворачиваются, ориентируясь преимущественно в направлении поля. С магнитным моментом атома $\mathfrak{M}$ связан момент импульса электронной оболочки $\mathfrak{M}/\Gamma$, где Γ — гиромагнитное отношение. Магнитный момент тела равен $V\mathbf{I}$, где V — объем тела, а $\mathbf{I}$ — вектор намагничивания. Поэтому в результате намагничивания момент импульса электронных оболочек тела увеличивается на $\mathbf{L}_{\text{эл}} = V\mathbf{I}/\Gamma$. Но повороты атомов и магнитных моментов осуществляются под действием столкновений, т. е. *внутренних сил*, которые не могут изменить общий момент импульса тела. Отсюда следует, что кристаллическая решетка тела должна получить такой же момент импульса, но противоположного знака, т. е. $\mathbf{L}_{\text{реш}} = -V\mathbf{I}/\Gamma$. Если до намагничивания тело находилось в состоянии покоя, то в результате намагничивания оно должно прийти во вращение. Если Θ — момент инерции тела, то угловая скорость вращения $\boldsymbol{\omega}$ определяется уравнением

$$\Theta\boldsymbol{\omega} = -\frac{V}{\Gamma}\,\mathbf{I}, \tag{78.1}$$

а само вращение называется *магнитомеханическим явлением*. Оно вполне аналогично вращению скамьи Жуковского, когда сидящий на ней человек поворачивает ось раскрученного велосипедного колеса, которое он держит в руках (см. т. I, § 34, п. 7). Роль велосипедного колеса играют электронные оболочки атомов, роль скамьи Жуковского и сидящего на ней человека — кристаллическая решетка тела.

2. Для оценки эффекта предположим, что тело массы M имеет форму цилиндрика радиуса r и намагничивается до насыщения. Если магнитный момент атома равен одному магнетону Бора $\mathfrak{M}_{\text{B}} = e\hbar/(2mc)$, то магнитный момент тела будет $VI = MN\mathfrak{M}_{\text{B}}/A$, где N — постоянная Авогадро, а A — атомная масса. Допустим, что магнитный момент атома обусловлен орбитальным движением электрона и, следовательно, $\Gamma = -e/(2mc)$. Подставляя эти данные в формулу (78.1) и принимая

во внимание, что $\Theta = (1/2)Mr^2$, получим

$$\omega = \frac{2N\hbar}{Ar^2}.$$

Для железного цилиндрика радиуса $r = 1$ мм ($A = 56$) эта формула дает $\omega = 2{,}25 \cdot 10^{-3}$ рад/с.

Эффект очень мал. Для его усиления Эйнштейн и де Гааз, впервые наблюдавшие этот эффект в 1915 г., воспользовались явлением резонанса. В их опытах небольшой железный цилиндрик подвешивался на тончайшей кварцевой нити и помещался внутри соленоида, по обмотке которого пропускался переменный ток, периодически намагничивавший и размагничивавший цилиндрик (рис. 177).

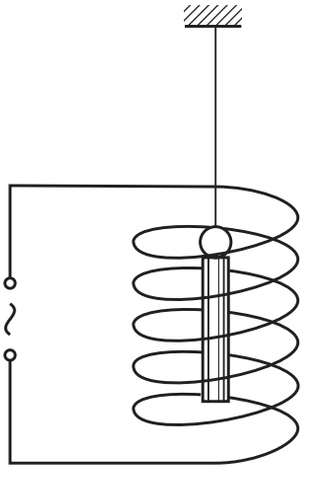


Рис. 177

Повороты цилиндрика отмечались с помощью маленького зеркальца, скрепленного с ним. Уравнение крутильных колебаний цилиндрика записывается в виде

$$\frac{d}{dt}\left(L_{\text{реш}} + L_{\text{эл}}\right) = -f\varphi - \alpha\dot{\varphi},$$

или

$$\Theta\ddot{\varphi} + \alpha\dot{\varphi} + f\varphi = -\frac{dL_{\text{эл}}}{dt} = -\frac{V}{\Gamma}\dot{I},$$

где φ — угол отклонения цилиндрика из положения равновесия, f — модуль кручения нити, α — постоянная, учитывающая сопротивление воздуха и прочие тормозящие силы, которые предполагаются пропорциональными скорости. Если ввести собственную частоту колебаний цилиндрика ω_0 и коэффициент затухания γ по формулам $\omega_0^2 = f/\Theta$, $\gamma = \alpha/(2\Theta)$, то получится

$$\ddot{\varphi} + 2\gamma\dot{\varphi} + \omega_0^2\varphi = -\frac{V}{\Theta\Gamma}\dot{I}. \tag{78.2}$$

Это уравнение вынужденных крутильных колебаний (см. § 122). Величина, стоящая в правой части уравнения (78.2), играет роль внешней силы. Она возникает в результате намагничивания и перемагничивания цилиндрика и предполагается известной. Эта величина, как и намагничивание I, периодически меняется во времени с периодом, равным периоду переменного тока T. На опыте менялась частота переменного тока $\omega = 2\pi/T$, пока колебания цилиндрика не становились наиболее интенсивными. Это происходило при резонансе, т. е. тогда, когда $\omega = \omega_0$. Зависимость между I и H (или, что то же, между B и H) нелинейна. В этом случае правую часть уравнения (78.2) можно разложить в ряд Фурье. Для нахождения решения вблизи резонанса достаточно сохранить в этом разложении только член с основной частотой ω (см. § 128). Исследуя вынужденные крутильные колебания цилиндрика, можно было найти гиромагнитное отношение Γ.

3. Существует явление, обратное магнитомеханическому. Оно называется *гиромагнитным* и заключается в том, что при вращении парамагнитных и ферромагнитных тел они намагничиваются. Объяснение этого явления, в сущности, уже было дано в конце предыдущего параграфа. В магнитном поле электронная оболочка атома приходит во вращение относительно кристаллической решетки с угловой скоростью $\boldsymbol{\Omega} = -\Gamma\mathbf{B}$. При наличии такого относительного вращения столкновения между атомами приводят к намагничиванию среды. Получилась бы та же намагниченность, если бы относительное вращение было создано не магнитным полем, а любым другим способом. Отсюда следует, что если решетку привести во вращение с частотой $\boldsymbol{\omega}$, равной, но противоположно направленной $\boldsymbol{\Omega}$, то намагниченность будет такой же. Иными словами, вращение тела с частотой $\boldsymbol{\omega}$ вызывает то же намагничивание, как и магнитное поле с напряженностью

$$\mathbf{B}_{\text{эфф}} = \frac{\boldsymbol{\omega}}{\Gamma}. \tag{78.3}$$

Явление экспериментально наблюдалось Барнетом (1873–1956) в 1914 г. Для того чтобы составить представление о масштабе явления, допустим, что гиромагнитное отношение связано с орбитальным движением электронов [$\Gamma = -e/(2mc)$], и предположим, что частота вращения составляет 100 оборотов в секунду ($\omega = 2\pi \cdot 100$ рад/с). Тогда

$$B_{\text{эфф}} = \frac{2mc}{e}\,\omega \approx 7 \cdot 10^{-5}\ \text{Гс}.$$

Для сравнения заметим, что земное магнитное поле на поверхности Земли меняется в пределах 0,28–0,70 Гс.

4. Исследования магнитомеханического и гиромагнитного явлений показали, что гиромагнитное отношение Γ всегда отрицательно. Тем самым было подтверждено, что магнетизм обусловлен движением отрицательных электрических зарядов (электронов). Числовые значения величины Γ, как и следовало ожидать, оказались заключенными между пределами $e/(2mc)$ и $e/(mc)$. Весьма существенно, что для всех исследованных ферромагнетиков (железо, никель, кобальт, ряд сплавов) гиромагнитное отношение оказалось равным не $-e/(2mc)$, а $-e/(mc)$. Это показывает, что *магнетизм ферромагнетиков обусловлен одним только спином электронов, а не их орбитальным движением.*

В настоящее время разработаны значительно более точные методы измерения магнитных моментов электронов и атомных ядер, а также гиромагнитных отношений. Особенно широкие применения получили методы, основанные на *магнитном резонансе* (электронном и ядерном). Это явление будет рассмотрено в т. V, ч. 1 нашего курса — в атомной физике.

§ 79. Формальная теория ферромагнетизма Вейсса

1. Первая количественная теория ферромагнетизма была разработана французским физиком Вейссом (1865–1940) в 1907 г., хотя до

Вейсса сходные идеи высказывались русским физиком Б.Л. Розингом еще в 1892 г. Но работы Розинга были забыты. Теория Вейсса носит полуфеноменологический характер. Вейсс предположил, что атомы ферромагнетика, как и парамагнетика, обладают магнитными моментами и взаимодействуют между собой с силами, зависящими от угла между этими магнитными моментами. Эти силы стремятся установить магнитные моменты соседних атомов параллельно друг другу. В результате ориентации магнитных моментов атомов в определенном направлении и создается намагничивание ферромагнетика. В теории Вейсса силы взаимодействия между атомами формально сводятся к некоторому «эффективному» магнитному полю, которое и ориентирует атомы ферромагнетика. Эффективное поле складывается из обычного макроскопического поля в веществе $\mathbf{H}$ и некоторого гипотетического «*молекулярного поля*». Последнее, согласно предположению Вейсса, пропорционально намагниченности ферромагнетика $\mathbf{I}$, так что эффективное поле может быть представлено в виде

$$\mathbf{B}_{\text{эфф}} = \mathbf{H} + b\mathbf{I}, \tag{79.1}$$

где b — некоторая положительная постоянная, характеризующая свойства различных ферромагнетиков. Она называется *постоянной Вейсса*. Исходя из этих предположений, нетрудно рассчитать намагничивание ферромагнетика I. Для этого надо только в теории Ланжевена поле $\mathbf{H}$ заменить на эффективное поле $\mathbf{B}_{\text{эфф}}$. Это дает $I = n\mathfrak{M}L(x)$, где $x = \mathfrak{M}(H + bI)/kT$. Разрешая последнее соотношение относительно I и замечая, что $n\mathfrak{M}$ есть намагничивание насыщения I_s, получаем систему двух уравнений:

$$I = I_s L(x), \quad I = \frac{kTn}{I_s b}\, x - \frac{H}{b},$$

из которой нетрудно вычислить намагничивание I. Эту систему удобно исследовать графически, откладывая по горизонтальной оси величину x, а по вертикальной — намагничивание I (рис. 178). Первое уравнение (79.2) представится кривой Ланжевена OA_0A, а второе — прямой CA, пересекающей вертикальную ось в точке C с ординатой $OC = -H/b$.

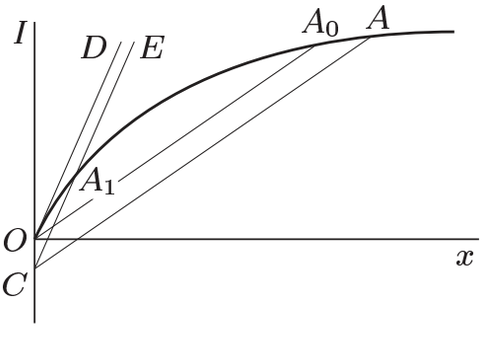


Рис. 178

2. Допустим сначала, что наклон прямой CA меньше наклона кривой Ланжевена в начале координат, т. е. $kTn/(I_s b) < I_s(dL/dx)_{x=0}$, или $T < T_{\text{к}}$, где введено обозначение

$$T_{\text{к}} = \frac{I_s^2 b}{kn}\left(\frac{dL}{dx}\right)_{x=0}. \tag{79.2}$$

Тогда прямая пересечет кривую Ланжевена в точке A, ордината которой и представит намагничивание ферромагнетика I. Будем теперь

уменьшать магнитное поле H до нуля. При этом точка C будет подниматься к точке O, а точка A — перемещаться к точке A_0. Когда магнитное поле H обратится в нуль, ферромагнетик останется намагниченным — его намагничивание представится ординатой точки A_0.

Ферромагнетик будет спонтанно намагничен и в том случае, когда он вообще не вносился ни в какое магнитное поле, так как благодаря гипотетическому взаимодействию между атомами, введенному Вейссом, состояние спонтанного намагничивания «энергетически выгодно». В самом деле, допустим, что вначале магнитные моменты ориентированы хаотически. Если в результате флуктуации магнитные моменты двух или нескольких атомов установятся параллельно в некотором направлении, то появится вейссовское внутреннее поле, которое заставит магнитные моменты остальных атомов предпочесть то же направление. Намагничивание, вообще говоря, не достигнет насыщения, так как этому препятствует тепловое движение. Следует учесть, что благодаря анизотропии кристаллической решетки в ней имеются *оси легкого намагничивания*, вдоль которых и предпочитают выстраиваться магнитные моменты атомов кристалла. Ориентированным атомам нелегко изменить направление, в котором они выстроились, так как для этого им пришлось бы пройти через ряд менее предпочтительных направлений. По этой причине спонтанное намагничивание, раз возникнув в кристалле, сохраняется в нем и в дальнейшем. Таким образом, при $T < T_{\text{К}}$ ферромагнетик должен быть *спонтанно намагничен*. Энергии теплового движения недостаточно, чтобы разрушить это намагничивание. Величина $T_{\text{К}}$ называется *температурой* или *точкой Кюри*.

Ниже точки Кюри из-за наличия спонтанного намагничивания обычное определение магнитной восприимчивости и проницаемости с помощью соотношений $\mathbf{I} = \varkappa\mathbf{H}$ и $\mathbf{B} = \mu\mathbf{H}$ лишено смысла. Величины $\varkappa$ и μ в этом случае определяются соотношениями

$$\varkappa = \frac{dI}{dH}, \quad \mu = \frac{dB}{dH}$$

и являются функциями напряженности поля H.

3. Предположим теперь, что наклон прямой CA больше наклона кривой Ланжевена в точке O. Это будет тогда, когда температура тела выше точки Кюри ($T > T_{\text{К}}$). Тогда при отсутствии магнитного поля прямая CA займет положение OD, т. е. пересечет кривую Ланжевена только в начале координат, где намагничивание равно нулю. Спонтанного намагничивания не возникнет: намагничивание разрушается тепловым движением. Чтобы намагнитить тело, необходимо приложить магнитное поле. Рассчитаем возникающее при этом намагничивание. Прямая CA в рассматриваемом случае займет положение CE и пересечет кривую Ланжевена в точке A_1, ордината которой численно равна намагниченности тела. Ордината $OC = -H/b$, как показывает анализ экспериментальных данных, мала, а потому мал и участок OA_1 кривой Ланжевена. Заменив этот участок прямой линией, можно написать

$L(x) = (dL/dx)_{x=0}x$, так что

$$I = I_s \left(\frac{dL}{dx}\right)_{x=0} x.$$

Решая это уравнение совместно со вторым уравнением (79.2) и воспользовавшись обозначением (79.3), получим

$$\mathbf{I} = \varkappa \mathbf{H}, \tag{79.3}$$

где

$$\varkappa = \frac{T_{\text{К}}}{b(T - T_{\text{К}})} = \frac{\text{const}}{T - T_{\text{К}}}. \tag{79.4}$$

Намагничивание пропорционально полю, т. е. выше точки Кюри ферромагнетик превращается в парамагнетик, причем зависимость магнитной восприимчивости от температуры определяется *законом Кюри–Вейсса* (79.5).

Можно сказать, что *ниже точки Кюри вещество находится в «упорядоченном состоянии», характеризующемся спонтанным намагничиванием. В точке Кюри происходит фазовое превращение его в «неупорядоченное состояние», в котором намагничивание отсутствует.*

4. Таким образом, теория Вейсса объясняет основные факты из области ферромагнетизма: спонтанную намагниченность, существование температуры Кюри, закон Кюри–Вейсса. Можно было бы предполагать, что поле $b\mathbf{I}$ сводится к обычному взаимодействию магнитных диполей, совершенно аналогичному взаимодействию электрических диполей диэлектрика. Если бы это было так, то, по аналогии с формулой (35.13), молекулярное поле Вейсса, во всяком случае по порядку величины, определялось бы выражением $(4\pi/3)\mathbf{I}$, т. е. постоянная Вейсса была бы порядка единицы. К тому же заключению приводят следующие элементарные соображения. Так как магнитное поле диполя быстро убывает с расстоянием, то поле в центре атома, создаваемое окружающими атомами, было бы порядка поля, создаваемого ближайшим атомом, т. е. $\mathfrak{M}/a^3$, где a — среднее расстояние между атомами. Так как $\mathfrak{M}/a^3$ порядка $\mathbf{I}$, то мы приходим к прежней оценке. Но поле такого порядка совершенно недостаточно, чтобы достигнуть насыщения намагничивания ферромагнетика, так как в парамагнетиках, где ситуация совершенно аналогична, насыщение не наблюдается.

С еще большей убедительностью мы придем к тому же заключению, если на основе экспериментальных данных оценим величину постоянной Вейсса b. Это можно сделать с помощью формулы (79.3). В этой формуле в качестве функции Ланжевена возьмем $L_{1/2}(x) = \operatorname{th} x \approx x$ в соответствии с экспериментально установленным фактом, что ферромагнетизм обусловлен спинами электронов (см. предыдущий параграф). Тогда получим

$$T_{\text{К}} = \frac{I_s^2 b}{kn}. \tag{79.5}$$

Измерив температуру Кюри и намагничивание при насыщении I_s, можно по этой формуле вычислить постоянную b. В результате для b получаются значения порядка 10^3–10^4. Таким образом, внутренние поля, необходимые для объяснения ферромагнетизма, в тысячи и десятки тысяч раз больше магнитных полей, которые могут создать ориентированные магнитные моменты атомов ферромагнетика. Следовательно, ферромагнетизм не может быть объяснен магнитным взаимодействием атомов.

Прямой опыт, поставленный в 1927 г. Я.Г. Дорфманом (1898–1974), доказал, что молекулярное поле, ориентирующее атомы ферромагнетика, не может быть магнитной природы. Между полюсами сильного электромагнита помещалась никелевая фольга толщиной около 20 мкм, расположенная параллельно магнитному полю. Через фольгу перпендикулярно к ее поверхности пропускался пучок быстрых электронов (β-лучей) от радиоактивного источника. После прохождения через фольгу след пучка регистрировался на фотопластинке. Сначала опыт производился с выключенным электромагнитом. При включении магнитного поля электромагнита след пучка смещался в сторону. Если бы молекулярное поле Вейсса было магнитной природы, то смещение вызывалось бы эффективным полем $\mathbf{B}_{\text{эфф}} = \mathbf{H} + b\mathbf{I}$. В этом случае ожидаемое смещение было бы 10 мм. В действительности смещение получилось 0,3 мм и соответствовало действию поля 10^4 Гс, т. е. величине магнитной индукции в образце.

5. Физическая природа молекулярного поля Вейсса была установлена в 1927 г. Я.И. Френкелем (1894–1952) и почти одновременно с ним — Гейзенбергом (1901–1976) на основе квантовой механики. Подробное изложение этого вопроса выходит за пределы нашего руководства. Кратко суть дела заключается в следующем. Если бы к электронам и атомным ядрам ферромагнетика применить обычный закон Кулона, но рассмотреть движение электронов на основе уравнений квантовой механики, то результат оказался бы таким же, какой получился бы в классической механике при наличии, помимо кулоновских сил, некоторых добавочных сил взаимодействия между электронами. Эти добавочные силы, которые надо ввести в классическую теорию, чтобы согласовать ее с квантовой, называются *обменными силами*. Уже из приведенной формулировки видно, что обменные силы не могут быть истолкованы классически, так как в исходные уравнения квантовой механики никакие обменные силы не входят. Они добавляются в классическом рассмотрении для согласования его с квантовым. Обменные силы являются *короткодействующими*, т. е. действуют на расстояниях порядка атомных. При определенных условиях, относящихся к электронному строению атомов, структуре кристаллической решетки и т. п., обменные силы стремятся установить спины электронов соседних атомов параллельно друг другу, чем и объясняется намагничивание ферромагнетиков. Обменное взаимодействие, как показывает расчет, может быть с достаточной точностью учтено введением некоторого эквивалентного молекулярного поля $b\mathbf{I}$. Тем самым теория Вейсса получает настоящее

физическое обоснование, без каких-либо специальных допущений и гипотез.

6. Спонтанное намагничивание, предсказываемое теорией Вейсса, находится в кажущемся противоречии с тем фактом, что даже ниже температуры Кюри железо и другие ферромагнетики, как правило, не намагничены, хотя и существуют постоянные магниты, остаточное намагничивание и пр. Вейсс устранил это противоречие, введя предположение, подтвердившееся дальнейшими исследованиями, что ферромагнетики ниже точки Кюри распадаются в магнитном отношении на множество чрезвычайно малых макроскопических областей. Каждая из таких областей спонтанно намагничена до значения, соответствующего по теории положению точки A_0 на рис. 178. Эти области спонтанного намагничивания получили название *доменов*. Их ни в коем случае не следует смешивать с мельчайшими кристалликами, из которых обычно состоят ферромагнитные тела (поликристаллы). В обычных условиях направления магнитных моментов доменов хаотически распределены в пространстве, так что тело в целом представляется макроскопически ненамагниченным. При включении внешнего магнитного поля домены, ориентированные по полю, растут за счет доменов, ориентированных против поля, т. е. происходит смещение границ доменов. Такое смещение в слабых полях носит *обратимый характер*. В более сильных полях происходит переориентация магнитных моментов в пределах всего домена. Перемагничивание прибретает *необратимый характер* — появляется гистерезис и остаточное намагничивание.

Распад ферромагнетика на домены происходит потому, что этот процесс «энергетически выгоден». Если бы весь ферромагнетик был намагничен в одном определенном направлении, то в этом случае получился бы минимум энергии обменного взаимодействия электронов. Однако такому ферромагнетику соответствовала бы значительная энергия возбуждаемого им магнитного поля. При дроблении ферромагнетика на домены и появлении доменов различной ориентации магнитное поле, возбуждаемое ферромагнетиком, ослабляется. Уменьшается и соответствующая ему энергия. Что касается энергии обменного взаимодействия электронов, то благодаря короткодействующему характеру обменных сил эта энергия остается неизменной для всех электронов, за исключением электронов на границах доменов, где она возрастает из-за различной ориентации спинов электронов соседних доменов. Обменная энергия атомов, расположенных на границах доменов, может рассматриваться как *поверхностная энергия*, поскольку она пропорциональна полной площади поверхностей, вдоль которых граничат домены. По мере дробления доменов поверхностная, а с ней и полная энергия обменного взаимодействия возрастают, а энергия магнитного поля убывает. Дробление доменов прекращается, когда сумма магнитной и обменной энергий достигает минимума. Условием этого минимума и определяется размер доменов.

7. Доменная структура ферромагнетиков доказана экспериментально. Наиболее прямым доказательством является получение так называемых *порошковых фигур*. Тщательно отполированная поверхность ферромагнетика покрывается тонким слоем жидкости, в которой взвешены мельчайшие частицы ферромагнитного порошка (например, Fe_2O_3). Частицы оседают преимущественно в местах максимальной неоднородности магнитного поля, т. е. на границах между доменами. Благодаря этому осевший порошок очерчивает границы доменов. Получившиеся «порошковые фигуры» хорошо видны в микроскоп небольшого увеличения.

Другой, косвенный, метод основан на наблюдении *эффекта Баркгаузена* (1881–1956). Если к ферромагнитному образцу подносить и удалять сильный магнит, то образец будет намагничиваться и перемагничиваться. Перемагничивание осуществляется резким поворотом или опрокидыванием направления вектора намагничивания в пределах всего домена. Такие скачкообразные повороты и опрокидывания вектора намагничивания домена называются *скачками Баркгаузена*. Для их наблюдения ферромагнитный образец в виде длинного стерженька или проволоки помещается в катушке, концы которой через усилитель присоединены к громкоговорителю или к осциллографу. При каждом повороте или опрокидывании магнитного момента домена скачкообразно меняется магнитный поток через катушку, и в последней возбуждается индукционный ток. В громкоговорителе слышен своеобразный шорох, а на осциллограмме осциллографа появляется беспорядочно меняющийся всплеск. В этом и состоит эффект Баркгаузена. Из осциллограмм эффекта Баркгаузена, а также с помощью порошковых фигур оказалось возможным приблизительно определить объем и линейные размеры доменов. Объем домена оказался порядка 10^{-6} см^3. Длина доменов может достигать 2–3 мм.

8. В заключение отметим, что в зависимости от структуры кристалла обменные силы могут вызывать не только параллельную, но и *антипараллельную ориентацию* электронных спинов соседних атомов. В простейшем случае (в отсутствие намагничивающего поля) электронные спины как бы образуют две пространственные подрешетки с противоположно направленными спинами, вставленные одна в другую. Если величина намагничивания обеих подрешеток одинакова, то их магнитные моменты компенсируют друг друга, так что кристалл в целом не обладает магнитным моментом. Такие тела называются *антиферромагнетиками*, а соответствующее им упорядочение электронных спинов — *антиферромагнетизмом*. Явление антиферромагнетизма было теоретически предсказано Неелем (р. 1904) в 1932 г. и независимо от него Л.Д. Ландау в 1933 г. К антиферромагнетикам относятся некоторые соединения марганца (MnO, MnS, MnF_2), железа (FeF_2, $FeCl_2$, FeO), хрома (CrSb, Cr_2O_3) и многих других элементов. При низких температурах антиферромагнетики имеют ничтожно малую магнитную восприимчивость, которая увеличивается с повышением температуры. При некоторой температуре, называемой *антиферро-*

магнитной точкой Кюри или *точкой Нееля*, магнитное упорядочение спинов разрушается и антиферромагнетик превращается в парамагнетик. С дальнейшим повышением температуры, как у всякого парамагнетика, магнитная проницаемость уменьшается. Таким образом, она достигает максимума в антиферромагнитной точке Кюри.

Если величина намагничивания обеих подрешеток неодинакова, то возникает *нескомпенсированный антиферромагнетизм.* В этом случае тело обладает магнитным моментом, который может быть довольно значительным. Такие тела называются *ферримагнетиками.* По своим магнитным свойствам они аналогичны ферромагнетикам. Если при этом ферримагнетики обладают еще полупроводниковыми свойствами, то они называются *ферритами.*

§ 80. Сверхпроводники и их магнитные свойства

1. Явление сверхпроводимости было открыто в 1911 г. Камерлинг-Оннесом (1853–1926) после того, как в 1908 г. им же был ожижен гелий и стала доступной область температур вблизи абсолютного нуля. Исследуя сопротивление ртути постоянному току при приближении к абсолютному нулю, Камерлинг-Оннес обнаружил, что при температуре 4,12 K (по современным измерениям, 4,15 K) сопротивление скачкообразно уменьшалось до нуля или, во всяком случае, до неизмеримо малой величины. Дальнейшие исследования показали, что аналогично ведут себя и многие другие металлы. Это явление назвали *сверхпроводимостью*, а вещества, где оно наблюдается, — *сверхпроводниками.* Температура $T_к$, при которой происходит скачкообразное уменьшение сопротивления, называется *температурой перехода в сверхпроводящее состояние* или *критической температурой.* Состояние сверхпроводника выше критической температуры называется *нормальным*, а ниже — *сверхпроводящим.* В настоящее время известно, что примерно половина металлических элементов переходит в сверхпроводящее состояние. Сверхпроводимость обнаружена также более чем у тысячи сплавов и соединений металлов. Из чистых металлов наибольшей температурой перехода обладает ниобий (9,3 K), а среди сплавов и химических соединений рекордным [1)] является соединение ниобия с германием Nb_3Ge ($T_к = 23{,}2$ K).

Наиболее точным методом для количественных оценок сопротивления сверхпроводников в сверхпроводящем состоянии является создание индукционного тока в *сверхпроводящем кольце.* Кольцо из исследуемого материала при температуре выше критической вносится в магнитное поле. Затем оно охлаждается до температуры ниже критической и тем

[1)] В 1987 г. были открыты высокотемпературные сверхпроводники с критической температурой $T_к$, превосходящей температуру кипения жидкого азота при атмосферном давлении (77,4 K). К 1990 г. наивысшее надежно установленное значение $T_к$ равно примерно 125 K. Речь идет при этом о довольно сложных слоистых соединениях меди и кислорода. (*Примеч. ред.*)

самым переводится в сверхпроводящее состояние. После этого магнитное поле выключается и в кольце возбуждается индукционный ток. В обычном нормальном металле индукционный ток быстро затухает. По быстроте этого затухания можно судить о сопротивлении кольца. Если же кольцо сверхпроводящее, то ток продолжается практически неограниченно долго. Наблюдая такой ток (по величине возбуждаемого им магнитного поля), можно оценить верхний предел сопротивления кольца. Таким путем найдено, что удельное сопротивление свинца в сверхпроводящем состоянии меньше $4 \cdot 10^{-23}$ Ом · см, т. е. по крайней мере в 10^{17} раз меньше удельного сопротивления меди при комнатной температуре. При переходе через критическую температуру скачок сопротивления составляет по крайней мере 14 порядков. Это позволяет с полным основанием считать, что в сверхпроводящем состоянии электрическое сопротивление постоянному току действительно исчезает.

2. Сверхпроводник ниже критической температуры не оказывает никакого электрического сопротивления только тогда, когда ток *постоянный.* В случае переменных токов сопротивление отлично от нуля и тем больше, чем выше частота переменного тока. Этот факт проще всего понять на основе так называемой *двухжидкостной модели сверхпроводника.* Согласно этой модели электроны, создающие электрический ток в сверхпроводнике, разделяются на две группы: *сверхпроводящие* и *нормальные.* Сверхпроводник как бы пропитан двумя электрическими жидкостями, откуда и произошло название модели. Сверхпроводящие электроны при своем движении не встречают никаких сил сопротивления, они движутся по инерции, а потому для поддержания сверхпроводящего тока не требуется никакого электрического поля. Напротив, нормальные электроны испытывают столкновения с атомами решетки, как в обычных металлах, и для тока нормальных электронов справедлив обычный закон Ома. При включении электрического поля ускоряются как сверхпроводящие, так и нормальные электроны. Когда ток становится постоянным, электрическое поле в сверхпроводнике должно исчезнуть, так как в противном случае сверхпроводящие электроны непрерывно ускорялись бы, а создаваемый ими ток неограниченно возрастал бы. Но при отсутствии электрического поля нормальные электроны не могут создать никакого тока. Ток создается только сверхпроводящими электронами, а потому и не возникает никакого электрического сопротивления. Если же сила тока меняется, то в сверхпроводнике должно существовать электрическое поле, ускоряющее как сверхпроводящие, так и нормальные электроны. Появляется ток нормальных электронов и связанные с ним электрическое сопротивление и джоулево тепло. Правда, для обычных переменных токов, с которыми имеет дело электротехника (частота $\nu \sim 50$ Гц), на долю тока нормальных электронов приходится ничтожная часть. Однако с повышением частоты эта доля непрерывно возрастает. Когда частота ν становится достаточно высокой, начинают проявляться квантовые эффекты, еще больше увеличивающие потери энергии на джоулево тепло. Это происходит тогда, когда энергии

кванта $h\nu$ достаточно для перевода сверхпроводящего электрона на более высокий энергетический уровень, т. е. в группу нормальных электронов. Частоту ν легко оценить, заметив, что, когда температура достигает критической, энергия теплового движения kT становится достаточной для такого перехода. Поэтому должно быть $h\nu \gtrsim kT_{\text{к}}$. Так как $T_{\text{к}} \sim 1$ К, то отсюда получаем $\nu \gtrsim 10^{11}$ Гц, т. е. указанный процесс начинает происходить в области инфракрасных частот. С изложенной точки зрения понятно, почему сверхпроводники, даже при температуре ниже критической, в оптическом отношении ведут себя как обычные металлы.

3. Перейдем теперь к рассмотрению магнитных свойств сверхпроводников. До 1933 г. считалось, что в магнитном отношении сверхпроводники ведут себя просто как идеальные проводники, т. е. тела, электропроводность которых бесконечно велика. В идеальном проводнике, как известно (см. § 71), сохраняется магнитный поток через любой «жидкий» контур, движущийся вместе с проводником. Если проводник твердый (а все сверхпроводники твердые) и неподвижен, то и всякий «жидкий» контур в нем также неподвижен и недеформируем. Так как контур можно взять бесконечно малым и любым, то из сохранения потока $\int \mathbf{B}\, d\mathbf{S}$ следует и сохранение вектора $\mathbf{B}$. Таким образом, индукция $\mathbf{B}$ в твердом идеальном проводнике не может изменяться.

Применим доказанную теорему к сверхпроводникам. Будем характеризовать внешние условия, в которых находится сверхпроводник, температурой T и внешним приложенным магнитным полем B_e (создаваемым, например, электромагнитом), откладывая эти параметры на осях координат (рис. 179). Переведем сверхпроводник (имеющий, на-

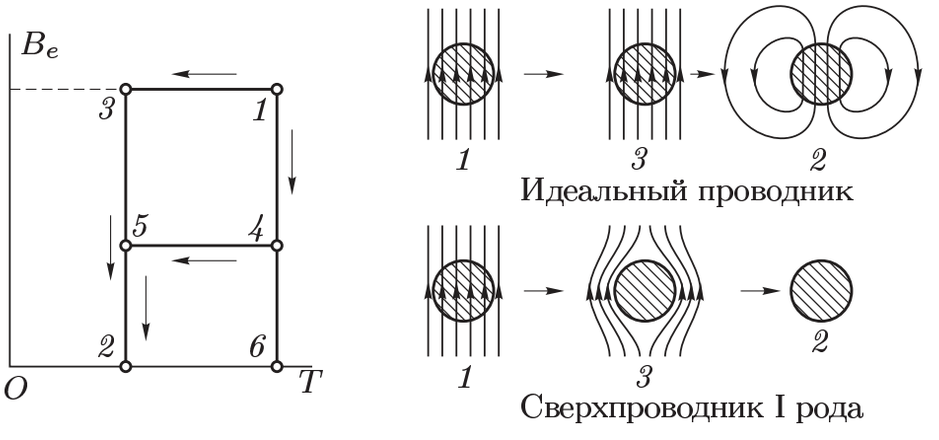


Рис. 179

пример, форму шара) из несверхпроводящего состояния *1* в сверхпроводящее состояние *2* по пути *132*. Сначала, не меняя магнитного поля B_e, охладим тело до точки *3*, переведя его в сверхпроводящее состояние. Все сверхпроводники не ферромагнитны, так что их магнитная проницаемость очень близка к единице. Поэтому при указанном охлаждении поле внутри и вне сверхпроводника не изменится. В точке *3*, сохраняя температуру постоянной, выключим внешнее приложенное

поле B_e, т. е. перейдем в точку *2*. В силу доказанной теоремы магнитное поле внутри сверхпроводника не изменится. Вне сверхпроводника останется неоднородное магнитное поле, так как должно выполняться граничное условие — непрерывность нормальных составляющих вектора индукции на границе тела. Произведем теперь переход в ту же точку *2* по пути *1452*. В наших рассуждениях ничего не изменится — достаточно в качестве исходной взять просто точку *4*, где внешнее приложенное поле B_e слабее, чем в точке *1*. В точке *2* внутри и вне тела получится такая же картина поля, что и раньше, только поле всюду будет слабее. В частности, если переход произвести по пути *1462*, то магнитное поле в конечном состоянии всюду будет равно нулю. Таким образом, заданием внешних параметров T и B_e состояние сверхпроводника не определяется однозначно. Сверхпроводник может перейти в любое из бесчисленного множества состояний в зависимости от «пути перехода», т. е. от способа, каким достигнуты значения параметров T и B_e.

4. Реальные сверхпроводники ведут себя иначе. Вальтер Мейсснер (1882–1974) совместно с Оксенфельдом показал в 1933 г., что *в сверхпроводящем состоянии магнитное поле внутри сверхпроводника равно нулю* ($\mathbf{B} = \mathbf{H} = 0$). Это значит, что *при охлаждении сверхпроводника ниже критической температуры магнитное поле вытесняется из него* (*эффект Мейсснера–Оксенфельда*). Поэтому картина изменения магнитного поля при изменении внешних параметров T и B_e, изображенная в верхнем ряду рис. 179, не соответствует реальному поведению сверхпроводника. Действительное поведение показано на том же рисунке в нижнем ряду. Состояние сверхпроводника *однозначно* определяется параметрами T и B_e и *не зависит от пути перехода* в это состояние. Позднее было показано, что не все сверхпроводники ведут себя в соответствии с результатом, установленным Мейсснером и Оксенфельдом. Сверхпроводники, для которых эффект Мейсснера–Оксенфельда имеет место в описанной форме, называются *сверхпроводниками I рода*, а сверхпроводники, где явление протекает иначе, — *сверхпроводниками II рода*. Более точное определение будет дано ниже (см. п. 9). Сейчас же заметим, что к сверхпроводникам I рода относятся все чистые металлы, за исключением ниобия, ванадия и технеция, которые являются сверхпроводниками II рода. Сначала мы будем предполагать, что имеем дело со сверхпроводниками I рода.

Так как в сверхпроводнике нет магнитного поля, то в его объеме не могут течь и электрические токи, т. е. внутри сверхпроводника $\mathbf{j} = 0$. Это непосредственно следует из теоремы о циркуляции $\operatorname{rot} \mathbf{H} = (4n/c)\mathbf{j}$. Все токи должны течь по поверхности сверхпроводника. Эти поверхностные токи возбуждают магнитное поле, компенсирующее внутри проводника внешнее приложенное поле. Таков механизм вытеснения магнитного поля из сверхпроводника, о котором говорится в эффекте Мейсснера–Оксенфельда.

Эффект Мейсснера–Оксенфельда очень наглядно проявляется в *парении магнита* над поверхностью сверхпроводника. На тарелку из

сверхпроводника (например, свинцовую), охлажденную до температуры ниже критической, опускается небольшой магнит. При этом в тарелке возбуждаются незатухающие индукционные токи. Отталкивая магнит, эти токи и заставляют его «парить» над тарелкой на определенной высоте. Явление наблюдается и в том случае, когда магнит кладется на тарелку, температура которой выше критической, а затем охлаждением тарелка приводится в сверхпроводящее состояние. Дело в том, что вытеснение магнитного поля из сверхпроводника также сопровождается изменениями магнитных потоков, а следовательно, и возбуждением индукционных токов. Эти токи определяются только взаимным расположением магнита и тарелки и совсем не зависят от того, каким способом было достигнуто это расположение. Поэтому явление будет выглядеть так же, как и при первой постановке опыта.

5. В действительности ток у границы сверхпроводника течет не по поверхности, а *в слое конечной толщины.* В тот же слой должно проникать и магнитное поле, как это видно из уравнения $\operatorname{rot}\mathbf{H} = (4\pi/c)\mathbf{j}$. Наиболее простая оценка *глубины проникновения магнитного поля* в сверхпроводник была дана братьями Фрицем и Гансом Лондонами, разработавшими первую феноменологическую теорию сверхпроводимости. Приведем эту оценку. Будем предполагать, что имеем дело с полями, медленно меняющимися во времени. Так как сверхпроводники не ферромагнитны, то можно пренебречь разницей между $\mathbf{B}$ и $\mathbf{H}$ и записать фундаментальные уравнения электродинамики в виде

$$\operatorname{rot}\mathbf{B} = \frac{4\pi}{c}\,\mathbf{j}, \quad \operatorname{rot}\mathbf{E} = -\frac{1}{c}\,\dot{\mathbf{B}}, \tag{80.1}$$

причем мы будем также пренебрегать разницей между частной и полной производными по времени. Предполагая, что токи создаются движением только сверхпроводящих электронов, напишем далее $\mathbf{j} = -n_s e\mathbf{v}_s$, где n_s — концентрация таких электронов. После дифференцирования по времени получим $d\mathbf{j}/dt = -n_s e\dot{\mathbf{v}}_s$. Ускорение электрона $\dot{\mathbf{v}}_s$ найдется из уравнения $m\dot{\mathbf{v}}_s = -e\mathbf{E}$, если пренебречь действием магнитного поля. Тогда

$$\frac{d\mathbf{j}}{dt} = -\frac{c^2}{4\pi\Lambda^2}\,\mathbf{E}, \tag{80.2}$$

где введено обозначение

$$\Lambda = \sqrt{\frac{mc^2}{4\pi n_s e^2}}. \tag{80.3}$$

Продифференцировав первое уравнение (80.1) по t, исключим из уравнений (80.1) и (80.2) величины $\mathbf{E}$ и $d\mathbf{j}/dt$. Тогда получим

$$\dot{\mathbf{B}} = -\Lambda^2 \operatorname{rot}\operatorname{rot}\dot{\mathbf{B}}.$$

Этому уравнению удовлетворяет $\mathbf{B} = \text{const}$. Но такое решение не согласуется с эффектом Мейсснера–Оксенфельда, так как внутри сверхпроводника должно быть $\mathbf{B} = 0$. Лишнее решение получилось потому,

что при выводе дважды применялась операция дифференцирования по времени. Чтобы автоматически исключить это решение, Лондоны ввели гипотезу, что в последнем уравнении производную $\mathbf{B}$ следует заменить самим вектором $\mathbf{B}$. Это дает

$$\mathbf{B} = -\Lambda^2 \operatorname{rot} \operatorname{rot} \mathbf{B}. \tag{80.4}$$

Для определения глубины проникновения магнитного поля внутрь сверхпроводника допустим, что последний ограничен плоскостью и занимает все полупространство по одну сторону от нее. Направим ось Z внутрь сверхпроводника нормально к его границе. Пусть магнитное поле параллельно оси X, так что $B_y = B_z = 0$. Тогда

$$\operatorname{rot} \mathbf{B} = \frac{\partial B_x}{\partial z}\, \mathbf{e}_y, \quad \operatorname{rot} \operatorname{rot} \mathbf{B} = -\frac{\partial^2 B_x}{\partial z^2}\, \mathbf{e}_x$$

и уравнение (80.4) дает

$$\frac{\partial^2 \mathbf{B}}{\partial z^2} - \frac{\mathbf{B}}{\Lambda^2} = 0.$$

Решение этого уравнения, обращающееся в нуль при $z \to \infty$, имеет вид

$$\mathbf{B} = \mathbf{B}_0 e^{-z/\Lambda}. \tag{80.5}$$

Постоянная интегрирования $\mathbf{B}_0$ дает поле на поверхности сверхпроводника. На протяжении длины Λ магнитное поле убывает в e раз. Величина Λ и принимается за меру глубины проникновения магнитного поля в металл.

Для получения численной оценки примем, что на каждый атом металла приходится один сверхпроводящий электрон, полагая $n_s \approx 5 \cdot 10^{22}$ см$^{-3}$. Тогда по формуле (80.3) найдем $\Lambda \approx 2{,}5 \cdot 10^{-6}$ см, что по порядку величины совпадает со значениями, полученными непосредственными измерениями. Концентрация сверхпроводящих электронов убывает по мере приближения к критической температуре $T_к$. В соответствии с этим при возрастании температуры от 0 до $T_к$ глубина проникновения Λ монотонно растет и обращается в бесконечность при $T = T_к$ (см. рис. 180, где представлена зависимость Λ от T для олова).

6. *При увеличении напряженности внешнего возбуждающего поля $\mathbf{B}_e$ выше определенного предела сверхпроводимость разрушается, т. е. сверхпроводник переходит в нормальное состояние, а магнитное поле проникает внутрь сверхпроводника.* Магнитное поле, при котором это происходит, называется *критическим полем* и обозначается $\mathbf{B}_к$. Критическое поле зависит от температуры. Его легко определить из термодинамических соображений. Чтобы освободиться от осложняющих обстоятельств, не имеющих отношения к существу вопроса, будем предполагать, что сверхпроводящее тело имеет форму длинного цилиндра, а внешнее приложенное поле однородно и направлено параллельно оси цилиндра. Тогда магнитное поле внутри цилиндра (когда он находится в нормальном состоянии), а также намагничивание последнего

будут также однородными. Для определенности предположим, что поле $\mathbf{B}_e$ создается длинным соленоидом, по боковой поверхности которого

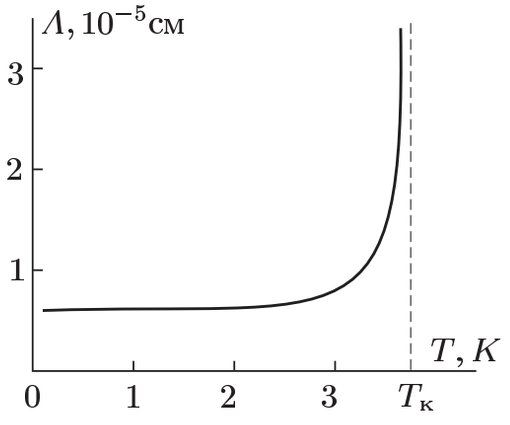


Рис. 180

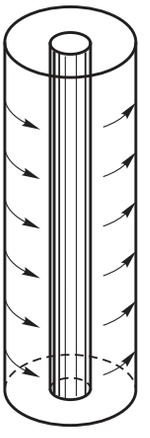


Рис. 181

циркулирует постоянный ток I, а ось цилиндра совпадает с осью соленоида (рис. 181) Длины цилиндра и соленоида для простоты примем одинаковыми и равными l. Влияние краев цилиндра и соленоида при достаточно большом l можно не принимать во внимание. Магнитную проницаемость сверхпроводника будем считать равной единице как в сверхпроводящем, так и в нормальном состоянии. Пусть ψ_s и ψ_n — объемные плотности свободной энергии сверхпроводника в сверхпроводящем и нормальном состоянии при отсутствии магнитного поля. Поскольку температура ниже критической, реализуется сверхпроводящее состояние; оно является более устойчивым, а потому должно быть $\psi_s < \psi_n$.

Критическое магнитное поле $\mathbf{B}_\text{к}$ можно определить как такое поле $\mathbf{B}_e$, когда при заданной температуре сверхпроводящая и нормальная фазы находятся в равновесии друг с другом. Тогда полная свободная энергия всей системы будет одна и та же независимо от того, в какой фазе находится сверхпроводящий цилиндр. Пусть цилиндр квазистатически переходит из сверхпроводящего в нормальное состояние, и притом так, что во время перехода температура T, внешнее приложенное поле $\mathbf{B}_e$, а с ним и ток I в соленоиде поддерживаются постоянными ($\mathbf{B}_e = \mathbf{B}_\text{к}$). Так как магнитное поле в сверхпроводник не проникает (мы пренебрегаем глубиной проникновения), то свободная энергия сверхпроводящего цилиндра будет $V_1\psi_s$, где V_1 — объем цилиндра. Сюда надо добавить энергию магнитного поля, локализованную вне цилиндра, т. е. $(V - V_1)B_\text{к}^2/(8\pi)$, где V — объем соленоида. Полная свободная энергия системы в начальном состоянии будет

$$\Psi_1 = V_1\psi_s + \frac{1}{8\pi}(V - V_1)B_\text{к}^2.$$

Во время перехода будет меняться магнитный поток Φ через поперечное сечение цилиндра и возбуждаться электродвижущая сила индукции $\mathscr{E}^\text{инд} = -(1/c)\,d\Phi/dt$. Для поддержания силы тока в соленоиде I

постоянной должна быть затрачена элементарная работа

$$-\mathscr{E}^{\text{инд}} I\,dt = \frac{1}{c}\,I\,d\Phi.$$

Поскольку ток I постоянен, полная затраченная работа во время перехода будет

$$A = \frac{1}{c}\,I\Phi = \frac{1}{c}\,ISB_e,$$

где S — площадь поперечного сечения соленоида. А так как $B_e = B_{\text{к}} = 4\pi I/l$, то $A = lSB_{\text{к}}^2/(4\pi) = V_1 B_{\text{к}}^2/(4\pi)$. В результате для свободной энергии системы в конечном состоянии получаем

$$\Psi_2 = \Psi_1 + A = V_1\left(\psi_s + \frac{1}{8\pi}\,B_{\text{к}}^2\right) + \frac{V}{8\pi}\,B_{\text{к}}^2.$$

Результат, разумеется, не зависит от способа перехода системы в конечное состояние, поскольку свободная энергия есть функция состояния. Но в конце перехода сверхпроводящий цилиндр перешел в нормальное состояние, а магнитное поле заполняет весь объем соленоида. Поэтому для Ψ_2 можно написать также

$$\Psi_2 = V_1\psi_n + \frac{V}{8\pi}\,B_{\text{к}}^2.$$

Сравнивая оба выражения для Ψ_2, находим

$$\psi_s + \frac{1}{8\pi}\,B_{\text{к}}^2 = \psi_n. \tag{80.6}$$

Этим соотношением и определяется критическое поле, при котором начинается разрушение сверхпроводимости.

Сверхпроводимость разрушается также электрическим током, когда последний превосходит известный предел (*критический ток*). Однако этот эффект является простым следствием только что рассмотренного. Действительно, ток, протекающий по сверхпроводнику, возбуждает магнитное поле. Когда оно достигает критического значения, начинается разрушение сверхпроводимости.

7. Состояние сверхпроводника однозначно определяется заданием внешнего приложенного поля $\mathbf{B}_e$ и температуры T (при условии, что давление, объем и прочие параметры поддерживаются постоянными), т. е. точкой на плоскости (T, B_e). На соответствующей диаграмме состояния (рис. 182) кривая равновесия делит плоскость на две части: одной соответствует сверхпроводящая фаза, а другой — нормальная. Кривая равновесия имеет параболическую форму. На ней располагаются точки $(T_{\text{к}}, B_{\text{к}})$. Из кривой равновесия видно, как магнитное поле понижает температуру перехода в сверхпроводящее состояние. Ранее введенную критическую температуру мы обозначаем теперь через $T_{\text{к}0}$, чтобы подчеркнуть, что это есть температура перехода в сверхпроводящее состояние при отсутствии магнитного поля ($B_e = 0$). Аналогичный смысл имеет обозначение $B_{\text{к}0}$.

8. До сих пор предполагалось, что сверхпроводящий образец имеет форму длинного цилиндра, а внешнее возбуждающее поле $\mathbf{B}_e$ однородно и параллельно оси образца. В этом случае магнитное поле на

Рис. 182

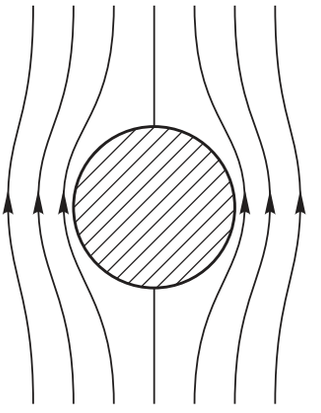


Рис. 183

поверхности образца имеет одинаковое значение во всех точках. При возрастании магнитного поля оно во всех этих точках достигнет критической величины $B_к$ одновременно. Когда это произойдет, сверхпроводящее состояние будет разрушено во всем объеме образца. Последний целиком перейдет в нормальное состояние.

Сложнее ведут себя образцы другой формы, так как в этих случаях критическое поле $\mathbf{B}_к$ в различных точках поверхности образца достигается при различных значениях $\mathbf{B}_e$. Рассмотрим простейший случай, когда образец имеет форму шара и внесен в однородное магнитное поле $\mathbf{B}_e$ (рис. 183). Максимального значения магнитное поле на поверхности шара достигает на экваторе, где оно составляет $(3/2)B_e$ (см. задачу 8). Поэтому разрушение сверхпроводимости $(3/2)B_e = B_к$. Но при таком значении поля B_e в нормальное состояние не может перейти весь шар. В противном случае (так как $\mu = 1$) магнитное поле во всем пространстве стало бы однородным и равным $\mathbf{B}_e = (3/2)\mathbf{B}_к$, т. е. меньше критического, а потому шар должен был бы вернуться в сверхпроводящее состояние. Разрушение сверхпроводимости в магнитном поле происходит путем «распада» шара на чередующиеся тонкие слои, а в случае тел более сложной формы — на малые области (*домены*) из сверхпроводящей и нормальной фаз, подобные доменам, на которые распадается ферромагнетик. Такое состояние распада сверхпроводника на сверхпроводящие и нормальные домены, связанное с его формой и обусловленное магнитным полем, называется *промежуточным состоянием*. Его существование было доказано прямыми опытами.

9. Одним из существенных факторов, определяющих поведение сверхпроводника, является *поверхностная энергия*, связанная с наличием границ раздела между нормальной и сверхпроводящей фазами. Эта энергия аналогична энергии поверхностного натяжения на границе раздела двух жидкостей. Однако в случае сверхпроводников поверхностная энергия может быть и положительной и отрицательной.

Сверхпроводники, для которых поверхностная энергия положительна, называются *сверхпроводниками* I *рода*, а сверхпроводники с отрицательной поверхностной энергией — *сверхпроводниками* II *рода*. Долгое время считалось, что все сверхпроводники являются сверхпроводниками I рода. Возможность существования сверхпроводников II рода была теоретически предсказана в 1957 г. А.А. Абрикосовым (р. 1928).

Допустим, что сверхпроводящий образец имеет форму длинного цилиндра, а внешнее магнитное поле прилагается параллельно его оси. В таком случае никакого промежуточного состояния, связанного с формой образца, не возникнет. Однако при возрастании магнитного поля $\mathbf{B}_e$ сверхпроводники I и II рода будут вести себя по-разному. Если внешнее приложенное поле $\mathbf{B}_e$ достигает критического значения $\mathbf{B}_к$, то в случае сверхпроводников I рода сверхпроводящая фаза становится термодинамически неустойчивой и весь образец переходит в более устойчивую — нормальную — фазу. Дробление сверхпроводника на малые сверхпроводящие и нормальные домены не наблюдается, так как на образование границ между ними требуется дополнительная затрата энергии. В соответствии с этим на диаграмме состояния сверхпроводника I рода существуют только две области: *сверхпроводящая* и *нормальная* (рис. 182). Не то будет в случае сверхпроводников II рода. Так как здесь поверхностная энергия отрицательна, то имеется возможность понижения полной свободной энергии системы путем дробления образца на сверхпроводящие и нормальные домены. Такое дробление не связано с формой образца, а является *внутренним свойством* сверхпроводников II рода. Кроме того, в отличие от промежуточного состояния, где доменная структура сравнительно грубая и видна невооруженным глазом, домены в сверхпроводниках II рода значительно мельче (порядка 10^{-5} см и меньше). Описанное состояние сверхпроводника II рода, в котором он существует в виде сверхпроводящих и нормальных доменов, называется *смешанным состоянием.* На диаграмме состояния сверхпроводника II рода (рис. 184) различают три области: *сверхпроводящую*, *область смешанного состояния* и *нормальную область.* Границы между этими областями определяют «*нижнее*» $B_{к1}$ и «*верхнее*» $B_{к2}$ *критические поля.* Сверхпроводники II рода (сверхпроводящие сплавы) нашли применения в качестве соленоидов, предназначенных для получения сильных магнитных полей (около 100 кГс). Сверхпроводники I рода для этой цели не годятся из-за низких значений критических магнитных полей, разрушающих сверхпроводимость.

Рис. 184

10. Физическая природа сверхпроводимости была понята только в 1957 г. Еще задолго до этого Ландау была создана *теория сверхтекучести* гелия II. Оказалось, что сверхтекучесть — это *макроскопический квантовый эффект.* Однако перенести теорию Ландау на явление сверхпроводимости мешало то обстоятельство, что атомы гелия, обладая нулевым спином, подчиняются *статистике Бозе* (1894–1974)–*Эйнштейна.* Электроны же, обладая половинным спином, подчиняются *принципу Паули* и *статистике Ферми–Дирака* (см. т. II, § 82). Для таких частиц невозможна *бозе–эйнштейновская конденсация*, необходимая для возникновения сверхтекучести. Решающий шаг в понимании сверхпроводимости был сделан американскими физиками Бардином (р. 1908), Купером (р. 1930) и Шриффером (р. 1931). Их теория в математическом отношении была усовершенствована советским физиком Н.Н. Боголюбовым (р. 1909). Основная идея заключается в следующем.

Между электронами металла существует кулоновское отталкивание. Однако оно в значительной степени ослаблено экранирующим действием положительных ионов решетки. Между тем взаимодействие электронов с колебаниями решетки приводит к слабому притяжению между ними. Это притяжение есть *квантовый эффект.* Оно при определенных условиях может превзойти кулоновское отталкивание между электронами, и тогда электроны группируются в пары (*куперовские пары*). Эти пары, обладая нулевым спином, ведут себя как *бозе-частицы* и переносят электрический ток в сверхпроводниках. Ограничимся этим замечанием, так как здесь невозможно входить в теорию сверхпроводимости по существу.

ЗАДАЧИ

1. Сверхпроводящие катушки с самоиндукциями L_1 и L_2 соединены параллельно и включены в цепь гальванической батареи с электродвижущей силой $\mathscr{E}$ (рис. 185). Найти установившиеся токи в катушках I_1 и I_2 и ток в общей цепи I, если коэффициентом взаимной индукции катушек можно пренебречь.

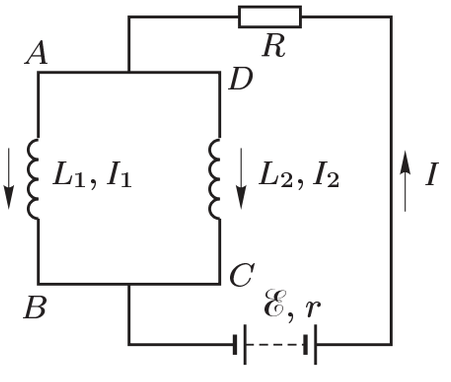


Рис. 185

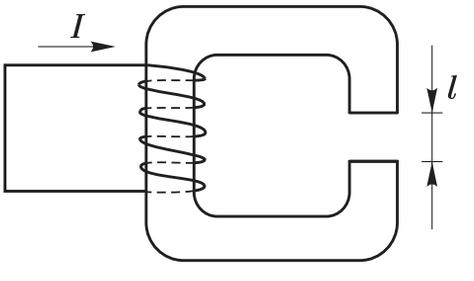


Рис. 186

Решение. По закону Ома $I = \mathscr{E}/(R + r)$. До включения батареи магнитный поток через сверхпроводящий контур $ABCD$ был равен нулю.

Он сохранится таковым и после включения тока, т. е. $L_1 I_1 - L_2 I_2 = 0$, откуда

$$I_1 / I_2 = L_2 / L_1. \tag{80.7}$$

Присоединив сюда уравнение $I_1 + I_2 = I$, найдем токи I_1 и I_2 в отдельности.

2. В замкнутой сверхпроводящей обмотке электромагнита возбужден постоянный ток I (рис. 186). В магнитопроводе из мягкого железа с магнитной проницаемостью μ имеется небольшой зазор толщины l, так что рассеянием магнитного поля в зазоре можно пренебречь. Длина магнитопровода (вместе с зазором) равна L. Чему будет равен ток I_0 в обмотке электромагнита, если в зазор вставить пластину толщины l из того же железа?

Ответ. $I_0 = \dfrac{I}{1 + (\mu - 1) l / L}$.

3. Длинный сверхпроводящий цилиндр (из сверхпроводника I рода) внесен в постоянное однородное магнитное поле $\mathbf{B}$, направленное параллельно оси цилиндра. Определить силу f, действующую на единицу площади боковой поверхности цилиндра.

Ответ. $f = B^2/(8\pi)$ (давление, направленное к оси цилиндра).

4. Кольцо из тонкой проволочки помещено в однородное магнитное поле $\mathbf{B} = 10$ Гс, перпендикулярное к плоскости кольца, и охлаждением переведено в сверхпроводящее состояние. Найти силу тока в кольце после выключения магнитного поля, если радиус кольца $R = 5$ см, а радиус проволоки $a = 1$ мм.

Указание. Самоиндукция тонкого проволочного кольца (если ток течет по его поверхности) в гауссовой системе дается выражением

$$L = 4\pi R [\ln (8R/a) - 2].$$

Ответ. $I = c\pi R^2 B / L = 9{,}4 \cdot 10^{10}$ СГСЭ-ед. $= 31$ А.

5. Над плоской поверхностью сверхпроводника I рода параллельно этой поверхности подвешен тонкий прямолинейный провод на расстоянии h от плоскости, по которому течет постоянный ток I. Найти линейную плотность сверхпроводящего тока i, текущего по поверхности сверхпроводника.

Указание. Применить метод зеркальных изображений.

Ответ. $i = Ih/(\pi r^2)$, где r — расстояние между проводом и точкой наблюдения. Сверхпроводящий ток параллелен току I и течет в противоположном направлении.

6. Над плоской поверхностью сверхпроводника I рода на изолирующем слое толщины $h = 5$ мм лежит тонкое сверхпроводящее кольцо радиуса $R = 10$ см, по которому течет постоянный ток I. При каком токе I кольцо начнет парить над сверхпроводником, если масса кольца $m = 1$ г?

Ответ. $I \geqslant c\sqrt{mgh/(2\pi R)} = 8{,}4 \cdot 10^{10}$ СГСЭ-ед. $= 25$ А.

7. На какой высоте h постоянный магнитик с магнитным моментом $\mathfrak{M} = 10^3$ Гс $\cdot$ см^3 и массой $m = 10$ г будет парить в горизонтальном положении над плоской горизонтальной поверхностью сверхпроводника I рода? Магнитик считать точечным диполем.

Указание. Применить метод зеркальных изображений. Воспользоваться выражением для взаимной потенциальной энергии двух точечных диполей.

Ответ. $h = \dfrac{1}{2}\left(\dfrac{3\mathfrak{M}^2}{mg}\right)^{1/4} \approx 2{,}1$ см.

8. Шар радиуса a из сверхпроводника I рода внесен в постоянное однородное магнитное поле $\mathbf{B}_0$. Определить магнитное поле $\mathbf{B}$ вне шара, если поле $\mathbf{B}_0$ еще разрушает сверхпроводимость в шаре. Найти также поверхностную плотность тока сверхпроводимости r.

Р е ш е н и е. Введем предположение, оправдываемое последующим расчетом, что вне шара на однородное поле $\mathbf{B}_0$ накладывается поле точечного диполя $\mathfrak{M}$, помещенного в центре шара (рис. 187). Тогда

$$\mathbf{B} = \mathbf{B}_0 + \frac{3(\mathfrak{M}r)}{r^5}\, r - \frac{\mathfrak{M}}{r^3}. \qquad (80.8)$$

Момент $\mathfrak{M}$ определится из условия, чтобы нормальная составляющая вектора $\mathbf{B}$ на поверхности шара обращалась в нуль:

$$\mathbf{B}_r = \mathbf{B}_0 \cos\vartheta + \frac{2\mathfrak{M}}{a^3}\cos\vartheta = 0.$$

Этому условию при любых значениях угла ϑ удовлетворяет

$$\mathfrak{M} = -\frac{a^3}{2}\,\mathbf{B}_0. \qquad (80.9)$$

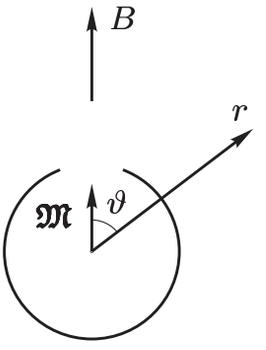


Рис. 187

На экваторе, т. е. при $r = a$, $\vartheta = 90^\circ$, получаем $\mathbf{B} = (3/2)\mathbf{B}_0$. Поверхностный ток сверхпроводимости течет вдоль параллелей $\vartheta = \text{const}$, причем по теореме о циркуляции $i = (3c/8\pi)B_0 \sin\vartheta$.

9. Бесконечно длинный цилиндр радиуса a из сверхпроводника I рода внесен в постоянное однородное магнитное поле $\mathbf{B}_0$, перпендикулярное к оси цилиндра. Определить магнитное поле $\mathbf{B}$ вне цилиндра, если поле $\mathbf{B}_0$ еще не разрушает сверхпроводимость в цилиндре. Найти также поверхностную плотность тока сверхпроводимости i.

О т в е т.

$$\mathbf{B} = \mathbf{B}_0 - \frac{2a^2(\mathbf{B}_0\mathbf{r})}{r^4}\,\mathbf{r} + \frac{a^2}{r^4}\,\mathbf{B}_0, \qquad (80.10)$$

где $\mathbf{r}$ — радиус-вектор, перпендикулярный к оси цилиндра, проведенный от оси к точке наблюдения. Поле на поверхности цилиндра максимально при $\vartheta = 90^\circ$ и равно $\mathbf{B} = 2\mathbf{B}_0$ (см. рис. 187). Поверхностный ток течет параллельно оси цилиндра, причем $i = (c/2\pi)B_0 \sin\vartheta$.

10. Сверхпроводящий шарик летит по направлению к соленоиду вдоль его оси. Поле в центре соленоида $B = 1000$ Гс. Какова должна быть начальная скорость v шарика, чтобы он смог пролететь через соленоид насквозь? Диаметр соленоида много больше диаметра шарика. Плотность материала шарика $\rho = 8$ г/см^3.

Р е ш е н и е. Пользуясь результатом решения задачи 8, легко показать, что потенциальная энергия в центре соленоида равна $\frac{1}{2}(\mathfrak{M}\mathbf{B}) = \frac{1}{4}B^2R^3$. Применяя далее уравнение энергии, находим

$$v \geqslant B\sqrt{3/(4\pi\rho)} \approx 170 \text{ см/с}.$$

11. Для получения сверхсильных магнитных полей в катушках применяются обмотки из сверхпроводников II рода. Концы обмоток (пока проволока находится в нормальном состоянии) присоединяются к источнику тока. Затем катушки охлаждаются до гелиевых температур и переходят в сверхпроводящее (или смешанное) состояние. После этого катушки отключаются от источника тока и замыкаются накоротко.

Сверхпроводящий короткозамкнутый многослойный соленоид с внутренним радиусом $a = 2$ см, наружным $b = 4$ см и длиной $l = 20$ см создает внутри поле $B_0 = 100$ кГс. Оценить количество жидкого гелия, которое испарится

при переходе соленоида из сверхпроводящего (или смешанного) в нормальное состояние. Теплота испарения гелия $q = 0{,}7$ кал/см^3.

Р е ш е н и е. Пренебрегаем краевыми эффектами. Ток в обмотке аппроксимируем током с постоянной объемной плотностью, что в оценочных вычислениях допустимо. Применяя теорему о циркуляции, находим магнитное поле B на различных расстояниях r от оси соленоида:

$$B = \begin{cases} B_0 & \text{при } r \leqslant a, \\ B_0(b-r)/(b-a) & \text{при } a \leqslant r \leqslant b. \end{cases}$$

После этого легко находится магнитная энергия W системы:

$$W = \frac{lB_0^2}{48}\,[3a(a+b) - b(b-a)] \approx 1{,}17 \cdot 10^{11} \text{ эрг} \approx 2{,}8 \cdot 10^3 \text{ кал}.$$

Количество испарившегося гелия $W/q \approx 4 \cdot 10^3$ см$^3 = 4$ л.

Глава IV
УРАВНЕНИЯ МАКСВЕЛЛА

§ 81. Ток смещения

1. Основные уравнения электромагнитного поля в неподвижных средах, применимые не только к постоянным, но и к переменным электромагнитным полям, были установлены Максвеллом. К уравнениям Максвелла можно прийти путем последовательного обобщения опытных фактов. Надо решить, какие из полученных ранее уравнений могут быть сохранены, какие должны быть отброшены и какие надо обобщить. Есть один руководящий принцип, который позволяет продвинуться в этом направлении. Следует исключить из числа основных такие уравнения, в основе которых лежит представление о непосредственном действии на расстоянии. К ним относятся законы Кулона, Био и Савара и пр. Эти законы несовместимы с экспериментально подтвержденным представлением о конечной скорости распространения взаимодействий, а потому не могут оставаться верными во всех случаях. Можно сохранить только такие уравнения, которые не противоречат представлениям теории поля. Так мы и поступали во всем предшествовавшем изложении. Мы выдвинули в качестве гипотезы, что теорема Гаусса (13.4), уравнение (58.1) и закон электромагнитной индукции (66.1) являются общими законами электродинамики. То обстоятельство, что они удовлетворяют требованиям теории поля, следует из того, что их можно представить в дифференциальной форме (13.5), (58.2) и (66.4). К основным уравнениям электродинамики мы присоединим и закон сохранения электрического заряда. В дифференциальной форме он имеет вид

$$\frac{\partial \rho}{\partial t} + \operatorname{div} \mathbf{j} = 0. \tag{81.1}$$

Если электромагнитное поле стационарно, то это уравнение переходит в

$$\operatorname{div} \mathbf{j} = 0. \tag{81.2}$$

2. Теорема о циркуляции

$$\oint_L \mathbf{H}\, d\mathbf{l} = \frac{4\pi}{c}\, \mathbf{I} \tag{81.3}$$

также может быть преобразована в дифференциальную форму:

$$\operatorname{rot} \mathbf{H} = \frac{4\pi}{c}\, \mathbf{j}, \tag{81.4}$$

а потому удовлетворяет требованиям теории поля. Однако она не может входить в число основных уравнений электродинамики. Действительно, дивергенция всякого ротора тождественно равна нулю. Поэтому, взяв дивергенцию от обеих частей уравнения (81.4), получим $\operatorname{div}\mathbf{j} = 0$. Но это соотношение справедливо только для стационарных токов. В общем случае оно противоречит уравнению (81.1). Сомневаться в справедливости уравнения (81.1) нет оснований, так как оно выражает *закон сохранения электрического заряда.* Отсюда следует, что уравнения (81.3) и (81.4) могут быть верны только для стационарных токов. Для переменных электромагнитных полей они должны быть обобщены.

Чтобы прийти к обобщенным уравнениям, воспользуемся следующим наводящим рассуждением. Поскольку дивергенция левой части уравнения (81.4) тождественно равна нулю, в правой части этого уравнения должен стоять вектор, дивергенция которого также всегда равна нулю. В случае стационарных электромагнитных полей этот вектор должен переходить в $\mathbf{j}$. Легко указать вектор, удовлетворяющий этим условиям. Дифференцируя по времени соотношение $\operatorname{div}\mathbf{D} = 4\pi\rho$, получаем

$$\frac{\partial\rho}{\partial t} - \frac{1}{4\pi}\operatorname{div}\dot{\mathbf{D}} = 0,$$

или ввиду уравнения (81.1)

$$\operatorname{div}\left(\mathbf{j} + \frac{1}{4\pi}\dot{\mathbf{D}}\right) = 0. \tag{81.5}$$

Величину

$$\mathbf{j}_{\text{см}} = \frac{1}{4\pi}\dot{\mathbf{D}} \tag{81.6}$$

Максвелл назвал *током* (точнее, *плотностью тока*) *смещения*, а сумму $\mathbf{j} + \mathbf{j}_{\text{см}}$ — *полным током.* Таким образом,

$$\operatorname{div}(\mathbf{j} + \mathbf{j}_{\text{см}}) = 0, \tag{81.7}$$

т. е. *полный ток всегда соленоидален.* Поэтому противоречие с уравнением (81.1) устранится, если в уравнении (81.4) ток проводимости $\mathbf{j}$ заменить полным током, т. е. написать

$$\operatorname{rot}\mathbf{H} = \frac{4\pi}{c}(\mathbf{j} + \mathbf{j}_{\text{см}}). \tag{81.8}$$

Так и поступил Максвелл.

Приведенные рассуждения ни в какой мере не могут служить доказательством уравнения (81.8). На них следует смотреть только как на один из бесконечного множества способов устранения математического противоречия между уравнениями (81.1) и (81.4). А что таких способов бесконечно много, видно уже из того, что не возникает новых математических противоречий, если в правой части уравнения (81.8) добавить произвольный вектор, дивергенция которого равна нулю. Настоящим доказательством уравнения (81.8) могут служить только опытные факты, подтверждающие это уравнение.

3. К необходимости обобщения уравнений (81.3) и (81.4) можно прийти также с помощью других соображений. Приведем два примера.

Пример 1. Пусть в неограниченную однородную среду помещен металлический шар, которому сообщен электрический заряд Q (рис. 188). Если среда проводящая, то появятся электрические токи, текущие в радиальных направлениях. Они будут возбуждать магнитное поле. При попытке указать его направление возникает следующая трудность. Вектор $\mathbf{B}$ не может иметь радиальной составляющей. Система сферически симметрична. Если бы радиальная составляющая вектора $\mathbf{B}$ существовала, то она была бы одной и той же во всех точках всякой сферы S, концентрической с поверхностью шара. Радиальная составляющая $\mathbf{B}$ на сфере S была бы всюду направлена либо от центра, либо к центру шара. В обоих случаях поток вектора $\mathbf{B}$ через сферу S был бы отличен от нуля, что противоречит уравнению (58.1). Следовательно, вектор $\mathbf{B}$ должен быть перпендикулярен к радиусу, проведенному из центра шара к рассматриваемой точке. А это также невозможно, так как все направления, перпендикулярные к радиусу, ничем не выделены — все они совершенно равноправны. Единственная возможность, допускаемая симметрией шара, состоит в том, что векторы $\mathbf{B}$ и $\mathbf{H}$ всюду должны равняться нулю. Но в таком случае должен равняться нулю и ток $\mathbf{j}$, как это непосредственно следует из уравнения (81.4). Значит, уравнение (81.4) и эквивалентное ему уравнение (81.3) в рассматриваемом случае не могут быть верными.

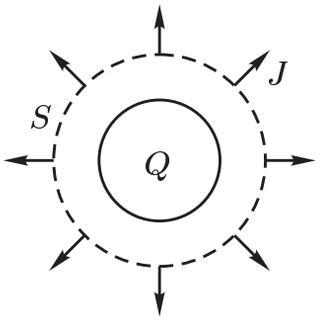


Рис. 188

Для устранения возникшего противоречия необходимо допустить, что магнитные поля возбуждаются не только токами проводимости, а еще чем-то. К току проводимости I надо что-то добавить, чтобы уничтожить возбуждаемое им магнитное поле. Эта добавка и есть ток смещения. Его значение $I_{\text{см}}$ определится из условия $I + I_{\text{см}} = 0$. Полный ток проводимости, текущий от заряженного шара, связан с зарядом Q соотношением $I = -dQ/dt$, а потому $I_{\text{см}} = dQ/dt$. По закону Кулона $Q = r^2 D$. Дифференцируя это выражение и разделив результат на поверхность сферы $4\pi r^2$, найдем плотность тока смещения:

$$\mathbf{j}_{\text{см}} = \frac{\dot{\mathbf{Q}}}{4\pi r^2} = \frac{1}{4\pi}\dot{\mathbf{D}}.$$

Это выражение совпадает с (81.6).

4. Пример 2. Соединим проводом обкладки плоского заряженного конденсатора (рис. 189). По проводу потечет электрический ток. Допущение, что в этом случае применима формула (81.3), снова приводит к трудностям. Циркуляция вектора $\mathbf{H}$, стоящего в левой части уравнения (81.3), зависит только от формы и расположения контура L. Она — величина вполне определенная. Между тем ток I, стоящий в правой части того же уравнения, таким свойством не обладает. Для

определения I надо мысленно натянуть на контур L какую-то поверхность S и найти пронизывающий ее ток. Однако сила переменного тока может меняться вдоль провода. В этих случаях величина I будет зависеть от того, в каком месте поверхность S пересекается с проводом. С особой отчетливостью указанная неопределенность проявится, если поверхность S провести между обкладками конденсатора, нигде не пересекая провода. Тогда $I = 0$. Для устранения неопределенности к току I в уравнении (81.3) надо добавить какое-то слагаемое $I_{\text{см}}$, чтобы сумма $I + I_{\text{см}}$ не зависела от выбора вспомогательной поверхности S. Это слагаемое и есть ток смещения.

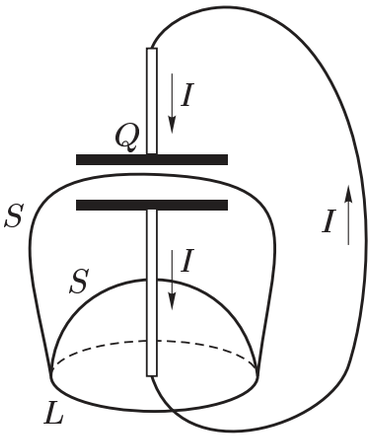


Рис. 189

Независимость полного тока $I + I_{\text{см}}$ от формы поверхности, натянутой на один и тот же контур L, эквивалентна утверждению, что полный ток через любую замкнутую поверхность всегда равен нулю. Токи, удовлетворяющие этому условию, не совсем удачно называются *замкнутыми.* Замкнутость токов не следует понимать в смысле замкнутости линий тока. Если линии тока замкнуты, то и сами токи также замкнуты. Обратное справедливо не всегда: линии тока в случае замкнутых токов не обязательно должны быть сами замкнутыми. Таким образом, формальное содержание гипотезы Максвелла сводится к утверждению, что *полные токи всегда замкнуты.* Токи проводимости, если они не замкнуты, замыкаются токами смещения.

Из условия замкнутости полного тока можно получить и выражения для тока смещения и его плотности. Обратимся снова к примеру с конденсатором. Идеализируя систему, можно сказать, что по проводу течет только ток проводимости, а через конденсатор — только ток смещения. Ток смещения дополняет ток проводимости до замкнутого тока. Поэтому ток проводимости в проводе должен быть равен току смещения в конденсаторе: $I_{\text{см}} = \dot{Q}$, где Q — заряд на той пластине конденсатора, к которой течет ток. Очевидно, $Q = S\sigma = SD/(4\pi)$. Дифференцируя по времени и разделив на S, снова получаем

$$\mathbf{j}_{\text{см}} = \frac{1}{4\pi}\dot{\mathbf{D}}.$$

Таким образом, тремя различными способами мы приходим к одному и тому же выражению для плотности тока смещения.

5. Токи смещения существуют только там, где меняется электрическое поле (точнее, электрическое смещение $\mathbf{D}$). Поэтому физическое содержание гипотезы Максвелла о токах смещения сводится к утверждению, что *переменные электрические поля являются источниками магнитных полей.* Это открытие принадлежит всецело Максвеллу.

Оно вполне аналогично открытию электромагнитной индукции, согласно которому переменные магнитные поля возбуждают поля электрические.

6. Ток смещения в диэлектрике состоит из двух существенно различных слагаемых. По определению вектора электрической индукции $\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P}$, а потому

$$\mathbf{j}_{\text{см}} = \frac{1}{4\pi}\dot{\mathbf{E}} + \dot{\mathbf{P}}. \tag{81.9}$$

Величина $\dot{\mathbf{P}}$ называется *плотностью тока поляризации.* Вектор поляризации определяется выражением $\mathbf{P} = \sum e_i \mathbf{r}_i$. Суммирование ведется по всем связанным зарядам, находящимся в единице объема вещества. Дифференцируя это выражение по времени, получим

$$\mathbf{j}_{\text{пол}} \equiv \dot{\mathbf{P}} = \sum e_i \mathbf{v}_i,$$

где $\mathbf{v}_i$ — скорость движения i-го заряда. Таким образом, *ток поляризации есть электрический ток, обусловленный движением связанных зарядов.* Последние принципиально ничем не отличаются от свободных зарядов. Поэтому нет ничего неожиданного в том, что токи поляризации возбуждают магнитное поле. Принципиально новое содержится в утверждении, что и вторая часть тока смещения $(1/4\pi)\dot{\mathbf{E}}$, которая не связана ни с каким движением зарядов, а обусловлена только изменениями электрического поля по времени, также является источником магнитного поля. *Даже в вакууме всякое изменение электрического поля во времени возбуждает в окружающем пространстве магнитное поле.* Открытие этого обстоятельства — наиболее существенный и решающий шаг, сделанный Максвеллом при построении своей электродинамики.

ЗАДАЧИ

1. Пространство между обкладками длинного цилиндрического конденсатора заполнено однородным диэлектриком со слабой электрической проводимостью. Когда конденсатор заряжен, в диэлектрике от одной обкладки к другой течет электрический ток. Пренебрегая краевыми эффектами, найти напряженность магнитного поля между обкладками.

Ответ. $H = 0$.

2. Заряженный и отключенный от источника электричества плоский конденсатор медленно разряжается объемными токами проводимости, возникающими в диэлектрике между обкладками из-за наличия слабой электрической проводимости. Пренебрегая краевыми эффектами, вычислить напряженность магнитного поля внутри конденсатора.

Решение. Если σ — поверхностная плотность электричества на положительной обкладке, то $D = 4\pi\sigma$ и, следовательно, $j_{\text{см}} = (1/4\pi)\dot{D} = \dot{\sigma}$. По закону сохранения электрического заряда $j = -\dot{\sigma}$. Следовательно, $j + j_{\text{см}} = j_{\text{полн}} = 0$. Магнитное поле в конденсаторе равно нулю.

3. Заряженный и отключенный от источника электричества плоский конденсатор, состоящий из двух одинаковых дисков радиуса R, пробивается электрической искрой вдоль своей оси. Считая разряд квазистационарным

и пренебрегая краевыми эффектами, вычислить мгновенное значение напряженности магнитного поля внутри конденсатора (в зависимости от расстояния r до его оси), если сила тока в электрической искре в рассматриваемый момент времени равна I.

Решение. В силу симметрии магнитные силовые линии будут коаксиальными окружностями с общей осью, совпадающей с осью конденсатора. Поле H найдется по формуле

$$\oint H\,dl = 2\pi r H = \frac{4\pi}{c}\,(I + i_{\text{см}}),$$

где $i_{\text{см}} = Ir^2/R^2$ — ток смещения, пронизывающий круг радиуса r. В результате получим

$$H = \frac{2I}{cr}\left(1 - \frac{r^2}{R^2}\right).$$

4. Плоский конденсатор состоит из двух одинаковых металлических дисков, пространство между которыми заполнено однородным диэлектриком с диэлектрической проницаемостью ε. Расстояние между внутренними поверхностями дисков равно d. Между обкладками конденсатора поддерживается переменное напряжение $V = V_0 \sin \omega t$. Пренебрегая краевыми эффектами, найти магнитное поле в пространстве между обкладками конденсатора.

Ответ. $H = \dfrac{\varepsilon\omega r}{2cd}\,V_0 \cos \omega t$, где r — расстояние от оси конденсатора. Магнитные силовые линии имеют форму коаксиальных окружностей с общей осью, совпадающей с осью конденсатора.

§ 82. Система уравнений Максвелла

1. Дополнив основные факты из области электромагнетизма установлением магнитных действий токов смещения, Максвелл смог написать *систему фундаментальных уравнений электродинамики.* Таких уравнений четыре. В интегральной форме они имеют вид

$$\oint\limits_L \mathbf{H}\,d\mathbf{l} = \frac{4\pi}{c}\int\limits_S \left(\mathbf{j} + \frac{1}{4\pi}\,\frac{\partial \mathbf{D}}{\partial t}\right) d\mathbf{S}, \tag{82.1}$$

$$\oint\limits_L \mathbf{E}\,d\mathbf{l} = -\frac{1}{c}\int\limits_S \frac{\partial \mathbf{B}}{\partial t}\,d\mathbf{S}, \tag{82.2}$$

$$\oint\limits_S (\mathbf{D}\,d\mathbf{S}) = 4\pi \int \rho\,dV, \tag{82.3}$$

$$\oint\limits_S (\mathbf{B}\,d\mathbf{S}) = 0. \tag{82.4}$$

В дифференциальной форме:

$$\operatorname{rot} \mathbf{H} = \frac{4\pi}{c}\,\mathbf{j} + \frac{1}{c}\,\frac{\partial \mathbf{D}}{\partial t}, \tag{82.1a}$$

$$\operatorname{rot}\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}, \tag{82.2a}$$

$$\operatorname{div}\mathbf{D} = 4\pi\rho, \tag{82.3a}$$

$$\operatorname{div}\mathbf{B} = 0. \tag{82.4a}$$

В число фундаментальных не включено уравнение непрерывности, выражающее закон сохранения электрического заряда, так как это уравнение является следствием уравнений (82.1) и (82.3). Действительно, возьмем бесконечно малый контур L, натянем на него произвольную конечную поверхность S, а затем стянем этот контур в точку, оставляя поверхность S конечной. В пределе циркуляция $\oint \mathbf{H}\, d\mathbf{l}$ обратится в нуль, S превратится в замкнутую поверхность, а уравнение (82.1) перейдет в

$$\oint\left(\mathbf{j} + \frac{1}{4\pi}\dot{\mathbf{D}}\right) d\mathbf{S} = 0.$$

Интеграл $\oint (\mathbf{j}\, d\mathbf{S})$ есть ток I, вытекающий из объема V, ограниченного поверхностью S. Кроме того, записав уравнение (82.3) в виде

$$\oint (\mathbf{D}\, d\mathbf{S}) = 4\pi q$$

и дифференцируя его по времени, получим

$$\frac{\partial q}{\partial t} = \frac{1}{4\pi}\oint (\dot{\mathbf{D}}\, d\mathbf{S}).$$

В результате получится уравнение

$$\frac{\partial q}{\partial t} = -I,$$

выражающее закон сохранения электрического заряда. Тот же закон можно получить из дифференциальных уравнений (82.1а) и (82.3а). Достаточно взять дивергенцию от обеих частей уравнения (82.2а) и воспользоваться уравнением (82.3а). Тогда получится уравнение (81.1).

Уравнения Максвелла показывают, что *источниками электрического поля могут быть либо электрические заряды, либо магнитные поля, меняющиеся во времени. Магнитные же поля могут возбуждаться либо движущимися электрическими зарядами* (*электрическими токами*), *либо переменными электрическими полями.* Уравнения не симметричны относительно электрического и магнитного полей. Это связано с тем, что в природе существуют электрические заряды, но, насколько известно в настоящее время, нет зарядов магнитных. Стремление достигнуть симметрии уравнений электродинамики заставило Дирака выдвинуть гипотезу о существовании магнитных зарядов — *единичных магнитных полюсов*, или *монополей.* Логических возражений против такой гипотезы нет. Если бы она оправдалась, то потребовалось бы обобщение уравнений Максвелла. К источникам магнитного поля добавились бы магнитные заряды, а к источникам электрического поля — магнитные токи, обусловленные движением

таких зарядов. Справедливость же самих уравнений Максвелла была бы ограничена теми областями пространства, в которых нет магнитных зарядов и магнитных токов. Однако многочисленные попытки экспериментально обнаружить магнитные монополи не привели к положительному результату.

2. Уравнения Максвелла в интегральной форме справедливы и в тех случаях, когда существуют *поверхности разрыва*, на которых свойства среды или напряженности электрического и магнитного полей меняются скачкообразно. Поэтому в этой форме уравнения Максвелла обладают большей общностью, чем в дифференциальной форме, которая предполагает, что все величины в пространстве и во времени меняются непрерывно. Можно, однако, достигнуть полной математической эквивалентности обеих форм уравнений Максвелла. Для этого надо дифференциальные уравнения дополнить *граничными условиями*, которым должно удовлетворять электромагнитное поле на границе раздела двух сред. Эти условия содержатся в интегральной форме уравнений Максвелла. Они были выведены в соответствующих местах курса и имеют вид

$$D_{2n} - D_{1n} = 4\pi\sigma, \tag{82.5}$$

$$B_{1n} = B_{2n}, \tag{82.6}$$

$$E_{1t} = E_{2t}, \tag{82.7}$$

$$[\mathbf{n}\mathbf{H}_2] - [\mathbf{n}\mathbf{H}_1] = \frac{4\pi}{c}\,\mathbf{i}. \tag{82.8}$$

Здесь σ — поверхностная плотность электрических зарядов, а $\mathbf{i}$ — поверхностная плотность тока проводимости на рассматриваемой границе раздела. В частном случае, когда поверхностных токов нет, последнее условие переходит в

$$H_{1t} = H_{2t}. \tag{82.9}$$

Подчеркнем еще раз, что рассуждения, с помощью которых мы пришли к уравнениям Максвелла, ни в коей мере не могут служить их доказательством. Существенно новые принципы никогда не содержатся в старой теории и не могут быть выведены из нее логически. В этом смысле нельзя вывести и уравнения Максвелла. На них следует смотреть как на *основные аксиомы электродинамики, полученные путем обобщения опытных фактов.*

3. Фундаментальные уравнения Максвелла в форме (82.1)–(82.4) или (82.1а)–(82.4а) не составляют еще полной системы уравнений электромагнитного поля. Среди них два уравнения векторных и два скалярных. Если их записать в координатной форме, то получится всего восемь уравнений, связывающих 16 величин: пятнадцать составляющих векторов $\mathbf{E}$, $\mathbf{D}$, $\mathbf{B}$, $\mathbf{H}$, $\mathbf{j}$ и скаляр ρ. Ясно, что для 16 величин восьми уравнений недостаточно. Фундаментальные уравнения Максвелла не содержат никаких постоянных, характеризующих свойства среды, в которой возбуждено электромагнитное поле. Необходимо дополнить эти

уравнения такими соотношениями, в которые входили бы величины, характеризующие индивидуальные свойства среды. Эти соотношения называются *материальными уравнениями.*

Принципиальный способ получения материальных уравнений дают молекулярные теории поляризации, намагничивания и электрической проводимости среды. В основе таких теорий лежат какие-то идеализированные модели среды. Применяя к ним уравнения классической или квантовой механики, а также методы статистической физики, можно установить связь между векторами $\mathbf{P}$, $\mathbf{I}$, $\mathbf{j}$, с одной стороны, и векторами $\mathbf{E}$ и $\mathbf{B}$ — с другой. Таким путем, в зависимости от характера среды и электромагнитного поля, получаются более или менее сложные соотношения, которые и дополняют фундаментальные уравнения Максвелла до полной системы уравнений электродинамики.

Наиболее просты материальные уравнения в случае слабых электромагнитных полей, сравнительно медленно меняющихся в пространстве и во времени. В этом случае для изотропных неферромагнитных и несегнетоэлектрических сред материальные уравнения могут быть записаны в виде

$$\mathbf{D} = \varepsilon \mathbf{E}, \tag{82.10}$$

$$\mathbf{B} = \mu \mathbf{H}, \tag{82.11}$$

$$\mathbf{j} = \lambda \mathbf{E}, \tag{82.12}$$

где ε, μ, λ — постоянные, характеризующие электромагнитные свойства среды. Они называются *диэлектрической* и *магнитной проницаемостью* и *электрической проводимостью среды.* Такими материальными уравнениями пользовался сам Максвелл. Разумеется, он не связывал величины ε, μ, λ с атомными и молекулярными константами вещества, а рассматривал их как постоянные, вводимые в теорию *феноменологически.* Электронная теория показала, что справедливость таких материальных уравнений связана с выполнением двух условий. Во-первых, за времена порядка собственных периодов внутриатомных и внутримолекулярных колебаний электромагнитное поле должно меняться мало. Во-вторых, поле должно меняться мало на протяжении межатомных и межмолекулярных расстояний. Это и есть та «медленность» изменения полей, о которой говорилось выше.

Иногда уравнения (82.10)–(82.12) также включают в систему уравнений Максвелла. Мы не будем этого делать, так как эти уравнения не обладают той общностью и фундаментальностью, которая свойственна уравнениям Максвелла. Под уравнениями Максвелла мы будем понимать только четыре уравнения: (82.1)–(82.4) или (82.1а)–(82.4а).

Когда поля стационарны ($\partial \mathbf{D}/\partial t = \partial \mathbf{B}/\partial t = 0$), уравнения Максвелла распадаются на две группы независимых уравнений. Первую группу составляют уравнения электростатики

$$\operatorname{rot} \mathbf{E} = 0, \quad \operatorname{div} \mathbf{D} = 4\pi\rho, \tag{82.13}$$

вторую — уравнения магнитостатики

$$\operatorname{rot}\mathbf{H}=\frac{4\pi}{c}\mathbf{j},\quad \operatorname{div}\mathbf{B}=0. \tag{82.14}$$

В этом случае электрическое и магнитное поля независимы друг от друга. Источниками электрического поля будут только электрические заряды, источниками магнитного поля — только токи проводимости.

§ 83. Скорость распространения электромагнитных возмущений

1. Из уравнений максвелловской электродинамики следует существование принципиально нового физического явления, предсказанного самим Максвеллом. Это — *электромагнитные волны*, или *возмущения*, распространяющиеся в пространстве с определенной скоростью. Убедимся в этом на простейшем примере. Рассмотрим бесконечно протяженную однородную диэлектрическую среду с диэлектрической и магнитной проницаемостями ε и μ. Поместим в нее бесконечную равномерно заряженную плоскость, которую примем за координатную плоскость XY (рис. 190). Пока плоскость вместе с зарядами на ней неподвижна, электрическое поле в окружающем пространстве будет нормально к плоскости и равно $E=2\pi\sigma/\varepsilon$, как это следует из теоремы Гаусса и соображений симметрии (σ — поверхностная плотность заряда, см. § 6 и 13). Приведем теперь плоскость вместе с зарядами на ней в движение в направлении оси X с произвольно меняющейся скоростью. Тогда, как будет показано ниже, появится магнитное поле и поперечная составляющая электрического поля, т. е. составляющая, параллельная заряженной плоскости. Это — *переменные поля*, которые и будут интересовать нас в настоящем параграфе. Что касается нормальной составляющей вектора $\mathbf{E}$, то она останется без изменений, так как рассуждения, с помощью которых была получена формула (6.1), сохраняют силу и в рассматриваемом случае, если только величину E заменить на D_n. Нормальная составляющая вектора $\mathbf{E}$ есть просто статическое электрическое поле заряженной плоскости, накладывающееся на переменное электромагнитное поле движущихся зарядов. Так как статическое поле нас не интересует, то можно совсем не обращать внимания на наличие нормальной составляющей поля $\mathbf{E}$, что мы и будем делать. Можно было бы совсем избавиться от нормальной составляющей, поместив бесконечно близко от заряженной плоскости вторую такую же неподвижную плоскость, заряженную противоположно. Электрическое поле такой плоскости было бы чисто статическим и уничтожило бы нормальную составляющую,

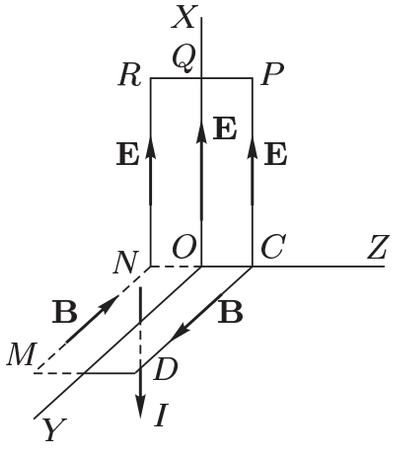


Рис. 190

о которой мы говорили. В то же время новая заряженная плоскость, поскольку заряды на ней неподвижны, не оказала бы никакого влияния на переменное поле электромагнитной волны.

2. Исследуем теперь, какое электромагнитное поле возбуждается заряженной плоскостью *благодаря ее движению*. Заряды, движущиеся вместе с плоскостью, эквивалентны электрическому току I, текущему параллельно оси X (на рис. 190 ток течет в отрицательном направлении оси X). Электрический ток возбуждает магнитное поле, силовые линии которого обвиваются вокруг тока. В случае поверхностного тока, текущего по бесконечной плоскости, магнитное поле будет параллельно оси Y: при $z > 0$ оно направлено в положительную, а при $z < 0$ — в отрицательную сторону оси Y. В силу симметрии магнитные поля по разные стороны заряженной плоскости на одинаковых расстояниях от нее одинаковы по абсолютной величине, но противоположны по направлению. При переходе через заряженную плоскость магнитное поле, в согласии с формулой (82.8), испытывает скачок непрерывности, связанный с поверхностным током.

Магнитные потоки через прямоугольные, симметрично расположенные неподвижные контуры $OCPQ$ и $NOQR$ одинаковы по абсолютной величине, но противоположны по знаку. Так как эти потоки меняются во времени, то по закону электромагнитной индукции возникнет электрическое поле, циркуляция которого вдоль рассматриваемых контуров отлична от нуля. Электрическое поле будет параллельно оси X, так как вдоль оси Z, как выяснено выше, переменное электрическое поле не возбуждается, а электрического поля вдоль оси Y не будет в силу симметрии. Так как магнитный поток через контур $NCPR$ равен нулю, то будет равна нулю и циркуляция вектора $\mathbf{E}$ по тому же контуру. Отсюда следует, что переменное электрическое поле по разные стороны заряженной плоскости на одинаковых расстояниях от нее будет одинаково не только по величине, но и по направлению. На самой плоскости поле $\mathbf{E}$, вообще говоря, будет иметь другое значение. В противном случае циркуляции вектора $\mathbf{E}$ по контурам $OCPQ$ и $NOQR$ обратились бы в нуль, а, следовательно, магнитные потоки, пронизывающие эти контуры, все время оставались бы постоянными.

Электрическое поле $\mathbf{E}$, возбужденное переменным магнитным полем, в свою очередь создаст переменный электрический поток (поток вектора $\mathbf{D}$) через прямоугольный контур $CNMD$ (см. рис. 190). Иначе говоря, возникнет ток смещения, который также будет возбуждать магнитное поле, параллельное оси Y. Направление этого магнитного поля определяется правилом Ленца: оно будет препятствовать всяким изменениям уже существующего магнитного поля. Если остановить заряженную плоскость, то ток I прекратится. Однако возбужденное им электромагнитное поле останется. Электрическое и магнитное поля будут взаимно поддерживать друг друга: всякое изменение магнитного поля возбуждает поле электрическое, и наоборот. Таким образом, по разные стороны заряженной плоскости после ее остановки останутся

два электромагнитных поля, симметрично расположенных относительно этой плоскости. Как будет показано ниже, они не останутся на месте, а будут распространяться от заряженной плоскости в противоположных направлениях. Это и есть *электромагнитные волны*, или *электромагнитные возмущения.* Из предыдущего следует, что они поперечны как относительно вектора $\mathbf{E}$, так и относительно вектора $\mathbf{B}$.

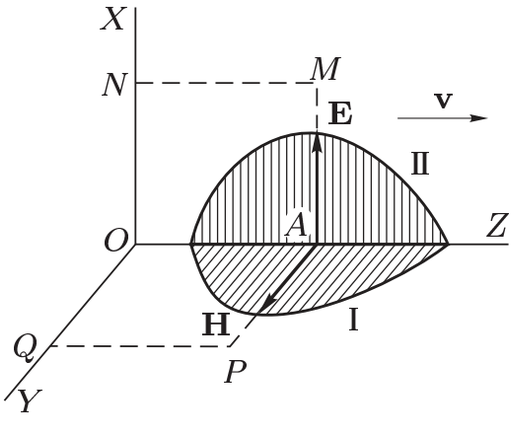


Рис. 191

3. Возьмем одно из этих возмущений, например возмущение, расположенное справа от заряженной плоскости (рис. 191). Электрическое поле в рассматриваемый момент представлено кривой I, расположенной в вертикальной плоскости, а магнитное поле — кривой II, расположенной в горизонтальной плоскости. Введем предположение, оправдываемое последующими расчетами, что эта картина электромагнитного поля без изменения формы перемещается вправо с какой-то скоростью $\mathbf{v}$. Возьмем два неподвижных прямоугольных контура $OAMN$ и $OQPA$ и запишем уравнения Максвелла в виде

$$\oint\limits_{OAMN} \mathbf{E}\,d\mathbf{l} = -\frac{1}{c}\,\frac{\partial \Phi_{\text{м}}}{\partial t}, \qquad \oint\limits_{OQPA} \mathbf{H}\,d\mathbf{l} = \frac{1}{c}\,\frac{\partial \Phi_{\text{эл}}}{\partial t},$$

где $\Phi_{\text{м}}$ — магнитный поток, а $\Phi_{\text{эл}}$ — поток вектора $\mathbf{D}$ через соответствующие контуры. Возьмем для простоты длину стороны AM равной единице. Тогда, так как на контуре $OAMN$ поле $\mathbf{E}$ отлично от нуля только на стороне AM, первое из этих уравнений принимает вид

$$E \equiv E_x = -\frac{1}{c}\,\frac{\partial \Phi_{\text{м}}}{\partial t}.$$

Аналогично, второе уравнение преобразуется к виду

$$H \equiv H_y = -\frac{1}{c}\,\frac{\partial \Phi_{\text{эл}}}{\partial t}.$$

Согласно нашему предположению, за время dt электромагнитное поле переместится на расстояние $v\,dt$. Магнитный поток $vB\,dt$ выйдет за пределы контура $OAMN$, а электрический поток $vD\,dt$ — за пределы контура $OQPA$. Вследствие этого потоки $\Phi_{\text{м}}$ и $\Phi_{\text{эл}}$ через указанные контуры изменятся на $d\Phi_{\text{м}} = -vB\,dt$, $d\Phi_{\text{эл}} = -vD\,dt$. Отсюда $\partial\Phi_{\text{м}}/\partial t = -vB$, $\partial\Phi_{\text{эл}}/\partial t = -vD$, и из предыдущих уравнений получаем

$$E = \frac{v}{c}\,B, \quad H = \frac{v}{c}\,D. \tag{83.1}$$

До сих пор не были использованы материальные уравнения $\mathbf{D} = \varepsilon\mathbf{E}$ и $\mathbf{B} = \mu\mathbf{H}$. Если их принять во внимание, то можно исключить D и B.

Это дает

$$E = \frac{v}{c}\,\mu H, \quad H = \frac{v}{c}\varepsilon E. \tag{83.1a}$$

Отсюда

$$v = \frac{c}{\sqrt{\varepsilon\mu}}. \tag{83.2}$$

Для скорости v получилось вполне определенное конечное выражение, отличное от нуля. Это оправдывает введенное выше допущение о характере изменения электромагнитного возмущения во времени и пространстве. Действительно, предположив, что электромагнитное возмущение распространяется без изменения формы, мы не фиксировали скорость этого распространения, а определили ее из требования, чтобы возмущение удовлетворяло уравнениям Максвелла. Это возможно только тогда, когда v определяется выражением (83.2). Возмущение не может стоять на месте, так как тогда величина v была бы равна нулю. Оно не может распространяться и мгновенно, так как тогда наше рассуждение привело бы к результату $v = \infty$. Таким образом, уравнения Максвелла допускают решения в виде электромагнитных возмущений, распространяющихся со скоростью $v = c/\sqrt{\varepsilon\mu}$.

4. Формулы (83.1) можно представить в векторной форме:

$$\mathbf{E} = -\frac{1}{c}\,[\mathbf{vB}], \quad \mathbf{H} = \frac{1}{c}\,[\mathbf{vD}]. \tag{83.3}$$

Здесь $\mathbf{v}$ дает не только скорость распространения возмущения по величине, но и ее направление. Те же самые соотношения справедливы и для возмущения, распространяющегося влево. В обоих случаях векторы $\mathbf{E}$, $\mathbf{B}$, $\mathbf{v}$ находятся в правовинтовом соотношении между собой (рис. 192). Если изменить направление одного из векторов $\mathbf{E}$ или $\mathbf{B}$ на противоположное, то направление распространения возмущения также изменится на противоположное.

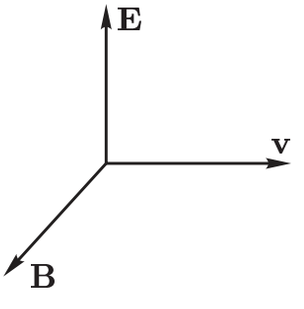


Рис. 192

Формулы (83.3) описывают так называемое *плоское возмущение*, т. е. такое возмущение, в котором электромагнитное поле одно и то же во всех точках плоскости, перпендикулярной к его распространению (эта плоскость называется *фронтом волны*). Для более полной характеристики его называют также *бегущим возмущением* (или *бегущей волной*), поскольку оно распространяется только в одном направлении. В нашем примере электрический вектор был параллелен оси X, а магнитный — оси Y. Конечно, возможно и возмущение с электрическим вектором вдоль оси Y и магнитным вдоль оси X. Ввиду линейности и однородности уравнений Максвелла суперпозиция таких возмущений будет также решением этих уравнений. Эта суперпозиция представляет также *плоскую бегущую электромагнитную волну*, для которой справедливы соотношения (83.3).

Электрический и магнитный векторы в такой волне всегда перпендикулярны между собой, а также к направлению ее распространения. Однако не обязательно, чтобы электрический вектор все время лежал в одной и той же плоскости, его направление в пространстве может изменяться. То же относится и к магнитному вектору. В частном случае, когда электрический (а следовательно, и магнитный) вектор во всех точках пространства лежит в одной плоскости, волна называется *линейно поляризованной.*

5. Рассмотренный нами мысленный пример возбуждения электромагнитных волн наиболее прост в теоретическом отношении. Совсем не обязательно брать бесконечную заряженную плоскость и приводить ее в движение. Мы поступили так только потому, чтобы возмущение получилось плоским. Для возбуждения же электромагнитных волн существенно только наличие *электрических зарядов, движущихся ускоренно.* В передающих радиостанциях электромагнитные волны возбуждаются быстропеременными электрическими токами, текущими по системе проводов (в антеннах). Волны, возбуждаемые таким образом, конечно, не будут плоскими. Только на больших расстояниях от излучающей антенны небольшие участки таких волн могут рассматриваться как приблизительно плоские. Вот почему на примере плоских возмущений выявляются все наиболее важные особенности электромагнитных волн.

6. Из формул (83.1а) следует

$$\varepsilon E^2 = \mu H^2. \tag{83.4}$$

Это означает, что *в бегущей плоской электромагнитной волне электрическая энергия в любой момент равна магнитной.* Аналогичное положение имеет место в случае упругой бегущей волны, где полная энергия также распределяется поровну между кинетической и потенциальной (см. т. I, § 82). Таким свойством обладают все возмущения, подчиняющиеся *принципу суперпозиции.* Для электромагнитных волн это свойство можно было бы доказать тем же методом, каким оно было получено для упругих волн в механике (см. т. I, § 82, п. 2). Надо только рассмотреть возмущения, получающиеся из начального состояния, в котором отлично от нуля либо только электрическое, либо только магнитное поле.

7. В вакууме $\varepsilon = \mu = 1$, и формула (83.2) дает $v = c$. Опыты Вильгельма Вебера (1804–1891) и Рудольфа Кольрауша (1809–1858) по измерению электродинамической постоянной c, а также все последующие измерения показали, что *электродинамическая постоянная равна скорости света в вакууме.* Таким образом, *в вакууме электромагнитные волны должны распространяться со скоростью света.* Это обстоятельство, по-видимому, навело Максвелла на мысль об электромагнитной теории света. Согласно этой теории *свет есть частный случай электромагнитных волн.* От всех остальных электромагнитных волн свет отличается только количественно — длиной волны. Сам Максвелл не предпринял попыток получить электромагнитные волны

на опыте, хотя он и был не только величайшим теоретиком, но и первоклассным экспериментатором. Максвеллу не суждено было дожить до экспериментального подтверждения своего открытия, он умер в 1879 г. в возрасте 48 лет. На опыте электромагнитные волны впервые были получены Герцем в 1887–1888 гг. Он показал, что электромагнитные волны распространяются, отражаются, преломляются, огибают препятствия, интерферируют. Он измерил скорость распространения электромагнитных волн и показал, что их свойства правильно описываются уравнениями Максвелла. Начиная с этого момента, теория Максвелла быстро получила всеобщее признание. (Опыты Герца будут описаны в § 142 после того, как мы познакомимся с переменными токами и электромагнитными колебаниями.) Открытие Максвелла и Герца вскоре привело к изобретению радио А.С. Поповым (1859–1906).

8. Уравнение Ньютона $md^2r/dt^2 = \mathbf{F}(\mathbf{r})$ в случае консервативных сил не меняется, если изменить знак времени, т. е. заменить t на $-t$. С этой симметрией уравнений механики связано следующее свойство консервативных систем. *Если в некоторый момент времени изменить на противоположные направления скоростей всех материальных точек замкнутой консервативной системы, то система повторит свое движение в обратном порядке.*

Аналогичной симметрией обладают и уравнения Максвелла в непроводящих (непоглощающих) средах. Действительно, в таких средах уравнения Максвелла (с учетом материальных уравнений) записываются в виде

$$\operatorname{rot}\frac{\mathbf{B}}{\mu} = \frac{\varepsilon}{c}\frac{\partial \mathbf{E}}{\partial t}, \quad \operatorname{rot}\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}, \qquad \operatorname{div}(\varepsilon\mathbf{E}) = 4\pi\rho, \quad \operatorname{div}\mathbf{B} = 0. \tag{83.5}$$

Если в этих уравнениях изменить знаки у времени t и вектора $\mathbf{B}$, то они перейдут в

$$\operatorname{rot}\left(\frac{-\mathbf{B}}{\mu}\right) = \frac{\varepsilon}{c}\frac{\partial \mathbf{E}}{\partial(-t)}, \quad \operatorname{rot}\mathbf{E} = -\frac{1}{c}\frac{\partial(-\mathbf{B})}{\partial(-t)}, \qquad \operatorname{div}(\varepsilon\mathbf{E}) = 4\pi\rho, \quad \operatorname{div}\mathbf{B} = 0. \tag{83.6}$$

т. е. в уравнения, тождественные с исходными уравнениями (83.5). Таким образом, *уравнения Максвелла в непроводящих средах не меняются при одновременном изменении знаков у t и $\mathbf{B}$. То же самое имеет место при одновременном изменении знаков у t и $\mathbf{E}$.*

Возьмем теперь произвольное решение уравнений (83.5):

$$\mathbf{E}(t, \mathbf{r}) = \mathbf{F}(t, \mathbf{r}), \quad \mathbf{B}(t, \mathbf{r}) = \boldsymbol{\Phi}(t, \mathbf{r}). \tag{83.7}$$

Тогда

$$\mathbf{E}_1(t, \mathbf{r}) = \mathbf{F}(-t, \mathbf{r}), \quad \mathbf{B}_1(t, \mathbf{r}) = -\boldsymbol{\Phi}(-t, \mathbf{r}) \tag{83.8}$$

будет также решением тех же уравнений, так как оно удовлетворяет уравнениям (83.6), тождественным с уравнениями (83.5). Заданием

во всем пространстве векторов $\mathbf{E}$ и $\mathbf{B}$ в начальный (произвольно выбранный) момент времени $t = 0$ решение системы уравнений (83.5) определяется однозначно. Решение (83.7) удовлетворяет начальным условиям

$$\mathbf{E}_{t=0} = \mathbf{F}(0, \mathbf{r}), \quad \mathbf{B}_{t=0} = \boldsymbol{\Phi}(0, \mathbf{r}),$$

а решение (83.8) — начальным условиям

$$\mathbf{E}_{1,t=0} = \mathbf{F}(0, \mathbf{r}), \quad \mathbf{B}_{1,t=0} = -\boldsymbol{\Phi}(0, \mathbf{r}).$$

Эти решения, очевидно, связаны соотношениями

$$\mathbf{E}_1(t, \mathbf{r}) = \mathbf{E}(-t, \mathbf{r}), \quad \mathbf{B}_1(t, \mathbf{r}) = -\mathbf{B}(-t, \mathbf{r}),$$

выражающими следующий результат. *Если в некоторый момент времени изменить на противоположное направление магнитного вектора во всех точках пространства, то при отсутствии проводящих тел электромагнитное поле повторит свою историю в обратном порядке, с той только разницей, что вектор* $\mathbf{B}$ *в соответствующие моменты времени будет направлен противоположно по сравнению с его направлением в исходном процессе.*

§ 84. Энергия и поток энергии

1. Уравнения Максвелла необходимо дополнить соотношениями, выражающими закон сохранения энергии. Пусть среда, в которой возбуждается электромагнитное поле, неподвижна. При изменении электромагнитного поля и прохождении электрического тока в единице объема совершается элементарная внешняя работа

$$\delta A^{\text{внеш}} = \frac{1}{4\pi} (\mathbf{E}\, d\mathbf{D} + \mathbf{H}\, d\mathbf{B}) + (\mathbf{j}\mathbf{E})\, dt. \tag{84.1}$$

Отдельные слагаемые этого выражения были получены в электростатике и в учении о магнитных полях постоянных токов. Отметим особо, что при получении работы намагничивания $(1/4\pi)\mathbf{H}\, d\mathbf{B}$ была использована теорема о циркуляции без учета тока смещения. Однако это не обязательно должно отразиться на общности окончательного результата $(1/4\pi)\mathbf{H}\, d\mathbf{B}$, так как и исходное выражение для элементарной работы (69.1) получено также без учета тока смещения. Поэтому отмеченное обстоятельство не препятствует распространению выражения (84.1) и на случай переменных электромагнитных полей. Выражение (84.1) должно рассматриваться как один из постулатов макроскопической теории электричества.

2. Работа (84.1) идет на приращение внутренней энергии за вычетом теплоты, уходящей из единицы объема среды вследствие теплопроводности. От теплопроводности можно отвлечься, предполагая, что она равна нулю. Это не сказывается на общности окончательного результата. Таким образом, если u — внутренняя энергия единицы объема

среды, то $\delta A^{\text{внеш}} = du$, или

$$\frac{\partial u}{\partial t} = \frac{1}{4\pi}(\mathbf{E}\dot{\mathbf{D}} + \mathbf{H}\dot{\mathbf{B}}) + (\mathbf{jE}). \tag{84.2}$$

Под u мы понимаем плотность *всей* внутренней энергии, а не только электромагнитную часть ее. Поэтому уравнение (84.2) справедливо для любых сред, в том числе ферромагнитных и сегнетоэлектрических. Оно учитывает не только джоулево тепло (слагаемое $\mathbf{jE}$), но и тепло ферромагнитного и диэлектрического гистерезиса. Используя уравнения Максвелла (82.1а) и (82.2а), преобразуем правую часть уравнения (84.2) к виду

$$\mathbf{E}\left(\frac{1}{4\pi}\dot{\mathbf{D}} + \mathbf{j}\right) + \frac{1}{4\pi}\mathbf{H}\dot{\mathbf{B}} = \frac{c}{4\pi}(\mathbf{E}\operatorname{rot}\mathbf{H} - \mathbf{H}\operatorname{rot}\mathbf{E}).$$

В силу известного векторного тождества

$$\mathbf{E}\operatorname{rot}\mathbf{H} - \mathbf{H}\operatorname{rot}\mathbf{E} = -\operatorname{div}[\mathbf{EH}].$$

Поэтому, если ввести обозначение

$$\mathbf{S} = \frac{c}{4\pi}[\mathbf{EH}], \tag{84.3}$$

то уравнение (84.2) примет вид

$$\frac{\partial u}{\partial t} + \operatorname{div}\mathbf{S} = 0. \tag{84.4}$$

Для физической интерпретации этого уравнения сравним его с уравнением непрерывности

$$\frac{\partial \rho}{\partial t} + \operatorname{div}\mathbf{j} = 0, \tag{84.5}$$

в котором ρ означает плотность вещества или электричества, а $\mathbf{j}$ — плотность потока вещества или электрического тока.

Формальная аналогия между уравнениями (84.4) и (84.5) приводит к представлению, что энергия течет в пространстве подобно жидкости, причем вектор $\mathbf{S}$ играет роль *плотности потока электромагнитной энергии.* Математически это представление выражается более наглядно и непосредственно, если уравнение (84.4) записать в интегральной форме:

$$\frac{\partial}{\partial t}\int\limits_V u\,dV = \oint S_n\,dF, \tag{84.4а}$$

где V — произвольный объем среды, ограниченный замкнутой поверхностью F, а $\mathbf{n}$ — внутренняя нормаль к этой поверхности. В такой форме уравнение означает, что приращение внутренней энергии в объеме V происходит за счет электромагнитной энергии, втекающей в этот объем из окружающего пространства через поверхность F.

Представление о течении энергии сохраняется также при учете теплопроводности и упругих свойств среды, но к плотности потока

электромагнитной энергии в этом случае надо добавить *плотность потока теплоты* и *плотность упругой энергии.*

Общее представление о потоке энергии в пространстве впервые было введено Н. А. Умовым (1846–1915) в 1874 г. Поэтому вектор плотности потока энергии без конкретизации ее физической природы называется *вектором Умова.* Конкретные выражения для этого вектора были получены Умовым, естественно, только для упругих сред и вязких жидкостей. Через 11 лет идеи Умова были разработаны Пойнтингом (1852–1914) применительно к электромагнитной энергии. Пойнтингом было получено и выражение (84.3), а потому вектор $\mathbf{S}$ называется *вектором Пойнтинга.* Формула (84.4), или (84.4а), выражает закон сохранения энергии в электродинамике и носит название *теоремы Умова–Пойнтинга.* Отметим замечательную простоту и общность формулы (84.3). Вектор Пойнтинга выражается только через напряженности полей $\mathbf{E}$ и $\mathbf{H}$ и не содержит никаких величин, характеризующих индивидуальные свойства среды, в которой течет электромагнитная энергия.

3. Допустим теперь, что нет гистерезиса. Тогда вектор $\mathbf{D}$ будет однозначной функцией вектора $\mathbf{E}$, а вектор $\mathbf{B}$ — однозначной функцией вектора $\mathbf{H}$. Теплоты гистерезиса не будет, и элементарная работа $\delta A^{\text{внеш}}$ пойдет только на выделение джоулева тепла $(\mathbf{jE})\,dt$ и на приращение электромагнитной энергии. Поэтому в соответствии с формулой (84.1) выражение

$$dw = \frac{1}{4\pi}\,(\mathbf{E}\,d\mathbf{D} + \mathbf{H}\,d\mathbf{B}) \tag{84.6}$$

должно быть истолковано как приращение плотности электромагнитной энергии w. Сама величина w представится интегралом

$$w = \frac{1}{4\pi}\int(\mathbf{E}\,d\mathbf{D} + \mathbf{H}\,d\mathbf{B}). \tag{84.7}$$

В частном случае, когда справедливы соотношения $\mathbf{D} = \varepsilon\mathbf{E}$, $\mathbf{B} = \mu\mathbf{H}$, отсюда получаем

$$w = \frac{1}{8\pi}\,(\varepsilon E^2 + \mu H^2), \tag{84.8}$$

или

$$w = \frac{1}{8\pi}\,(\mathbf{ED} + \mathbf{HB}). \tag{84.9}$$

4. Приведем примеры на течение электромагнитной энергии.

Пример 1. Поток энергии в плоской электромагнитной волне (см. рис. 192). Поскольку волна распространяется в одном направлении, электромагнитная энергия должна распространяться в том же направлении. Поток энергии численно равен количеству энергии, переносимому в одну секунду через квадратный сантиметр, перпендикулярный к направлению распространения волны. Если v — скорость распространения волны, то $S = vw$. Согласно соотношению (83.4) в плоской бегущей волне плотность электрической энергии равна плотности магнитной, а потому $w = (\varepsilon/4\pi)E^2$. В силу того же соотношения $\sqrt{\varepsilon}E = \sqrt{\mu}H$. Поэтому $w = (1/4\pi)\sqrt{\varepsilon\mu}EH$. Если учесть еще

формулу (83.2), то получится

$$S = vw = \frac{c}{4\pi} EH.$$

Так как волна распространяется в направлении вектора [**EH**], то это выражение совпадает с выражением (84.3).

Пример 2. Выделение джоулева тепла в проводе. По цилиндрическому проводу радиуса r течет постоянный ток I (рис. 193). Магнитное поле обвивается вокруг тока и на поверхности провода равно $H = 2I/(cr) = 2j\pi r/c$. Электрическое поле $\mathbf{E}$ параллельно оси провода. Таким образом, вектор Пойнтинга $\mathbf{S}$ направлен внутрь провода нормально к его боковой поверхности. Следовательно, электромагнитная энергия втекает внутрь провода из окружающего пространства. Площадь боковой поверхности провода равна $2\pi rl$, где l — его длина. Количество электромагнитной энергии, ежесекундно вступающей в провод, будет

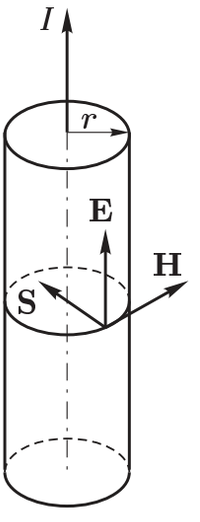


Рис. 193

$$S \cdot 2\pi rl = \frac{c}{4\pi} EH \cdot 2\pi rl = \pi r^2 ljE = VjE,$$

где $V = \pi r^2 l$ — объем провода. Но такое же количество теплоты выделяется в проводе при прохождении электрического тока. Таким образом, *электромагнитная энергия из окружающего пространства вступает внутрь провода и в нем превращается в джоулево тепло.*

Пример 3. Зарядка конденсатора. Допустим, что конденсатор плоский с круговыми обкладками, а зарядка производится квазистатически. Отвлечемся от неоднородностей электромагнитного поля вблизи краев конденсатора. Тогда расчет производится так же, как в предыдущем примере. Надо только плотность тока проводимости $\mathbf{j}$ заменить на плотность тока смещения $\dot{\mathbf{D}}/(4\pi)$. Сделав это, найдем, что из окружающего пространства ежесекундно в конденсатор втекает электромагнитная энергия $(V/4\pi)(\mathbf{E}\dot{\mathbf{D}})$, где V — объем конденсатора. Приращение энергии конденсатора за время dt будет

$$dW = \frac{V}{4\pi} (\mathbf{E}\dot{\mathbf{D}})\, dt = \frac{V}{4\pi} (\mathbf{E}\, d\mathbf{D}).$$

Интегрируя это выражение в предположении, что справедливо соотношение $\mathbf{D} = \varepsilon\mathbf{E}$, получим

$$W = \frac{V}{8\pi} (\mathbf{ED}).$$

Рассуждение существенно не изменится, если принять во внимание неоднородности электрического и магнитного полей вблизи краев конденсатора. Вблизи краев у поля $\mathbf{E}$ имеется радиальная слагающая, а у поля $\mathbf{H}$ — слагающая, параллельная оси конденсатора. Это приводит к появлению дополнительных потоков энергии, параллельных оси

конденсатора. Однако, ввиду симметрии, поток энергии, направленный вверх, будет компенсирован потоком, направленным вниз. Поэтому полный поток энергии, вступающей в конденсатор, определяется только осевой слагающей вектора $\mathbf{E}$ и азимутальной слагающей вектора $\mathbf{H}$. Если пренебречь изменениями этих слагающих вблизи краев конденсатора, то получится результат, найденный выше.

Таким же путем можно получить выражение для магнитной энергии соленоида, по которому течет постоянный ток. Это рекомендуется сделать читателю.

5. Относительно приведенного вывода выражения для плотности потока электромагнитной энергии необходимо сделать следующее принципиальное замечание. Уравнения (84.4) и (84.4а) не изменятся, если вектор $\mathbf{S}$ заменить на $\mathbf{S} + \operatorname{rot}\mathbf{a}$, где $\mathbf{a}$ — произвольный вектор. Действительно, дивергенция ротора любого вектора тождественно равна нулю. Тождественно равен нулю и поток вектора $\operatorname{rot}\mathbf{a}$ через произвольную замкнутую поверхность. Таким образом, приведенными рассуждениями вектор плотности потока электромагнитной энергии определяется не однозначно. Это не создает никаких трудностей в применениях закона сохранения энергии, так как при этом речь идет всегда о потоках энергии через *замкнутые поверхности.* Трудности возникают, во всяком случае в дорелятивистской физике, когда представление о потоке энергии применяется к *незамкнутым поверхностям.* В дорелятивистской физике хотя и вводилось представление о локализации энергии в пространстве, но на саму энергию никогда не смотрели как на какую-то субстанцию, которой присущи определенная масса и движение. Теория относительности устранила это принципиальное различие между массой и энергией и показала, что *всякая энергия обладает массой*: *масса равна энергии, деленной на квадрат скорости света.* Но поток вещества и связанный с ним поток массы являются величинами вполне определенными и однозначными. Поэтому той же определенностью и однозначностью должен обладать и поток энергии.

Вместе с тем поток энергии должен быть связан с *потоком импульса.* Определим плотность потока импульса $\mathbf{g}$ для электромагнитного поля и ограничимся полем в вакууме, чтобы в дальнейшем избежать усложнений, вносимых наличием вещества. Пусть w — плотность электромагнитной энергии. Соответствующая ей плотность массы будет w/c^2. Если энергия движется со скоростью v, то с этим движением связан импульс, значение которого в единице объема среды дается выражением $\mathbf{g}_{\text{эл}} = w\mathbf{v}/c^2$. Этот импульс называется *электромагнитным импульсом* или *импульсом электромагнитного поля.* Замечая, что wv есть плотность потока $\mathbf{S}$ электромагнитной энергии, получаем для плотности электромагнитного импульса

$$\mathbf{g}_{\text{эл}} = \frac{\mathbf{S}}{c^2} = \frac{1}{4\pi c}\,[\mathbf{EH}]. \tag{84.10}$$

Понятие электромагнитного импульса было введено Максом Абрагамом (1875–1922) еще до возникновения теории относительности.

6. Рассмотрим теперь случай, когда электромагнитное поле $\mathbf{E}, \mathbf{H}$ статическое, т. е. не меняется во времени. Согласно формуле (84.3) должен существовать поток электромагнитной энергии во всех точках пространства, где векторы $\mathbf{E}$ и $\mathbf{H}$ не коллинеарны. Так как для стационарных процессов $\partial u/\partial t = 0$, то из уравнения (84.4) следует $\operatorname{div}\mathbf{S} = 0$. Это значит, что энергия течет подобно несжимаемой жидкости: во всякий объем пространства втекает столько же энергии, сколько и вытекает. Непосредственно такое течение электромагнитной энергии не проявляется ни в каких физических явлениях. Однако оно приводит к следствиям, допускающим экспериментальную проверку, если учесть, что с потоком электромагнитной энергии связан электромагнитный импульс и его момент. Пример, приводимый ниже, показывает, что это действительно так.

Рассмотрим заряженный цилиндрический конденсатор, помещенный в постоянное однородное магнитное поле $\mathbf{H}$, параллельное оси конденсатора (рис. 194 *а*). Поля $\mathbf{E}$ и $\mathbf{H}$ взаимно перпендикулярны. Век-

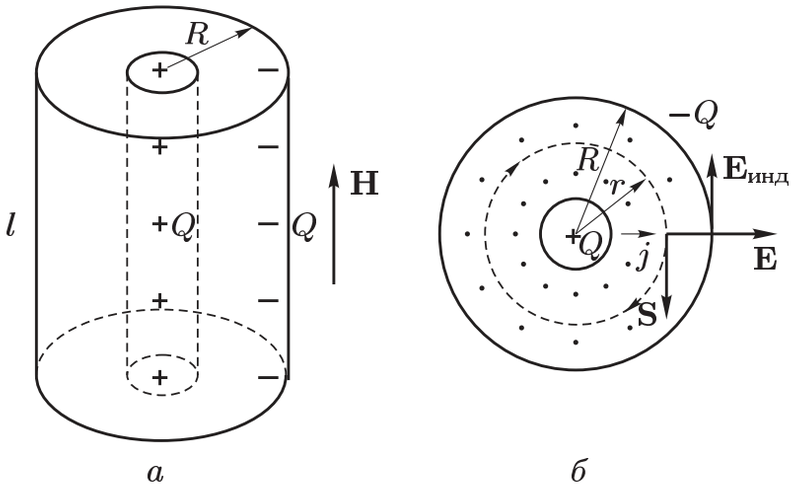


Рис. 194

тор Пойнтинга $\mathbf{S}$ направлен по касательным к коаксиальным окружностям, центры которых расположены на оси конденсатора, а плоскости перпендикулярны к этой оси (рис. 194 *б*). Таким образом, происходит непрерывная циркуляция электромагнитной энергии вдоль этих окружностей. Полный электромагнитный импульс системы, ввиду ее аксиальной симметрии, равен нулю. Однако момент этого импульса $\mathbf{L}_{\text{эл}}$ относительно оси системы отличен от нуля. Именно

$$L_{\text{эл}} = \int r g_{\text{эл}}\, dV = \frac{1}{4\pi c} \int r E H\, dV.$$

Подставив сюда $E = 2Q/(lr)$, $dV = 2\pi r l\, dr$ и выполнив интегрирование, получим

$$L_{\text{эл}} = \frac{QH}{c} \int_0^R r\, dr = \frac{QH}{2c} R^2,$$

где l — длина конденсатора, Q — его заряд, а R — радиус наружной обкладки. Для упрощения вычислений радиусом внутренней обкладки мы пренебрегаем, предполагая его малым. Как видно из рисунка, вектор $\mathbf{L}_{\text{эл}}$ направлен против поля $\mathbf{H}$, когда внутренняя обкладка заряжена положительно, и по полю $\mathbf{H}$, когда она заряжена отрицательно. Будем предполагать, что положительный заряд находится на внутренней обкладке, и напишем

$$\mathbf{L}_{\text{эл}} = -\frac{QR^2}{2c}\,\mathbf{H}. \tag{84.11}$$

Допустим теперь, что в конденсаторе от внутренней обкладки к наружной течет радиальный электрический ток, медленно разряжающий конденсатор. Для этого надо предположить, что конденсатор заполнен слабо проводящей средой, диэлектрическая и магнитная проницаемости которой практически не отличаются от единицы. После разряда электромагнитный момент импульса (84.10) исчезнет, и конденсатор должен получить такой же по величине и по направлению механический момент импульса. Если конденсатор был подвешен на нити, то после разряда нить должна закрутиться. Что это действительно так, показывает следующее рассуждение.

Если $\mathbf{j}$ — объемная плотность радиального тока, то на элемент объема диэлектрика действует амперова сила $d\mathbf{F} = (1/c)[\mathbf{jH}]\,dV$. Момент $d\mathbf{M}$ этой силы относительно оси конденсатора численно равен $(1/c)rjH\,dV$ и направлен против поля $\mathbf{H}$, как это видно из рис. 194 *б*. Таким образом, в векторной форме $d\mathbf{M} = -(1/c)rj\mathbf{H}\,dV$. Подставив сюда $dV = 2\pi rl\,dr$, $j = I/(2\pi rl)$, после интегрирования получим

$$\mathbf{M} = -\frac{I\mathbf{H}}{c}\int_0^R r\,dr = -\frac{IR^2}{2c}\,\mathbf{H}.$$

Под действием момента $\mathbf{M}$ механический момент импульса конденсатора будет изменяться в соответствии с уравнением

$$d\mathbf{L} = \mathbf{M}\,dt = -\frac{R^2\mathbf{H}}{2c}\,I\,dt = -\frac{R^2\mathbf{H}}{2c}\,dQ.$$

Выполнив интегрирование, найдем, что после разрядки момент импульса конденсатора будет

$$\mathbf{L} = -\frac{QR^2}{2c}\,\mathbf{H}, \tag{84.12}$$

в согласии с прежним результатом.

К тому же результату мы придем, если, оставляя конденсатор заряженным, выключим магнитное поле $\mathbf{H}$. При выключении магнитного поля возникнет кольцевое электрическое поле. На окружности радиуса R напряженность этого поля определится по закону электромагнитной индукции:

$$2\pi R E_{\text{инд}} = -\frac{1}{c}\,\frac{d\Phi}{dt} = -\frac{1}{c}\,\pi R^2\,\frac{dH}{dt},$$

откуда

$$E_{\text{инд}} = -\frac{R}{2c}\,\frac{dH}{dt}.$$

На отрицательный заряд $-Q$ такое поле будет действовать с моментом

$$\mathbf{M} = \frac{Q}{2c}\,R^2\frac{d\mathbf{H}}{dt}.$$

Поэтому

$$d\mathbf{L} = \mathbf{M}\,dt = \frac{Q}{2c}\,R^2\,d\mathbf{H}.$$

Интегрируя в пределах от $\mathbf{H}$ до 0, получаем прежний результат (84.12).

Разобранный здесь эффект крайне мал, но в принципе может наблюдаться.

ЗАДАЧИ

1. Плоский воздушный конденсатор, обкладками которого являются два одинаковых диска, заряжен до высокой разности потенциалов и затем отключен от источника напряжения. В центре конденсатора происходит пробой (проскакивает электрическая искра), в результате чего конденсатор разряжается. Считая разряд квазистационарным и пренебрегая неоднородностью поля на краях конденсатора, определить полный поток электромагнитной энергии, вытекающей из пространства между обкладками. Обсудить явление с точки зрения сохранения и превращения энергии.

О т в е т. Поток энергии равен нулю (см. задачу 3 к § 81).

2. Плоскому конденсатору емкости C, обкладками которого являются два одинаковых диска, сообщен заряд Q. Затем конденсатор был отключен от источника электричества. После этого пластины были соединены длинным цилиндрическим проводом, проходящим вне конденсатора, и конденсатор разрядился. Пренебрегая неоднородностью поля на краях конденсатора, показать непосредственным расчетом, что полный поток электромагнитной энергии из конденсатора равен полному потоку электромагнитной энергии, втекающей внутрь провода. Проанализировать явление с точки зрения представления о движении, превращении и сохранении энергии.

Указание. См. примеры 2 и 3 из текста настоящего параграфа.

О т в е т. Электрическая энергия вытекает из конденсатора через его края, втекает внутрь провода и там превращается во внутреннюю (тепловую) энергию.

3. Плоский воздушный конденсатор, состоящий из двух одинаковых дисков, заряжен электричеством и помещен внутри соленоида, создающего однородное постоянное магнитное поле $B = 1000$ Гс. Магнитное поле создается батареей, посылающей постоянный ток в обмотку электромагнита. Электрическое поле между пластинами конденсатора равно $E = 10000$ В/см. Объем воздушного пространства между пластинами конденсатора равен $V = 100$ см^3. Конденсатор пробивается электрической искрой вдоль его оси и в результате этого разряжается. Как изменится механический импульс системы после пробоя? Обсудить результат с точки зрения закона сохранения импульса.

Р е ш е н и е. Ввиду осевой симметрии полный электромагнитный импульс поля равен нулю. В результате разряда конденсатора он измениться не может. Поэтому не может измениться и полный механический импульс системы. Но в результате разряда электромагнитный импульс, локализированный

в конденсаторе, уменьшается на $(V/4\pi c)[\mathbf{EH}]$, а электромагнитный импульс поля вне конденсатора увеличивается на такую же величину. В соответствии с этим конденсатор приобретает механический импульс $(V/4\pi c)[\mathbf{EH}]$, равный 10^{-4} г $\cdot$ см/с. Соленоид получит такой же, но противоположно направленный импульс. Искру можно рассматривать как ток проводимости. Если бы все электрическое поле конденсатора было локализовано только внутри него, то магнитное поле искры вне конденсатора было бы полностью компенсировано магнитным полем тока смещения (см. задачу 3 к § 81). На самом деле часть тока смещения проходит вне конденсатора и создает там магнитное поле. Это магнитное поле действует на токи, текущие в соленоиде, и меняет импульс последнего.

4. В предыдущей задаче конденсатор не пробивается, а разрывается цепь батареи, питающей соленоид. Как в результате этого изменится механический импульс системы?

Ответ. Так же, как в предыдущей задаче.

§ 85. Международная система единиц (СИ)

1. В механике строго научная система единиц (СГС), в которой за основные величины приняты *длина*, *масса* и *время*, была разработана на основе законов Ньютона. В худшем положении оказалась электродинамика, основные принципы которой (уравнения Максвелла) были установлены и получили признание только в конце XIX века. До этого времени уже получили широкое распространение случайно выбранные практические единицы: *вольт*, *ампер*, *ом* и их производные, никак не связанные с системой единиц в механике. Естественно было ввести единую систему единиц для механических и электромагнитных величин. Здесь физика и электротехника пошли разными путями. Физика не вводила новых основных величин, а рассматривала электрические и магнитные величины как *производные механических.* Построенные по такому принципу системы единиц называются *абсолютными.* К таким системам относится и *гауссова система СГС*, которая в настоящем курсе принята за основную. Электротехника, сохранив механические величины, не захотела жертвовать и практическими электрическими единицами: вольтом, ампером, омом и пр. Последнее условие — довольно жесткое. Удовлетворить ему оказалось возможным только ценой *существенного ухудшения* системы единиц. Это относится и к так называемой *Международной системе единиц* (сокращенно *СИ*), разработанной за последнее время и рекомендованной в качестве основной. Ниже изложены основы построения системы СИ, а затем отмечены ее принципиальные недостатки.

В СИ изменен масштаб основных механических величин: вместо сантиметра введен *метр*, вместо грамма — *килограмм.* Особой выгоды в этом нет, так как все равно невозможно удовлетворить требованию, чтобы величина единицы была всегда одинаково удобна. Одна и та же единица в одних случаях будет слишком велика, в других слишком мала. Этот вопрос удовлетворительно решается введением приставок «микро», «милли», «мега» и т. д., а также степеней 10. Но, разумеется, изменение масштабов основных величин принципиально

ничего не меняет и в этом смысле никаких возражений не вызывает. Принципиальными являются два момента.

Во-первых, к трем основным механическим величинам — длине, времени и массе — в СИ добавлена в качестве независимой четвертая, чисто электрическая, величина, имеющая *самостоятельную размерность.* Такой величиной выбрана *сила электрического тока*, а ее единицей — *ампер.* Количество электричества есть величина производная с единицей *ампер-секунда*, называемой *кулоном.*

Во-вторых, уравнения Максвелла в СИ записываются в так называемой *рационализированной форме*, т. е. в форме, не содержащей никаких числовых множителей. В интегральной форме эти уравнения пишутся так:

$$\oint (\mathbf{H}\, d\mathbf{l}) = I + \int \left(\frac{\partial \mathbf{D}}{\partial t}\, d\mathbf{S} \right), \tag{85.1}$$

$$\oint (\mathbf{E}\, d\mathbf{l}) = -\int \left(\frac{\partial \mathbf{B}}{\partial t}\, d\mathbf{S} \right), \tag{85.2}$$

$$\oint (\mathbf{D}\, d\mathbf{S}) = q, \tag{85.3}$$

$$\oint (\mathbf{B}\, d\mathbf{S}) = 0, \tag{85.4}$$

а в дифференциальной так:

$$\operatorname{rot} \mathbf{H} = \mathbf{j} + \frac{\partial \mathbf{D}}{\partial t}, \tag{85.1 a}$$

$$\operatorname{rot} \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \tag{85.2 a}$$

$$\operatorname{div} \mathbf{D} = \rho, \tag{85.3 a}$$

$$\operatorname{div} \mathbf{B} = 0. \tag{85.4 a}$$

Идея «рационализации» уравнений Максвелла принадлежит Хевисайду (1850–1925). Хевисайд исходил из того, что уравнения Максвелла — это *фундаментальные уравнения*, а потому целесообразно освободить их от численных множителей типа 4π. Для этого достаточно изменить величины единиц электрического заряда, а также напряженностей электрического и магнитного полей. Практической выгоды от такой рационализации нет. Исчезая из одних формул, численные коэффициенты появляются в других, так что общее число коэффициентов практически остается неизменным. Идея Хевисайда была поддержана Лоренцем. В *системе Хевисайда–Лоренца* уравнения Максвелла выглядят так же, как и в гауссовой системе, с тем единственным отличием,

что безразмерный множитель 4π в них всюду заменен другим безразмерным множителем — единицей. Поэтому никакими принципиальными преимуществами перед гауссовой системой система Хевисайда–Лоренца не обладает. Обе системы одинаково хороши. «Рационализация» в СИ идет дальше: опускается не только безразмерный коэффициент 4π, но и *размерная величина* — скорость света в вакууме c.

2. Уравнениями (85.1)–(85.4) в СИ уже предопределены размерность и единицы векторов $\mathbf{D}$ и $\mathbf{H}$, а именно:

$$[\mathbf{D}] = \text{Кл}/\text{м}^2 = \text{А}\cdot\text{с}/\text{м}^2, \quad [\mathbf{H}] = \text{А}/\text{м}.$$

Связь с механикой устанавливается посредством *силовых векторов* $\mathbf{E}$ и $\mathbf{B}$. Последние определяются соотношениями

$$\mathbf{F}_e = q\mathbf{E}, \tag{85.5}$$

$$\mathbf{F}_m = q[\mathbf{v}\mathbf{B}], \tag{85.6}$$

где $\mathbf{F}_e$ и $\mathbf{F}_m$ — силы, действующие на заряд q в электрическом и магнитном полях. Отсюда получаем

$$[\mathbf{E}] = \frac{\text{Н}}{\text{Кл}} = \frac{\text{кг}\cdot\text{м}}{\text{А}\cdot\text{с}^3} = \frac{\text{В}}{\text{м}}, \quad [\mathbf{B}] = \frac{\text{Н}\cdot\text{с}}{\text{Кл}\cdot\text{м}} = \frac{\text{Н}}{\text{А}\cdot\text{м}} = \frac{\text{кг}}{\text{А}\cdot\text{с}^2} = \frac{\text{Вб}}{\text{м}^2}.$$

Единица индукции $\mathbf{B}$ называется *тесла*. Единицы $\mathbf{E}$, $\mathbf{D}$ и $\mathbf{H}$ не получили специальных названий. В соответствии с размерностью их называют *вольт на метр*, *кулон на квадратный метр* и *ампер на метр*. Размерности всех четырех векторов $\mathbf{E}, \mathbf{D}, \mathbf{B}, \mathbf{H}$ разные. Даже в вакууме векторы $\mathbf{E}$ и $\mathbf{D}$ с одной стороны, $\mathbf{B}$ и $\mathbf{H}$, с другой, в СИ — величины разные. В вакууме они связаны соотношениями

$$\mathbf{D} = \varepsilon_0\mathbf{E}, \quad \mathbf{B} = \mu_0\mathbf{H}. \tag{85.7}$$

Величины ε_0 и μ_0 называются *электрической* и *магнитной постоянными*.

Из уравнения (85.3) следует, что в СИ электрическое поле точечного заряда q в вакууме определяется формулой

$$D = \frac{q}{4\pi r^2} \quad \text{или} \quad E = \frac{q}{4\pi\varepsilon_0 r^2}.$$

Поэтому закон Кулона в вакууме должен писаться так:

$$F = \frac{q_1 q_2}{4\pi\varepsilon_0 r^2}. \tag{85.8}$$

Аналогично, теорема о циркуляции (85.1) для прямого тока в вакууме дает

$$H = \frac{I}{2\pi r}, \quad B = \frac{\mu_0 I}{2\pi r}.$$

Следовательно, для силы взаимодействия двух тонких параллельных проводов, по которым текут токи I_1 и I_2, получаем

$$F = \mu_0 \frac{I_1 I_2}{2\pi r}\, l, \tag{85.9}$$

где l — длина участка одного из проводов, к которому приложена сила $\mathbf{F}$.

Формулы (85.8) и (85.9) позволяют определить постоянные ε_0 и μ_0. Действительно, пусть $q_1 = q_2 = 1$ Кл, $r = 1$ м. Тогда по формуле (85.8) находим $F = 1/(4\pi\varepsilon_0)$ ньютонов. Вычислим ту же силу по закону Кулона $F = q_1 q_2 / r^2$ в гауссовой системе единиц. В этом случае $r = 100$ см, а по определению кулона $q_1 = q_2 = 10c$ СГСЭ-ед., где $c \approx 10^8$ м/с — скорость света в вакууме в м/с. Таким образом, $F = c^2 \cdot 10^{-2}$ дин $= c^2 \cdot 10^{-7}$ Н. Путем сравнения с предыдущим результатом находим

$$\varepsilon_0 = \frac{10^7}{4\pi c^2} \approx 8{,}854 \cdot 10^{-12} \ \text{Ф/м}. \tag{85.10}$$

Аналогично поступаем для определения магнитной постоянной μ_0. В формуле (85.9) полагаем $I_1 = I_2 = 1$ А, $l = r = 1$ м. Тогда $F = \mu_0/(2\pi)$ Н. Воспользуемся теперь для вычисления той же силы формулой в гауссовой системе единиц $F = 2I_1 I_2 l / r(100c)^2$. (В знаменателе мы написали $100c$, а не c, так как предполагаем, что скорость c измеряется в м/с, а не в см/с, как должно делаться в гауссовой системе.) Полагая в этой формуле $l = r = 100$ см, $I_1 = I_2 = 10c$ СГСЭ-ед., получим $F = 2 \cdot 10^{-2}$ дин $= 2 \cdot 10^{-7}$ Н. Сравнивая полученный результат с предыдущим, находим

$$\mu_0 = 4\pi \cdot 10^{-7} \approx 1{,}256 \cdot 10^{-6} \ \text{Г/м} \tag{85.11}$$

Сами постоянные ε_0 и μ_0 никакого реального физического смысла не имеют, а являются только *размерными коэффициентами*, искусственно введенными для перевода значений по существу одних и тех же величин ($\mathbf{E}$ и $\mathbf{D}$, $\mathbf{B}$ и $\mathbf{H}$ в вакууме) из одних единиц в другие. Однако эти коэффициенты связаны соотношением

$$\varepsilon_0 \mu_0 = \frac{1}{c^2}, \tag{85.12}$$

где скорость света должна измеряться в м/с. Только комбинация $\varepsilon_0\mu_0$ имеет реальный физический смысл.

3. Материальные уравнения в средах в СИ записываются в виде

$$\mathbf{D} = \varepsilon_0 \varepsilon \mathbf{E}, \tag{85.13}$$

$$\mathbf{B} = \mu_0 \mu \mathbf{H}, \tag{85.14}$$

$$\mathbf{j} = \lambda \mathbf{E}. \tag{85.15}$$

Диэлектрическая и магнитная проницаемости ε и μ безразмерные и равны соответствующим величинам в гауссовой системе. В системе СИ их называют *относительными*, а произведения $\varepsilon_0\varepsilon$ и $\mu_0\mu$ — *абсолютными проницаемостями среды*. Электрическая проводимость λ имеет размерность

$$[\lambda] = \left[\frac{I}{ES}\right] = \left[\frac{I}{\varphi l}\right] = \frac{\text{А}}{\text{В} \cdot \text{м}} = \frac{1}{\text{Ом} \cdot \text{м}}.$$

Плотность энергии в СИ

$$w = \frac{1}{2}(\mathbf{ED} + \mathbf{HB}), \tag{85.16}$$

вектор Пойнтинга

$$\mathbf{S} = [\mathbf{EH}]. \tag{85.17}$$

Электрический дипольный момент определяется прежним выражением $\mathbf{p} = q\mathbf{l}$ и измеряется в кулон-метрах.

По-прежнему определяется и вектор поляризации $\mathbf{P}$, как дипольный момент единицы объема среды. Как и вектор $\mathbf{D}$, он имеет размерность заряда, деленного на площадь. Поэтому вектор $\mathbf{D}$ определяется выражением

$$\mathbf{D} = \varepsilon_0\mathbf{E} + \mathbf{P}. \tag{85.18}$$

Множитель ε_0 введен сюда для уравнивания размерностей обоих слагаемых в правой части. Вектор $\mathbf{P}$ связан с $\mathbf{E}$ соотношением $\mathbf{P} = \varepsilon_0\alpha\mathbf{E}$, где α — безразмерная величина, называемая *поляризуемостью среды.* Она связана с ε соотношением

$$\varepsilon = 1 + \alpha. \tag{85.19}$$

Отсюда видно, что поляризуемость α в СИ в 4π раз больше ее значения в гауссовой системе.

Поляризуемость молекулы β определяют формулой $\mathbf{p} = \varepsilon_0\beta\mathbf{E}$. Величина β имеет размерность объема и связана с α соотношением $\alpha = N\beta$, где N — число молекул в единице объема среды.

Магнитный момент $\mathfrak{M}$ витка с током определяется выражением $\mathfrak{M} = I\mathbf{S}$. Поэтому вектор намагничивания $\mathbf{I}$, или магнитный момент единицы объема магнетика, имеет ту же размерность, что и вектор $\mathbf{H}$. В соответствии с этим вектор $\mathbf{H}$ определяется выражением

$$\mathbf{H} = \frac{\mathbf{B}}{\mu_0} - \mathbf{I}. \tag{85.20}$$

Магнитная восприимчивость $\varkappa$ безразмерна, причем по определению $\mathbf{I} = \varkappa\mathbf{H}$. Она связана с магнитной проницаемостью μ соотношением

$$\mu = 1 + \varkappa. \tag{85.21}$$

Значит, магнитная восприимчивость $\varkappa$ в СИ в 4π раз больше, чем в гауссовой системе.

4. Можно указать однообразный систематический прием для перевода электродинамических формул из гауссовой системы в СИ и обратно. Для этого каждой физической величине в гауссовой системе сопоставляется определенный «*переводной*» *коэффициент.* После замены каждой величины такой же величиной, умноженной на соответствующий переводной коэффициент, уравнения гауссовой системы переходят в уравнения СИ. Задача нахождения таких коэффициентов *не однозначна.* Действительно, если найден какой-либо один набор коэффициентов, то после умножения их на одну и ту же величину получится

другой набор, также пригодный для выполнения требуемого преобразования. Так как уравнения механики в обеих системах единиц записываются одинаково, то нет надобности вводить переводные коэффициенты для чисто механических величин. Коэффициенты нужны только для электрических и магнитных величин. Умножение любой величины на произвольную механическую величину оставляет переводной коэффициент без изменения. Так, электрическому полю $\mathbf{E}$ и потенциалу φ сопоставляется один и тот же переводной коэффициент, так как в силу соотношения $\mathbf{E} = -\operatorname{grad}\varphi$ поле $\mathbf{E}$ получается из потенциала путем деления на механическую величину — длину. Одинаковые коэффициенты будут иметь заряд q, его плотность ρ, ток I, его плотность $\mathbf{j}$, поляризация $\mathbf{P}$ и т. д. Скорость света c в СИ явно не входит. Она заменяется на $1/\sqrt{\varepsilon_0\mu_0}$. Относительные диэлектрическая и магнитная проницаемости, как величины, одинаковые в обеих системах, не преобразуются.

После этих разъяснений перейдем к нахождению системы переводных коэффициентов e, d, ..., сопоставляемых величинам $\mathbf{E}$, $\mathbf{D}$, ... согласно следующей схеме:

$\mathbf{E}$	$\mathbf{D}$	$\mathbf{B}$	$\mathbf{H}$	I, q	c
e	d	b	h	i	$1/\sqrt{\varepsilon_0\mu_0}$

Для решения задачи можно было бы воспользоваться системой уравнений Максвелла. Однако проще исходить из выражений для плотности энергии, ее потока и силы:

$$w_e = \frac{1}{8\pi}(\mathbf{ED}), \quad w_m = \frac{1}{8\pi}(\mathbf{HB}), \quad \mathbf{S} = \frac{c}{4\pi}[\mathbf{EH}], \quad \mathbf{F} = q\mathbf{E}.$$

Преимущество такого подхода состоит в том, что в написанных соотношениях слева стоят чисто механические величины, которые не должны преобразовываться. После умножения на переводные коэффициенты эти соотношения переходят в

$$w_e = \frac{ed}{8\pi}(\mathbf{ED}), \quad w_m = \frac{hb}{8\pi}(\mathbf{HB}), \quad \mathbf{S} = \frac{eh}{4\pi\sqrt{\varepsilon_0\mu_0}}[\mathbf{EH}], \quad \mathbf{F} = eiq\mathbf{E}.$$

Переводные коэффициенты надо подобрать так, чтобы эти соотношения переходили в

$$w_e = \frac{1}{2}(\mathbf{ED}), \quad w_m = \frac{1}{2}(\mathbf{HB}), \quad \mathbf{S} = [\mathbf{EH}], \quad \mathbf{F} = q\mathbf{E}.$$

Для этого должно быть

$$\frac{ed}{4\pi} = \frac{hb}{4\pi} = \frac{eh}{4\pi\sqrt{\varepsilon_0\mu_0}} = ie = 1.$$

Так как один из этих коэффициентов можно выбрать произвольно, то мы наложим дополнительное «условие симметрии» $e/\sqrt{\varepsilon_0} = h/\sqrt{\mu_0}$.

После этого найдем

$$e=\sqrt{4\pi\varepsilon_0},\quad h=\sqrt{4\pi\mu_0},\quad d=\sqrt{\frac{4\pi}{\varepsilon_0}},\quad b=\sqrt{\frac{4\pi}{\mu_0}},\quad i=\frac{1}{\sqrt{4\pi\varepsilon_0}}.$$

Легко проверить, что все фундаментальные уравнения Максвелла с помощью найденных коэффициентов преобразуются правильно. Возьмем, например, уравнение (82.1а). После умножения его на переводные коэффициенты получим

$$\sqrt{4\pi\mu_0}\,\mathrm{rot}\,\mathbf{H}=4\pi\frac{\sqrt{\varepsilon_0\mu_0}}{\sqrt{4\pi\varepsilon_0}}\,\mathbf{j}+\sqrt{\varepsilon_0\mu_0}\sqrt{\frac{4\pi}{\varepsilon_0}}\,\frac{\partial\mathbf{D}}{\partial t},$$

а после сокращения

$$\mathrm{rot}\,\mathbf{H}=\mathbf{j}+\frac{\partial\mathbf{D}}{\partial t},$$

а это есть уравнение Максвелла (85.1а) в системе СИ.

Пользуясь найденными коэффициентами, легко найти переводные коэффициенты и для остальных физических величин. Некоторые из них приведены в табл. 1. Коэффициенты обратного преобразования от системы СИ к гауссовой равны обратным значениям коэффициентов, служащих для прямого преобразования. Пусть, например, соотношение

Таблица 1. Таблица перевода выражений и формул из гауссовой системы в систему СИ и обратно

Наименование	Гауссова система	Система СИ
Скорость света	c	$1/\sqrt{\varepsilon_0\mu_0}$
Напряженность электрического поля	$\mathbf{E}$	$\sqrt{4\pi\varepsilon_0}\,\mathbf{E}$
Потенциал	φ	$4\pi\varepsilon_0\varphi$
Электрическое смещение	$\mathbf{D}$	$\sqrt{4\pi/\varepsilon_0}\,\mathbf{D}$
Заряд, плотность заряда, ток, плотность тока, поляризованность	$q,\ \rho,\ I,\ \mathbf{j},\ \mathbf{P}$	$\frac{1}{\sqrt{4\pi\varepsilon_0}}\ (q,\ \rho,\ I,\ j,\ \mathbf{P})$
Магнитная индукция, магнитный поток	$\mathbf{B},\ \Phi$	$\sqrt{4\pi/\mu_0}\ (\mathbf{B},\ \Phi)$
Напряженность магнитного поля	$\mathbf{H}$	$\sqrt{4\pi\mu_0}\,\mathbf{H}$
Магнитный момент, намагниченность	$\mathfrak{M},\ \mathbf{I}$	$\sqrt{\mu_0/4\pi},\ (\mathfrak{M},\ \mathbf{I})$
Электрическая проницаемость, магнитная проницаемость (относительные)	$\varepsilon,\ \mu$	$\varepsilon,\ \mu$
Электрическая поляризуемость, магнитная восприимчивость	$\alpha,\ \varkappa$	$(4\pi)^{-1}\ (\alpha,\ \varkappa)$
Удельная проводимость	λ	$\lambda/(4\pi\varepsilon_0)$
Сопротивление	R	$4\pi\varepsilon_0 R$
Емкость	C	$(4\pi\varepsilon_0)^{-1}c$
Индуктивность	L	$(4\pi/\mu_0)L$

системы СИ $\mathbf{D} = \varepsilon_0\mathbf{E} + \mathbf{P}$ требуется перевести в гауссову систему. Произведя соответствующую замену, получим

$$\sqrt{\frac{\varepsilon_0}{4\pi}}\,\mathbf{D} = \varepsilon_0 \frac{1}{\sqrt{4\pi\varepsilon_0}}\,\mathbf{E} + \sqrt{4\pi\varepsilon_0}\,\mathbf{P},$$

или после сокращения $\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P}$.

5. Приведем еще табл. 2 для перевода численных значений физических величин из СИ в гауссову систему и обратно. Множители 3 (кроме входящих в показатели степени) при точных расчетах следует заменить на 2,99792458 в соответствии с точным значением скорости света. Например, в строке «электрическая индукция» вместо $12\pi\cdot10^5$ при точных расчетах следует брать значение $2{,}997924582 \cdot 4\pi \cdot 10^5$. В тех случаях, когда для гауссовых единиц существует общепринятое наименование, оно приведено в таблице. В остальных случаях приведено только число таких единиц.

Таблица 2. Таблица перевода численных значений физических величин из системы СИ в гауссову систему

Наименование	Обозначение	Система СИ	Гауссова система
Длина	l	1 м (метр)	10^2 см
Масса	m	1 кг (килограмм)	10^3 г
Время	t	1 с (секунда)	1 с
Сила	$\mathbf{F}$	1 Н (ньютон)	10^5 дин
Работа, энергия	A, $\mathscr{E}$	1 Дж (джоуль)	10^7 эрг
Мощность	P	1 Вт (ватт)	10^7 эрг/с
Давление	$\mathscr{P}$	1 Па (паскаль)	10 дин/см^2
Сила электрического тока	I	1 А (ампер)	$3 \cdot 10^9$
Электрический заряд	q	1 Кл (кулон)	$3 \cdot 10^9$
Напряженность электрического поля	$\mathbf{E}$	1 В/м (вольт на метр)	$\frac{1}{3} \cdot 10^{-4}$
Электрический потенциал	φ	1 В (вольт)	$\frac{1}{3} \cdot 10^{-2}$
Поляризованность	$\mathbf{P}$	1 Кл/м^2 (кулон на квадратный метр)	$3 \cdot 10^5$
Электрическое смещение	$\mathbf{D}$	1 Кл/м^2 (кулон на квадратный метр	$12\pi \cdot 10^5$
Электрическая емкость	C	1 Ф (фарад)	$9 \cdot 10^{11}$ см
Электрическое сопротивление	R	1 Ом (ом)	$\frac{1}{9} \cdot 10^{-11}$ с $\cdot$ см$^{-1}$

Продолжение табл. 2

Наименование	Обозначение	Система СИ	Гауссова система
Удельное электрическое сопротивление	ρ	1 Ом · м (ом-метр)	$\frac{1}{9}\cdot 10^{-9}$ с
Электрическая проводимость	$\Lambda = \frac{1}{R}$	1 См (сименс)	$9\cdot 10^{11}$ см · $с^{-1}$
Удельная электрическая проводимость	λ	1 См/м (сименс на метр)	$9\cdot 10^{9}$ $с^{-1}$
Магнитный поток	Φ	1 Вб (вебер)	10^{8} Мкс
Магнитная индукция	$\mathbf{B}$	1 Т (тесла)	10^{4} Гс
Напряженность магнитного поля	$\mathbf{H}$	1 А/м (ампер на метр)	$4\pi\cdot 10^{-3}$ Э
Намагниченность	$\mathbf{I}$	1 А/м (ампер на метр)	$\frac{1}{4\pi}\cdot 10^{4}$ Гс
Индуктивность	L	1 Г (генри)	10^{9} см

6. Постоянство СИ — чисто инженерное. При расчетах по формулам этой системы окончательные численные результаты сразу получаются в практических единицах — вольтах, амперах, омах, джоулях и пр. Недостатки СИ более существенны и носят *принципиальный характер.* Для описания электромагнитного поля СИ вводит *четыре* вектора $\mathbf{E}$, $\mathbf{D}$, $\mathbf{B}$, $\mathbf{H}$ *не только в материальных средах, но и в вакууме.* По существу СИ не отличается от электротехнической системы, предложенной Джорджи в начале двадцатого столетия. В то время уравнения Максвелла были мало распространены, в особенности в электротехнике, а их физическое толкование еще не установилось. Преобладали механические воззрения на природу электромагнитного поля. Считалось, что вакуум (по тогдашней терминологии «эфир»), как среда, в которой возбуждается электромагнитное поле, принципиально не отличается от остальных материальных сред.

Система Джорджи возникла под сильным влиянием этих старых воззрений. Согласно последним, векторы $\mathbf{E}$ и $\mathbf{D}$, $\mathbf{B}$ и $\mathbf{H}$ в вакууме отличаются друг от друга не только численным множителем, но и *по существу.* Они находятся примерно в том же отношении, как растягивающее усилие и смещение в теории упругости.

Но физика уже давно отказалась от того принципиального различия между $\mathbf{E}$ и $\mathbf{D}$, $\mathbf{B}$ и $\mathbf{H}$, которое возникло на почве механической теории эфира. Она установила, что электромагнитное поле в вакууме описано полностью, если задан *один электрический вектор* $\mathbf{E}$ *и один магнитный вектор* $\mathbf{B}$. Совпадение численных значений $\mathbf{E}$ и $\mathbf{D}$, а также $\mathbf{B}$ и $\mathbf{H}$ в вакууме является для физика *не результатом произвольного соглашения, а выражением действительного тождества этих величин.*

Напротив, введение размерных множителей ε_0 и μ_0 в СИ представляется ему *искусственным приемом*, с помощью которого формулы приводятся к виду, удобному для инженерных расчетов в электротехнике. Дух отживших физических представлений витает и над СИ. В частности, он повлиял на терминологию: первоначально величины ε_0 и μ_0 назывались *диэлектрической* и *магнитной проницаемостями вакуума.* Только полная бессодержательность таких понятий заставила отказаться от этих терминов и заменить их нейтральными терминами «*электрическая*» и «*магнитная*» *постоянные.*

От этого, конечно, величины ε_0 и μ_0 не сделались содержательными. Эти ненужные величины в СИ засоряют физику и загромождают формулы.

До появления теории относительности ко всякой физически рациональной системе единиц необходимо было предъявлять требование, чтобы в ней векторы **E** и **D**, а также **B** и **H** имели *одинаковую размерность*. Размерности электрического и магнитного полей могли и не совпадать. Теория относительности усилила это требование. Она показала, что деление электромагнитного поля на электрическое и магнитное *относительно*, т. е. зависит от выбора системы отсчета. Оказалось, что векторы **E** и **B** объединяются в один антисимметричный тензор второго ранга, а векторы **D** и **H** — в другой. Естественно, что компоненты одного и того же тензора должны иметь *одинаковые размерности*. После этого стало почти абсолютной необходимостью, чтобы имели одинаковую размерность *все четыре* вектора **E**, **B**, **D** и **H**. Этому требованию СИ не удовлетворяет. В ней надо вводить размерные множители для уравнивания размерностей компонент обоих тензоров. Напротив, гауссова система СГС ему удовлетворяет, хотя она и была создана задолго до теории относительности. В этом отношении СИ не более логична, чем, скажем, система, в которой длина измеряется в метрах, ширина — в парсеках, а высота — в световых годах.

Из всех предложенных систем единиц гауссова система СГС до настоящего времени остается наилучшей в физике. Для физика значительно легче выполнить все вычисления в гауссовой системе и лишь в конце, если это требуется, сделать пересчет к практическим единицам, чем все время быть обремененным грузом искусственно введенных ненужных величин и понятий (ε_0 и μ_0, абсолютная и относительные проницаемости, **E** и **D**, **B** и **H** в вакууме и пр.), возникших на почве СИ.

Однако приходится считаться с широким распространением, которое получила СИ.

Глава V

ДВИЖЕНИЕ ЗАРЯЖЕННЫХ ЧАСТИЦ В ЭЛЕКТРОМАГНИТНЫХ ПОЛЯХ

§ 86. Движение в постоянных и однородных полях

1. Вопросы о движении заряженных частиц в электромагнитных полях были уже частично рассмотрены в виде задач к § 4 и 57. Изучим теперь эти вопросы более систематически. При этом мы исключим из рассмотрения весь материал, относящийся к электронной и ионной оптике. Его предполагается включить в следующий том после изложения геометрической оптики. Не будем также касаться ускорителей и масс-спектрометрии, так как этот материал более уместно изложить в разделе атомной и ядерной физики.

Простейшим случаем является движение частиц в постоянных и однородных электромагнитных полях.

2. В постоянном электрическом поле на частицу с зарядом e действует сила $\mathbf{F} = e\mathbf{E}$. Если движение не релятивистское, а поле однородно, то частица движется с постоянным ускорением $\mathbf{a} = e\mathbf{E}/m$. Такое движение вполне аналогично движению частицы в постоянном однородном гравитационном поле. В общем случае траекторией движения будет парабола. Для релятивистских движений масса частицы m возрастает со скоростью v, а ускорение убывает. Разбор этого случая дается в задаче, помещенной в конце этого параграфа.

3. В постоянном магнитном поле на заряженную частицу действует сила $\mathbf{F}_m = (e/c)[\mathbf{vB}]$. Эта сила перпендикулярна к скорости $\mathbf{v}$, а потому работы не производит. Она только искривляет траекторию, но не изменяет абсолютную величину скорости частицы. Не меняется, следовательно, и релятивистская масса частицы m.

Допустим теперь, что магнитное поле не только постоянно, но и однородно. Если скорость частицы направлена вдоль поля $\mathbf{B}$, то сила $\mathbf{F}_m$ обратится в нуль. Частица будет двигаться прямолинейно с постоянной скоростью, магнитное поле не влияет на движение частицы, если последнее происходит вдоль поля.

Если же частица движется перпендикулярно к магнитному полю, то ее скорость, оставаясь постоянной по абсолютной величине, меняется по направлению. Сила $\mathbf{F}_m = (e/c)[\mathbf{vB}]$ также постоянна по абсолютной величине и нормальна к траектории частицы. Отсюда следует, что траекторией частицы будет окружность, плоскость которой перпендикулярна к магнитному полю. Направление вращения частицы по окружности показано на рис. 195 (магнитное поле направлено к читателю). Если заряд e положителен, то направления вектора $\mathbf{B}$ и угловой скорости вращения $\boldsymbol{\omega}$ противоположны. Если же заряд e отрицателен,

то эти направления совпадают. Ускорение частицы направлено к центру окружности, по которой она вращается, и равно $\omega^2\rho$, где ρ — радиус этой окружности. Значение угловой скорости ω найдется из уравнения движения $m\omega^2\rho = |e|Bv/c$. Так как $v = \omega\rho$, то отсюда получаем $\omega = |e|B/(mc)$, или в векторной форме

$$\boldsymbol{\omega} = -\frac{e\mathbf{B}}{mc}. \tag{86.1}$$

Величина ω называется *циклотронной частотой*, а ρ — *циклотронным* или *ларморовским радиусом* [1]. Заметим, что формула (86.1) справедлива как для нерелятивистских, так и для релятивистских движений, если только под m понимать релятивистскую массу частицы.

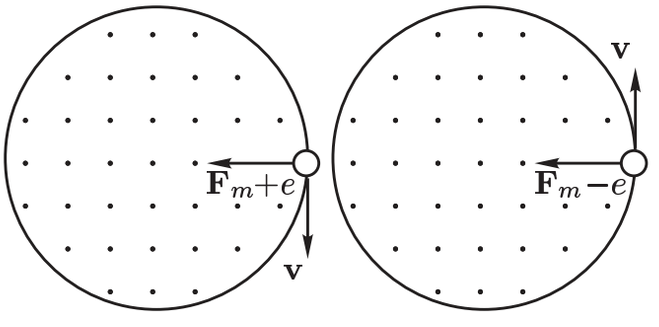


Рис. 195

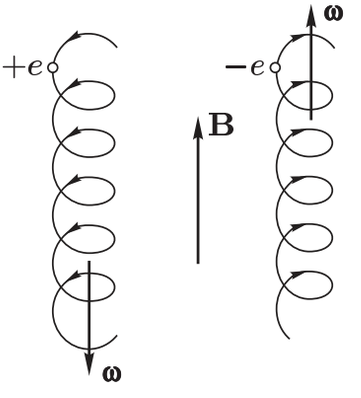


Рис. 196

При рассмотрении общего случая, когда скорость $\mathbf{v}$ направлена под углом к магнитному полю, ограничимся нерелятивистскими скоростями. Представим скорость $\mathbf{v}$ в виде $\mathbf{v} = \mathbf{v}_{\parallel} + \mathbf{v}_{\perp}$, где $\mathbf{v}_{\parallel}$ — скорость вдоль поля, а $\mathbf{v}_{\perp}$ — перпендикулярно к нему. Движения с этими скоростями независимы. Первое есть равномерное прямолинейное движение вдоль поля со скоростью $\mathbf{v}_{\parallel}$, второе — равномерное вращение по окружности вокруг поля с угловой скоростью (86.1). Радиус этой окружности равен $\rho = \mathbf{v}_{\perp}/\omega$. В результате сложения обоих движений возникает движение по спирали, ось которой параллельна магнитному полю (рис. 196).

4. Допустим теперь, что на постоянное однородное магнитное поле $\mathbf{B}$ наложено постоянное же однородное электрическое поле $\mathbf{E}$. Будем предполагать, что $E \ll B$. Только при выполнении этого условия, как будет видно из дальнейшего, движение может происходить с нерелятивистскими скоростями. Движение описывается уравнением

$$m\dot{\mathbf{v}} = e\left(\mathbf{E} + \frac{1}{c}[\mathbf{vB}]\right). \tag{86.2}$$

[1] Величину ω часто называют также *ларморовской частотой*. Однако, во избежание путаницы, мы будем называть ларморовской частотой величину, вдвое меньшую, т. е. $|e|B/(2mc)$ (см. § 76).

Введем систему отсчета, равномерно движущуюся относительно исходной со скоростью $\mathbf{v}_д$. В новой системе относительная скорость частицы $\mathbf{v}'$ удовлетворяет уравнению

$$m\dot{\mathbf{v}}' = e\left(\mathbf{E} + \frac{1}{c}[\mathbf{v}'\mathbf{B}] + \frac{1}{c}[\mathbf{v}_д\mathbf{B}]\right).$$

Допустим сначала, что вектор $\mathbf{E}$ перпендикулярен к магнитному полю $\mathbf{B}$. Подберем скорость $\mathbf{v}_д$ так, чтобы выполнялось условие

$$\mathbf{E} + \frac{1}{c}[\mathbf{v}_д\mathbf{B}] = 0,$$

из которого следует

$$\mathbf{v}_д = c\frac{[\mathbf{EB}]}{B^2}, \tag{86.3}$$

а потому

$$m\dot{\mathbf{v}}' = \frac{e}{c}[\mathbf{v}'\mathbf{B}].$$

В новой системе отсчета из уравнения относительного движения электрическое поле исключилось. Его влияние компенсировано скоростью $\mathbf{v}_д$. Движение частицы происходит так, как если бы было только одно магнитное поле, т. е. по спирали. В исходной системе отсчета магнитное поле заставляет частицу равномерно вращаться по спирали. На это вращение накладывается медленное равномерное движение со скоростью $\mathbf{v}_д$, определяемой формулой (86.3). Такое движение называется *электрическим дрейфом.*

Значение скорости электрического дрейфа дается выражением $\mathbf{v}_д = cE/B$. При $E > B$ это выражение приводило бы к результату $\mathbf{v}_д > c$, что не имеет смысла. Это показывает, что нерелятивистское рассмотрение справедливо только при выполнении условия $E \ll B$, что и предполагалось выше.

Допустим теперь, что постоянное электрическое поле направлено под углом к магнитному. Разложим поле $\mathbf{E}$ на составляющую $\mathbf{E}_{\|}$ вдоль $\mathbf{B}$ и на составляющую $\mathbf{E}_{\perp}$, перпендикулярную к $\mathbf{B}$: $\mathbf{E} = \mathbf{E}_{\|} + \mathbf{E}_{\perp}$. Тогда движение частицы представится в виде суперпозиции трех движений: 1) равноускоренного движения в направлении $\mathbf{B}$ ускорением $\mathbf{a}_{\|} = (e/m)\mathbf{E}_{\|}$; 2) равномерного вращения по окружности вокруг $\mathbf{B}$ с угловой скоростью (86.1); 3) электрического дрейфа со скоростью

$$\mathbf{v}_д = \frac{c}{B^2}[\mathbf{E}_{\perp}\mathbf{B}] = \frac{c}{B^2}[\mathbf{EB}]. \tag{86.3а}$$

Направление и скорость электрического дрейфа не зависят от знака заряда и массы частицы: положительные и отрицательные частицы дрейфуют совершенно одинаково. В результате сложения движений 1) и 2) возникает движение по спирали. Ось спирали направлена параллельно магнитному полю, однако шаг спирали из-за наличия ускорения $\mathbf{a}_{\|}$ будет меняться во времени. Такая картина движения сохранится до тех пор, пока скорость частицы из-за наличия того же ускорения не возрастет настолько, что движение перейдет в релятивистское.

5. Заметим еще, что такое же движение в постоянном однородном магнитном поле возникает и в том случае, когда на частицу будет действовать не электрическое поле, а любая постоянная сила $\mathbf{F}$, например сила тяжести. Роль поля $\mathbf{E}$ будет играть вектор $\mathbf{F}/e$. В результате возникнет дрейфовое движение со скоростью

$$\mathbf{v}_{\text{д}} = \frac{c}{B^2 e}\,[\mathbf{F}\mathbf{B}]. \tag{86.4}$$

Однако теперь направление действующей силы $\mathbf{F}$ не зависит от знака заряда частицы, а потому скорости дрейфа положительных и отрицательных частиц будут направлены в противоположные стороны.

ЗАДАЧА

Исследовать релятивистское движение заряженной частицы в постоянном однородном электрическом поле.

Р е ш е н и е. Движение происходит в плоскости, параллельной электрическому полю $\mathbf{E}$ и начальной скорости частицы $\mathbf{v}_0$. Примем эту плоскость за координатную плоскость XY, направив ось X параллельно полю $\mathbf{E}$. Тогда уравнения движения можно записать в виде

$$\dot{p}_x = eE, \quad \dot{p}_y = 0,$$

и, следовательно,

$$p_x = eEt, \quad p_y = p_0 = \text{const},$$

причем за начальный принят момент времени, когда импульс $\mathbf{p}$ направлен параллельно оси Y. Релятивистская масса частицы найдется из соотношения

$$(mc)^2 = (m_0 c)^2 + p^2.$$

Используя его, находим уравнения для координат частицы:

$$\sqrt{(m_0c)^2 + p_0^2 + (eEt)^2}\,\frac{dx}{dt} = eEct,$$

$$\sqrt{(m_0c)^2 + p_0^2 + (eEt)^2}\,\frac{dy}{dt} = p_0 c,$$

или

$$dx = \frac{ct\,dt}{\sqrt{\tau^2 + t^2}}, \quad dy = \frac{p_0}{eE}\,\frac{c\,dt}{\sqrt{\tau^2 + t^2}},$$

где введено обозначение

$$\tau^2 = \frac{(m_0c)^2 + p_0^2}{(eE)^2}. \tag{86.5}$$

После интегрирования получаем

$$x = c\sqrt{\tau^2 + t^2}, \quad y = \frac{p_0 c}{eE}\,\operatorname{arcsh}\frac{t}{\tau}. \tag{86.6}$$

Этими уравнениями и определяется движение. Найдя t из второго уравнения и подставив в первое, получим уравнение траектории:

$$x = c\tau\,\operatorname{ch}\frac{eEy}{p_0 c}. \tag{86.7}$$

Это — цепная линия. При $|eEy|/(p_0c) \ll 1$ она, как и следовало ожидать, переходит в параболу

$$x = c\tau\left[1 + \frac{eEy}{2(p_0c)^2}\right]. \tag{86.8}$$

§ 87. Дрейф заряженной частицы в неоднородном магнитном поле при наличии слабого электрического поля

1. В общем случае, когда магнитное и электрическое поля неоднородны и меняются во времени, движение частицы приобретает весьма сложный и запутанный характер. Проинтегрировать уравнения движения в аналитической форме в этом случае не удается. Для расчета движения приходится обращаться к сложным и утомительным численным методам. Есть, однако, случай, когда можно нарисовать сравнительно простую и обозримую картину движения, не обращаясь к численным методам расчета. Это будет тогда, когда магнитное поле сильное, а его изменения в пространстве и во времени происходят медленно. На магнитное поле может накладываться электрическое, но оно должно быть слабым по сравнению с магнитным. При этих условиях задачу можно приближенно решать по методу последовательных приближений.

В нулевом приближении полностью пренебрегают электрическим полем, а также пространственно-временными неоднородностями магнитного поля. Движение частицы представляется как быстрое вращение по ларморовскому кружку, центр которого перемещается вдоль магнитной силовой линии. Электрическое поле и пространственно-временные неоднородности магнитного поля учитываются в первом приближении. Они проявляются в том, что центр ларморовского кружка получает дополнительное медленное движение. Такое движение называется *дрейфом*, а центр самого ларморовского кружка — *ведущим центром* частицы. Параметры движения — циклотронная частота ω, радиус ларморовского кружка ρ, продольная $v_{\parallel}$ и поперечная $v_{\perp}$ скорости частицы при этом будут медленно меняться. Медленность означает, что за циклотронный период $T = 2\pi/\omega$ изменения этих параметров будут малы по сравнению со значениями самих параметров. Для этого необходимо, чтобы на протяжении ларморовского кружка и в течение циклотронного периода магнитное поле изменялось мало. Так как магнитное поле предполагается сильным, то размеры ларморовского кружка будут малы. Быстрые вращения по такому кружку часто не представляют интереса. Чтобы их исключить, достаточно усреднить движение частицы по временам порядка циклотронного периода. Тогда вместо движения самой частицы останется усредненное, или сглаженное, движение ее ведущего центра. Теория, рассматривающая движение частицы в такой постановке, называется *дрейфовой.* За последние два-три десятилетия дрейфовая теория получила многочисленные применения при анализе движения космических частиц в межзвездных и межпланетных магнитных полях, а также в различного рода магнитных ловушках, предназначенных для удержания и нагрева плазмы с целью получения управляемых термоядерных реакций. Ниже в упрощенной форме излагаются основы дрейфовой теории и ее простейшие результаты.

2. Задача дрейфовой теории — определить скорость плавного движения ведущего центра, обусловленного электрическим полем и пространственно-временными неоднородностями магнитного поля. В силу уравнения Максвелла $\partial\mathbf{B}/\partial t = -c\,\mathrm{rot}\,\mathbf{E}$ временные неоднородности магнитного поля можно исключить, выразив их через соответствующие пространственные неоднородности поля электрического. Магнитное поле $\mathbf{B}$, как уже сказано, предполагается сильным. Величины, пропорциональные $\mathbf{B}$, считаются величинами *нулевого порядка.* Это самые большие величины, но в дрейфовую теорию они не входят, так как проявляются только в быстрых вращениях по циклотронным окружностям, которые выпадают в результате усреднения по циклотронным периодам. Члены, пропорциональные электрическому полю и первым пространственным производным магнитного поля, считаются величинами *первого порядка малости.* Влиянием первых и высших производных вектора $\mathbf{E}$, вторых и высших производных вектора $\mathbf{B}$ будем пренебрегать. В этом приближении скорость плавного движения ведущего центра будет представляться линейной функцией напряженности электрического поля $\mathbf{E}$ и первых пространственных производных вектора $\mathbf{B}$. Понятно, что в принятом приближении все слагаемые этой линейной функции *независимы* и могут быть вычислены *независимо друг от друга.* Изменение магнитного поля в пространстве складывается из изменения его по абсолютной величине и из изменения по направлению. В соответствии с этим дрейфовое движение ведущего центра можно разложить на три движения: 1) *дрейф под действием электрического поля, или электрический дрейф*; 2) *дрейф, вызванный изменениями магнитного поля только по абсолютной величине*; 3) *дрейф, вызванный изменениями магнитного поля только по направлению* (т. е. искривлением магнитных силовых линий). Скорости всех этих дрейфов в линейном приближении можно вычислить *независимо*, что и будет сделано ниже.

3. Электрический дрейф. Как выяснено выше, в первом приближении неоднородность электрического поля учитывать не надо. Не надо учитывать и неоднородность магнитного поля, так как она может сказаться на скорости электрического дрейфа лишь во втором приближении. Поэтому можно воспользоваться результатом (86.3а) и для неоднородных полей. Однако в этом случае формула (86.3а) уже не будет точной, а дает скорость электрического дрейфа только в первом приближении.

Происхождение электрического дрейфа и аналогичного дрейфа, вызванного любой малой возмущающей силой $\mathbf{F}$, наложенной на магнитное поле, легко понять из следующих соображений. Предполагая магнитное поле $\mathbf{B}$ однородным, рассмотрим проекцию траектории частицы на плоскость, перпендикулярную к этому полю. Примем эту плоскость за плоскость рисунка, направив магнитное поле к читателю (рисунки 197 и 198). В отсутствие электрического поля проекцией траектории частицы будет окружность ларморовского радиуса ρ. Для

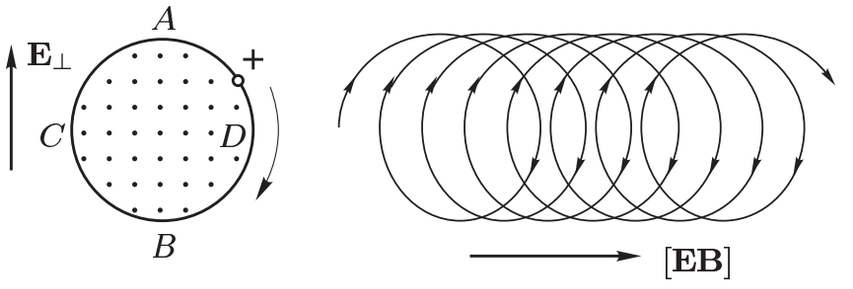


Рис. 197

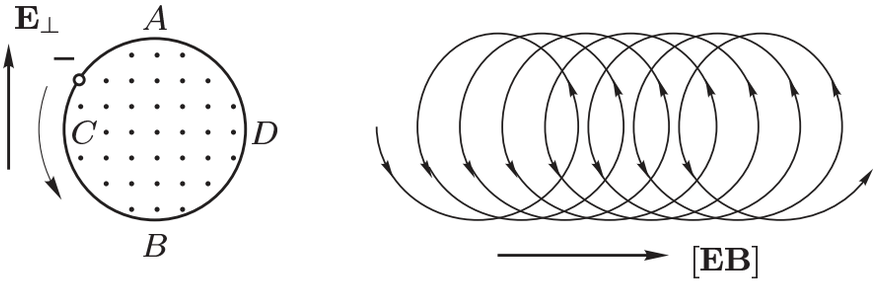


Рис. 198

определенности будем иметь в виду движение положительно заряженной частицы (рис. 197). Наложим теперь электрическое поле, составляющая $\mathbf{E}_\perp$ которого направлена вверх. Так как при перемещении частицы вверх поле $\mathbf{E}_\perp$ совершает над ней положительную работу, то скорость частицы $\mathbf{v}_\perp$ в верхнем положении A будет больше, чем в нижнем положении B. Кроме того, в точке A силы электрического и магнитного полей действуют на частицу в противоположные стороны, а в точке B — в одну и ту же сторону. Оба эти обстоятельства приводят к тому, что радиус кривизны r проекции траектории в верхней части увеличивается, а в нижней уменьшается, как это видно из выражений

$$\frac{1}{r} = \frac{eB}{mv_\perp c} - \frac{eE_\perp}{mv_\perp^2} \quad \text{(в точке } A\text{)},$$

$$\frac{1}{r} = \frac{mB}{mv_\perp c} + \frac{eE_\perp}{mv_\perp^2} \quad \text{(в точке } B\text{)}.$$

В результате окружность перейдет в незамкнутую кривую, двигаясь по которой проекция частицы будет медленно перемещаться вправо (рис. 197, справа). Это перемещение и есть электрический дрейф. Для отрицательно заряженной частицы аналогичное перемещение представлено на рис. 198. В обоих случаях частица дрейфует вправо, т. е. направление электрического дрейфа, как это и должно быть, *не зависит от знака заряда частицы.*

Скорость электрического дрейфа легко определить из следующих соображений. В положениях C и D (рис. 197) частица движется с одной и той же скоростью v_C, которая больше скорости частицы в нижнем положении B и меньше скорости ее в верхнем положении A на одну и ту же величину u. Величина u и есть скорость электрического дрейфа: $u = v_\text{д}$. Действительно, если ввести систему отсчета, движущуюся вправо

со скоростью u, то в этой системе скорости частицы в положениях A и B сравняются, и частица будет равномерно вращаться по окружности. Величину u определим из уравнения энергии: $m(v_C+u)^2/2-mv_C^2/2 = eE_\perp\rho$. Подставляя сюда $\rho = mcv_C/(eB)$ и пренебрегая квадратом скорости u, получим

$$v_д = u = \frac{E_\perp}{B}\,c.$$

что совпадает с формулой (86.3а).

4. Дрейф, вызванный изменениями магнитного поля только по числовому значению. Для определения мгновенной скорости дрейфа введем локальную систему отсчета, направив ось Z вдоль магнитного поля $\mathbf{B}$. Единичный вектор вдоль $\mathbf{B}$ обозначим через $\mathbf{h}$. Направление главной нормали N к магнитной силовой линии примем за ось Y. Ось X направим в отрицательную сторону бинормали $\mathbf{b} = [\mathbf{hN}]$ (рис. 199).

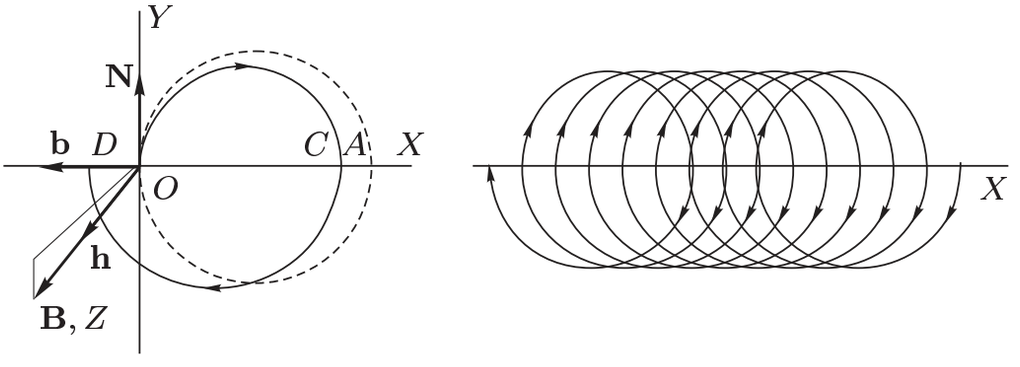


Рис. 199

Максимальное изменение числового значения магнитного поля происходит в направлении главной нормали $\mathbf{N}$. В случае, который нас более всего интересует, когда частица движется в области пространства, где не текут электрические токи, поле $\mathbf{H} \equiv \mathbf{B}$ в направлении $\mathbf{N}$ будет возрастать. Для доказательства возьмем в соприкасающейся плоскости две бесконечно близкие магнитные силовые линии BC и AD и два бесконечно коротких отрезка AB и CD, перпендикулярных к этим силовым линиям (рис. 200). Циркуляция вектора $\mathbf{H}$ по бесконечно малому контуру $BCDA$ будет $H_1l_1 - H_2l_2$, где l_1 и l_2 — длины сторон BC и AD, а H_1 и H_2 — напряженности магнитного поля на этих сторонах. Так как в отсутствие электрических токов эта циркуляция равна нулю, то магнитное поле H будет больше на более короткой стороне. Отсюда и следует наше утверждение.

Сейчас речь идет о влиянии на скорость дрейфа изменения магнитного поля, но не его направления. Поэтому можно отвлечься от кривизны магнитных силовых линий и считать их прямолинейными. Можно также отвлечься от наличия продольной составляющей скорости $\mathbf{v}_{||}$, так как движение вдоль магнитного поля не влияет на силы, действующие на частицу. Иными словами, во всех промежуточных расчетах под

$\mathbf{v}$ следует понимать поперечную скорость $\mathbf{v}_\perp$. Ради краткости опустим всюду значок $\perp$, за исключением окончательных формул, в которых поперечная скорость по-прежнему будет обозначаться через $\mathbf{v}_\perp$. Кроме того, во всех промежуточных расчетах заряд частицы e будем считать положительным.

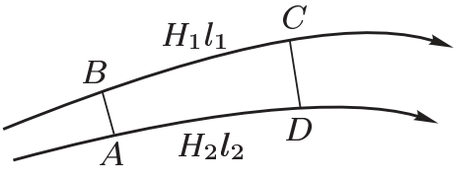


Рис. 200

Положительно заряженная частица будет вращаться по часовой стрелке, как указано на рис. 199. Выйдя из точки O, частица двигалась бы по окружности радиуса $\rho = v/\omega_B$, если бы магнитное поле $\mathbf{B}$ было постоянно. Эта окружность изображена на рис. 199 пунктиром. Она пересекает ось X в точке A на расстоянии $OA = 2\rho$ от точки O. В действительности магнитное поле, а с ним и кривизна траектории возрастают с возрастанием координаты y. Поэтому частица вернется к оси X в какой-то точке C, расположенной левее A. При движении в нижней половине плоскости XY, наоборот, кривизна траектории будет меньше, а потому частица пересечет ось X в точке D, расположенной левее O. Таким образом, за один оборот положительно заряженная частица сместится влево на отрезок OD. Отрицательно заряженная частица сместилась бы в противоположном направлении. При медленном изменении магнитного поля в пространстве траектория частицы за один оборот мало отличается от окружности, и смещение OD будет мало по сравнению с радиусом ρ. Путь частицы за время многих оборотов изображен на рис. 199 справа. Частица быстро вращается по окружности, центр которой медленно перемещается параллельно оси X. Такое перемещение и есть дрейф.

Найдем теперь скорость дрейфа $v_д$, усреднив движение частицы по быстрому ларморовскому вращению. Уравнение движения частицы $\dot{\mathbf{v}} = [\mathbf{v}\omega_B]$ запишем в координатной форме:

$$\dot{v}_x = v_y\omega_B, \quad \dot{v}_y = -v_x\omega_B.$$

Поскольку дрейф происходит параллельно оси X и отсутствует в направлении оси Y, скорость дрейфа $v_д = \overline{v}_x$ найдется из требования, что среднее значение v_y, а следовательно, и $\dot{v}_y = -v_x\omega_B$ равно нулю, т. е. $\overline{v_x\omega_B} = 0$. Разложим ω_B в ряд по степеням y и оборвем это разложение на линейном члене:

$$\omega_B = \omega_0 + \left(\frac{d\omega_B}{dy}\right)_{y=0} y.$$

Тогда

$$\overline{v}_x\omega_0 + \frac{d\omega_B}{dy}\,\overline{yv_x} = 0.$$

Так как величина $d\omega_B/dy$ предполагается малой, то среднее значение произведения $yv_x = y\dot{x}$ достаточно вычислить в нулевом приближении,

т. е. считать при вычислении, что частица вращается по окружности

$$x = \rho\cos\omega_0 t, \quad y = -\rho\sin\omega_0 t.$$

Тогда

$$\overline{y\dot{x}} = \rho^2\omega_0\overline{\sin^2\omega_0 t} = \frac{\rho^2\omega_0}{2},$$

и, следовательно,

$$v_{\text{д}} = \overline{v}_x = -\frac{\rho^2}{2}\,\frac{d\omega_B}{dy}.$$

В векторной форме:

$$\mathbf{v}_{\text{д}} = \frac{\rho^2}{2}\,\frac{d\omega_B}{dy}\,\mathbf{b}, \tag{87.1}$$

или с учетом соотношений $\rho = v/\omega_B$ и $\omega_B = eB/(mc)$

$$\mathbf{v}_{\text{д}} = \frac{mcv_\perp^2}{2eB^2}\,\frac{\partial B}{\partial N}\,\mathbf{b}. \tag{87.2}$$

В этом виде формула верна и для положительно заряженной, и для отрицательно заряженной частицы. *Положительно заряженная частица дрейфует в положительном направлении бинормали к магнитной силовой линии, отрицательно заряженная — в отрицательном направлении.*

Придадим теперь формуле (87.2) более наглядную геометрическую форму, предполагая, что пространство, в котором движется частица, свободно от электрических токов. Имея в виду, что поле $\mathbf{B}$ в рассматриваемой точке направлено вдоль оси Z, запишем формулу (87.2) в виде

$$\mathbf{v}_{\text{д}} = \frac{mcv_\perp^2}{2eB^2}\,\frac{\partial B_z}{\partial y}\,\mathbf{b}.$$

Так как при отсутствии электрических токов $\operatorname{rot}\mathbf{B} = 0$, то $\partial B_z/\partial y = \partial B_y/\partial z$, и, следовательно,

$$\mathbf{v}_{\text{д}} = \frac{mcv_\perp^2}{2eB^2}\,\frac{\partial B_y}{\partial z}\,\mathbf{b}.$$

Как видно из рис. 201, $\operatorname{tg}\alpha = B_y/B_z$, где α — угол между касательной к магнитной силовой линии и осью Z. В начале координат O касательная горизонтальна, а потому в его малой окрестности $\operatorname{tg}\alpha$ можно заменить на α. По той же причине дифференцирование по дуге магнитной силовой линии можно заменить дифференцированием по координате z. Радиус кривизны R силовой линии определяется соотношением

$$\frac{1}{R} = \frac{d\alpha}{ds} = \frac{\partial}{\partial z}\left(\frac{B_y}{B_z}\right) = \frac{1}{B_z}\,\frac{\partial B_y}{dz} - \frac{B_y}{B_z^2}\,\frac{\partial B_z}{\partial z}, \quad \text{или} \quad \frac{1}{R} = \frac{1}{B}\,\frac{\partial B_y}{\partial z},$$

так как в точке O $B_z = B$, $B_y = 0$. В результате получается

$$\mathbf{v}_{\text{д}} = \frac{mcv_\perp^2}{2eBR}\,\mathbf{b}. \tag{87.3}$$

5. Дрейф, вызванный искривлением магнитных силовых линий. Для расчета скорости этого дрейфа введем локальную систему отсчета, вращающуюся вокруг центра кривизны C магнит-

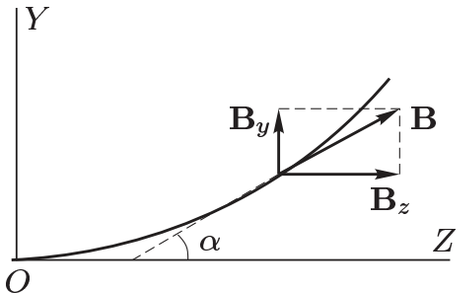


Рис. 201

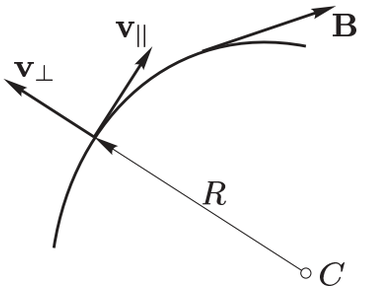


Рис. 202

ной силовой линии (рис. 202). Ось вращения направим параллельно бинормали $\mathbf{b}$ этой силовой линии (она перпендикулярна к плоскости рисунка). Числовое значение угловой скорости вращения $\mathbf{\Omega}$ определим из условия $\mathbf{v}_{\|} = [\mathbf{\Omega R}]$. Тогда в рассматриваемой локальной системе отсчета частица будет иметь только поперечную скорость $\mathbf{v}_{\perp}$, которая играет роль относительной скорости $\mathbf{v}_{\text{отн}}$. При рассмотрении движения в локальной системе отсчета к действующим силам надо добавить две силы инерции: кориолисову и центробежную. Силу инерции, вызванную неравномерностью вращения, учитывать не надо, так как она может влиять на дрейф частицы лишь во втором или высшем порядке малости. Кориолисова сила инерции $2m[\mathbf{v}_{\text{отн}}\mathbf{\Omega}] = 2m\Omega[\mathbf{v}_{\perp}\mathbf{b}]$ направлена вдоль силовой линии, а потому она будет изменять только продольную составляющую $v_{\|}$ скорости $\mathbf{v}$. Центробежная сила инерции $mv_{\|}^2\mathbf{R}/R^2$ вызовет дрейф в перпендикулярном направлении. Скорость этого дрейфа, согласно формуле (86.4), определяется выражением

$$\mathbf{v}_{\text{д}} = \frac{c}{B^2 e}\left[\frac{mv_{\|}^2}{R^2}\,\mathbf{RB}\right] = -\frac{mcv_{\|}^2}{BeR}\,[\mathbf{Nh}].$$

Если $\mathbf{N}$ — единичный вектор главной нормали к магнитной силовой линии, то $\mathbf{R} = -R\mathbf{N}$. Поэтому, полагая в последнем соотношении $\mathbf{B} = B\mathbf{h}$ и учитывая, что $\mathbf{b} = [\mathbf{hN}]$, получим

$$\mathbf{v}_{\text{д}} = \frac{mcv_{\|}^2}{BRe}\,\mathbf{b}. \tag{87.4}$$

Поскольку дрейф, выражаемый этой формулой, вызывается центробежной силой инерции, он называется *центробежным дрейфом.*

6. Теперь можно обратиться к общему случаю, т. е. к случаю произвольного электромагнитного поля, для которого справедливо дрейфовое приближение. Так как все дрейфы, которые были рассмотрены, в первом приближении независимы, то в общем случае их надо просто сложить. Таким путем для скорости сглаженного движения ведущего

центра в пространстве, где не текут электрические токи, получаем

$$\mathbf{V} = v_{\|}\mathbf{h} + \frac{c}{B^2}\,[\mathbf{EB}] + \frac{mc}{eBR}\left(v_{\|}^2 + \frac{1}{2}\,v_{\perp}^2\right)\mathbf{b}. \tag{87.5}$$

Здесь произведены небольшие изменения в обозначениях. Под $\mathbf{h}$ мы теперь понимаем единичный вектор касательной к магнитной силорой линии, проходящей *через ведущий центр*, а не через саму частицу. Величины же $v_{\|}$ и $v_{\perp}$ означают *усредненные скорости* частицы вдоль этого *нового вектора* $\mathbf{h}$ и перпендикулярно к нему. Точно так же значения полей $\mathbf{B}$ и $\mathbf{E}$ мы берем *в точке нахождения ведущего центра*, а не частицы. Такая замена совершенно не затрагивает все слагаемые в правой части формулы (87.5), за исключением первого, так как она меняет эти слагаемые только в первом или высшем порядке малости. Но эта замена существенна для слагаемого $v_{\|}\mathbf{h}$, так как она нулевого порядка малости. Если бы сохранить прежний смысл вектора $\mathbf{h}$, то в это слагаемое надо было бы ввести поправку первого порядка малости. Если же понимать $\mathbf{h}$ в новом смысле, как мы сделали, то такая поправка не нужна.

Итак, *в сильном, но слабо неоднородном магнитном поле при наличии слабого электрического поля заряженная частица быстро вращается по ларморовской окружности. Центр ларморовской окружности движется вдоль магнитной силовой линии со скоростью* $v_{\|}$ *и испытывает дрейф перпендикулярно к магнитному полю. Дрейф вызывается электрическим полем и неоднородностями магнитного поля. Скорость электрического дрейфа определяется выражением* $(c/B^2)[\mathbf{EB}]$. *Направление этого дрейфа не зависит от знака заряда частицы. Дрейф, вызываемый неоднородностями магнитного поля, происходит в направлении бинормали к магнитной силовой линии, причем положительно заряженные частицы дрейфуют в положительном направлении бинормали, а отрицательно заряженные — в противоположном направлении. Скорость этого «магнитного» дрейфа определяется выражением* $\dfrac{mc}{eBR}\left(v_{\|}^2 + \dfrac{1}{2}\,v_{\perp}^2\right)\mathbf{b}$.

§ 88. Адиабатический инвариант

1. При движении заряженной частицы в неоднородном магнитном поле или при изменении самого магнитного поля ларморовский радиус частицы ρ, а также ее поперечная скорость $v_{\perp}$ изменяются. Исследуем характер этого изменения, предполагая, что магнитное поле слабо неоднородно и меняется во времени медленно.

Рассмотрим сначала случай, когда частица движется перпендикулярно к магнитному полю $\mathbf{B}$, а само поле $\mathbf{B}$ однородно и меняется только во времени. Предположим, что электрического поля нет, за исключением поля, обусловленного изменениями $\mathbf{B}$ во времени. В этих условиях дрейф частицы отсутствует, как это видно из формулы (87.5). Если бы магнитное поле было постоянно, то частица двигалась бы

по окружности радиуса $\rho = v/\omega$. При изменении магнитного поля траектория частицы перестает быть замкнутой. Однако если магнитное поле меняется медленно, то за циклотронный период $T = 2\pi/\omega$ отклонения траектории от окружности будут малы. Переменное магнитное поле индуцирует поле электрическое. Вследствие этого ларморовский радиус ρ и скорость частицы v изменяются в соответствии с уравнением $m\dot{v} = eE_s$, где E_s — проекция электрического поля $\mathbf{E}$ на направление траектории. Изменениями величины E_s за циклотронный период можно пренебречь, а ввиду незначительности отклонения траектории от ларморовской окружности вместо E_s можно взять проекцию вектора $\mathbf{E}$ на направление ларморовской окружности (последняя на рис. 203 изображена штриховой линией). Тогда E_s определится из уравнения

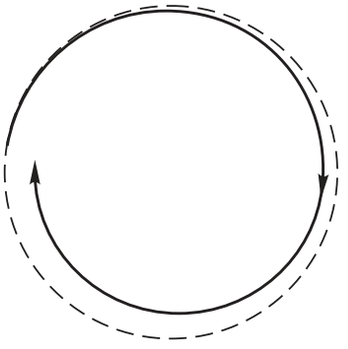


Рис. 203

$$\oint E_s\, ds = \frac{1}{c}\frac{d\Phi}{dt} = \frac{\pi\rho^2}{c}\frac{dB}{dt}.$$

Здесь производная $d\Phi/dt$ взята с плюсом, а не с минусом, так как речь идет о проекции на направление движения частицы, а оно при $e > 0$ противоположно положительному направлению обхода контура. Написанное уравнение дает

$$E_s = \frac{\rho}{2c}\frac{dB}{dt} = \frac{mv}{2eB}\frac{dB}{dt}.$$

Уравнение движения частицы переходит в

$$m\frac{dv}{dt} = \frac{mv}{2B}\frac{dB}{dt}.$$

Отсюда после интегрирования получим

$$v^2/B = \text{const}.$$

Таким образом, величина v^2/B при движении частицы сохраняется неизменной. Однако это справедливо только при *медленных изменениях* магнитного поля, т. е. величина v^2/B является не точным интегралом движения, а *адиабатическим инвариантом.*

Если у частицы есть продольная скорость $v_\parallel$, то эта скорость не окажет никакого влияния на поперечное движение. В нашем рассуждении полную скорость v надо просто заменить ее поперечной составляющей $v_\perp$, что дает

$$v_\perp^2/B = \text{const}.$$

Адиабатическим инвариантом будет величина $v_\perp^2/B$.

2. Рассмотрим теперь случай, когда магнитное поле, в котором движется частица, постоянно, но слабо неоднородно. Допустим, что дрейф в боковом направлении отсутствует, т. е. ведущий центр частицы движется вдоль магнитной силовой линии. Последнюю ради простоты будем считать прямолинейной (она на рис. 204 изображена штриховой

линией). Частица может двигаться как в сторону схождения, так и в сторону расхождения магнитных силовых линий. При движении на частицу действует сила Лоренца $\mathbf{F} = (e/c)[\mathbf{vB}]$. Она имеет слагающую вдоль центральной силовой линии. Эта слагающая будет замедлять продольное движение частицы, если оно происходит в направлении схождения магнитных силовых линий, и ускорять в противоположном случае. А так как полная кинетическая энергия частицы сохраняется, то возрастание продольной скорости $v_{\parallel}$ будет сопровождаться убыванием поперечной $v_{\perp}$, и наоборот. Покажем, что при этом соотношение (88.1) остается справедливым. Для этого достаточно перейти к системе отсчета, движущейся со скоростью $\mathbf{v}_{\parallel}$. Поступательная сила инерции — $m\dot{\mathbf{v}}_{\parallel}$, появляющаяся в этой системе, параллельна магнитному полю $\mathbf{B}$, а потому не оказывает никакого влияния на скорость дрейфа. Однако в движущейся системе отсчета магнитное поле $\mathbf{B}$ становится переменным, и к нему полностью применимы рассуждения, которыми мы пользовались при доказательстве формулы (88.1).

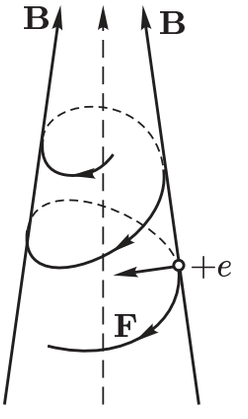


Рис. 204

3. Допустим теперь, что есть дрейф в поперечном направлении со скоростью $\mathbf{v}_{\text{д}}$. Этот дрейф может вызываться либо электрическим полем, либо неоднородностями магнитного поля, либо и тем и другим. Он может вызываться и любыми другими силами. Для последующего изложения все это не имеет значения. Перейдем в систему отсчета, движущуюся со скоростью $\mathbf{v}_{\text{д}}$. Силой инерции — $m\dot{\mathbf{v}}_{\text{д}}$ в этом случае можно пренебречь, как величиной второго или высшего порядка малости. После этого рассматриваемый случай сведется к уже рассмотренным. Отсюда следует, что адиабатическим инвариантом будет величина $\mathbf{v}'^2_{\perp}/B$, где $\mathbf{v}'$ — скорость частицы в движущейся системе отсчета. Но $v'_{\perp}$ есть скорость вращения частицы по ларморовской окружности, если отвлечься от дрейфового движения последней и от движения вдоль магнитной силовой линии. Таким образом, адиабатическая инвариантность $v_{\perp}^2/B$ сохраняется, если понимать под $v_{\perp}$ скорость вращения частицы по ларморовской окружности в только что указанном смысле.

4. Частица, вращающаяся по ларморовскому кружку, обладает магнитным моментом $\mathfrak{M}$, величина которого, как легко вычислить, равна $mv_{\perp}^2/(2B)$. Этот момент направлен против поля независимо от знака заряда частицы, а потому в векторной форме

$$\mathfrak{M} = -\frac{mv_{\perp}^2}{2B}\,\mathbf{h}. \tag{88.1}$$

Аналогично, магнитный поток через ларморовский кружок равен

$$\Phi = \pi\frac{m^2c^2v_{\perp}^2}{e^2B}. \tag{88.2}$$

С учетом соотношения (88.1) отсюда следует, что обе величины $\mathfrak{M}$ и Φ являются адиабатическими инвариантами.

5. С адиабатической инвариантностью величины (88.1) связано явление *отражения заряженных частиц от областей сильного магнитного поля.* Пусть магнитное поле не меняется во времени, а частица движется в сторону схождения магнитных силовых линий. Тогда, как было выяснено в п. 2, ее продольная скорость будет уменьшаться, а поперечная — увеличиваться. Если $\mathbf{v}$ — полная скорость, а α — угол наклона ее к магнитной силовой линии, то $v_\perp = v \sin\alpha$. А так как при движении в постоянном магнитном поле числовое значение скорости $\mathbf{v}$ не изменяется, то из (88.1) следует

$$\frac{\sin^2\alpha}{B} = \frac{\sin^2\alpha_0}{B_0} = \text{const}. \tag{88.3}$$

Здесь индексом нуль обозначены значения B и α в каком-либо положении частицы, условно принимаемом за исходное. Из полученного соотношения следует, что в область, определяемую условием $B > B_0/\sin^2\alpha_0$, частица проникнуть не может, так как в противном случае $|\sin\alpha|$ должен был бы превзойти единицу. По мере продвижения в область более сильного поля угол α будет возрастать. Если при этом он все время остается меньше 90°, то частица будет продолжать движение вперед. Это будет тогда, когда для всех точек траектории выполняется условие

$$\sin\alpha_0 < \sqrt{B_0/B}. \tag{88.4}$$

Если же угол α в некоторой точке достигает 90°, то частица начнет двигаться в обратном направлении, т. е. отразится. Явление напоминает полное отражение света в оптике. Точка, где произойдет отражение, определяется уравнением

$$\sin\alpha_0 = \sqrt{B_0/B}. \tag{88.5}$$

§ 89. Определение удельного заряда частицы

1. *Удельным зарядом частицы* называется отношение заряда e этой частицы к ее массе m. При определении этого отношения заряд e принято выражать в единицах СГСМ, а массу m — в граммах. В соответствии с этим в настоящем параграфе применяется система СГСМ. Удельный заряд можно определить, исследуя движение частицы в поперечных электрическом и магнитном полях. Такие исследования производились Дж. Дж. Томсоном и его сотрудниками в конце прошлого и начале настоящего столетия с целью установления природы катодных и анодных лучей в трубках с разреженными газами (давление порядка нескольких сотых мм рт. ст.). Они привели к открытию *электрона* и *изотопов*, т. е. химических элементов, ядра которых имеют одинаковые заряды, но различные массы.

2. Исследуем сначала движение частицы в поперечном электрическом поле заряженного конденсатора. Направим ось X параллельно пластинам конденсатора, а ось Z — перпендикулярно. Пусть частица перед входом в конденсатор двигалась вдоль оси X (рис. 205). В дальнейшем под действием электрического поля конденсатора она отклонится в направлении оси Z и будет двигаться в плоскости ZX. Уравнения движения частицы имеют вид

$$m\frac{d^2x}{dt^2} = eE_x, \quad m\frac{d^2z}{dt^2} = eE_z. \quad (89.1)$$

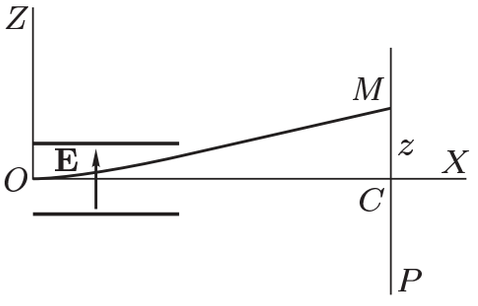


Рис. 205

Будем предполагать, что угол наклона траектории частицы к оси X на протяжении всего движения мал и, следовательно, $v_z \ll v_x$. В этом случае величиной v_z^2 написать можно пренебречь по сравнению с v_x^2 и написать

$$v_x = \sqrt{v^2 - v_z^2} \approx v\left[1 - \frac{1}{2}\left(\frac{v_z}{v}\right)^2\right] \approx v,$$

где v — полная скорость частицы. В том же приближении скорость v можно считать постоянной, а движение параллельно оси X — равномерным. Исключив время t с помощью соотношения $dx = v\,dt$, получим уравнение траектории в дифференциальной форме:

$$\frac{d^2z}{dx^2} = \frac{e}{mv^2}E_z. \quad (89.2)$$

В конденсаторе электрическое поле однородно и равно $\mathbf{E}$, за исключением малой области вблизи его краев, где E_z меняется от E до 0. Выйдя из конденсатора, частица движется свободно, т. е. прямолинейно и равномерно, и попадает на фотопластинку P в точке M, отклонившись от оси X на расстояние $z = CM$. Отклонение z найдется двукратным интегрированием уравнения (89.2) и равно

$$z = A\frac{e}{mv^2}, \quad (89.3)$$

где A — постоянная прибора:

$$A = \int_0^L dx \int_0^x E_z(x')\,dx',$$

а через L обозначено расстояние OC от начала координат O до фотопластинки. Введя обозначение $f(x) = \int_0^x E_z(x')\,dx'$, можем написать $A = \int_0^L f(x)\,dx$. После этого, выполнив интегрирование по частям, получим

$$A = \int_0^L (L - x) E_x \, dx. \tag{89.4}$$

Если пренебречь неоднородностью поля на краях конденсатора, то интегрирование легко выполняется и приводит к результату

$$A = El(L - l/2),$$

где l — длина конденсатора.

3. Рассмотрим теперь движение заряженной частицы поперек магнитного поля в аналогичных условиях. Поле предполагается однородным направленным параллельно оси Y (рис. 206). Исключение составляют только края области, занимаемой магнитным полем, где существуют отступления от однородности. Однако, как и в случае электрического поля, влияние неоднородностей магнитного поля предполагается малым. Пусть частица перед входом в магнитное поле по-прежнему двигалась в направлении оси X, а в дальнейшем отклонялась от этого направления мало. Тогда при вычислении силы Лоренца $\mathbf{F} = e[\mathbf{vB}]$ скорость $\mathbf{v}$ можно считать направленной всюду параллельно оси X, т. е. вдоль единичного вектора $\mathbf{i}$: $\mathbf{v} = v\mathbf{i}$. В этом приближении

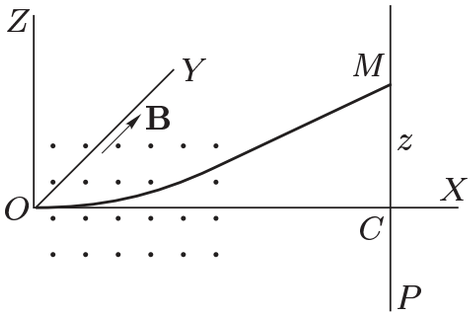


Рис. 206

$$\mathbf{F} = ev[\mathbf{iB}] = ev([\mathbf{ij}]B_y + [\mathbf{ik}]B_z) = ev(B_y\mathbf{k} - B_z\mathbf{j}).$$

Так как основное поле параллельно оси Y, то составляющей B_z можно пренебречь. Тогда частица будет отклоняться в направлении оси Z в соответствии с уравнением

$$m\frac{d^2z}{dt^2} = evB_y.$$

Это уравнение может быть получено из второго уравнения (89.1), если в нем E_z заменить на vB_y. Поэтому, не производя дальнейших вычислений, можно написать сразу

$$z = C\frac{e}{mv}, \tag{89.5}$$

где C — постоянная прибора:

$$C = \int_0^L (L - x) B_y \, dx. \tag{89.6}$$

4. Таким образом, отклонения в поперечном электрическом поле пропорциональны $e/(mv^2)$, а в поперечном магнитном поле — $e/(mv)$.

Поэтому, измерив эти отклонения, можно вычислить не только, удельный заряд e/m, но и скорость частицы v. Практически удобно конденсатор поместить в магнитное поле, чтобы частица подвергалась одновременному действию электрического и магнитного полей. Сами поля $\mathbf{E}$ и $\mathbf{B}$ могут быть либо параллельными, либо перпендикулярными друг к другу. В ранних исследованиях опыты производились с катодными лучами (электронами) и с анодными лучами (ионами).

Если электрическое поле конденсатора $\mathbf{E}$ перпендикулярно к магнитному полю $\mathbf{B}$, то эти поля будут отклонять частицу в одном и том же или в прямо противоположных направлениях (на рисунках 205 и 206 в направлении оси Z). Удобно напряжение на конденсаторе подобрать таким, чтобы для частиц с определенной скоростью эти отклонения компенсировали друг друга, т. е. чтобы частица проходила через прибор без отклонения. Затем надо выключить одно из полей и измерить получающееся отклонение. Тогда

$$z = A\frac{e}{mv^2} = -C\frac{e}{mv}.$$

Измерив z, отсюда легко вычислить скорость v и удельный заряд e/m. Таким путем Дж. Дж. Томсон в 1897 г. впервые измерил e/m для катодных лучей.

Одна из трудностей в этих исследованиях состояла в том, что частицы в электронных и ионных пучках обладали большим разбросом скоростей. Для исключения влияния этого разброса электрическое и магнитное поля выбирались параллельными друг другу. Допустим, что они направлены вдоль оси Z. Тогда электрическое поле будет отклонять частицу в направлении оси Z, а магнитное — в направлении оси Y. Для этих отклонений можно написать

$$z = A\frac{e}{mv^2}, \quad y = C\frac{e}{mv}.$$

Эти соотношения дают в параметрической форме уравнение кривой на фотопластинке, на которую попадают частицы с одинаковым удельным зарядом e/m, но с различными скоростями v. Параметром служит скорость v. Исключив этот параметр, представим уравнение той же кривой в виде

$$z = \frac{A}{C^2}\,\frac{m}{e}\,y^2.$$

Это — парабола. Измеряя отношение z/y^2, можно вычислить удельный заряд e/m. Этим *методом парабол* Дж. Дж. Томсон в 1912 г. открыл изотопы нерадиоактивных элементов (неон).

Развитие вакуумной техники и разработка источников электронов и ионов позволили производить подобные измерения в более определенных и лучше контролируемых условиях. Например, для получения пучка электронов с определенной скоростью v можно использовать явление *термоэлектронной эмиссии.* Источником электродов служит

раскаленная вольфрамовая нить. Электроны, испущенные этой нитью, ускоряются до определенной энергии приложенным напряжением и одновременно коллимируются с помощью отверстий или щелей, а затем подвергаются отклонению в электрических и магнитных полях.

В нашу задачу не входит изложение современных методов измерения удельных зарядов. Этим занимается *масс-спектрометрия* и *масс-спектрография.* Укажем только на основной результат, установленный еще в конце прошлого столетия. Оказалось, что в случае анодных лучей удельный заряд e/m зависит от состава газа в трубке и составляет 10^4 СГСМ-ед. заряда/г или меньше. Для катодных лучей эта величина много больше, а именно $e/m = 1{,}759 \cdot 10^7$ СГСМ-ед. заряда/г, и не зависит от состава газа в трубке. Установление этого факта Дж. Дж. Томсоном в 1897 г. означало открытие электрона.

ЗАДАЧА

В одном из ранних методов определения удельного заряда электрона электроны, вырванные из алюминиевого диска K, ускорялись разностью потенциалов V, приложенной между K и щелью S (рис. 207). Пройдя через щель S, электронный пучок попадал в однородное магнитное поле, перпендикулярное к плоскости рисунка. Вся система помещалась в вакууме. Изменяя напряженность магнитного поля, добивались того, чтобы ток на коллекторе C, регистрируемый гальванометром Г, был максимален. Измерив магнитное поле B в этот момент, можно вычислить e/m. Провести этот расчет, если расстояние между щелью S и коллектором C равно $d = 10$ см, угол между прямой, проведенной от S к C, и начальным направлением электронного пучка $\alpha = 30°$, $V = 1000$ В, $B = 10{,}6$ Гс.

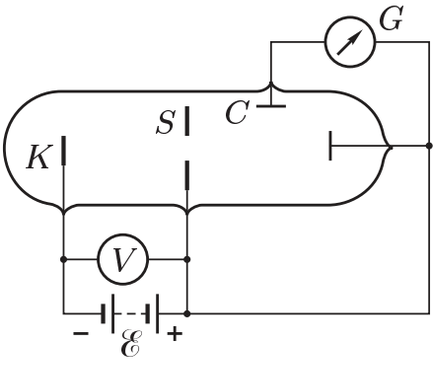


Рис. 207

Ответ. Удельный заряд $\dfrac{e}{m} = \dfrac{8V}{B^2 d^2} \sin^2 \alpha = 1{,}78 \cdot 10^7$ СГСМ-ед. заряда/г.

§ 90. Измерение элементарного заряда методом масляных капель

1. Опыты по измерению удельного заряда e/m укрепили представление об атомистической природе электричества. Дж. Дж. Томсон и его ученики Таунсенд (1868–1957) и Чарльз Вильсон (1869–1959) произвели первые измерения и самого *элементарного заряда*, т. е. наименьшего электрического заряда, встречающегося в природе. Однако их методами нельзя было получить точные результаты. Точные измерения были выполнены Робертом Милликеном (1868–1953) в классических опытах в 1908–1916 гг. Эти опыты принесли также неопровержимое

доказательство *атомизма электричества.* Милликен измерял электрический заряд малых капелек масла. Схема его установки показана на рис. 208. В тщательно изготовленный плоский конденсатор через отверстие в верхней пластине могут попадать мелкие капельки масла, получаемые с помощью специального распылителя. С целью предохранения капелек от конвекционных потоков воздуха конденсатор заключен в защитный кожух, температура и давление воздуха в котором поддерживаются постоянными. На пластины конденсатора можно было накладывать постоянное напряжение от источника в несколько тысяч вольт. В ходе опыта это напряжение можно было менять. При распылении капельки масла заряжаются и, попадая в конденсатор, движутся под действием собственного веса и приложенного электрического поля. Движение отдельной капельки можно наблюдать с помощью микроскопа через специальное окошко.

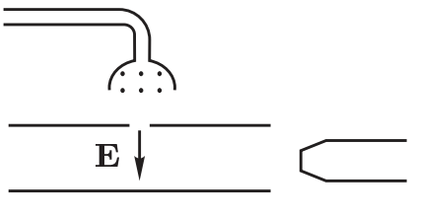


Рис. 208

Аналогичной установкой пользовался А.Ф. Иоффе (1880–1960) в 1912 г. В его опытах вместо капелек масла применялись цинковые пылинки, а также капельки ртути.

2. Допустим сначала, что электрического напряжения на конденсаторе нет. Тогда капля, попавшая в кондесатор, будет падать вниз под действием собственного веса, встречая при этом падении силу сопротивления kv, пропорциональную скорости капли v. Установившаяся скорость падения v_g в поле тяжести определится уравнением

$$kv_g = (m - m_0)g, \tag{90.1}$$

где m — масса капли, а m_0 — масса вытесненного ею воздуха. Последняя введена для учета архимедовой подъемной силы. Если капля заряжена, то при наложении электрического поля E ее движение изменится. Поле E подбирают таким, чтобы капля стала подниматься вверх. Если v_E — установившаяся скорость капли при подъеме вверх, а q — ее заряд, то

$$kv_E = qE - (m - m_0)g.$$

Из этих уравнений находим

$$q = \frac{k(v_g + v_E)}{E}. \tag{90.2}$$

Освещением рентгеновскими лучами можно слегка ионизовать воздух между пластинами конденсатора. Тогда заряд капли, а с ним и скорость установившегося движения ее в том же электрическом поле могут скачкообразно измениться. Если капля по-прежнему поднимается вверх с установившейся скоростью v'_E, то ее новый заряд будет

$$q' = \frac{k(v_g + v'_E)}{E},$$

и, следовательно,

$$\frac{q'}{q} = \frac{v_g + v'_E}{v_g + v_E}.$$

Измеряя скорости установившегося движения одной и той же капли в одном и том же электрическом поле, можно сравнивать заряды q и q'. Если капля мала, а электричество имеет атомистическое строение, то можно ожидать, что заряд капли будет состоять из небольшого количества элементарных зарядов e. В таком случае отношение q'/q будет отношением небольших целых чисел. Это и наблюдалось в опытах Милликена и Иоффе.

При другом способе обработки наблюдений вычисляются скачки скорости Δv_E одной и той же капли при ее перезарядке. Согласно формуле (90.2) они связаны с изменением заряда капли Δq соотношением

$$\Delta q = k\frac{\Delta v_E}{E}. \tag{90.3}$$

Если электричество имеет атомистическое строение, то величина Δq должна принимать только определенные дискретные значения, равные целому числу элементарных зарядов. В частности, сам элементарный заряд e будет равен наименьшему из этих значений Δq (за исключением, конечно, значения $\Delta q = 0$). Поэтому изменения установившейся скорости капли Δv_E при перезарядке должны носить скачкообразный характер и быть кратными определенной наименьшей величине. Это также подтвердили наблюдения.

3. Для количественного определения заряда капли по формуле (90.2) необходимо знать коэффициент k. Его можно вычислить по формуле Стокса $k = 6\pi\eta a$, где η — вязкость воздуха (см. т. I, § 101). Из-за малости капли прямое измерение ее радиуса a с помощью микроскопа невозможно. Микроскоп дает лишь *дифракционное изображение* капли в виде яркой звездочки, получающейся вследствие рассеяния света на капле. Форма и размер этого дифракционного изображения не имеют никакого сходства с действительными формой и размерами рассеивающей капли. Для определения радиуса a можно воспользоваться той же формулой Стокса. Подставляя в формулу (90.1)

$$k = 6\pi\eta a, \quad m - m_0 = \frac{4\pi}{3}\, a^3(\rho - \rho_0),$$

где ρ — плотность масла, а ρ_0 — воздуха, находим

$$a = \sqrt{\frac{9\eta v_g}{2(\rho - \rho_0)}}, \tag{90.4}$$

и, следовательно,

$$q = \frac{9\pi\eta}{E}\sqrt{\frac{2\eta v_g}{\rho - \rho_0}}\,(v_g + v_E). \tag{90.5}$$

Наименьшее значение q или Δq, вычисленное по формуле (90.5) и будет равно элементарному заряду e.

4. На деле оказалось, что в случае очень малых капель вычисления по формуле (90.5) приводили к аномально большим значениям элементарного заряда e, которые были тем больше, чем меньше размер капли. Милликен объяснил этот результат неприменимостью формулы Стокса к очень малым капелькам. Дело в том, что формула Стокса выводится в предположении, что вязкая среда, в которой движется шар, является *сплошной.* В случае газов для выполнения этого условия необходимо, чтобы средняя длина свободного пробега молекулы газа λ была мала по сравнению с размерами шара ($\lambda \ll a$). Кённингам в 1910 г., применяя кинетическую теорию газов, ввел поправку в формулу Стокса и получил

$$k = \frac{6\pi\eta a}{1 + A\lambda/a}, \tag{90.6}$$

где A — постоянная. Эту формулу можно также обосновать, записав знаменатель в виде $f(\lambda/a)$, а затем разложить его по степеням λ/a, оборвав это разложение на линейном члене относительно λ/a. Поскольку λ обратно пропорциональна давлению газа $\mathscr{P}$, формулу можно также привести к виду

$$k = \frac{6\pi\eta a}{1 + B/(\mathscr{P}a)}. \tag{90.7}$$

Постоянную B можно вычислить газокинетически, однако надежнее измерить ее экспериментально. Подстановка выражения (90.7) в формулу (90.1) приводит к кубическому уравнению относительно радиуса капли a:

$$\frac{\eta v_g}{1 + B/(\mathscr{P}a)} = \frac{2}{9}(\rho - \rho_0)a^2. \tag{90.8}$$

Так как дробь $B/(\mathscr{P}a)$ является малой поправкой, то это уравнение можно решить методом последовательных приближений. В нулевом приближении поправкой $B/(\mathscr{P}a)$ пренебрегаем совсем и получаем для a прежнее выражение (90.4). Его будем теперь обозначать через a_0, т. е.

$$a_0 = \sqrt{\frac{9\eta v_g}{2(\rho - \rho_0)}}. \tag{90.9}$$

Затем в знаменателе уравнения (90.8) дробь $B/(\mathscr{P}a)$ заменим на $B/(\mathscr{P}a_0)$ и из полученного таким путем квадратного уравнения найдем a в первом приближении. Далее можно было бы найти a во втором приближении и т. д. Однако это вряд ли имеет смысл, так как исходная формула (90.7) верна лишь с точностью до линейных членов относительно $B/(\mathscr{P}a)$. Однако, как показал Милликен, точность первого приближения уже достаточна.

Из изложенного ясно, что в первом приближении в формуле (90.5) вязкость η надо просто заменить на $\dfrac{\eta}{1 + B/(\mathscr{P}a_0)}$. Это дает

$$q = \frac{q_0}{[1 + B/(\mathscr{P}a_0)]^{3/2}},$$

где q_0 — заряд капли, вычисленный в нулевом приближении, т. е. по формуле (90.5), а q — заряд, вычисленный в первом приближении, который принимается за истинный заряд капли. В частности, для элементарного заряда e получаем

$$\left(\frac{e_0}{e}\right)^{2/3} = 1 + \frac{B}{\mathscr{P}a_0}. \tag{90.10}$$

Будем производить измерения при различных давлениях $\mathscr{P}$ и откладывать по оси абсцисс $1/(\mathscr{P}a_0)$, а по оси ординат $(e_0/e)^{2/3}$. Тогда должна получиться прямая линия (рис. 209). И действительно, Милликен убедился, что экспериментальные точки точно ложатся на одну прямую. Это доказывает правильность исходных положений, на которых основывались вычисления. Продолжив прямую (90.10) до пересечения с осью ординат, найдем, что в точке пересечения $e = e_0$. Величина e_0, соответствующая этой точке пересечения, и есть элементарный заряд e. Такое значение e мы получили бы по формуле (90.5), если бы производили измерения с большими каплями, когда применимость формулы Стокса не вызывает сомнений. По наклону прямой (90.10) можно определить и постоянную B. По современным данным, $e = 4{,}803 \cdot 10^{-10}$ СГСЭ-ед. $= 1{,}602 \cdot 10^{-19}$ Кл. Милликен получил несколько меньшее значение, так как он пользовался заниженным значением для вязкости воздуха. Зная e и удельный заряд e/m, можно вычислить массу электрона $m = 9{,}11 \cdot 10^{-28}$ г.

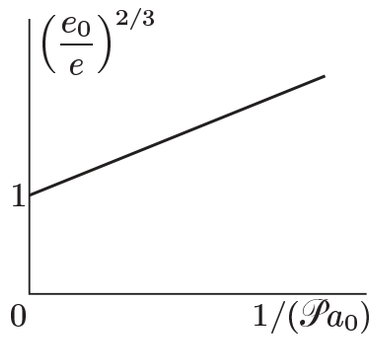


Рис. 209

§ 91. Электромагнитная масса

1. В § 84 было установлено, что закон сохранения импульса в электродинамике приводит к заключению, что электромагнитное поле должно обладать импульсом, плотность которого дается выражением

$$\mathbf{g} = \frac{1}{4\pi c}[\mathbf{EH}]. \tag{91.1}$$

К такому заключению мы пришли, сопоставляя выражение для плотности потока электромагнитной энергии $\mathbf{S} = (c/4\pi)[\mathbf{EH}]$ с формулой Эйнштейна $\mathscr{E} = mc^2$. К тому же заключению в самом общем виде приводит и классическая электродинамика, не использующая теорию относительности. Мы убедились в этом в § 84 на частном примере.

2. Вычислим теперь электромагнитный импульс, связанный с равномерно движущимся зарядом. При вычислении будем предполагать, что скорость заряда v мала по сравнению со скоростью света c. Кроме того, предположим, что заряд равномерно распределен по поверхности шара радиуса a. В системе отсчета, в которой шар неподвижен, поля $\mathbf{E}$

и $\mathbf{B}$ внутри шара равны нулю. Значит, электрическое поле будет равно нулю и во всякой системе отсчета, относительно которой заряженный шар движется прямолинейно и равномерно. Это непосредственно следует из формулы преобразования электрического поля (66.6). Магнитное поле внутри шара также равно нулю, но этот результат нам пока не понадобится. Таким образом, для вычисления полного электромагнитного импульса $\mathbf{L}_{\text{эл}}$ выражение (91.1) надо проинтегрировать по объему всего бесконечного пространства вне шара.

Мы предполагаем, что заряд e движется в вакууме. В вакууме поля $\mathbf{B}$ и $\mathbf{H}$ тождественно совпадают между собой и определяются выражением

$$\mathbf{H} = \frac{e}{c}\left[\mathbf{v}\frac{\mathbf{r}}{r^3}\right] = \frac{1}{c}[\mathbf{v}\mathbf{E}_0],$$

где $\mathbf{E}_0$ — статическое (кулоновское) поле заряда e. Электрическое поле $\mathbf{E}$ движущегося заряда отлично от статического поля $\mathbf{E}_0$. Однако при принятой нами точности расчета этим различием можно пренебречь. Действительно, поле $\mathbf{E}$ можно представить в виде ряда по степеням v/c. Первый член этого ряда есть статическое поле $\mathbf{E}_0$. Оно вносит в электромагнитное количество движения $\mathbf{L}_{\text{эл}}$ слагаемое, пропорциональное v/c. Последующие члены дают слагаемые порядка $(v/c)^2$, $(v/c)^3$ и т. д., которыми мы пренебрегаем. Итак, в принятом приближении

$$\mathbf{g} = \frac{1}{4\pi c}[\mathbf{E}_0\mathbf{H}] = \frac{1}{4\pi c^2}[\mathbf{E}_0[\mathbf{v}\mathbf{E}_0]] = \frac{1}{4\pi c^2}\{E_0^2\mathbf{v} - (\mathbf{E}_0\mathbf{v})\mathbf{E}_0\}.$$

Примем направление скорости $\mathbf{v}$ за ось X. Тогда можно написать

$$(\mathbf{E}_0\mathbf{v})\mathbf{E}_0 = (E_{0x}v_x)(E_{0x}\mathbf{i} + E_{0y}\mathbf{j} + E_{0z}\mathbf{k}).$$

При интегрировании этого выражения целесообразно предварительно усреднить его по всем направлениям пространства. Тогда останется $\overline{E_{0x}^2}v_x\mathbf{i} = E_0^2\mathbf{v}/3$, и, следовательно,

$$\mathbf{L}_{\text{эл}} = \frac{4\mathbf{v}}{3c^2}\int \frac{E_0^2}{8\pi}\,dV,$$

или

$$\mathbf{L}_{\text{эл}} = \frac{4}{3}\frac{W_{\text{эл}}}{c^2}\mathbf{v}, \tag{91.2}$$

где $W_{\text{эл}}$ — электростатическая энергия заряда e. Заметим, что результат (91.2) справедлив для любого сферически симметричного, а не только для поверхностного распределения заряда в шаре. Действительно, исходя из уравнений Максвелла, нетрудно доказать, что электрическое и магнитное поля равномерно движущегося заряда связаны соотношением $\mathbf{H} = (1/c)[\mathbf{v}\mathbf{E}]$, а этого достаточно, чтобы получить формулу (91.2), если ограничиться при этом членами первого порядка по v.

Таким образом, благодаря наличию электромагнитного поля, к импульсу электрона добавляется электромагнитный импульс. При $v \ll c$

его можно представить в виде $L_{\text{эл}} = m_{\text{эл}}\mathbf{v}$, где

$$m_{\text{эл}} = \frac{4}{3}\,\frac{W_{\text{эл}}}{c^2}. \tag{91.3}$$

Эта величина называется *электромагнитной массой.* Если электричество распределено по поверхности шара, то $W_{\text{эл}} = e^2/(2a)$ и, следовательно,

$$m_{\text{эл}} = \frac{2e^2}{3ac^2}. \tag{91.4}$$

3. К понятию электромагнитной массы можно также прийти на основании следующих соображений. Энергия магнитного поля шара, заряженного по поверхности, определяется выражением

$$W_m = \int \frac{H^2}{8\pi}\,dV = \frac{e^2v^2}{8\pi}\int \frac{\sin^2\vartheta}{r^4}\,dV,$$

где ϑ — угол между направлениями скорости $\mathbf{v}$ и радиуса $\mathbf{r}$, проведенного из центра шара. Элемент объема представим в виде $dV = 2\pi r^2 \sin\vartheta\,d\vartheta$ и получим

$$W_m = \frac{e^2v^2}{4}\int\limits_a^\infty \frac{dr}{r^2}\int\limits_0^\pi \sin^3\vartheta\,d\vartheta = \frac{e^2v^2}{3a} = \frac{m_{\text{эл}}}{2}\,v^2.$$

Благодаря наличию магнитного поля энергия шара увеличилась на величину W_m. Это увеличение можно трактовать как *увеличение кинетической энергии* или как *возрастание массы* шара на величину электромагнитной массы.

Недостаток второго вывода состоит в том, что в нем не выяснено влияние электрического поля, энергия которого также возрастает со скоростью $\mathbf{v}$. Однако если сопоставить второй вывод с первым, то можно прийти к заключению, что в принятом приближении электрическая энергия на величину электромагнитной массы не влияет. В этом приближении электромагнитная масса связана с энергией, идущей на *возбуждение только магнитного поля.*

4. В высших приближениях не только усиливается влияние энергии магнитного поля на электромагнитную массу, но появляется и влияние электрической энергии. Сама электромагнитная масса начинает зависеть от скорости электрона $\mathbf{v}$, а также от характера распределения электрического заряда внутри этой частицы и деформаций, возникающих при ее движении. Физики начала нашего столетия, вводя различные произвольные предположения относительно распределения заряда и деформаций электрона, пытались решить вопрос о природе массы электрона и ее зависимости от скорости. Считалось, что масса всякой частицы состоит из «истинной массы», не зависящей от скорости, и «кажущейся» электромагнитной массы, меняющейся со скоростью. Исследуя зависимость массы электрона от скорости, ученые пытались отделить истинную массу от кажущейся. В частности, был

поставлен вопрос, не является ли вся масса электрона кажущейся, т. е. электромагнитной. Эти исследования уже давно сохранили один только исторический интерес. Исследуя зависимость массы от скорости, нельзя решить вопрос о физической природе массы, так как теория относительности, в полном согласии с опытом, показала, что всякая масса, какова бы ни была ее природа, должна меняться со скоростью в соответствии с формулой

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}}, \tag{91.5}$$

причем между массой тела и его энергией должно существовать соотношение

$$\mathscr{E} = mc^2. \tag{91.6}$$

Этим вопрос о природе массы электрона, конечно, не снимается. Не снимается и гипотеза об электромагнитной массе электрона. То обстоятельство, что в формулу (91.3) входит числовой коэффициент $4/3$, а в формуле Эйнштейна (91.6) такого коэффициента нет, в классической теории объясняли тем, что электростатическая энергия $W_{\text{эл}}$ не есть полная энергия электрона. Необходимость введения дополнительной энергии очевидна из того, что при наличии одних только электростатических сил электрон не может находиться в равновесии: под действием кулоновских сил отталкивания он должен был бы разлететься на части. Чтобы этого не было, по классическим представлениям, необходимы дополнительные силы неэлектростатического происхождения и связанная с ними энергия.

Если принять, что вся масса электрона электромагнитного происхождения, и опустить в формуле (91.4) числовой множитель $2/3$, то получится

$$a = \frac{e^2}{mc^2} = 2{,}818 \cdot 10^{-13} \text{ см}. \tag{91.7}$$

Эта величина называется *классическим радиусом электрона.* На формулу (91.7) нельзя смотреть как на выражение, определяющее «истинные размеры электрона», так как никаких других независимых способов определения размеров электрона не существует. На классический радиус электрона (91.7) следует смотреть как на некоторую *характерную длину, ограничивающую снизу область применимости классической теории поля.* Классическая теория поля неприменима также для *очень сильных полей*, порядка

$$E_0 = \frac{e}{a^2} = \frac{m^2c^4}{e^3} = 6 \cdot 10^{15} \text{ СГСЭ-ед.} = 1{,}8 \cdot 10^{18} \text{ В/см}.$$

5. В действительности размеры области, в которой классические представления уже неприменимы, ограничены *квантовыми эффектами* и примерно в сто раз превышают классический радиус электрона. Эти размеры легко оценить элементарно с помощью квантовомеханического *соотношения неопределенностей между энергией и временем.* Для читателя, имеющего

некоторые представления об основных понятиях квантовой механики, приведем эту оценку. Если частица наблюдается в течение времени Δt, то, согласно квантовой механике, о ее энергии можно говорить лишь с точностью $\Delta\mathscr{E}$, удовлетворяющей условию $\Delta\mathscr{E}\Delta t \gtrsim \hbar$, где $\hbar = h/2\pi = 1{,}05 \cdot 10^{-27}$ эрг $\cdot$ с — постоянная Планка, деленная на 2π. Это условие и называется соотношением неопределенностей между энергией и временем. Если время Δt достаточно мало, то в вакууме на короткое время могут рождаться *электрон-позитронные пары.* Для этого неопределенность энергии $\Delta\mathscr{E}$ должна быть не меньше собственной энергии электрон-позитронной пары, т. е. $2mc^2$. (Массы электрона и позитрона одинаковы). Соответствующий промежуток времени от момента рождения до момента исчезновения электрон-позитронной пары будет не больше $\Delta t \sim \hbar/(2mc^2)$. За это время электрон-позитронная пара может пройти расстояние не больше $\Lambda_к = c\Delta t \sim \hbar/(mc) = 0{,}38 \cdot 10^{-9}$ см. (Последняя величина называется *комптоновской длиной волны.*) Таким образом, любой «точечный заряд» на расстояниях $\Lambda_к$ как бы окружен областью «виртуальных» электрон-позитронных пар (*«поляризация вакуума»*). Это означает, что классическая картина «точечного заряда» на таких расстояниях уже неприменима. Комптоновская длина волны $\Lambda_к = \hbar/(mc)$ и ограничивает снизу область применимости классической теории поля. Ее отношение к классическому радиусу электрона равно

$$\frac{\Lambda_к}{a} = \frac{\hbar c}{e^2} = 137.$$

Точно так же классическая теория поля неприменима уже при полях, примерно в 137 раз меньших, чем E_0. Действительно, пусть электрическое поле $E_{кв}$ на длине $\Lambda_к$ создает такую разность потенциалов, что $e\Lambda_к E_{кв} \gtrsim 2mc^2$. Отсюда, опуская коэффициенты, получаем

$$E_{кв} \gtrsim \frac{m^2c^3}{e\hbar} = \frac{e^2}{\hbar c}\,E_0 = \frac{1}{137}\,E_0.$$

В таком поле виртуально возникшая электрон-позитронная пара будет разорвана, т. е. произойдет *пробой вакуума* — явление невозможное с точки зрения классических представлений.

Глава VI
ЭЛЕКТРОЛИТЫ

§ 92. Электролиз и электролитическая диссоциация

1. *Электролитами* в широком смысле слова называются вещества, химически разлагающиеся на составные части, когда по ним проходит, электрический ток. К таким веществам относятся многие соединения металлов с металлоидами в расплавленном состоянии, а также твердые вещества, проводящие ток, например *твердые растворы* и *ионные кристаллы.* Однако основными представителями электролитов, имеющими главнейшие научно-технические применения, являются *водные растворы неорганических кислот*, *солей* и *оснований.* На них мы и сосредоточим основное внимание. Далеко не всякие водные растворы — электролиты. Например, раствор сахара в воде не проводит электрического тока и электролитом не является. Многие неводные растворы (например, растворы, в которых растворителями служат спирты) также обладают электролитическими свойствами, но у них эти свойства выражены значительно слабее, чем у водных растворов.

Разложение электролита на его составные части под действием электрического тока называется *электролизом.* Продукты разложения выделяются на *электродах*, т. е. проводящих телах надлежащей геометрической формы, погруженных в электролит и соединенных с полюсами гальванической батареи или другого источника постоянного тока. Эти продукты, как правило, вступают в химические реакции с электродами или растворителем. Тогда говорят о *вторичных реакциях* при электролизе. Последние долгое время затемняли собою природу явления электролиза и его законы. При изучении электролиза надо прежде всего исключить осложняющее влияние вторичных реакций и сосредоточить внимание на реакциях *первичных.* Так мы и поступим.

2. Первая гипотеза для объяснения электролиза была предложена Гротгусом (1785–1822) в 1805 г. Согласно этой гипотезе молекулы растворенного вещества состоят из двух частей, из которых одна заряжена положительно, а другая — отрицательно. В растворе, на который наложено электрическое поле, по какой-то причине некоторые из молекул растворенного вещества распадаются на эти составные части. Последние переходят в раствор в виде заряженных частиц, получивших позднее название *ионов.* Положительные ионы под действием электрического поля двигаются к отрицательному электроду (катоду), а отрицательные ионы — к положительному (аноду). Несмотря на такое движение, вне электродов в каждом элементе объема всегда будет находиться одинаковое количество положительных и отрицательных

ионов. Если выключить электрическое поле, то положительные ионы соединятся с отрицательными в нейтральные молекулы и никакого химического разложения растворенного вещества наблюдаться не будет. Не то произойдет вблизи электродов. Достигнув электродов, ионы будут передавать им свои заряды и выделяться в виде нейтральных атомов, образующих продукты разложения растворенного вещества.

3. Возникает вопрос: почему нейтральные молекулы в растворе распадаются на ионы? До Клаузиуса считали, что это происходит под действием наложенного электрического поля. Таким образом, считалось, что электрическое поле производит разложение растворенного вещества на составные части. Клаузиус в 1857 г. первый указал, что если бы это было так, то для каждого химического соединения потребовалась бы некоторая минимальная напряженность электрического поля $E_{\text{мин}}$, чтобы преодолеть химическое сродство. Чем больше сродство, тем больше было бы поле $E_{\text{мин}}$. При $E < E_{\text{мин}}$ в растворе не было бы ионов, и он не мог бы проводить электрический ток. На самом деле все происходит иначе. Опыты показали, что *электролиты подчиняются закону Ома* $\mathbf{j} = \lambda\mathbf{E}$, а потому *любое сколь угодно слабое электрическое поле вызывает электролиз во всяком электролите.*

На разложение растворенного вещества электролита требуется дополнительная затрата энергии. Она вовсе не мала, как показывает следующий пример. При электролизе соляной кислоты HCl на электродах выделяются водород и хлор — вещества, обладающие большим химическим сродством друг к другу. Они соединяются со взрывом, причем при образовании одного грамма соляной кислоты выделяется около 600 калорий теплоты. Таким образом, если бы химическое разложение электролита производилось электрическим полем, то энергия электрического тока тратилась бы не только на выделение джоулевой теплоты, но и на указанное химическое разложение. Между тем опыты показали, что *для электролитов выполняется закон Джоуля–Ленца, согласно которому вся энергия тока тратится на джоулеву теплоту.*

4. Имея в виду изложенные соображения, Клаузиус в 1857 г. предложил иное объяснение возникновения ионов в растворах. Согласно кинетической теории газов, одним из главных основоположников которой был сам Клаузиус, атомы внутри молекулы совершают беспорядочное тепловое движение. Обычно энергии этого теплового движения недостаточно, чтобы преодолеть химические силы притяжения, удерживающие вместе противоположно заряженные части молекулы. В растворах, однако, под действием растворителя силы химического сродства ослабляются, и молекула, по крайней мере на короткое время, *диссоциирует*, т. е. расщепляется на противоположно заряженные ионы. При встрече положительного иона с отрицательным ионы могут *рекомбинировать*, т. е. соединиться в нейтральную молекулу. Другие нейтральные молекулы, наоборот, могут диссоциировать на ионы. В результате непрерывно идущих процессов диссоциации и рекомбинации

устанавливается *статистическое равновесие*, в котором доля диссоциированных молекул в среднем остается неизменной во времени. При наложении электрического поля положительные и отрицательные ионы устремляются в противоположные стороны в соответствии с воззрениями Гротгуса. Возникает электрический ток, и происходит выделение продуктов разложения на электродах. Таким образом, электрическое поле не имеет никакого отношения к химическому разложению молекул растворенного вещества на составные части. *Роль электрического поля сводится только к тому, чтобы разделить уже существующие положительные и отрицательные ионы и собрать их на различных электродах.* В этом и состоит сущность явления электролиза.

5. К аналогичным выводам из совсем других соображений пришел Сванте Аррениус (1859–1927), являющийся творцом *теории электролитической диссоциации.* Опыты показали, что водные растворы солей, кислот и оснований имеют более высокое осмотическое давление, чем это следует из закона Вант-Гоффа (1852–1911). Понижение упругости насыщенного пара, а также понижение температуры замерзания и повышение точки кипения для этих растворов больше, чем это следует из закона Рауля (см. т. II, § 124–126). На этом основании Аррениус и пришел к выводу, что молекулы растворенного вещества в водных растворах диссоциированы на ионы, и обосновал этот вывод на большом экспериментальном материале. По теории Аррениуса ионы, отличаясь от нейтральных атомов и молекул наличием электрического заряда, обладают и совершенно иными свойствами. Например, газообразный водород почти нерастворим в воде, а ионы водорода могут содержаться в водных растворах в очень больших концентрациях. Ионы натрия химически не действуют на воду, тогда как металлический натрий бурно реагирует с ней с образованием едкого натра. Хлор обладает характерным запахом и цветом, тогда как его ионы в растворе не имеют ни запаха, ни цвета. Ионы SO_4^{2-}, пока они заряжены отрицательно, могут существовать только в растворах, нейтральные молекулы такого состава неизвестны.

6. Возникновение диссоциации молекул в растворе можно пояснить на примере молекулы NaCl. Это полярная молекула, построенная из положительно заряженных ионов натрия Na^+ и отрицательно заряженных ионов хлора Cl^-. Молекулы воды также полярные и обладают большими дипольными моментами. Когда молекула NaCl оказывается между двумя молекулами воды, то эти три полярные молекулы стремятся ориентироваться преимущественно так, чтобы быть обращенными друг к другу разноименно заряженными концами, как показано на схематическом рис. 210. Но в таком случае молекулы воды ослабят связь между ионами Na^+ и Cl^-, и энергии теплового движения может оказаться достаточно, чтобы молекула

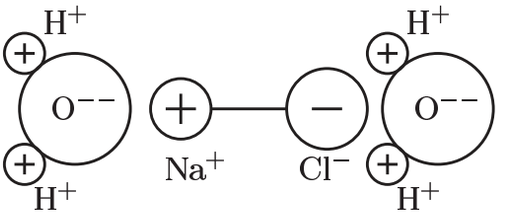


Рис. 210

NaCl распалась на ионы. Когда ионы Na^+ и Cl^- разойдутся достаточно далеко, то воду, разделяющую их, можно рассматривать как сплошную среду, ослабляющую силу притяжения между указанными ионами в ε раз. Это ослабление уменьшает вероятность рекомбинации рассматриваемых ионов. Отсюда ясно, что наибольшей диссоциирующей способностью должны обладать растворители с наибольшей диэлектрической проницаемостью ε. Такова прежде всего вода, для которой $\varepsilon = 81$.

7. *Коэффициентом* или *степенью диссоциации* α называется отношение числа диссоциировавших молекул к общему числу молекул растворенного вещества. Величина α зависит от концентрации растворенного вещества. Эта зависимость в общих чертах может быть получена из следующих простых соображений. Пусть n — число молекул растворенного вещества в единице объема раствора. Из этого числа $n\alpha$ молекул диссоциированы, а остальные $n(1-\alpha)$ молекул не диссоциированы. Диссоциированные молекулы могут рекомбинировать. Среднее число процессов рекомбинации в единице объема в единицу времени пропорционально $(n\alpha)^2$ и может быть представлено в виде $A(n\alpha)^2$. Среднее число обратных процессов, т. е. новых актов диссоциации, будет пропорционально числу имеющихся недиссоциированных молекул и представляется выражением $Bn(1-\alpha)$. В установившемся состоянии число прямых процессов в среднем должно равняться числу обратных процессов, т. е. $A(n\alpha)^2 = Bn(1-\alpha)$. Отсюда

$$\frac{\alpha^2}{1-\alpha} = \frac{B}{An} = \frac{K}{n}. \tag{92.1}$$

Коэффициенты A и B, а следовательно, и их отношение $K = B/A$ зависят от температуры и давления раствора. Если концентрация молекул растворенного вещества, а с ней и концентрации ионов в растворе малы, то ионы можно рассматривать как *независимые невзаимодействующие частицы*, подобно молекулам идеального газа. В этих условиях коэффициент K не зависит от концентрации n, а отношение $\alpha^2/(1-\alpha)$ обратно пропорционально концентрации. Это положение называется *законом разведения Оствальда* (1853–1932). Когда $n \to 0$, то $\alpha^2/(1-\alpha) \to \infty$, а следовательно, $\alpha \to 1$. Таким образом, *в бесконечно разбавленных растворах все молекулы растворенного вещества диссоциированы.*

Закон разведения Оствальда хорошо выполняется для *слабых электролитов*, т. е. таких электролитов, у которых степень диссоциации мала ($\alpha \ll 1$). Для *сильных электролитов*, т. е. электролитов, у которых величина α порядка единицы ($\alpha \sim 1$), закон Оствальда не согласуется с опытными данными. Исключение, конечно, составляет случай сильно разбавленных растворов, для которых этот закон приводит к правильному результату $\alpha = 1$. Современная теория растворов принимает, что молекулы сильного электролита диссоциированы полностью, а все особенности поведения раствора пытается объяснить взаимодействием ионов между собой и с молекулами растворителя.

§ 93. Примеры электролиза

1. Молекулы серной кислоты H_2SO_4 при растворении в воде диссоциируют согласно уравнению

$$H_2SO_4 \rightleftarrows 2H^+ + SO_4^{2-}.$$

Если в раствор опущены платиновые электроды, соединенные с полюсами гальванической батареи, то положительные ионы водорода H^+ устремятся к катоду, а отрицательные ионы SO_4^{2-} — к аноду. Ионы H^+, нейтрализуясь электронами катода, превращаются в нейтральные атомы, а затем в молекулы H_2. Ионы SO_4^{2-}, отдавая отрицательные заряды аноду, превращаются в радикалы SO_4, вступающие в реакцию с водой согласно уравнению

$$2SO_4 + 2H_2O = 2H_2SO_4 + O_2.$$

Таким образом, количество серной кислоты в растворе не меняется. Однако на катоде выделяется водород, а на аноде — кислород. Окончательный итог электролиза сводится к разложению воды на кислород и водород.

2. Опустим в водный раствор медного купороса $CuSO_4$ медный анод и катод из какого-либо проводящего материала, не реагирующего с медным купоросом, например угля. Ионы Cu^{2+} пойдут к катоду, там будет выделяться чистая медь. Ионы SO_4^{2-} после нейтрализации на аноде вступают с ним в химическое соединение согласно уравнению $Cu + SO_4 = CuSO_4$. Таким образом, количество медного купороса в растворе меняться не будет. В результате электролиза происходит как бы перенос меди с анода на катод. Описанный процесс применяется для рафинирования (очистки) меди.

3. Рассмотрим, наконец, электролиз водного раствора поваренной соли NaCl. Этот пример интересен тем, что в зависимости от материала электрода можно получить и первичные, и вторичные продукты электролиза. Если катод ртутный, то выделяющийся на нем металлический натрий будет растворяться в ртути, образуя амальгаму, из которой натрий может быть легко выделен. Таким путем получается первичный продукт электролиза. Если же катод сделан из платины, то выделяющийся на нем металлический натрий будет входить в соединения с водой, образуя едкий натр в соответствии с уравнением $2Na + 2H_2O = 2Na(OH) + H_2$, а на катоде будет выделяться вторичный продукт электролиза — водород.

4. Стекло при обычных температурах является непроводником электричества. Однако если его нагреть до температуры в несколько сот градусов, то оно начинает заметно проводить электричество. Наглядной демонстрацией этого может служить опыт, схема которого приведена на рис. 211. В цепь городского тока через реостат R включена стеклянная палочка, на концы которой A и B намотана голая медная проволока. При комнатной температуре сопротивление

стеклянной палочки составляет многие миллионы ом. Поэтому при замыкании рубильника через цепь потечет ничтожный ток (микроамперы или еще меньше), и никакого накала лампочки $Л$ не будет. Но если палочку AB нагреть на газовой горелке до температуры 300–400 °С, то ее сопротивление упадет до нескольких десятков омов, и нить лампочки раскалится. Если после этого убрать горелку и одновременно закоротить лампочку ключом K, то общее сопротивление цепи уменьшится, а ток возрастет. Стеклянная палочка будет нагреваться электрическим током и раскалится до яркого свечения, в результате чего ее сопротивление еще больше уменьшится, а ток возрастет. В конце концов палочка расплавится.

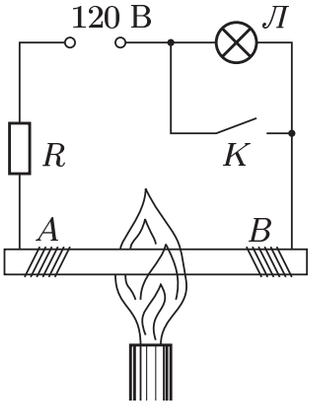


Рис. 211

Чем же объясняется проводимость стекла? Стекло представляет собой сильно переохлажденную жидкость, обладающую громадной вязкостью. Оно является также электролитом, в котором имеются положительные ионы натрия Na^+. При нагревании, когда стекло размягчается и его вязкость сильно уменьшается, ионы в стекле приобретают заметную подвижность. Они-то и являются переносчиками тока в стекле. Доказательством этого может служить следующий демонстрационный опыт. В тигль, подогреваемый газовой горелкой (рис. 212), помещается расплавленная чилийская селитра $NaNO_3$, в которую погружается примерно на одну треть баллон вакуумной лампочки накаливания (газонаполненная не годится). Нить лампочки накаливается постоянным током. Угольный электрод a присоединяется к положительному концу нити и погружается в расплав селитры. Он служит анодом, по отношению к которому все точки нити лампочки имеют более низкий потенциал. Расплав селитры частично диссоциирован на ионы Na^+ и NO_3^-. Под действием разности потенциалов между анодом и нитью лампочки ионы Na^+ движутся в селитре в направлении от анода a к баллону лампочки. Затем они проникают внутрь лампочки через ее стенки. На этих стенках они нейтрализуются электронами, испускаемыми накаленной нитью, и превращаются в нейтральные атомы Na. Испаряясь, последние осаждаются на более холодных частях внутренней поверхности стеклянного баллона лампочки, где образуется хорошо видимый зеркальный слой натрия.

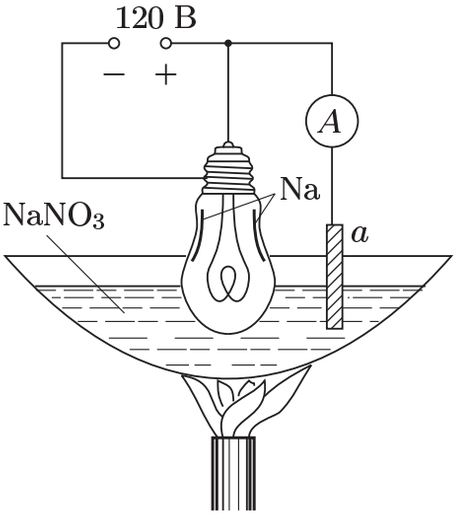


Рис. 212

§ 94. Законы электролиза Фарадея и элементарный заряд

1. Электролиз впервые наблюдался в 1800 г. Никольсоном (1753–1815) и Карлейлем, разложившими воду током от вольтова столба. Через семь лет Дэви (1778–1829) выделил и открыл натрий путем электролиза влажных кусков едкой щелочи. В дальнейшем электролиз наблюдался на многих других электролитах. Открытие количественных законов электролиза Фарадеем последовало в 1833 г. после того, как он ввел подразделение продуктов и реакций при электролизе на *первичные* и *вторичные*. *Законы Фарадея* определяют количества первичных продуктов, выделяющихся на электродах при электролизе. Их легко получить из рассмотрения механизма явления электролиза.

Каждый ион содержит целое число элементарных зарядов e, и поэтому заряд иона по абсолютной величине представляется выражением νe, где ν — целое число, равное валентности соответствующего химического элемента или соединения. Пусть на электроде, выделилось n таких ионов. Тогда их заряд по абсолютной величине будет $n\nu e$. Если эти ионы выделились на катоде, то их заряд нейтрализуется электронами, подтекающими к катоду по проводам, соединяющим его с источником тока. Если же они выделятся на аноде, то такое же количество электронов по проводам утечет от анода. В обоих случаях через цепь пройдет количество электричества $q = n\nu e$. Пусть M — масса вещества, выделившегося на электроде, а m — масса иона (атома или молекулы). Тогда $n = M/m$, и, следовательно, $M = mq/(\nu e)$. Умножив числитель и знаменатель последней дроби на постоянную Авогадро N, получим

$$M = \frac{A}{\nu}\,\frac{q}{F}, \tag{94.1}$$

где $A = Nm$ — атомная масса, а F — постоянная, определяемая выражением

$$F = Ne. \tag{94.2}$$

Эта постоянная называется *постоянной Фарадея.*

Для выяснения физического смысла постоянной Фарадея введем понятие *моль*. Так называют количество вещества, масса которого в граммах равна молярной массе A атома (а в случае химического соединения — молекулы), деленной на валентность ν. Если в формуле (94.1) положить $M = A/\nu$, то получится $q = F$. Это означает, что постоянная Фарадея есть заряд, переносимый при электролизе одним молем ионов.

Формула (94.1) содержит оба закона электролиза, открытых Фарадеем. Согласно первому закону *масса вещества, выделяющегося при электролизе на каждом электроде, пропорциональна количеству прошедшего электричества* q. Второй закон утверждает, что *эта масса пропорциональна химическому эквиваленту* A/ν.

2. Измерив на опыте M и q, можно по формуле (94.1) вычислить постоянную Фарадея F. Таким путем найдено $F = 2{,}8926 \times \times 10^{14}$ СГСЭ/моль$= 9{,}6487 \cdot 10^4$ Кл/моль. Зная F и N, можно по формуле (94.2) вычислить и элементарный заряд e. Именно таким путем английский ученый Стоней впервые оценил величину элементарного заряда в докладе, сделанном на съезде Британской ассоциации в 1874 г. Работа Стонея была опубликована только в 1881 г., после того как Гельмгольц в речи, посвященной памяти Фарадея, которую он прочел 5 апреля того же года в Королевском обществе в Лондоне, изложил (по-видимому, независимо от Стонея) те же результаты, полученные таким же путем. В этой знаменитой речи Гельмгольц сказал:

«Если мы признаем существование атомов химических элементов, то мы не можем избежать и дальнейшего заключения, что электричество, как положительное, так и отрицательное, разделено на определенные элементарные количества, которые ведут себя как атомы электричества».

§ 95. Скорости ионов и электрическая проводимость электролитов

1. При движении в электролите ион испытывает силу сопротивления, которая в случае установившегося движения пропорциональна скорости иона. На этом основании ион в электролите можно характеризовать *подвижностью*, определение которой было дано формулой (42.14). Здесь более удобно пользоваться той подвижностью, которая в § 42 была обозначена малой буквой b. Подвижности положительных и отрицательных ионов разные. Обозначая их через b^+ и b^-, а установившиеся скорости ионов — через $\mathbf{u}^+$ и $\mathbf{u}^-$, можем написать

$$\mathbf{u}^+ = b^+\mathbf{E}, \quad \mathbf{u}^- = -b^-\mathbf{E}. \tag{95.1}$$

Что касается концентрации, то она для положительных и отрицательных ионов одинакова, так как раствор электролита остается электрически нейтральным (точнее, квазинейтральным), и равна αn, где n — концентрация нейтральных молекул растворенного вещества, а α — степень диссоциации. Поэтому плотность тока в электролите определится выражением

$$\mathbf{j} = \alpha n(e^+\mathbf{u}^+ + e^-\mathbf{u}^-). \tag{95.2}$$

Так как заряд иона по абсолютной величине равен νe, где e — элементарный заряд, а ν — валентность иона, то

$$\mathbf{j} = \alpha n\nu e(b^+ + b^-)\mathbf{E}. \tag{95.3}$$

Отсюда для удельной электрической проводимости электролита получаем

$$\lambda = \alpha n\nu e(b^+ + b^-). \tag{95.4}$$

Концентрация n есть число нейтральных молекул растворенного вещества или, что то же самое, полное число ионов одного знака (как

свободных, так и связанных) в единице объема раствора. Введем так называемую *эквивалентную концентрацию растворенного вещества.* Эквивалентной концентрацией η называется число грамм-эквивалентов ионов одного знака, содержащихся в единице объема раствора. Так как грамм-эквивалент содержит N/ν частиц, то

$$\eta = \frac{n}{N/\nu} = \frac{n\nu}{N}. \tag{95.5}$$

Отсюда $n\nu = N\eta$. Кроме того, по формуле (94.2) $e = F/N$. С помощью этих соотношений формула (95.4) преобразуется к виду

$$\lambda = \alpha\eta F(b^+ + b^-). \tag{95.6}$$

Отношение удельной электрической проводимости λ к эквивалентной концентрации η растворенного вещества называется *эквивалентной* электрической проводимостью:

$$\Lambda = \frac{\lambda}{\eta} = \alpha F(b^+ + b^-). \tag{95.7}$$

Эквивалентная электрическая проводимость Λ_∞ при бесконечном разбавлении раствора найдется из этой формулы, если положить $\alpha = 1$ (см. § 92, п. 7). Это дает [1])

$$\Lambda_\infty = F(b^+ + b^-), \tag{95.8}$$

и, следовательно,

$$\alpha = \frac{\Lambda}{\Lambda_\infty}. \tag{95.9}$$

Измеряя эквивалентную электрическую проводимость, можно по этой формуле вычислить степень диссоциации растворенного вещества α. По формуле же (95.7) или (95.8) можно найти сумму подвижностей положительного и отрицательного ионов.

2. Чтобы определить значения самих подвижностей b^+ и b^-, к уравнению (95.8) надо присоединить еще одно уравнение. Его можно получить, измерив на опыте, как меняется при электролизе распределение концентрации растворенного вещества внутри электролита. Такие опыты были произведены Гитторфом (1824–1914) в 50-х годах прошлого столетия. Гитторф употреблял сосуды различной формы, в которых дном служили пористые перегородки (пергаментная бумага, животный пузырь или тонкая пористая пластинка из глины). В сосуды наливался раствор исследуемого электролита, и они собирались в один составной сосуд. Таким путем получался как бы один сосуд с электролитом, разделенный на несколько отделений пористыми перегородками.

[1]) Следуя Ф. Кольраушу (1840–1910), в электрохимии подвижностью иона в электролите принято называть произведение Fb. Формула (95.8) означает, что *эквивалентная электрическая проводимость электролита при бесконечном разбавлении раствора равна сумме подвижностей положительного и отрицательного ионов в смысле Кольрауша.*

Электролит в таком сосуде подвергался электролизу, причем ионы могли свободно переходить через пористые перегородки. После этого составной сосуд разбирался на отделения и измерялась концентрация растворенного вещества в каждом из них.

Допустим для простоты, что в составном сосуде имеется всего три одинаковых отделения A, B, K (рис. 213), содержавших в начале опыта по одинаковому числу N_0 молекул растворенного вещества. Пусть во время электролиза из отделения A через перегородку AB перешло в B ΔN^+ положительных ионов. Такое же количество положительных ионов перейдет из B в K через перегородку BK. В результате этих процессов число положительных ионов в отделении B не изменится, в отделении A оно уменьшится, а в отделении K увеличится на ΔN^+, так что числа положительных ионов в отделениях A и K будут равны соответственно

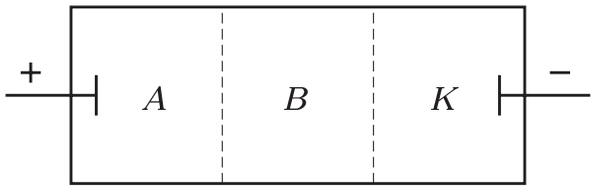


Рис. 213

$$N_A^+ = N_0 - \Delta N^+, \quad N_K^+ = N_0 + \Delta N^+.$$

Пусть, далее, через перегородки KB и BA за то же время прошло ΔN^- отрицательных ионов. От этого число отрицательных ионов в среднем отделении B не изменится, а в крайних отделениях A и K эти числа сделаются равными соответственно

$$N_A^- = N_0 + \Delta N^-, \quad N_K^- = N_0 - \Delta N^-.$$

В отделении A получится избыток отрицательных ионов над положительными, равный $N_A^- - N_A^+ = \Delta N^+ + \Delta N^-$. В отделении K, наоборот, число положительных ионов будет превышать число отрицательных на $N_K^+ - N_K^- = \Delta N^+ + \Delta N^-$, т. е. на ту же величину. До сих пор не было принято во внимание влияние электродов. В действительности избыточные ионы выделятся на электродах и будут нейтрализованы зарядами последних. В результате раствор всюду будет содержать по одинаковому количеству положительных и отрицательных ионов: в отделении A таких ионов останется $N_0 - \Delta N^+$, в отделении K — $(N_0 - \Delta N^-)$, в отделении B — N_0^-, т. е. число ионов не изменится. Таковы же будут и числа нейтральных молекул растворенного вещества в этих отделениях. Если $\Delta N^+ \neq \Delta N^-$, то концентрации электролита в отделениях A и K сделаются разными. Примем исходную концентрацию за единицу. Тогда в результате электролиза средние концентрации сделаются равными

$$c_A = \frac{N_0 - \Delta N^+}{N_0} = 1 - \frac{\Delta N^+}{N_0} \quad \text{(в отделении } A\text{)},$$

$$c_K = \frac{N_0 - \Delta N^-}{N_0} = 1 - \frac{\Delta N^-}{N_0} \quad \text{(в отделении } K\text{)}.$$

Концентрация электролита в отделении B не изменится. Из полученных соотношений следует

$$\frac{1-c_A}{1-c_K}=\frac{\Delta N^+}{\Delta N^-}. \tag{95.10}$$

Выразим теперь отношение $\Delta N^+/\Delta N^-$ через скорости и подвижности положительных и отрицательных ионов. Проведем в электролите произвольное сечение S. Тогда ΔN^+ можно определить как число положительных, а ΔN^- — отрицательных ионов, прошедших во время процесса через сечение S. Ясно, что ΔN^+ пропорционально току положительных, а ΔN^- — току отрицательных ионов, т. е. $\int n^+e^+u^+\,dS$ и $\int n^-e^-u^-\,dS$ соответственно. Поскольку электролит нейтрален ($|n^+e^+|=|n^-e^-|$), эти токи пропорциональны скоростям ионов u^+ и u^-. Следовательно,

$$\frac{u^+}{u^-}=\frac{b^+}{b^-}=\frac{1-c_A}{1-c_K}. \tag{95.11}$$

Вместо отношения скоростей или подвижностей ионов Гитторф ввел два отношения

$$\begin{aligned}\varkappa^+&=\frac{b^+}{b^++b^-}=\frac{1-c_A}{2-(c_A+c_K)},\\ \varkappa^-&=\frac{b^-}{b^++b^-}=\frac{1-c_K}{2-(c_A+c_K)},\end{aligned} \tag{95.12}$$

которые он назвал *числами переноса ионов*. Очевидно, $\varkappa^++\varkappa^-=1$.

3. Измеряя концентрации растворов до и после электролиза, Гитторф нашел, что в соляной кислоте ион водорода движется почти в 5 раз быстрее иона хлора. Измеряя же электрическую проводимость электролита, можно по формуле (95.8) вычислить сумму подвижностей b^++b^-, а затем и сами подвижности b^+ и b^-. Опыты показали, что *подвижность всякого иона в электролите не зависит от того, какие другие ионы находятся в растворе.* Этот результат был найден на опыте Фридрихом Кольраушем (1840–1910) и называется *законом Кольрауша о независимости движения ионов.* Он может рассматриваться как подтверждение гипотезы Аррениуса, согласно которой ионы в растворе уже не связаны между собой, а диссоциированы и движутся независимо друг от друга. Закон Кольрауша справедлив только для растворов малых концентраций. Если же концентрация ионов в растворе достигает заметной величины, то подвижности ионов начинают зависеть как от концентрации, так и от природы других ионов, содержащихся в растворе. С увеличением концентрации подвижность ионов уменьшается, хотя и не очень сильно.

Из табл. 3 видно, что подвижности ионов, вообще говоря, весьма малы. Наибольшей подвижностью обладает ион H^+, но и он в поле $E=1$ В/см перемещается всего на 12 см в час. Рассматривая таблицу, можно было бы думать, что радиус иона Li^+ больше радиуса иона Na^+,

а последний в свою очередь больше радиуса иона K^+. В действительности атомные радиусы щелочных металлов располагаются в обратной последовательности. Этот факт объясняют тем, что в растворе ионы *гидратированы*, т. е. окружены оболочкой воды, увлекаемой при их

Таблица 3. Значения подвижностей ионов в водных растворах слабой концентрации при температуре 18 °C

Ион	Подвижность, 10^{-3} см^2/(с · В)	Ион	Подвижность, 10^{-3} см^2/(с · В)
H^+	3,26	OH^-	1,80
Li^+	0,346	Cl^-	0,677
Na^+	0,450	NO_3^-	0,639
K^+	0,669	I^-	0,688

движении (см. т. II, § 122). Гидратация вызывается электрическими силами, с которыми ионы действуют на молекулы окружающей воды. Из всех щелочных ионов ион лития обладает наиболее сильным электрическим полем, а потому он гидратирован сильнее других ионов. Подвижности многовалентных ионов с разными зарядами, например Fe^{2+} и Fe^{3+}, не пропорциональны их зарядам, как может показаться на первый взгляд. Это также можно объяснить гидратацией, поскольку ионы большей валентности гидратированы сильнее.

4. Нернст (1864–1941) наглядно демонстрировал движение ионов, когда они окрашивали раствор. В U-образную трубку (рис. 214) снизу наливают фиолетовый раствор марганцовокислого калия $KMnO_4$ с концентрацией $\eta \approx 0{,}003$ грамм-эквивалента на 1 см^3. Поверх этого раствора наливают бесцветный раствор азотнокислого калия KNO_3 с той же эквивалентной концентрацией. Эти операции должны быть выполнены очень тщательно, чтобы граница между обоими растворами была резкой. Для этого в приборе предусмотрены специальные приспособления, которые мы не описываем. В растворе соль KNO_3 диссоциирует на ионы K^+ и NO_3^-, а соль $KMnO_4$ — на ионы K^+ и MnO_4^-. При пропускании электрического тока ионы K^+ в обоих растворах движутся к катоду. Достигнув последнего, они нейтрализуются, соединяются с водой и освобождают там водород. За движением этих ионов, поскольку они не дают окраски, мы следить не можем. Что касается ионов MnO_4^-, то они придают раствору фиолетовую окраску и движутся в направлении от катода к аноду. Поэтому в колене, где расположен катод, наблюдается понижение окрашенной границы, а в колене, где находится анод, — повышение. Скорость перемещения окрашенной границы равна скорости ионов MnO_4^- и может

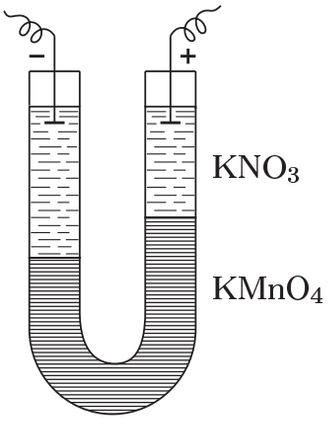


Рис. 214

быть непосредственно измерена. Если длина U-образной трубки равна 10 см, а приложенная разность потенциалов на электродах 50 В, то уже через 3 мин окрашенная граница (как в том, так и в другом колене) переместится на 0,45 см, так что разность окрашенных уровней в коленах трубки достигнет почти одного сантиметра.

§ 96. Гальванические элементы и аккумуляторы

1. В § 43 была изложена теория концентрационного элемента, в котором электрический ток возникает благодаря наличию градиента концентрации раствора. Было показано, что электродвижущая сила концентрационного элемента очень мала и практического значения не имеет. Тем не менее теория концентрационного элемента важна в принципиальном отношении. Благодаря простоте процессов, происходящих в концентрационном элементе, на нем проще всего уяснить возникновение электрического тока в гальванических элементах, усвоить понятие сторонних сил и обобщить закон Ома на случай наличия таких сил. Обращаемся теперь к краткому рассмотрению процессов, происходящих в гальванических элементах, применяющихся на практике.

2. Если кусок какого-либо металла, например цинка, погрузить в воду, то в первый момент такая система не будет находиться в равновесии. Кристаллическая решетка цинка состоит из положительных ионов Zn^{2+}. Под влиянием сильно полярных молекул воды ионы цинка, отрываясь от металла, начнут переходить в слой воды, примыкающий к поверхности погруженного куска. В результате этого цинк зарядится отрицательно, а вода — положительно. На границе металла и воды образуется тонкий поверхностный слой, называемый *двойным электрическим слоем*, в котором возникнет электрическое поле, направленное от воды к металлу. Это поле будет препятствовать переходу ионов металла в воду и способствовать возвращению их обратно в металл. Установится статистическое равновесие, в котором число ионов, переходящих из металла в воду, будет в среднем равно числу ионов, возвращающихся обратно из воды в металл. То же самое произойдет, если металл погрузить в водный раствор соли того же металла, например, цинк в раствор цинкового купороса $ZnSO_4$. Последний в растворе диссоциирует на ионы Zn^{2+} и SO_4^{2-}. Ионы цинка Zn^{2+}, получившиеся в растворе в результате электролитической диссоциации, конечно, ничем не отличаются от таких же ионов, перешедших в раствор от куска цинка. Повышение концентрации ионов Zn^{2+} в растворе, очевидно, облегчает переход этих ионов в металл и затрудняет обратный переход ионов Zn^{2+} из металла в раствор. Поэтому в растворе цинкового купороса цинк хотя и зарядится отрицательно, но слабее, чем в чистой воде. Равновесная разность потенциалов, которая установится между металлом и раствором, будет зависеть от температуры, давления, а также от концентрации раствора. Однако при одной и той же концентрации ионов металла в растворе она не будет зависеть от наличия в растворе других ионов, если только последние не осаждаются на металле, т. е. не участвуют в процессе

обмена ионами между раствором и металлом. Например, погрузив цинк сначала в раствор $ZnSO_4$, а затем в раствор $ZnCl_2$, мы получим в обоих случаях при одинаковой концентрации ионов Zn^{2+} одну и ту же разность потенциалов между цинком и раствором.

При погружении металла в раствор соли того же металла металл не всегда заряжается отрицательно. Например, если кусок меди погрузить в раствор медного купороса $CuSO_4$, то ионы Cu^{++} начнут осаждаться на меди, заряжая ее положительно. Двойной электрический слой на границе меди с раствором будет теперь обращен положительными зарядами к меди, а отрицательными — к раствору. Электрическое поле в таком слое будет направлено от меди к раствору. В установившемся состоянии медь окажется заряженной положительно, а раствор — отрицательно.

Итак, *при погружении металла в воду или в водный раствор, содержащий ионы того же металла, на границе металл–раствор образуется двойной электрический слой и возникает разность потенциалов между металлом и раствором.* Величина этой разности потенциалов для разных металлов разная. Она зависит от концентрации ионов рассматриваемого металла в растворе и сравнительно слабо зависит от температуры и давления.

3. При измерении разности потенциалов между металлическим электродом и электролитом, в который он погружен, встречается следующая трудность. Для включения вольтметра между электролитом и рассматриваемым электродом вольтметр надо присоединить к другому электроду, опущенному в тот же электролит. Однако между электролитом и вторым электродом возникает также разность потенциалов. Поэтому вольтметр укажет вовсе не разность потенциалов между интересующим нас металлом и электролитом, а разность потенциалов между двумя различными электродами, опущенными в один и тот же электролит. Однако при вычислении электродвижущей силы гальванического элемента достаточно знать разности потенциалов между его электродами. А эти разности не зависят от того, относительно какого тела измерены потенциалы самих электродов. Можно взять потенциал относительно любого тела, принимаемого за стандарт. Для этой цели чаще всего употребляется *нормальный водородный электрод*, состоящий из платинированной платины [1], опущенной в нормальный раствор, содержащий ионы водорода (например, в раствор серной кислоты), и омываемой током газообразного водорода при давлении 1 атм, причем водородный электрод и сочетаемый с ним другой электрод должны находиться при одинаковой температуре. (*Нормальным раствором* или *раствором нормальной концентрации* называют раствор, содержащий 1 грамм-эквивалент ионов металла или водорода в литре раствора.) Потенциал данного электрода относительно нормального водородного электрода называется *электродным потенциалом.* Если

[1] Так называют гладкую Pt, покрытую с помощью электролиза мелкодисперсной Pt.

при этом раствор, в который опущен электролит, нормальный, то электродный потенциал называют также *нормальным* и обозначают через $V_н$. В табл. 4 приведены значения $V_н$ для некоторых электродов. Перед электродом указаны также ионы, с которыми этот электрод обменивается при погружении в раствор. Такое указание необходимо, так как нормальный электродный потенциал зависит не только от вида металла, но и от величины заряда его ионов в растворе (например, Cu^{+} и Cu^{++}).

Таблица 4. Значения нормальных электродных потенциалов при 25 °C

Электрод	$V_н$, В	Электрод	$V_н$, В	Электрод	$V_н$, В
L^{+}/L	−3,045	Fe^{2+}/Fe	−0,440	H^{+}/H_2	0,0000
K^{+}/K	−2,925	Cd^{2+}/Cd	−0,403	Cu^{2+}/Cu	+0,337
Na^{+}/Na	−2,714	Co^{2+}/Co	−0,277	Cu^{+}/Cu	+0,521
Mg^{2+}/Mg	−2,37	Ni^{2+}/Ni	−0,250	Hg^{+}/Hg	+0,798
Al^{3+}/Al	−1,66	Sn^{2+}/Sn	−0,136	Ag^{+}/Ag	+0,799
Zn^{2+}/Zn	−0,763	Pb^{2+}/Pb	−0,126	Au^{3+}/Au	+1,50

Пользуясь приведенными данными, можно приближенно вычислить электродный потенциал $V_с$ и в том случае, когда концентрация электролита отлична от нормальной. Действительно, $V_с = V_н + (V_с - V_н)$. Разность потенциалов $(V_с - V_н)$, если концентрации не очень велики, можно вычислить с помощью теории концентрационного элемента. Надо только в формуле (44.13) положить равной нулю одну из подвижностей B^{+} или B^{-} (так как теперь в процессе установления равновесной разности потенциалов $V_с - V_н$ участвуют только ионы одного знака). Таким путем получаем $V_с - V_н = kT \ln c/(\nu e)$, где через νe обозначен заряд иона (e — элементарный заряд, а ν — целое число; для положительных ионов $\nu > 0$, для отрицательных $\nu < 0$). За единицу концентрации c принята концентрация нормального раствора. Следовательно,

$$V_с = V_н + \frac{kT}{\nu e} \ln c = V_н + \frac{RT}{\nu F} \ln c. \qquad (96.1)$$

4. Электрод, опущенный в раствор электролита, образует так называемый *полуэлемент.* Из двух таких полуэлементов образуется *гальванический элемент.* Примером может служить *элемент Даниэля.* Он состоит из цинкового электрода, погруженного в раствор $ZnSO_4$, и медного электрода, погруженного в раствор $CuSO_4$. Оба раствора отделены один от другого пористым цилиндром из необожженной глины. Стенки цилиндра не препятствуют движению ионов, но предохраняют растворы от быстрого перемешивания.

Разность потенциалов между электродами незамкнутого элемента называется *электродвижущей силой* последнего (см. § 44). Она приближенно равна разности электродных потенциалов соответствующих электродов. Допустим, например, что оба раствора $ZnSO_4$ и $CuSO_4$

в элементе Даниэля (1790–1845) имеют нормальные концентрации. Из таблицы нормальных электродных потенциалов находим, что электродный потенциал цинка равен −0,763 В, а меди +0,337 В. Поэтому электродвижущая сила элемента Даниэля будет $\mathscr{E} = 0{,}337 - (-0{,}763) = 1{,}10$ В. В этом расчете мы пренебрегли диффузионной разностью потенциалов, возникающей на границе двух растворов. Однако эта величина мала и обычно не превосходит 0,03 В (см. задачу 2 к § 44).

В *элементе Вольты* (1745–1827) оба электрода — цинковый и медный — погружены в один и тот же раствор серной кислоты H_2SO_4. Последний не содержит ни ионов цинка, ни ионов меди. Однако это имеет место только в первый момент после погружения. В дальнейшем же в раствор входят как ионы цинка, так и ионы меди, а потому этот случай принципиально не отличается от уже рассмотренного.

Если электроды элемента Даниэля соединить проводом, то электроны по этому проводу начнут переходить с отрицательного. электрода (цинка) на положительный (медь). От этого заряды обоих электродов по абсолютной величине уменьшатся. Это нарушит равновесие между электродами и электролитами, в которые они погружены. Ионы Zn^{2+} станут слабее удерживаться отрицательным (цинковым) электродом и начнут переходить в раствор, двигаясь при этом в направлении к положительному (медному) электроду. Со своей стороны положительный (медный) электрод станет слабее отталкивать ионы Cu^{2+}. Последние начнут переходить из раствора и осаждаться на положительном электроде. Что касается ионов SO_4^{2-}, то у отрицательного электрода образуется недостаток, а у положительного — избыток таких ионов. Короче, потечет электрический ток — во внешней цепи от меди к цинку, а во внутренней от цинка к меди. В результате этого цинк будет растворяться, а медь осаждаться на положительном электроде. Общее же число ионов SO_4^{2-} в растворе не изменится, увеличится только их концентрация у цинкового электрода, а у медного уменьшится.

5. Прохождение электрического тока в гальванических элементах может осложняться побочными продуктами, выделяющимися на электродах, а также появлением градиентов концентрации в электролите. Это ведет к возникновению дополнительной электродвижущей силы, направленной против электродвижущей силы самого элемента и ослабляющей величину последней. Такое явление называется *электролитической поляризацией*, или *поляризацией гальванических элементов*. Примером может служить элемент Вольты (цинковый и медный электроды в растворе серной кислоты). При замыкании элемента положительные ионы H^+ движутся к меди и осаждаются на ней. После длительной работы элемента медный электрод как бы заменяется водородным. Последний обладает способностью посылать в раствор ионы H^+ и создавать тем самым дополнительную электродвижущую силу, направленную противоположно электродвижущей силе самого элемента. Электродный потенциал водорода на 0,337 В ниже электродного потенциала меди (см. табл. 4). По этой причине электродвижущая сила элемента уменьшается примерно на такую же величину. Кроме

того, слой водорода, образующийся на медном электроде, увеличивает внутреннее сопротивление элемента. Все это ведет к уменьшению силы тока в цепи, т. е. к поляризации элемента.

Для устранения вредного влияния поляризации элементов можно было бы удалять водород и другие вредные продукты, выделяющиеся на электродах, путем механического перемешивания электролита. Но такой способ может быть эффективным только тогда, когда поляризация возникает из-за градиентов концентрации электролитов. В остальных случаях он мало эффективен. Более эффективны следующие два способа *деполяризации элементов*, которые и применяются на практике. Первый из них состоит в том, что в элементе применяются две жидкости, подобранные так, что во время работы элемента не происходит изменения состава электродов. Такой способ применяется, например, в элементе Даниэля. Здесь на медном электроде выделяется медь, и состав этого электрода во время работы элемента не меняется. На цинковом же электроде растворяется цинк, так что не меняется состав и этого электрода.

Второй способ деполяризации гальванических элементов — *химический.* Он состоит в том, что к электролиту добавляются вещества, активно взаимодействующие с веществами, вызывающими поляризацию. Так, для ослабления вредного влияния водорода, выделяющегося на катоде, *деполяризаторами* могут служить различные *окислители*, а в случае кислорода, выделяющегося на аноде, — *восстановители.* Химический способ применяется, например, в наиболее употребительном элементе — *элементе Лекланше* (1839–1882). Положительный электрод в этом элементе состоит из угольного стержня, окруженного спрессованной массой перекиси марганца MnO_2 и графита (для увеличения электрической проводимости). Эта масса в свою очередь окружена цинковым листом, который служит отрицательным электродом элемента. Все это погружено в водный раствор хлористого аммония (нашатыря) NH_4Cl. Электродвижущая сила элемента несколько меньше 1,5 В. На цинковом электроде ионы Zn^{++} переходят в раствор, заряжая цинк отрицательно. Ионы NH_4^+ переходят из раствора на второй электрод, заряжая его положительно. Затем происходят вторичные реакции:

$$2NH_4 = 2NH_3 + 2H,$$

$$2H + MnO_2 = H_2O + MnO,$$

$$2NH_3 + 2H_2O = 2NH_4(OH).$$

Перекись марганца MnO_2 является сильным окислителем и окисляет выделяющийся водород в воду, как указывает вторая из этих реакций. Практически элемент Лекланше выпускается в виде так называемых *сухих элементов*, в которых к раствору электролита добавляются *загустители* (крахмал, мука). Во избежание испарения воды элемент сверху заливается слоем смолы.

6. Электролитическая поляризация может возникать не только в гальванических элементах, но и при электролизе. Так, если в раствор серной кислоты опустить два платиновых электрода, соединенных с источником тока, то при электролизе на катоде будет осаждаться водород, а на аноде — кислород (см. § 93). По мере накопления этих газов будут повышаться их парциальные давления. Когда они станут равными атмосферному давлению, начнется выделение на электродах пузырьков водорода и кислорода. Если теперь отключить источник тока, то один из платиновых электродов окажется покрытым водородом, а другой — кислородом. В результате электроды поляризуются: водородный будет заряжен положительно, а кислородный — отрицательно. При соединении электродов через электролит потечет электрический ток, противоположный току при электролизе, а водород и кислород в виде ионов будут переходить обратно в раствор. Когда запас обоих газов израсходуется, ЭДС обратится в нуль и ток прекратится.

В частных случаях ЭДС поляризации может и не возникать. Так будет, например, когда два медных электрода опущены в раствор $CuSO_4$. При электролизе один из электродов растворяется, а на другом осаждается металлическая медь. Состав электродов не меняется, а потому поляризация не возникает. Точнее, появится поляризация, обусловленная изменением концентраций электролита у электродов (см. § 95). Но электродвижущая сила, возникающая при такой поляризации, очень мала.

Посмотрим теперь, как ЭДС поляризации сказывается на процессе электролиза. Допустим, например, что производится электролиз соляной кислоты HCl и при этом используются платиновые электроды. Пусть внешнее напряжение V, приложенное к электродам, невелико, например $V = 0{,}5$ В. По мере накопления на электродах водорода и хлора возрастает ЭДС поляризации. Когда она достигнет 0,5 В, ток прекратится. Допустим теперь, что напряжение V медленно возрастает. Тогда будет увеличиваться и количество газов на электродах. Когда парциальные давления газов сделаются равными атмосферному давлению, они начнут выделяться в виде пузырьков. Начиная с этого момента состав электродов меняться не будет, а электродвижущая сила поляризации достигнет предельного значения V_0, равного равновесной разности потенциалов между хлором и водородом. Она равна $V_0 = 1{,}36$ В. Ток через электролит I будет определяться эффективной разностью потенциалов $V - V_0$. Он равен

$$I = \frac{V - V_0}{R},$$

где R — сопротивление электролита между электродами. Для того чтобы электролиз продолжался непрерывно, необходимо приложить внешнее напряжение V, превышающее «пороговое напряжение» V_0. Это пороговое напряжение называется *напряжением разложения электролита.*

В действительности в зависимости от характера протекающих электрохимических реакций у электродов напряжение разложения электролита обычно превышает ЭДС поляризации. Это явление называется *перенапряжением при электролизе*. Например, электродный потенциал кислорода относительно водорода равен +1,23 В. Можно было бы ожидать, что напряжение разложения серной кислоты будет 1,23 В. На самом деле для этой величины опыт дает 1,64 В.

7. Разновидностью гальванических элементов являются *аккумуляторы*, или *вторичные элементы*. Вещества, расходуемые при возбуждении электрического тока, получаются внутри самих аккумуляторов путем электролиза. Наибольшее распространение получил *свинцовый*, или *кислотный*, *аккумулятор*. Он изготовляется из двух решетчатых свинцовых пластин, заполненных пастой из PbO и воды и опущенных в 30-процентную серную кислоту (с плотностью 1,2 г/см^3). В результате реакции

$$PbO + H_2SO_4 = PbSO_4 + H_2O$$

на пластинах образуется слой труднорастворимого сернокислого свинца $PbSO_4$. Для того чтобы аккумулятор стал источником тока, его необходимо зарядить. С этой целью через аккумулятор пропускают электрический ток. Ионы H^+, двигаясь к катоду, нейтрализуются на нем и вступают в реакцию

$$PbSO_4 + 2H = Pb + H_2SO_4.$$

Отрицательные ионы SO_4^{2-}, достигая анода, также нейтрализуются и сначала вступают в реакцию

$$PbSO_4 + SO_4 = Pb(SO_4)_2,$$

затем происходит обратимая реакция

$$Pb(SO_4)_2 + 2H_2O \rightleftarrows 2PbO_2 + 2H_2SO_4.$$

Таким образом, после зарядки катод аккумулятора превращается в губчатую массу металлического свинца, а анод — в перекись свинца PbO_2. Концентрация серной кислоты в аккумуляторе при зарядке повышается. Если выключить внешний источник тока, то получится гальванический элемент с анодом из PbO_2 и катодом из Pb. Если такой элемент не замкнут, то он может сохраняться в заряженном состоянии весьма долго. При соединении полюсов заряженного аккумулятора проводником через цепь потечет электрический ток, направление которого противоположно зарядному току. Аккумулятор начинает разряжаться. Отрицательные ионы SO_4^{2-} переходят из раствора на свинцовый катод, нейтрализуются и вступают в реакцию $Pb + SO_4 = PbSO_4$. Около анода происходит обратимая реакция

$$PbO_2 + 2H_2SO_4 \rightleftarrows Pb(SO_4)_2 + 2H_2O.$$

Положительные ионы H^+ переходят из раствора на анод (PbO_2), нейтрализуются там и вступают в реакцию

$$Pb(SO_4)_2 + 2H = PbSO_4 + H_2SO_4.$$

Концентрация серной кислоты понижается. В конце концов аккумулятор вернется в исходное состояние; обе пластины его будут состоять из $PbSO_4$, а концентрация H_2SO_4 примет свое начальное значение. Для того чтобы аккумулятор давал ток, его необходимо снова зарядить.

ЭДС свинцового аккумулятора в самом конце зарядки может достигать 2,7 В. При разрядке она уменьшается: сначала быстро примерно до 2,2 В, затем очень медленно приблизительно до 1,85 В. Дальнейшую разрядку производить нельзя, так как при этом пластины аккумулятора покрываются труднорастворимым слоем $PbSO_4$, отчего аккумулятор портится.

Максимальный заряд, который может отдать аккумулятор при разрядке, называется его *емкостью.* Емкость принято измерять в *амперчасах.* Она тем больше, чем больше поверхность электродов. Поэтому для увеличения емкости аккумулятор изготовляют не из сплошных свинцовых пластин, а из пластин, имеющих форму решеток, как описано выше.

Широкое распространение в настоящее время получили также *железо-никелевые*, или *щелочные*, *аккумуляторы.* Они имеют один электрод из железа, а другой — из никеля. Электролитом служит 20-процентный раствор едкого калия K(OH). В заряженном состоянии анодом в таких аккумуляторах является гидрат окиси никеля $Ni(OH)_3$, а катодом — железо. ЭДС щелочных аккумуляторов около 1,3 В. Их емкость при равном весе больше, чем у свинцовых аккумуляторов.

Глава VII

ЭЛЕКТРИЧЕСКИЕ ТОКИ В МЕТАЛЛАХ, ПОЛУПРОВОДНИКАХ И ВАКУУМЕ

§ 97. Инерция электронов в металлах

1. Как уже неоднократно говорилось, в металлах носителями тока являются *свободные электроны*, т. е. электроны, сравнительно слабо связанные с ионами кристаллической решетки металла. Ионы в металлах не участвуют в переносе электричества. Если бы это было не так, то прохождение электрического тока через металл сопровождалось бы электролизом и связанным с ним переносом вещества. На самом деле этого не наблюдается. К. Рике (1845–1915) в течение года пропускал электрический ток через три поставленные друг на друга цилиндра — медный, алюминиевый и снова медный. Несмотря на то, что общий заряд, прошедший через эти цилиндры в течение указанного времени, достигал примерно 3,5 миллиона кулонов, никакого проникновения металлов друг в друга обнаружено не было и масса цилиндров сохранялась с точностью до $\pm 0{,}03$ мг.

2. Еще более определенные заключения о природе носителей тока в металлах позволяют сделать опыты с возбуждением электрического тока *силами инерции.* Для уяснения идеи таких опытов рассмотрим тонкое проволочное кольцо, неравномерно вращающееся вокруг своей геометрической оси. При всяком ускорении вращения свободные электроны будут отставать, а при замедлении — опережать ионы кристаллической решетки кольца. Возникнет движение электронов относительно кристаллической решетки, т. е. *электрический ток.* Для количественного описания явления перейдем к системе отсчета, вращающейся вместе с кольцом. В этой системе отсчета появится сила инерции $\mathbf{F}_{\text{ин}}$, действующая на каждый свободный электрон. Разделив ее на заряд электрона e, получим стороннее поле $\mathbf{E}^{\text{стор}} = \mathbf{F}_{\text{ин}}/e$, которое и возбуждает электрический ток. Поскольку возбуждаемый ток — переменный, закон Ома надо писать в форме (42.6) или

$$\mathbf{j} = \tau_{\text{ин}} \frac{d\mathbf{j}}{dt} = \lambda(\mathbf{E}^{\text{стор}} + \mathbf{E}), \tag{97.1}$$

где $\mathbf{E}$ — электрическое поле, которое может появиться из-за смещения электронов относительно ионов. Приведя это уравнение обычным способом (см. § 44) к интегральной форме, получим

$$R\left(I + \tau_{\text{ин}} \frac{dI}{dt}\right) = \oint (\mathbf{E}^{\text{стор}} + \mathbf{E})\, d\mathbf{l},$$

где R — сопротивление кольца, а интегрирование ведется по контуру кольца. Первый интеграл $\oint \mathbf{E}^{\text{стор}}\, d\mathbf{l}$ есть электродвижущая сила $\mathscr{E}$ сторонних сил, действующих в контуре. Второй интеграл $\oint \mathbf{E}\, d\mathbf{l}$ равен $-L\, dI/dt$. Таким образом,

$$(L + R\tau_{\text{ин}})\frac{dI}{dt} + RI = \mathscr{E}. \tag{97.2}$$

В рассматриваемом нами случае электродвижущая сила $\mathscr{E}$ создается силами инерции. Силы инерции, перпендикулярные к оси провода, на силу тока I не влияют. Имеет значение только сила инерции, направленная вдоль оси провода. Она возникает из-за неравномерности вращения и равна $F_{\text{ин}} = -m\dot{v}$, где m — масса электрона, а v — линейная скорость вращения кольца. С учетом всего этого предыдущее уравнение принимает вид

$$(L + R\tau_{\text{ин}})\frac{dI}{dt} + RI = -\frac{m}{e}\, l\dot{v},$$

где l — длина кольца. Интегрируя полученное уравнение по времени от $t = t_1$ до $t = t_2$, найдем

$$(L + R\tau_{\text{ин}})(I_2 - I_1) + Rq = \frac{m}{e}\, l(v_1 - v_2),$$

где $q = \int I\, dt$ — количество протекшего электричества, а I_1, I_2, v_1, v_2 — значения силы тока I и скорости v в моменты времени t_1 и t_2 соответственно. Пусть до момента t_1 кольцо вращалось равномерно со скоростью $v_1 = v$, а в момент t_2 скорость v равна нулю и процесс установления тока в кольце к этому моменту закончился. Тогда $I_1 = I_2 = 0$ и, следовательно,

$$q = \frac{mlv}{eR}. \tag{97.3}$$

Измерив с помощью баллистического гальванометра количество протекшего электричества q, можно из этого уравнения определить удельный заряд e/m, а по направлению отклонения гальванометра судить о знаке заряда e.

Идея подобных опытов была высказана в 1913 г. русскими физиками Л. И. Мандельштамом (1879–1944) и Н. Д. Папалекси (1880–1947). Они же поставили качественные опыты и показали, что при крутильных колебаниях проволочной катушки вокруг ее геометрической оси действительно возникает переменный ток (см. задачу 3 к § 132). Приближавшаяся первая мировая война (опыты производились в Страсбурге) помешала Мандельштаму и Папалекси закончить начатое исследование. Опыт был предложен вновь Х. А. Лоренцем и осуществлен Толменом и Стюартом в 1916 г.

В опытах Толмена и Стюарта катушка из металлической проволоки приводилась в быстрое вращение вокруг своей геометрической оси (линейная скорость вращения достигала 300 м/с). Концы проволочной обмотки были соединены с очень чувствительным баллистическим гальванометром длинными гибкими проводами, скручивающимися во

время вращения катушки. При помощи специальных неподвижных катушек с электрическим током магнитное поле Земли в пределах катушки было настолько тщательно скомпенсировано, что при равномерном вращении последней гальванометр не обнаруживал никаких индукционных токов. Убедившись в этом, экспериментаторы быстро затормаживали вращение катушки, и тогда тотчас же гальванометр отклонялся. К этому опыту применимо уравнение (97.3), если в нем под l понимать длину проволоки в катушке (в опытах Толмена и Стюарта она достигала 500 м), а под R — общее сопротивление цепи, включая сопротивление гальванометра и соединительных проводов. Катушки изготовлялись из медной, алюминиевой и серебряной проволок. Направление отклонения гальванометра показало, что носителями тока в металлах являются отрицательные заряды. Удельный заряд e/m в пределах погрешностей измерений оказался таким же, как и у электронов в опытах с катодными лучами.

3. Кеттеринг и Скотт в 1944 г. произвели другой опыт, являющийся как бы обращением опыта Толмена и Стюарта. Они наблюдали изменение момента импульса проволочной катушки при изменении силы тока в ней. Подвешенная катушка, по которой протекал постоянный электрический ток, совершала крутильные колебания относительно своей геометрической оси $\varphi = a\sin\omega t$. Если ток обусловлен движением электронов, то сила тока представляется выражением $I = Sneu$, где S — площадь поперечного сечения провода, n — число свободных электронов в единице объема, а u — средняя упорядоченная скорость электронов относительно провода. С этим током связан момент количества движения относительно оси катушки $L_э = lSnmur$, где l — длина проволоки, а r — радиус витка. Таким образом, $L_э = (m/e)lrI$. Если сила тока изменится на ΔI, то изменится и момент импульса, связанный с движением электронов, а именно: $\Delta L_э = (m/e)lr\Delta I$. Однако момент импульса всей системы — кристаллической решетки и электронов — измениться не может. Действительно, ток меняется под действием электрического поля. Последнее действует не только на электроны, но и на ионы. Поскольку система электрически нейтральна, момент импульса, приобретенный электронами, будет равен по модулю и противоположен по знаку моменту импульса, приобретенному ионами кристаллической решетки. Отсюда следует, что при изменении тока I решетка приобретает момент импульса

$$\Delta L_р = -\Delta L_э = -\frac{m}{e}\,lr\Delta I.$$

Пусть изменение тока производится настолько быстро, что за время одного колебания катушка не успевает повернуться на заметный угол. Тогда можно считать, что угловая скорость решетки $\dot\varphi$ при крутильных колебаниях изменяется в соответствии с уравнением $\Theta\Delta\dot\varphi = \Delta L_р$, где Θ — момент инерции катушки относительно ее геометрической оси. Пусть изменение тока произошло в момент, когда катушка проходила через положение равновесия. В этот момент $\dot\varphi \equiv \omega a\cos\omega t = \omega a$. Поэтому амплитуда колебаний a получит приращение, определяемое

уравнением $\Theta\omega\Delta a = \Delta L_{\text{р}}$. Фактически в опыте производилось изменение на противоположное направления тока, так что $\Delta I = -2I$, и, следовательно,

$$\Delta a = \frac{\Delta L_{\text{р}}}{\Theta\omega} = \frac{2mlr}{e\Theta\omega}\, I = \frac{mlrT}{\pi e\Theta}\, I, \tag{97.4}$$

где $T = 2\pi/\omega$ — период крутильных колебаний катушки. На опыте измерялась величина Δa с помощью специального очень чувствительного устройства, которое мы не описываем.

Ввиду малости эффекта и необходимости защитить прибор от возмущений, значительно превосходящих самый эффект, опыт очень труден. Однако он был поставлен с использованием современной аппаратуры и притом настолько тщательно и в таких хороших условиях, что результаты получились весьма устойчивые. Опыт производился на специальной станции, расположенной вдали от дорог. Чувствительная часть прибора располагалась в идеально темном подвале, в котором поддерживалась постоянная температура и не было никаких потоков воздуха. В подвале были расположены три взаимно перпендикулярные системы катушек с общим центром. Они использовались для компенсации и контролирования различных составляющих земного магнитного поля. Наблюдатель делал отсчеты, находясь в верхнем помещении, при помощи особой системы телеуправления. Опыт производился с катушками из медной и алюминиевой проволок. Измеренное среднее значение удельного заряда e/m отличалось от общепринятого всего на 0,2 %.

§ 98. Явление Холла

1. Если исходить из модели «свободных» электронов (см. § 42), то плотность электрического тока в металле может быть выражена через их концентрацию n и подвижность по формуле

$$\mathbf{j} = enb\mathbf{E}. \tag{98.1}$$

Формула была бы справедлива и в том случае, если бы носителями тока были положительно заряженные частицы ($e > 0$). Если $e > 0$, то ток $\mathbf{j}$ направлен по полю $\mathbf{E}$; если же $e < 0$, то он направлен противоположно полю $\mathbf{E}$. Для определения двух величин n и b к уравнению (98.1) необходимо присоединить второе уравнение. С этой целью можно воспользоваться *эффектом Холла* (1855–1938).

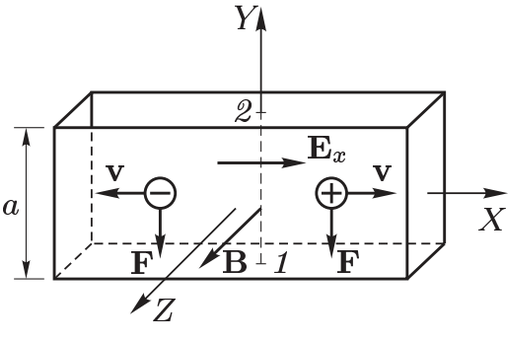


Рис. 215

Допустим, что вдоль длинной и тонкой металлической ленты течет постоянный ток с плотностью $\mathbf{j}$. Направление этого тока примем за ось X (рис. 215). Пусть перпендикулярно к плоскости ленты приложено постоянное однородное магнитное поле $\mathbf{B}$, направленное вдоль оси Z.

Если бы носителями тока были положительные частицы, то они двигались бы вместе с током вправо и сила Лоренца $(e/c)[\mathbf{vB}]$ отклоняла бы их вниз. Нижний край ленты стал бы заряжаться положительно, а верхний — отрицательно. Возникшее электрическое поле E_y препятствовало бы отклонению, вызываемому магнитным полем. Процесс накопления зарядов на нижнем и верхнем краях ленты продолжался бы до тех пор, пока не прекратилось бы течение зарядов поперек ленты. После этого поперек ленты между противоположными точками *1* и *2* установится положительная разность потенциалов $V_1 - V_2$. Если носителями тока являются отрицательные частицы, то вместе с током они будут перемещаться влево. Сила Лоренца $\mathbf{F} = (e/c)[\mathbf{vB}]$ будет отклонять такие частицы также вниз, так что теперь нижний край ленты зарядится отрицательно, а верхний — положительно. Таким образом, в этом случае поперечная разность потенциалов $V_1 - V_2$ будет отрицательной. Возникновение поперечной разности потенциалов в магнитном поле было экспериментально обнаружено Холлом в 1879 г., а само явление получило название *эффекта Холла.*

2. Рассчитаем приближенно разность потенциалов $V_1 - V_2$, пользуясь моделью свободных электронов. Между столкновениями движение частицы, переносящей ток, описывается уравнением

$$m\dot{\mathbf{v}} = e\left(\mathbf{E} - \frac{1}{c}\,[\mathbf{vB}]\right),$$

или в координатной форме

$$m\dot{v}_x = e\left(E_x + \frac{1}{c}\,v_y B\right), \quad m\dot{v}_y = e\left(E_y - \frac{1}{c}\,v_x B\right), \quad m\dot{v}_z = 0.$$

Допустим, что магнитное поле $\mathbf{B}$ слабое. Тогда поперечное электрическое поле E_y, а также члены, содержащие магнитное поле $\mathbf{B}$, в последних уравнениях можно рассматривать как малые поправки к основной силе eE_x. Для получения решения удобно применить метод последовательных приближений. В нулевом приближении оставляем только главный член eE_x и получаем

$$v_x = v_{0x} + \frac{e}{m}\,E_x t,$$

где $\mathbf{v}_0$ — скорость частицы непосредственно после последнего столкновения. Подставляя это выражение во второе уравнение, получаем уравнение первого приближения

$$m\dot{v}_y = e\left(E_y - \frac{v_{0x}}{c}\,B - \frac{eB}{mc}\,E_x t\right),$$

откуда

$$v_y = \frac{e}{m}\left(E_y t - \frac{v_{0x}}{c}\,Bt - \frac{eB}{2mc}\,E_x t^2\right).$$

Если τ — время свободного пробега, то среднее значение v_y между двумя последовательными столкновениями будет

$$\frac{1}{\tau}\int_0^\tau v_y\,dt = \frac{e}{2m}\left(E_y\tau - \frac{v_{0x}B}{c}\,\tau - \frac{eBE_x}{3mc}\overline{\tau^2}\right).$$

Это выражение надо подвергнуть вторичному усреднению по всем столкновениям или (что то же самое) по всем частицам. Если ввести предположение, которым мы пользовались уже в § 42, что после каждого столкновения все направления скорости $\mathbf{v}_0$ равновероятны, то при таком усреднении величина $v_{0x}\tau$ обратится в нуль, так что для скорости v_y, усредненной по всем частицам, получится

$$\overline{v}_y = \frac{e}{2m}\left(E_y\overline{\tau} - \frac{eB}{3mc}\,E_x\overline{\tau^2}\right).$$

В установившемся состоянии $\overline{v}_y = 0$, а потому

$$E_y = \frac{eB}{3mc}\,\frac{\overline{\tau^2}}{\overline{\tau}}\,E_x.$$

Исключим отсюда $E_x\overline{\tau^2}/\overline{\tau}$ с помощью формулы

$$j = \frac{ne^2}{2m}\,\frac{\overline{\tau^2}}{\overline{\tau}}\,E_x$$

(см. § 42). Тогда получим

$$E_y = RjB, \tag{98.2}$$

где

$$R = \frac{2}{3}\,\frac{1}{nec}. \tag{98.3}$$

Постоянная R называется *постоянной Холла.*

Учитывая, что предположения, введенные при наших вычислениях не вполне корректны, нет смысла сохранять коэффициент $2/3$. Отбросив этот коэффициент, получаем оценочную формулу [1])

$$R = \frac{1}{nec}. \tag{98.5}$$

Поперечная разность потенциалов $V_1 - V_2$, вызванная магнитным полем, определяется выражением

$$V_1 - V_2 = RjBa, \tag{98.6}$$

[1]) Более корректные вычисления, основанные на кинетическом уравнении Больцмана и классической статистике, приводят к результату

$$R = 3\pi/(8nec), \tag{98.4}$$

почти не отличающемуся от (98.3) и (98.4). Если же применять к электронам в металле статистику Ферми–Дирака, то результаты таких вычислений совпадают с формулой (98.4).

где a — ширина ленты. Зависимость такого типа и была экспериментально установлена Холлом.

Условие слабости магнитного поля, использованное в наших расчетах, сводится к требованию, чтобы поперечное электрическое поле E_y было слабым по сравнению с продольным полем E_x, т. е. $RjB \ll E_x$. Используя выражения для R и j, это условие нетрудно преобразовать к виду

$$\boldsymbol{\Omega}\overline{\tau} \ll 1, \tag{98.7}$$

где $\boldsymbol{\Omega} = |eB/(mc)|$ — циклотронная частота. Таким образом, время $\overline{\tau}$ должно быть мало по сравнению с циклотронным периодом $T = 2\pi/\boldsymbol{\Omega}$.

3. Для измерения коэффициента Холла чаще всего применяется *компенсационный метод.* К пластинке из исследуемого материала, по которой течет электрический ток, в точках *1* и *2* подводятся два контакта (рис. 216). В цепь включаются последовательно гальванометр G и компенсатор K, создающий напряжение, противоположное холловскому. Меняя это напряжение, добиваются того, чтобы ток через гальванометр обратился в нуль. После этого холловская разность потенциалов между точками *1* и *2* отсчитывается по компенсатору.

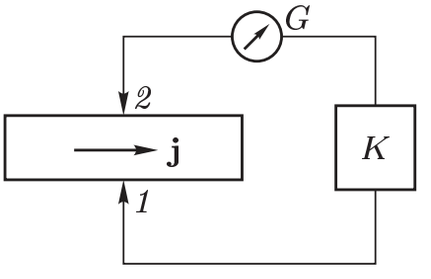


Рис. 216

В большинстве таблиц приводятся значения не величины R, а величины c^2R. (Такое значение коэффициент Холла получил бы, если бы пользоваться электромагнитной системой единиц.)

Таблица 5. Значения c^2R в гауссовой системе для некоторых металлов

Металл	$c^2R\cdot10^3$	A	ρ	z по (98.8)	Металл	$c^2R\cdot10^3$	A	ρ	z по (98.8)
Li	−1,70	7	0,53	0,83	Au	−0,736	197	19,3	1,43
Na	−2,50	23	0,97	0,98	Al	−0,343	27	2,7	3,0
Cu	−0,609	63,6	8,9	1,2	Pt	−0,230	195	21,5	4,1
Ag	−0,944	108	10,5	1,13					

Используя табл. 5, нетрудно показать, что холловская разность потенциалов $V_1 - V_2$ весьма мала. Возьмем, например, золотую ленту толщины $h = 0{,}1$ мм, по которой течет ток $I = 10$ А. Если $B = 10^4$ Гс, то формула (98.6) дает

$$V_1 - V_2 = R\frac{I}{ah}\,Ba = \frac{RIB}{h} = \frac{c^2RBI/c}{hc}.$$

Для перехода к практическим единицам умножаем эту величину на 300 и получаем

$$V_1 - V_2 = \frac{c^2RBI/c}{10^8h} = -\frac{0{,}736\cdot10^{-3}\cdot10^4\cdot1}{10^8\cdot10^{-2}} \approx 7\ \text{мкВ}.$$

Зная R, можно далее рассчитать концентрацию n электронов проводимости, а также число таких электронов z, приходящееся на один атом металла. Число атомов в единице объема определяется выражением $n_{\text{ат}} = N\rho/A$, где N — постоянная Авогадро, ρ — плотность металла, а A — атомная масса. Число z найдется делением концентрации электронов n на $n_{\text{ат}}$. С учетом формулы (98.4) это дает

$$z = \frac{A}{Rc\rho F} = \frac{A}{Rc^2\rho F/c}, \tag{98.8}$$

где F — постоянная Фарадея в гауссовой системе единиц. Если ее разделить на c, то получится то же число в электромагнитной системе: $F/c = 9650$. По формуле (98.8) и получены числа, приведенные в последнем столбце таблицы. Зная n и электрическую проводимость металла, можно вычислить по формуле (98.1) подвижность b. Для «нормальных металлов» значения b лежат в пределах около 5–50 см^2/(с·В). Столь низкие значения указывают на то, что электроны в металлах испытывают много соударений с кристаллической решеткой.

Для элементов первой группы периодической системы z приблизительно равно 1, т. е. числу валентных электронов. То же справедливо для алюминия — элемента третьей группы ($z \approx 3$). Помимо «нормальных металлов», для которых коэффициент Холла отрицателен, причем $c^2R \sim 10^{-3}$–10^{-4}, существуют металлы, обладающие *аномальными свойствами.* Так, коэффициент Холла висмута и всех металлов пятой группы периодической системы аномально велик (у висмута он примерно в 10^4 раз больше, чем у нормальных металлов) и резко меняется с температурой. У ферромагнетиков коэффициенты Холла в 10–100 раз больше, чем у нормальных металлов, и зависят от напряженности магнитного поля. Наиболее замечательно, однако, что коэффициент Холла может быть как положительным, так и отрицательным. Так, он отрицателен для щелочных металлов, Cu, Ag, Au, Mg, Ca, Hg, Al, Ga, In, Ti, Mn, Ni, Sn, Pd, Bi, Pt и положителен для Be, Zn, Cd, Tl, V, Cr, Fe, Co, Pb, Mo, Ru, As, Sb, Ta, W, Re, Ir. Дело обстоит так, как если бы носителями тока во второй из этих групп металлов были не отрицательные ($e < 0$), а положительные ($e > 0$) частицы. Такое заключение, однако, противоречит всей совокупности наших сведений о природе металлов. Это противоречие долгое время: являлось одной из основных трудностей в электронной теории металлов. Оно было вполне удовлетворительно разрешено квантовой теорией металлов (см. § 100).

§ 99. Применения статистики Ферми–Дирака к электронам в металлах

1. В § 42 мы говорили, что теория Друде, в которой электроны в металлах трактуются как частицы идеального газа, натолкнулась на трудности в вопросе о теплоемкости электронного газа. Эти трудности были успешно преодолены Зоммерфельдом в 1928 г. Зоммер-

фельд сохранил модель свободных электронов, но применил к ней вместо классической статистики квантовую статистику Ферми–Дирака (см. т. II, § 82).

В отличие от классической квантовая статистика принимает во внимание, что частица или система частиц, совершающая финитное движение в заданном силовом поле, может находиться лишь во вполне определенных *квантовых состояниях*, которым соответствуют определенные значения энергии. Эти значения называются *энергетическими уровнями* системы. При финитных движениях энергетические уровни *дискретны*, т. е. отделены один от другого конечными интервалами. В модели идеального электронного газа финитность и квантование движения обусловлены большими отталкивательными силами, действующими на электроны вблизи поверхности тела. Эти силы играют роль *непроницаемых стенок*, не позволяющих электронному газу выходить за пределы занимаемого им объема. При отсутствии таких «стенок» движение электронов было бы инфинитным и неквантованным.

В случае статистики Ферми–Дирака, которой подчиняется электронный газ, принимается во внимание также *принцип Паули*, согласно которому в каждом квантовом состоянии может находиться не более одного электрона. Соблюдение принципа Паули означает, что даже между «свободными» электронами существует какое-то взаимодействие, так что электроны нельзя считать вполне независимыми частицами. Однако это взаимодействие не является силовым. Оно — сугубо квантовый эффект, чуждый классическим представлениям.

С учетом всего изложенного выше для среднего числа электронов f, приходящихся на одно квантовое состояние, получается выражение

$$f = \frac{1}{\exp\dfrac{\varepsilon - \mu}{kT} + 1}, \tag{99.1}$$

где ε — кинетическая энергия электрона в рассматриваемом квантовом

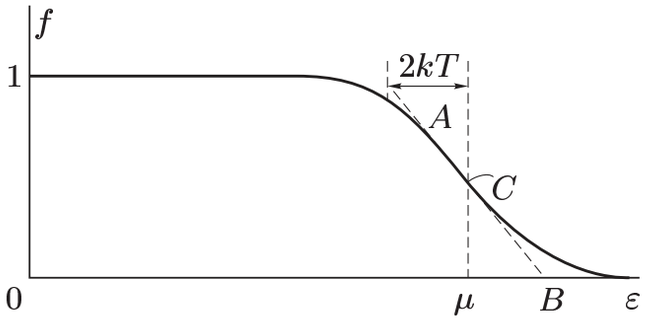


Рис. 217

состоянии, а μ — постоянная (химический потенциал электрона). Вид функции f представлен на рис. 217 сплошной линией.

2. При конкретных применениях формулы (99.1) необходимо знать число квантовых состояний, в которых может находиться электрон. Ответ на этот вопрос дает квантовая механика. Приведем его без

всяких обоснований. Введем *фазовое пространство* электрона, т. е. шестимерное пространство, по координатным осям которого откладываются прямоугольные пространственные координаты x, y, z и соответствующие им проекции импульса электрона p_x, p_y, p_z. Импульс электрона связан с его кинетической энергией ε соотношением $\varepsilon = p^2/2m$. Пусть V — объем, занимаемый электронным газом, а импульс электрона может принимать все значения от нуля до p. Объем соответствующего фазового пространства определяется выражением $\mathbf{\Omega} = (4\pi/3)p^3V$. Введем, далее, фазовые ячейки с фазовым объемом h^3 каждая. Число таких ячеек в рассматриваемой области фазового пространства будет $\mathbf{\Omega}/h^3$. Если бы электрон не обладал спином, то, как показывает квантовая механика, тому же выражению было бы равно и искомое число квантовых состояний электрона. Из-за наличия спина это выражение надо удвоить, так как в одной фазовой ячейке могут помещаться два электрона с противоположно направленными спинами. В дальнейшем будем предполагать, что $V = 1$. Тогда число квантовых состояний с импульсами от 0 до p представится выражением

$$Z = 2\frac{\mathbf{\Omega}}{h^3} = \frac{8\pi}{3h^3}\,p^3. \tag{99.2}$$

Число квантовых состояний с импульсами между p и $p + dp$ будет

$$dZ = \frac{8\pi p^2}{h^3}\,dp, \tag{99.3}$$

а с кинетической энергией между ε и $\varepsilon + d\varepsilon$

$$dZ = \frac{4\pi}{h^3}\,(2m)^{3/2}\varepsilon^{1/2}\,d\varepsilon. \tag{99.4}$$

3. Вычисления с использованием формулы (99.1) очень громоздки, хотя в принципе и просты. Однако основной интерес представляет случай *вырожденного* электронного газа, т. е. газа вблизи абсолютного нуля температур. Рассмотрим сначала полное вырождение, когда $T = 0$. В этом случае функция (99.1) переходит в

$$f = \begin{cases} 1, & \text{если } \varepsilon < \mu, \\ 1/2, & \text{если } \varepsilon = \mu, \\ 0, & \text{если } \varepsilon > \mu. \end{cases}$$

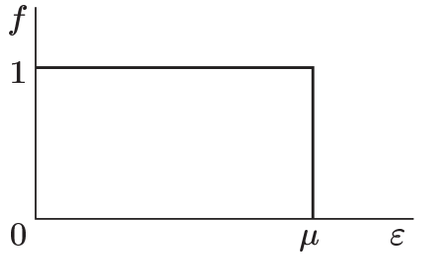


Рис. 218

Она представлена на рис. 218. Наибольшая энергия, которую может принимать электрон, $\varepsilon_{\text{макс}} = \mu$, а наибольший импульс — $p_{\text{макс}} = \sqrt{2m\mu}$. Величина μ называется *энергией* или *границей Ферми*. Число электронов в единице объема

$$n = Z_{\text{макс}} = \frac{8\pi}{3h^3}\,p^3_{\text{макс}} = \frac{8\pi}{3h^3}\,(2m\mu)^{3/2}, \tag{99.5}$$

откуда

$$\mu = \varepsilon_{\text{макс}} = \frac{h^2}{8m}\left(\frac{3n}{\pi}\right)^{2/3}. \tag{99.6}$$

Полная энергия газа

$$\mathscr{E} = \int \varepsilon\, dZ = \frac{3n}{2}\,\mu^{-3/2}\int_0^{\mu} \varepsilon^{3/2}\, d\varepsilon = \frac{3}{5}\,\mu n,$$

а средняя энергия, приходящаяся на один электрон,

$$\overline{\varepsilon} = \frac{3}{5}\,\mu = \frac{3h^2}{4m}\left(\frac{3n}{\pi}\right)^{2/3}. \tag{99.7}$$

Величина

$$T_g = \frac{\overline{\varepsilon}}{k} = \frac{3h^2}{4mk}\left(\frac{3n}{\pi}\right)^{2/3} \tag{99.8}$$

имеет размерность температуры и называется *температурой вырождения электронного газа*. Газ считается вырожденным, если его температура $T < T_g$. При сильном вырождении $T \ll T_g$.

Давление электронного газа

$$\mathscr{P} = \frac{2}{3}\,n\overline{\varepsilon} = \frac{h^2 n}{20m}\left(\frac{3n}{\pi}\right)^{2/3}. \tag{99.9}$$

Полагая для меди (см. табл. 5 на с. 417) $n = 10^{22}$ см$^{-3}$, по этим формулам находим $T_g \approx 5 \cdot 10^4$ К, $\mathscr{P} \approx 5 \cdot 10^4$ атм. Для всех металлов температура вырождения T_g много выше температуры плавления. Полученные цифры показывают, насколько мала кинетическая энергия теплового движения электронов металла по сравнению с «нулевой энергией», которой обладают электроны при абсолютном нуле температур. Практически электроны проводимости всех металлов находятся в состоянии сильного вырождения.

4. От полностью вырожденного перейдем к рассмотрению *сильно вырожденного* электронного газа. В этом случае почти все квантовые состояния ниже границы Ферми, для которых $\mu - \varepsilon \gg kT$, заполнены электронами, как и при абсолютном нуле температур. Исключение составляют только квантовые состояния вблизи самой границы Ферми. Здесь «прямоугольное распределение» Ферми, соответствующее абсолютному нулю температур, возмущается тепловым движением электронов, для которых $\varepsilon - \mu \sim kT$. Такие электроны непрерывно пересекают границу Ферми в прямом и обратном направлениях, причем в состоянии статистического равновесия средние числа прямых и обратных процессов одинаковы. Только такие электроны участвуют в тепловом движении. Ими, в частности, обусловлена теплоемкость электронного газа в металлах.

Выразим прежде всего концентрацию и кинетическую энергию электронного газа через энергию Ферми μ. Для этого надо вычислить

два интеграла:

$$n = \int f\, dZ \text{ и } \mathscr{E} = \int \varepsilon f\, dZ,$$

в которых интегрирование производится по всем значениям импульса p или энергии ε. Поскольку точное вычисление этих интегралов очень громоздко и требует специальных искусственных приемов, упростив задачу. Проведя в точке C (см. рис. 217) касательную AB к кривой Ферми, заменим этой касательной соответствующий участок кривой распределения Ферми. Иначе говоря, распределение Ферми аппроксимируем функцией

$$f = \begin{cases} 1, & \text{если } \varepsilon < \mu - 2kT, \\ \dfrac{1}{2} + \dfrac{\mu - \varepsilon}{4kT}, & \text{если } \mu - 2kT < \varepsilon < \mu + 2kT, \\ 0, & \text{если } \mu + 2kT < \varepsilon. \end{cases}$$

Тогда

$$n = \frac{3}{2}\, Z_{\text{макс}} \mu^{-3/2} \int\limits_0^{\mu - 2kT} \varepsilon^{1/2}\, d\varepsilon + \frac{3}{2}\, Z_{\text{макс}} \mu^{-3/2} \int\limits_{\mu - 2kT}^{\mu + 2kT} \left(\frac{1}{2} + \frac{\mu - \varepsilon}{4kT} \right) \varepsilon^{1/2}\, d\varepsilon,$$

где для сокращения введено выражение $Z_{\text{макс}}$, определяемое формулой (99.5). Взяв первый интеграл и учитывая, что $kT \ll \mu$, преобразуем результат с помощью разложения бинома. Обрывая разложение на членах второй степени по kT/μ, получим

$$\int \varepsilon^{1/2}\, d\varepsilon = \frac{2}{3}\, \mu^{3/2} \left(1 - \frac{2kT}{\mu} \right)^{3/2} \approx \frac{3}{2}\, \mu^{3/2} \left(1 - \frac{3kT}{\mu} + \frac{3}{2}\, \frac{k^2 T^2}{\mu^2} \right).$$

Во втором интеграле произведем замену переменной: $(\varepsilon - \mu)/(2kT) = x$. Тогда

$$\varepsilon^{1/2} = \mu^{1/2} \left(1 + \frac{2kTx}{\mu} \right)^{1/2} = \mu^{1/2} \left(1 + \frac{kTx}{\mu} \right),$$

$$\int \left(\frac{1}{2} + \frac{\mu - \varepsilon}{4kT} \right) \varepsilon^{1/2}\, d\varepsilon = kT \mu^{1/2} \int\limits_{-1}^{+1} (1 - x) \left(1 + \frac{kT}{\mu}\, x \right) dx =$$

$$= 2kT \mu^{1/2} \left(1 - \frac{1}{3}\, \frac{kT}{\mu} \right).$$

В результате получится

$$n = Z_{\text{макс}} \left(1 + \frac{1}{2}\, \frac{k^2 T^2}{\mu^2} \right). \tag{99.10}$$

Совершенно так же находим

$$\mathscr{E} = \frac{3}{5}\,\mu Z_{\text{макс}}\left(1+\frac{5}{2}\,\frac{k^2T^2}{\mu^2}\right). \tag{99.11}$$

Средняя энергия, приходящаяся на один электрон,

$$\overline{\varepsilon} = \frac{\mathscr{E}}{n} = \frac{3}{5}\,\mu\left(1+2\frac{k^2T^2}{\mu^2}\right), \tag{99.12}$$

а теплоемкость

$$c_v = \frac{\partial\overline{\varepsilon}}{\partial T} = \frac{12}{5}\,\frac{k^2}{\mu}\,T. \tag{99.13}$$

Поскольку для всех металлов вплоть до температуры плавления $kT/\mu \ll 1$, для теплоемкости c_v получается величина, малая по сравнению с k. Тем самым устраняются трудности, на которые натолкнулась классическая теория в вопросе о теплоемкости электронов в металлах.

5. Применим теперь теорию Зоммерфельда к выводу закона Видемана и Франца. Для этого воспользуемся формулой (42.24) и учтем, что перенос теплоты в металлах осуществляется только электронами вблизи границы Ферми, средняя кинетическая энергия которых $m\overline{v^2}/2 = \mu$. Подставив (99.13) в (42.26), получим

$$\frac{\chi}{\lambda} = \frac{16}{5}\left(\frac{k}{e}\right)^2 T. \tag{99.14}$$

Числовой коэффициент $16/5 = 3{,}2$ практически не отличается от коэффициента $\pi^2/3 = 3{,}29$, полученного Зоммерфельдом при строгих расчетах, а также от коэффициента 3, полученного в элементарной классической теории Друде. Однако классическая теория, приводя к практически правильному конечному результату (42.27), давала этому результату совершенно неправильное объяснение. По этой теории пропорциональность между χ/λ и T объяснялась тем, что средняя кинетическая энергия электрона равна $(3/2)kT$, т. е. пропорциональна абсолютной температуре. На самом деле закон Видемана и Франца (99.14) объясняется тем, что абсолютной температуре пропорциональна не средняя энергия, а *теплоемкость электрона.* Классическая теория допускала ошибку, завышая теплоемкость электронного газа. Однако эта ошибка случайно компенсировалась другой ошибкой. Скорость электронов, переносящих теплоту, определяется их кинетической энергией *вблизи границы Ферми*, тогда как классическая теория считала, что эта скорость порядка классической средней скорости теплового движения $\sqrt{kT/m}$. Тем самым скорость электронов, переносящих теплоту, сильно занижалась, а конечный результат (42.27) получался правильным.

6. Несмотря на успехи, теория Зоммерфельда обладает и существенными недостатками. В этой теории взаимодействие электронов с ионами кристаллической решетки, как и в классической теории, учитывается посредством *формально введенной* длины свободного пробега

электрона $\overline{l}$. Каково происхождение $\overline{l}$, как она зависит от температуры T — на эти вопросы теория Зоммерфельда не дает никакого ответа. Классическая теория (без достаточных на то оснований) считала, что процесс столкновения электронов с ионами решетки аналогичен столкновению твердых шариков. В такой модели $\overline{l}$ не зависит от температуры. Поэтому по формуле (42.25) следовало бы ожидать, что $\lambda \sim 1/T$. Если и в теории Зоммерфельда принять, что $\overline{l}$ не зависит от T, то λ также не должна была бы зависеть от T. На самом деле оба эти результата не согласуются с опытом. Для большинства чистых металлов при не очень низких температурах λ приблизительно обратно пропорциональна T: $\lambda \sim 1/T$.

Для решения вопроса надо рассмотреть на основе уравнений квантовой механики движение электрона *с учетом его взаимодействий с кристаллической решеткой*. Качественно основной результат можно понять без вычислений. Согласно квантовой механике движение электрона аналогично распространению какой-то волны в пространстве. В идеально прозрачной и однородной среде, например, световая волна распространяется без всякого ослабления и рассеяния в стороны. Так же вела бы себя и «электронная волна», если бы кристаллическая решетка, в которой она распространяется, была идеально правильной. В этом случае металл не оказывал бы электрическому току никакого сопротивления. На самом деле в реальной решетке *всегда есть примеси и возникают тепловые флуктуации*, нарушающие ее идеальную структуру. Благодаря этому электронная волна проходит через кристаллическую решетку не только в прямом направлении, но и рассеивается в стороны подобно тому, как рассеивается световой луч при распространении в мутной среде. Ослабление интенсивности волны происходит по экспоненциальному закону, т. е. $\sim \exp(-x/\overline{l})$, где $\overline{l}$ — постоянная, которая и играет роль длины свободного пробега электрона (см. т. II, § 88). Для полной длины свободного пробега, электрона $\overline{l}$ можно написать

$$\frac{1}{\overline{l}} = \frac{1}{\overline{l}_{\text{фл}}} + \frac{1}{\overline{l}_{\text{пр}}}, \tag{99.15}$$

где первый член справа обусловлен тепловыми флуктуациями, а второй — примесями. Величина $\overline{l}_{\text{пр}}$ от температуры не зависит, а для $\overline{l}_{\text{фл}}$ при обычных температурах расчет дает $\overline{l}_{\text{фл}} \sim T^{-1}$. В соответствии с этим и с формулой (42.25) получаем для удельного сопротивления

$$\rho \equiv \frac{1}{\lambda} = \alpha T + \rho_0, \tag{99.16}$$

где α и ρ_0 — постоянные. Для чистых металлов $\rho_0 = 0$, $\rho \sim T$. При низких температурах (которые, однако, выше критической температуры перехода в сверхпроводящее состояние) теория для чистых металлов дает $\rho \sim T^5$. Эти выводы теории согласуются с опытом.

7. В заключение рассмотрим парамагнетизм газа свободных электронов (см. § 77, п. 9), теория которого была дана Паули. Будем предполагать, что электронный газ полностью вырожден. При помещении в магнитное поле спин электрона может ориентироваться либо по, либо против поля. Этим ориентациям соответствуют два значения полной энергии электрона: $\varepsilon - \mathfrak{M}H$ и $\varepsilon + \mathfrak{M}H$, где $\mathfrak{M} = eh/(4\pi mc)$ — магнетон Бора, а ε по-прежнему означает кинетическую энергию электрона. Первая ориентация является более предпочтительной, поскольку ей соответствует меньшее значение полной энергии. Весь электронный газ можно рассматривать как два независимых газа, отличающихся друг от друга ориентациями спинов электронов. В распределении Ферми вместо кинетической энергии ε надо взять полную энергию $\varepsilon \mp \mathfrak{M}H$. От этого аргумент $\varepsilon - \mu$ заменится на $\varepsilon - (\mu \mp \mathfrak{M}H)$. Отсюда видно, что влияние магнитного поля сводится к смещению границы Ферми. Для газа с параллельной ориентацией спинов граница Ферми $\mu_1 = \mu + \mathfrak{M}H$, а с антипараллельной $\mu_2 = \mu - \mathfrak{M}H$. Далее, при рассмотрении каждого газа число состояний в формуле (99.2) надо уменьшить вдвое, т. е. вместо Z взять $Z/2$, поскольку в каждом газе осуществляется лишь одна из ориентации спина. С учетом всего этого для избытка концентрации одного газа над другим можно написать

$$n_1 - n_2 = \frac{Z}{2}(\mu + \mathfrak{M}H) - \frac{Z}{2}(\mu - \mathfrak{M}H),$$

где индекс 1 относится к параллельной, а индекс 2 — к антипараллельной ориентациям спинов. Предполагая, что $\mathfrak{M}H \ll \mu$, можем написать

$$n_1 - n_2 = \mathfrak{M}H\frac{\partial Z}{\partial \mu}.$$

Магнитный момент единицы объема будет

$$I = \mathfrak{M}(n_1 - n_2) = \mathfrak{M}^2\frac{\partial Z}{\partial \mu}\,H,$$

а магнитная восприимчивость электронного газа

$$\varkappa = \mathfrak{M}^2\frac{\partial Z}{\partial \mu} = \mathfrak{M}^2\frac{\partial n}{\partial \mu}.$$

Согласно формуле (99.5) $n \sim \mu^{3/2}$, и потому $dn/n = (3/2)\,d\mu/\mu$. С учетом этого

$$\varkappa = \frac{3}{2}\frac{\mathfrak{M}^2 n}{\mu}. \tag{99.17}$$

Подставив сюда значение для $\mathfrak{M}$ и воспользовавшись формулой (99.6), получим

$$\varkappa = \left(\frac{3n}{\pi^4}\right)^{1/3}\frac{e^2}{4mc^2}. \tag{99.18}$$

Формула (99.18) получена в предположении, что $T = 0$. Влияние температуры на магнитную восприимчивость электронного газа в состоянии сильного вырождения может сказаться лишь в виде малого

поправочного члена порядка $(kT/\mu)^2$ к основному эффекту, выражаемому формулой (99.18). Поэтому можно сказать, что парамагнетизм электронного газа не зависит от температуры. При выводе не учтено влияние магнитного поля на движение электронов. Если это учесть методами квантовой механики, то, как показал Л.Д. Ландау, выражение (99.18) надо уменьшить на одну треть. В таком виде формула удовлетворительно согласуется с опытом для щелочных металлов натрия и калия.

§ 100. Металлы и полупроводники

1. Удельная проводимость λ металлов при комнатной температуре меняется примерно в пределах от $6 \cdot 10^3$ до $6 \cdot 10^5$ $\text{Ом}^{-1} \cdot \text{см}^{-1}$. Твердые вещества с удельной проводимостью примерно от 10^4 до 10^{-10} $\text{Ом}^{-1} \cdot \text{см}^{-1}$ принято относить к классу так называемых *полупроводников*, а вещества с еще меньшей λ (приблизительно от 10^{-10} до 10^{-20} $\text{Ом}^{-1} \cdot \text{см}^{-1}$) — к классу *диэлектриков*. Носителями тока в полупроводниках и диэлектриках могут быть как электроны, так и ионы. В последнем случае говорят о *твердых электролитах*, так как прохождение электрического тока в этом случае сопровождается электролизом. Электролитический характер проводимости был установлен во многих солях (безводные $NaNO_3$, KNO_3, $AgNO_3$, LiH, NaCl, CsCl и пр.) и их расплавах. Однако в настоящее время полупроводниками принято называть только вещества, у которых носителями тока являются *электроны*. Только такие полупроводники нашли широкое применение в технике.

К полупроводникам относятся многие химические элементы (углерод в виде графита, бор, кремний, германий, фосфор, мышьяк, селен, серое олово, теллур, йод и др.), громадное количество сплавов и химических соединений. Почти все неорганические вещества окружающего нас мира — полупроводники. С чисто эмпирической точки зрения качественное различие между металлами и полупроводниками проявляется в характере зависимости удельной проводимости от температуры. С понижением температуры проводимость металлов возрастает и для чистых металлов стремится к бесконечности при приближении температуры к абсолютному нулю. У полупроводников, напротив, с понижением температуры проводимость убывает, вблизи абсолютного нуля полупроводник фактически становится изолятором. При высоких температурах проводимость полупроводников приближается к проводимости металлов. Такой ход проводимости объясняется тем, что концентрация носителей тока (электронов проводимости) в металлах практически не зависит от температуры, а в полупроводниках носители тока сами возникают в результате теплового движения.

2. Почему одни тела являются проводниками, а другие полупроводниками и изоляторами — на этот вопрос нельзя ответить в рамках модели газа свободных электронов. Необходимо учитывать взаимодействие

атомов между собой и с электронами. Представим себе, что кристаллическая решетка металла или полупроводника образуется в результате сближения изолированных атомов. Наружные, так называемые *валентные*, электроны атомов металла сравнительно слабо связаны с атомными ядрами, а такие же электроны полупроводников — значительно сильнее. При сближении атомов последние приходят во взаимодействие друг с другом. В результате этого валентные электроны отрываются от атомов металла и становятся *свободными электронами*, которые могут перемещаться по всему металлу, *коллективизируются*, по образному выражению Я.И. Френкеля. В полупроводниках, благодаря значительно более сильной связи валентных электронов с ядрами атомов, положение иное. Чтобы оторвать электрон от атома и превратить его в электрон проводимости, требуется сообщить ему некоторую энергию, называемую *энергией ионизации*. Такая энергия поставляется тепловыми колебаниями атомов решетки. Она может быть сообщена и иначе, например облучением полупроводника короткими электромагнитными волнами, потоком быстрых частиц, воздействием сильного электрического поля и т. д. Для разных полупроводников энергия ионизации валентного электрона лежит в пределах от 0,1 до 2 эВ, т. е. заметно выше средней кинетической энергии теплового движения атома $(3/2)kT \approx 0{,}04$ эВ. Несмотря на это, тепловое движение вызывает ионизацию атомов, так как из-за его хаотичности мгновенная кинетическая энергия атома может в несколько раз превышать ее среднее значение. Число атомов с энергией, равной или превышающей энергию ионизации, относительно очень мало. Поэтому в полупроводниках мала и концентрация свободных электронов, образующихся в результате ионизации. Однако с повышением температуры эта концентрация и связанная с ней электрическая проводимость повышаются. Процессы *ионизации*, конечно, сопровождаются обратными процессами *рекомбинации*, в результате которых свободные электроны вновь захватываются атомами. В состоянии равновесия средние числа актов ионизации и рекомбинации одинаковы, и устанавливается вполне определенная *равновесная* концентрация свободных электронов, зависящая от температуры полупроводника.

3. Для более детального анализа процессов, происходящих в металлах и полупроводниках, необходимо исследовать структуру энергетических уровней, которые могут занимать валентные электроны в этих телах. Энергетические уровни какого-либо валентного электрона в одном изолированном атоме представлены на схематическом рис. 219 *а*. Для простоты будем считать их *простыми*, т. е. *невырожденными*. Наинизший уровень, или уровень с наименьшей энергией e_1, называется *основным*, или *невозбужденным*. Все остальные уровни называются *возбужденными*. Рассмотрим теперь N тождественных атомов, удаленных друг от друга настолько далеко, что их взаимодействием можно полностью пренебречь. Энергетические уровни того же валентного электрона системы N невзаимодействующих атомов

получатся, если рис. 219 *а* повторить N раз — столько, сколько содержится атомов в системе (рис. 219 *б*). Теперь каждый простой уровень превращается в уровень кратности N. Будем сближать атомы друг с другом, чтобы они образовали кристаллическую решетку. Тогда из-за взаимодействия между атомами каждый кратный энергетический уровень расщепится на N простых уровней (рис. 220). Совокупность

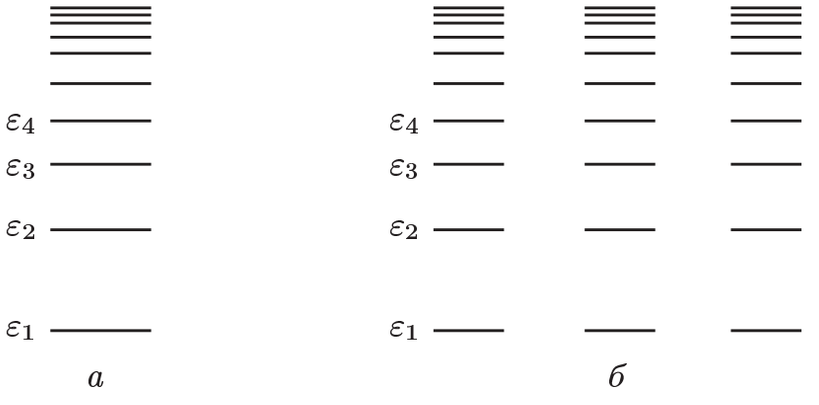


Рис. 219

энергетических уровней, на которые расщепляется кратный уровень, называется *энергетической зоной* или просто *зоной* кристалла. Зона, возникающая в результате расщепления N-кратного вырожденного

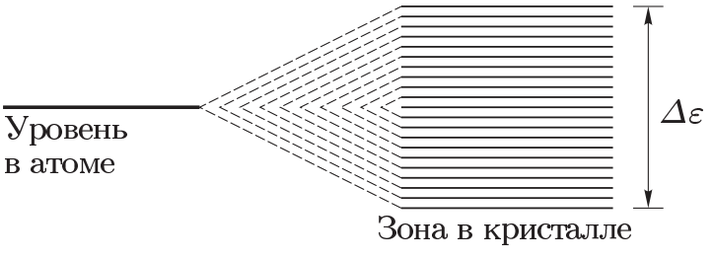


Рис. 220

основного уровня, называется *основной зоной*, а все остальные зоны — *зонами возбужденных уровней*. Ввиду того что N очень велико, расстояния между уровнями одной и той же зоны крайне малы, так что требуется ничтожная энергия, чтобы перевести электрон в пределах зоны с одного энергетического уровня на соседний. В этом смысле энергетические уровни каждой зоны ведут себя практически так, как если бы они были *непрерывны*. Однако соседние энергетические зоны, вообще говоря, разделены конечными интервалами энергии. Эти интервалы называются *запрещенными зонами*, так как энергия электрона не может принимать значения, лежащие в пределах таких интервалов. В противоположность запрещенным, зоны с дозволенными значениями энергии называются *разрешенными*. Энергетические зоны, разумеется, нельзя путать с пространственными зонами, т. е. областями пространства, в которых может находиться электрон.

Для правильного понимания изложенных результатов необходимо дополнить их следующим замечанием. Последовательный учет взаимодействия электронов между собой и с ионами кристаллической решетки требует решения «*задачи многих тел*» для систем с колоссальным количеством частиц. (В одном кубическом сантиметре меди содержится порядка 10^{22} электронов проводимости!) В строгой постановке такая задача неразрешима. Для возможности решения ее сильно упрощают и сводят к «*задаче одного тела*». Теория твердого тела, характеризующаяся таким упрощенным подходом, называется *зонной теорией.* В зонной теории считается, что электрон движется в постоянном электрическом поле, создаваемом ионами и всеми остальными электронами. Ионы, ввиду их относительно больших масс, считаются неподвижными. Электроны учитываются суммарно. Они как бы заменяются заряженной отрицательно электронной жидкостью, равномерно заполняющей пространство между ионами. Роль электронов в этой модели сводится только к тому, чтобы суммарно скомпенсировать положительные заряды ионов решетки, сделав последнюю *электрически нейтральной.* Электрическое поле в такой модели периодично в пространстве, причем периодами являются соответствующие *пространственные периоды* кристаллической решетки. В результате мы приходим к задаче о движении *одного* электрона в постоянном периодическом электрическом поле. Решение такой задачи в квантовой механике приводит к той же зонной структуре энергетических уровней, которая описана выше. Таким образом, такая структура характеризует возможные энергетические состояния *одного электрона*, находящегося в постоянном периодическом электрическом поле кристаллической решетки.

4. Внесем тело в постоянное электрическое поле, напряженность которого очень мала по сравнению с напряженностью внутриатомных и внутримолекулярных полей. Такое поле не меняет общий характер зонной структуры. Число энергетических уровней в зоне остается неизменным, но сами уровни смещаются, так как к энергии взаимодействия электрона с решеткой добавляется потенциальная энергия электрона во внешнем электрическом поле. Электроны, находящиеся в зоне, будут вести себя по-разному в зависимости от того, заполняют ли они все уровни зоны или некоторые уровни остаются свободными. При решении этого вопроса будем предполагать сначала, что абсолютная температура тела равна нулю.

Рассмотрим случай, когда *все* энергетические уровни зоны *заполнены электронами.* Если это имеет место в отсутствие электрического поля, то то же самое будет и после наложения слабого поля. Движение электрона в квантовой механике следует рассматривать как *процесс перехода* его *из одного возможного квантового состояния в другое.* Для возможности такого перехода необходимо, чтобы конечное квантовое состояние было *свободно*, т. е. не занято электроном. Но по условию все квантовые состояния зоны заполнены электронами. В этом случае между различными квантовыми состояниями зоны невозможны

никакие квантовые переходы, а потому электроны зоны *не могут* быть носителями электрического тока.

Рассмотрим теперь случай, когда *только часть* возможных квантовых уровней зоны *заполнена электронами*, а остальные уровни *свободны.* Если нет теплового движения или других источников, поставляющих энергию электронам, то заполненными окажутся все уровни с *самыми низкими* значениями энергии. Более высокие уровни окажутся свободными. То же будет и после наложения постоянного электрического поля **E**. Однако при этом произойдет *смещение* энергетических уровней. Уровни, бывшие наиболее низкими в отсутствие электрического поля, могут перестать быть таковыми после наложения поля. Получится другая система наиболее низких энергетических уровней. Начнутся квантовые переходы на эти уровни с прежних заполненных уровней. Они будут сопровождаться пространственными перемещениями электронов в направлении против электрического поля. Если электроны не могут уходить из тела, то этот процесс быстро прекратится, так как возникшие пространственные заряды создадут поле, которое внутри тела уничтожит внешнее приложенное поле. Если же смещающиеся электроны непрерывно отводить от тела (это происходит в замкнутой электрической цепи), то квантовые переходы электронов будут продолжаться непрерывно, пока в электрической цепи действует генератор, создающий в теле электрическое поле. Такие квантовые переходы и приводят к возникновению электрического тока в цепи. Таким образом, *для возможности электрического тока необходимо, чтобы энергетическая зона была заполнена электронами не целиком, а частично.*

5. В металлах основная энергетическая зона валентных электронов может быть отделена запрещенной зоной конечной ширины от вышележащей зоны возбужденных уровней (рис. 221 *а*). Но может быть и такой случай, когда ширина запрещенной зоны равна нулю, т. е. основная зона примыкает или даже перекрывается с ближайшей зоной возбужденных уровней (рис. 221 *б*). Этот случай сводится к предыдущему, так как обе зоны можно объединить в одну, рассматривая последнюю как основную зону. *В металлах основная зона всегда заполнена только частично.* Благодаря этому металлы и являются проводниками электрического тока.

В полупроводниках основная зона отделена от зоны возбужденных уровней конечным интервалом энергии $\Delta\varepsilon$ (рис. 221 *в*). Основную зону полупроводника принято называть *валентной,* а зону возбужденных уровней — *зоной проводимости. При абсолютном нуле температур валентная зона полностью заполнена электронами, а зона проводимости полностью свободна.* Поэтому при абсолютном нуле температур полупроводники не проводят электрического тока, т. е. являются *изоляторами. Изоляторы отличаются от полупроводников только большими значениями ширины запрещенной зоны* $\Delta\varepsilon$. Условно к изоляторам относят те полупроводники, для которых $\Delta\varepsilon$ превосходит примерно

2 эВ. Никакого качественного различия между полупроводниками и изоляторами нет. Различие — чисто количественное.

При повышении температуры электроны начинают обмениваться энергией с ионами кристаллической решетки. Благодаря этому электрон может получить добавочную кинетическую энергию порядка kT. Этой энергии может оказаться достаточно, чтобы некоторые электроны перевести из валентной зоны в зону проводимости. Такие электроны, перейдя в зону проводимости, начинают проводить электрический ток. Но проводимость возникает и по другой причине. В валентной зоне освобождаются квантовые состояния, не занятые электронами. Такие квантовые состояния получили весьма неудачное название *дырок*. Дырки также являются носителями электрического тока.

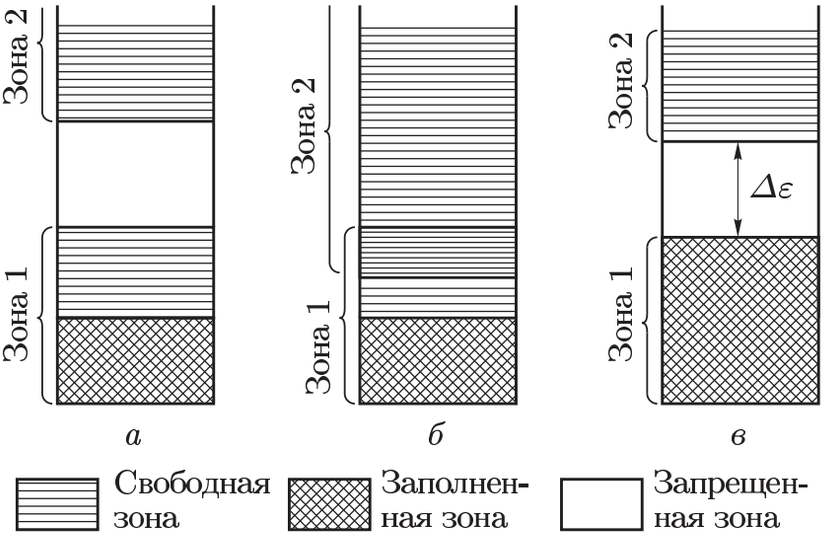


Рис. 221

Действительно, при наличии дырок электроны могут рекомбинировать с ними, т. е. совершать квантовые переходы из каких-то квантовых состояний в незаполненные состояния, т. е. дырки. Прежние заполненные состояния при этом освобождаются, т. е. превращаются в дырки. Последние в свою очередь могут рекомбинировать с другими электронами с образованием новых дырок и т. д. В результате этих процессов установится вполне определенная *равновесная концентрация* дырок, которая при отсутствии электрического поля будет одна и та же по всему объему полупроводника. При наличии электрического поля однородное распределение дырок в полупроводнике нарушится. Всякий квантовый переход электрона, поскольку его заряд отрицательный, сопровождающийся его перемещением против поля, уменьшает, а переход, связанный с перемещением в направлении поля, увеличивает потенциальную энергию системы. Поэтому первые переходы будут преобладать над вторыми. Это значит, что через полупроводник потечет электрический ток в направлении приложенного электрического поля $\mathbf{E}$. В незамкнутом полупроводнике ток будет продолжаться до тех

пор, пока возникшее электрическое поле не компенсирует внешнее приложенное поле **E**. Однако если непрерывно отводить перемещающиеся электроны (как это делается в замкнутой электрической цепи), то ток будет течь непрерывно. Конечный результат явления получается таким же, как если бы носителями тока были не электроны, а положительно заряженные частицы — дырки. Поэтому различают *электронную* и *дырочную проводимость* полупроводников.

Конечно, *истинными носителями тока* в металлах и полупроводниках являются *реальные электроны*, а не формально введенные дырки. Никаких дырок, как реально существующих положительно заряженных частиц, в действительности нет. Однако представление о дырках оказалось весьма плодотворным по следующим соображениям. Классическими законами для движения электронов можно пользоваться в тех случаях, когда концентрация этих частиц в соответствующей энергетической зоне мала. Этому условию удовлетворяют электроны в зоне проводимости полупроводника. Но в валентной зоне мы имеем дело с противоположным случаем. Там почти все состояния заполнены электронами, зато мала концентрация дырок. Здесь классические уравнения к движению электронов неприменимы. Электроны надо рассматривать квантовомеханически. Такое рассмотрение, благодаря *малой концентрации дырок*, приводит к поразительно простому результату. Оказывается, что в *электрическом поле дырки движутся так, как двигались бы при классическом рассмотрении положительно заряженные частицы*, обладающие определенной массой. Столь простой результат и оправдывает представление о дырках. Заметим в связи с этим, что благодаря малости концентрации *к электронам в зоне проводимости и к дыркам в валентной зоне применима классическая статистика Больцмана.*

6. Электрическая проводимость полупроводников, как электронная, так и дырочная, о которой говорилось выше, не связана с наличием примесей в полупроводнике. Поэтому ее называют *собственной электропроводностью* в отличие от *примесной* электрической проводимости, обусловленной наличием примесей атомов других химических элементов. Уже ничтожные количества примесей чрезвычайно сильно увеличивают электропроводность полупроводников. Так, добавление к чистому кристаллу кремния фосфора в количестве всего 0,001 атомного процента увеличивает электрическую проводимость этого кристалла более чем в сто тысяч раз. В металлах, как мы видели, наблюдается обратное: примеси всегда уменьшают электрическую проводимость металлов.

Такое поведение полупроводников объясняется тем, что при наличии примесей появляются *добавочные энергетические уровни*, располагающиеся в запрещенной зоне полупроводника. На схематическом рис. 222 *а* изображены энергетические зоны чистого полупроводника. Допустим, что добавочные уровни в запрещенной зоне появились вблизи нижнего края зоны проводимости (рис. 222 *б*). С этих уровней электроны будут переходить в зону проводимости. Если интервал

энергии $\Delta\varepsilon_1$, отделяющий добавочные уровни от зоны проводимости, мал по сравнению с шириной запрещенной зоны $\Delta\varepsilon$, то количество электронов в зоне проводимости, а с ним и электрическая проводимость полупроводника могут увеличиться на несколько порядков. Примеси такого типа, поставляющие электроны в зону проводимости, называются *донорами*, или *донорными примесями.* Добавочные энергетические уровни, которые они создают в запрещенной зоне, называются также *донорными уровнями.*

Примером донорной примеси могут служить атомы мышьяка, вводимые в кристаллическую решетку кремния. Кремний — четырех-, а мышьяк — пятивалентный атом. Это значит, что наружная оболочка

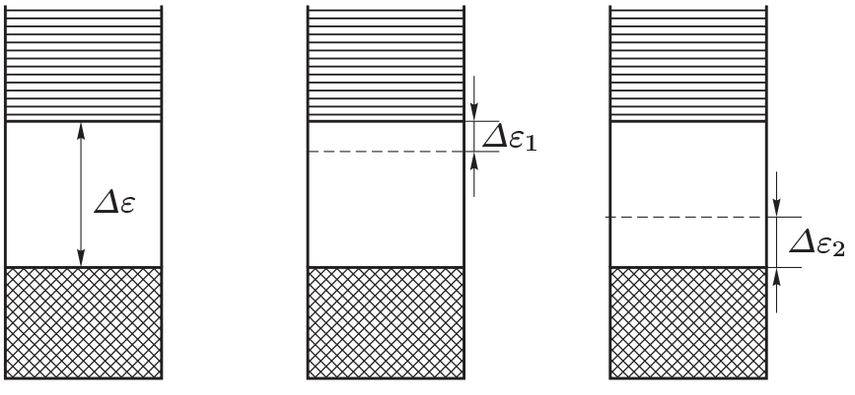


Рис. 222

атома кремния содержит четыре, а атома мышьяка — пять электронов. Пятый электрон может отщепиться от атома мышьяка в результате теплового движения. Получившийся положительный ион мышьяка может вытеснить из решетки один из атомов кремния и встать на его место. В результате этого между узлами решетки появится электрон проводимости.

Допустим теперь, что при введении атомов примеси добавочные уровни в запрещенной зоне появляются вблизи верхнего края валентной зоны (рис. 222 *в*). Тогда электроны из валентной зоны начнут переходить на эти добавочные уровни. В валентной зоне появятся дырки, а с ними и электропроводность полупроводника, но уже не электронная, а дырочная. Соответствующие примеси называются *акцепторами*, или *акцепторными примесями.* Дополнительные уровни, которые они создают в запрещенной зоне, также называются *акцепторными уровнями.*

Примером акцепторной примеси могут служить атомы бора или какого-либо другого элемента из третьей группы периодической системы. Наружная оболочка атома бора содержит три электрона. Атом бора может захватить недостающий четвертый электрон из какого-либо соседнего места кристалла. В этом месте образуется дырка, а образовавшийся отрицательный ион бора может вытеснить из кристаллической

решетки атом кремния и встать на его место. Так в кристалле кремния возникает дырочная проводимость.

Какой проводимостью обладает полупроводник — электронной или дырочной — об этом можно судить по знаку эффекта Холла. Полупроводники с донорной примесью называются *электронными*, или *полупроводниками n-типа* (от английского слова *negative* — отрицательный), а полупроводники с акцепторной примесью — *дырочными*, или *полупроводниками p-типа* (от английского слова *positive* — положительный). Могут быть и *смешанные полупроводники*, в которых носителями тока являются как электроны, так и дырки. Носители, которым принадлежит больший вклад в величину тока, называются *основными*, а прочие — *неосновными.*

ЗАДАЧА

Твердый водород является диэлектриком, плотность которого при нормальном давлении равна 0,076 г/см^3. Оценить плотность твердого водорода, при которой он становится металлом. Энергия ионизации атома водорода $\varepsilon_{\text{ион}} = 13{,}6$ эВ$= 2{,}18 \cdot 10^{-11}$ эрг.

Р е ш е н и е. Сжатие вещества сопровождается повышением уровня Ферми, которым при заданной плотности вещества определяется максимальная кинетическая энергия свободного электрона. Когда энергия Ферми μ становится равной энергии ионизации атома $\varepsilon_{\text{ион}}$, происходит ионизация атомов диэлектрика. Наружные электронные оболочки атомов разрушаются, валентные электроны коллективизируются, и диэлектрик становится металлом. Таким образом, концентрация атомов диэлектрика n, при которой он становится металлом, определяется условием $\mu \sim \varepsilon_{\text{ион}}$. Определив величину n из формулы (99.6) и умножив ее на массу атома водорода m_{H}, получим

$$\rho \approx \frac{\pi m_{\text{H}}}{3h^3} (8m\varepsilon_{\text{ион}})^{3/2} \approx 0{,}38 \text{ г/см}^3.$$

§ 101. Термоэлектронная эмиссия

1. То обстоятельство, что «свободные электроны» удерживаются внутри металла, указывает на то, что в поверхностном слое металла возникает задерживающее электрическое поле, препятствующее электронам выходить из металла в окружающий вакуум. Чтобы покинуть металл, электрон должен совершить некоторую работу, называемую *работой выхода.* Одна из причин возникновения работы выхода состоит в следующем. Если при тепловом движении электрон вылетит из металла, то он индуцирует на поверхности последнего заряды противоположного знака. Возникнет сила притяжения между электроном и поверхностью металла (так называемая *«сила электрического изображения»*, см. § 23), стремящаяся вернуть электрон обратно в металл. На преодоление этой силы требуется производство работы. Можно указать другую причину. Электроны, совершая тепловое движение, могут пересекать поверхность металла и удаляться от нее на небольшие расстояния (порядка атомных). Над поверхностью металла возникает электронная атмосфера, плотность которой быстро убывает при

удалении от металла. Под ней у поверхности металла остается слой положительно заряженных ионов. В результате образуется *двойной электрический слой*, действующий подобно конденсатору. Он не создает электрического поля во внешнем пространстве. Зато на преодоление электрического поля внутри самого двойного слоя также требуется производство работы. Вероятно, существуют другие причины возникновения работы выхода. Однако мы не будем входить в обсуждение этого вопроса, а ограничимся в дальнейшем лишь феноменологическим учетом работы выхода.

2. При повышении температуры металла увеличивается кинетическая энергия теплового движения электронов вблизи границы Ферми. Здесь она может стать настолько большой, что некоторые из электронов могут преодолевать задерживающий электрический потенциал на границе металла и выходить наружу. Если в окружающем вакууме существует электрическое поле, направленное к поверхности металла, то оно будет увлекать вышедшие электроны, и через вакуум потечет электрический ток. Этот ток называется *термоэлектронным*, а само явление — *термоэлектронной эмиссией.* Оно было открыто Эдисоном (1847–1931) в 1883 г.

Для наблюдения термоэлектронной эмиссии удобна вакуумная лампа с двумя электродами, называемая *вакуумным диодом.* Катодом лампы служит проволока (нить) из тугоплавкого металла (вольфрам, молибден и пр.), накаливаемая электрическим током. Анод чаще всего имеет форму металлического цилиндра, окружающего накаливаемый катод. Если диод включить в электрическую цепь, показанную на рис. 223, то при холодном катоде ток через цепь не пойдет. Если же нагреть катод до белого каления, то миллиамперметр покажет ток. При изменении полярности батареи ток прекращается. Это доказывает, что

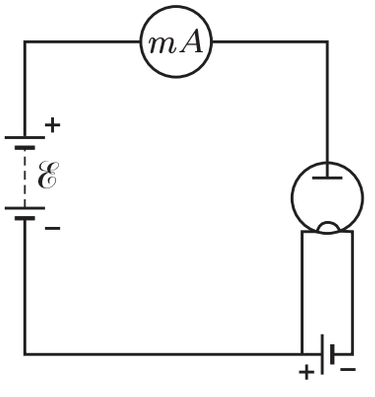


Рис. 223

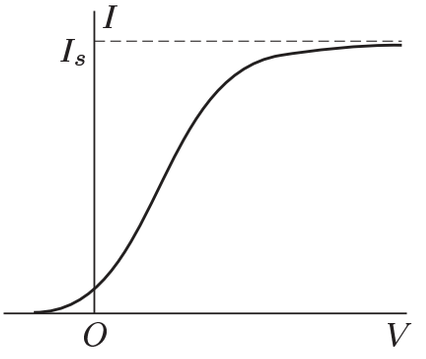


Рис. 224

носителями тока через вакуум являются *отрицательно заряженные частицы*, а именно *электроны*, так как никакими химическими превращениями вблизи электродов прохождение термоэлектронного тока не сопровождается.

Если, поддерживая температуру накаленного катода постоянной, менять напряжение V между анодом и катодом, то термоэлектронный

ток сначала будет возрастать (рис. 224). Однако это возрастание идет не пропорционально V, так что для вакуумного диода закон Ома не выполняется. По достижении определенного напряжения дальнейшее нарастание термоэлектронного тока практически прекращается. Последний достигает предельного значения I_s, называемого *током насыщения.* Наличие тока насыщения имеет весьма простое объяснение. Его величина определяется количеством термоэлектронов, которое в состоянии испариться с поверхности катода в единицу времени. Если электрическое поле настолько сильное, что оно отводит все электроны, испаряющиеся с поверхности катода, то дальнейшее увеличение напряженности поля уже не может привести к увеличению термоэлектронного тока. С этим и связано явление насыщения тока. Плотность термоэлектронного тока насыщения j_s определяет *эмиссионную способность* материала катода, т. е. максимальное количество электронов, которое может эмиттировать катод с единицы поверхности в единицу времени.

3. Плотность термоэлектронного тока насыщения зависит от материала катода и увеличивается с повышением температуры последнего. При вычислении этой плотности будем пользоваться моделью идеального электронного газа и применим к нему статистику Ферми–Дирака. Число квантовых состояний в единице объема металла с импульсами электронов внутри элемента объема импульсного пространства $dp_x dp_y dp_z$ с центром в точке (p_x, p_y, p_z) будет $dZ = (2/h^3)\, dp_x dp_y dp_z$, где коэффициент 2 учитывает спин электрона. Средняя концентрация электронов с такими импульсами определяется выражением

$$dn = \frac{2}{h^3}\,\frac{dp_x dp_y dp_z}{\exp\dfrac{\varepsilon-\mu}{kT}+1}. \tag{101.1}$$

Уточним теперь понятие работы выхода, введенное выше. Условимся называть этой величиной работу b, которую должен совершить электрон, чтобы с *уровня Ферми* выйти наружу металла. Для этого минимальная кинетическая энергия электрона на уровне Ферми должна быть $\varepsilon_{\text{мин}} = \mu + b$. Из металла могут выходить только такие электроны, кинетическая энергия которых на уровне Ферми превосходит $\varepsilon_{\text{мин}}$, т. е. $\varepsilon > \mu + b$. Только они играют роль при вычислении тока насыщения. Поэтому при дальнейших расчетах можно ограничиться рассмотрением лишь таких электронов.

Чтобы при эмиссии электронов кристаллическая решетка не разрушалась, из металла должна выходить ничтожная часть электронов. Для этого, как показывает формула (101.1), должно выполняться условие $\varepsilon - \mu \gg kT$, т. е. $b \gg kT$. Для таких электронов в знаменателе формулы (101.1) единицей можно пренебречь. Тогда формула (101.1) переходит в

$$dn = \left(\frac{2}{h^3}\exp\frac{\mu}{kT}\right)\exp\left\{-\frac{\varepsilon}{kT}\right\}\, dp_x dp_y dp_z. \tag{101.2}$$

Найдем долю электронов dn_z, z-составляющая импульса которых заключена между p_z и $p_z + dp_z$. Для этого предыдущее выражение надо проинтегрировать по p_x и p_y в пределах от $-\infty$ до $+\infty$. Так как $\varepsilon = (1/2m)(p_x^2+p_y^2+p_z^2)$, то в результате такого интегрирования получим

$$dn_z = \frac{4}{h^3}\,\pi m k T \exp\left\{-\frac{p_z^2/2m - \mu}{kT}\right\} dp_z. \tag{101.3}$$

Уход из металла электронов при термоэлектронной эмиссии, конечно, несколько нарушает равновесное распределение их по скоростям. Однако в нулевом приближении этим обстоятельством можно пренебречь, что и делается в дальнейшем. Число электронов рассматриваемой группы, падающих в единицу времени на единицу площади поверхности металла, определяется интегралом

$$\int v_z\, dn_z = \frac{1}{m}\int p_z\, dn_z,$$

где интегрирование ведется по всем электронам, для которых $p_z^2/2m \geqslant \mu + b$. Примем, что все эти электроны уходят из металла. Тогда плотность термоэлектронного тока насыщения j_s найдется умножением предыдущего интеграла на заряд электрона e. Введем новую переменную интегрирования

$$x = p_z^2/2m - \mu.$$

Тогда

$$j_s = \frac{4\pi e m}{h^3}\,(kT)^2 \int\limits_{b/kT}^{\infty} e^{-x}\, dx. \tag{101.4}$$

В результате получим

$$j_s = AT^2 e^{-b/kT}, \tag{101.5}$$

где постоянная A определяется выражением

$$A = \frac{4\pi m e k^2}{h^3} = 120\ \text{А}/(\text{см}^2 \cdot \text{К}), \tag{101.6}$$

т. е. *одинакова для всех металлов.* Такая одинаковость связана с использованием модели идеального электронного газа, положенной в основу расчета. В этой модели индивидуальные свойства металла характеризуются двумя параметрами: работой выхода электрона b и энергией Ферми μ. Однако последняя входит в формулу (101.4) только в качестве слагаемого переменной интегрирования x, так что результат не зависит от μ, а определяется только работой выхода b и температурой металла T.

Измеряя плотность термоэлектронного тока j_s, можно по формуле (101.5) вычислить как постоянную A, так и работу выхода b.

4. Формула (101.5) была впервые получена Ричардсоном (1879–1959) на основе термодинамических соображений с использованием некоторых эмпирических данных. Другой вывод, в котором была

установлена универсальность коэффициента A, был дан Дешманом в 1923 г. Поэтому формула получила название *формулы Ричардсона* (1879–1959)*–Дешмана.* Дешман в своем выводе еще не мог учитывать электронного спина. По этой причине значение универсального коэффициента A у него было вдвое меньше, так как учет спина удваивает число возможных квантовых состояний электрона. Результат Дешмана с точностью до нескольких процентов был подтвержден при экспериментальной проверке формулы (101.5) на особенно подходящих для этой цели тугоплавких металлах — вольфраме и молибдене. Любопытно, что учет электронного спина значительно ухудшал согласие теории с опытом. Этот результат можно объяснить уже в рамках модели идеального электронного газа при более последовательном квантовомеханическом рассмотрении. Дело в том, что по квантовой механике электроны, удовлетворяющие условию $\varepsilon > \mu + b$, не обязательно уходят из металла. Часть из них может *отразиться* обратно в металл от *потенциального барьера* на границе металла, который они должны преодолеть, чтобы выйти в вакуум. Однако мы не можем здесь входить в эти детали.

5. Заметим, что Ричардсон еще в 1901 г. получил другую формулу для плотности термоэлектронного тока насыщения. Он применил к электронному газу в металле классическую статистику Больцмана. Результат Ричардсона можно получить из наших формул, не производя новых вычислений. Надо только формулу (101.3) заменить соответствующим выражением из закона распределения скоростей Максвелла (см. т. II, § 72):

$$dn_z = n\left(\frac{1}{2\pi mkT}\right)^{1/2} \exp\left\{-\frac{p_z^2}{2mkT}\right\} dp_z,$$

где n — число свободных электронов в единице объема металла. Главное отличие этой формулы от (101.3) состоит в том, что в предэкспоненциальном выражении она содержит $\sqrt{T}$ в знаменателе, тогда как в формулу (101.3) температура T входит в числителе и притом в первой степени. Чтобы получить первую формулу Ричардсона, очевидно, достаточно произвести замену

$$\frac{4}{h^3}\,\pi mkT \to n\left(\frac{1}{2\pi mkT}\right)^{1/2}.$$

Это дает

$$j_s = A' T^{1/2} e^{-b/kT}, \tag{101.7}$$

где

$$A' = ne\sqrt{\frac{k}{2\pi m}}. \tag{101.8}$$

Формула (101.7) получает простое истолкование, если записать ее в виде

$$j_s = \frac{1}{4}\, ne\overline{v} e^{-b/kT},$$

где $\overline{v}$ — средняя тепловая скорость электрона по классической теории. Величина $1/(4n\overline{v})$ дает число электронов, ежесекундно ударяющихся о единицу поверхности металла (см. т. II, § 75). Из этих частиц надо взять те, которые

в состоянии преодолеть потенциальный барьер на границе металла. Для этого поток частиц надо умножить на экспоненциальный множитель $e^{-b/(kT)}$.

Практически обе формулы (101.5) и (101.7) дают совпадающие результаты, так как зависимость тока j_s от температуры в основном определяется экспоненциальным множителем, который меняется с температурой значительно сильнее, чем степенные множители T^2 и $T^{1/2}$. Действительно, логарифмируя (101.5), находим, что при изменении температуры ток j_s изменяется в соответствии с формулой

$$\frac{\Delta j_s}{j_s} = 2\frac{\Delta T}{T} + \frac{b\Delta T}{kT^2}.$$

Отношение первого слагаемого в правой части этой формулы ко второму равно $2kT/b$. Для вольфрама $b = 4{,}5$ эВ. Если взять $T = 2500$ К, то для этого отношения получится 0,10. В случае формулы (101.7) соответствующее отношение будет еще меньше, а именно 0,025. Поэтому формулу (101.5) часто пишут в упрощенном виде:

$$\ln j_s = \text{const} - \frac{b}{kT}. \tag{101.9}$$

6. Плотность термоэлектронного тока насыщения j_s, таким образом, определяется отношением $b/(kT)$. Для тугоплавких металлов W, Mo, Pt работа выхода относительно велика и составляет соответственно 4,52, 4,37, 4,8 эВ. Получить сильные термоэлектронные токи с катодами из этих металлов можно лишь при очень высоких температурах накала. Между тем на практике весьма существенно снизить рабочую температуру катода, так как при этом уменьшаются затраты энергии на накал катода и увеличивается срок службы лампы. В так называемых *сложных катодах* это достигается путем снижения работы выхода. К числу таких катодов относится *оксидный катод*, получивший широкое распространение в технике. При его изготовлении на металлическую подложку (керн) наносят углекислые соединения щелочноземельных металлов ($BaCo_3$, $SrCO_3$, $CaCO_3$). При прокаливании эти соли разлагаются (согласно уравнению $BaCO_3 \rightleftarrows BaO + CO_2$ и пр.), и подложка оказывается покрытой слоем окислов щелочноземельных металлов (BaO, SrO, CaO или их смесями). После этого катод *активируют*, т. е. в течение некоторого времени пропускают через лампу термоэлектронный ток при температуре катода около 1000 °С. В результате этого на поверхности катода образуется *моноатомный слой* ионов щелочноземельных металлов, который сильно понижает работу выхода и тем самым увеличивает эмиссионную способность катода. Накаливание катода лампы осуществляется либо пропусканием тока через подложку, либо посредством вспомогательной металлической спирали («подогревные катоды»). Рабочая температура современных оксидных катодов составляет 800–900 °С и даже ниже. Нормальная эмиссионная способность достигает 1 А/см^2, а при очень коротких импульсах тока (порядка 10^{-6}–10^{-5} с) она может доходить до 100 А/см^2 и выше. Для сравнения укажем, что рабочая температура вольфрамовых катодов 2200 °С, а снимаемые с них термоэлектронные токи на практике не превышают 0,1 А/см^2.

7. Если бы все электроны, испаряющиеся с поверхности катода, попадали на анод, то сила термоэлектронного тока I не зависела бы от приложенного напряжения V. На самом деле это не так. С возрастанием напряжения ток I растет (см. рис. 224). Объясняется это тем, что в пространстве между катодом и анодом образуется отрицательный пространственный заряд, создающий электрическое поле, направленное противоположно приложенному электрическому полю. Этот вопрос теоретически был исследован Ленгмюром (1881–1957) в 1913 г. Допустим, что электроды лампы плоские. Их можно считать бесконечно большими. Температуру катода T будем поддерживать постоянной. Начало координат поместим на поверхности катода, а ось X направим перпендикулярно к этой поверхности в сторону анода. Потенциал электрического поля φ будет функцией только координаты x. Он удовлетворяет уравнению Пуассона

$$\frac{d^2\varphi}{dx^2} = -4\pi\rho = 4\pi ne, \tag{101.10}$$

где e — абсолютное значение заряда электрона. Плотность тока по абсолютной величине $j = nev$. Если в межэлектродном пространстве создан достаточно высокий вакуум, то столкновения электронов можно не учитывать и определить скорость электрона из уравнения $mv^2/2 = e\varphi$. При этом потенциал катода мы принимаем равным нулю и пренебрегаем тепловыми скоростями, с которыми электроны покидают поверхность катода. Исключив из этих соотношений концентрацию электронов n и скорость v, придем к уравнению

$$\frac{d^2\varphi}{dx^2} = \frac{a^2}{\sqrt{\varphi}}, \tag{101.11}$$

где введено обозначение

$$a^2 = 4\pi j\sqrt{\frac{m}{2e}}. \tag{101.12}$$

Величина a, очевидно, постоянна, так как ток j от координаты x не зависит.

Сопоставляя уравнение (101.11) с (101.10), видим, что объемная плотность электричества ρ на поверхности катода обращается в бесконечность. Точка $x = 0$ является особой точкой уравнения (101.11), в которой оно теряет смысл. Этот недостаток теории является следствием принятой идеализации, в которой полностью пренебрегается тепловым разбросом скоростей электронов, уходящих с поверхности катода. Однако поверхностная плотность заряда на катоде остается конечной. Более того, она должна обращаться в нуль, как и напряженность электрического поля у поверхности катода. Если бы это было не так, то все электроны, эмиттируемые катодом, увлекались бы электрическим полем к аноду и термоэлектронный ток достигал бы насыщения при любых напряжениях на лампе. Таким образом, уравнение (101.11) надо

дополнить граничными условиями

$$\lim_{x\to 0}\varphi=\lim_{x\to 0}\frac{d\varphi}{dx}=0. \tag{101.13}$$

Для решения уравнения (101.1) умножаем его на производную $d\varphi/dx$ и интегрируем с учетом граничных условий (101.13). Таким путем получим

$$\left(\frac{d\varphi}{dx}\right)^2=4a^2\varphi^{1/2}.$$

При извлечении квадратного корня надо сохранить только знак плюс, так как потенциал φ должен возрастать от катода к аноду и, следовательно, должен быть положительным во всем межэлектродном пространстве. В результате после разделения переменных приходим к уравнению

$$\frac{d\varphi}{\varphi^{1/4}}=2a\,dx.$$

Интегрируя вторично, найдем

$$\frac{4}{3}\,\varphi^{3/4}=2ax, \tag{101.14}$$

или после подстановки значения a

$$j=\frac{1}{9\pi x^2}\sqrt{\frac{2e}{m}}\,\varphi^{3/2}. \tag{101.15}$$

В частности, если V — напряжение, а l — расстояние между электродами, то

$$j=CV^{3/2}, \tag{101.16}$$

где постоянная C определяется выражением

$$C=\frac{1}{9\pi l^2}\sqrt{\frac{2e}{m}}. \tag{101.17}$$

Формула (101.16) называется *законом трех вторых* Ленгмюра. Этот закон остается верным для электродов произвольной конфигурации. От формы электродов зависит только выражение для числового коэффициента C. Ленгмюр решил задачу для коаксиальных цилиндрических электродов, когда катодом служит внутренний, а анодом — наружный цилиндр. Та же задача была решена С.А. Богуславским (1883–1923), который, кроме того, рассмотрел противоположный случай, когда катодом был наружный, а анодом — внутренний цилиндр. Ленгмюр рассмотрел также случай сферических электродов. Во всех случаях, как и следовало ожидать, получился закон трех вторых.

При малых напряжениях закон трех вторых дает заниженные значения для термоэлектронного тока, так как этот закон не учитывает теплового разброса скоростей электронов. При больших напряжениях закон был бы точным, если бы эмиссионная способность катода была

бесконечно велика. Поскольку это не так, наблюдаются отступления от закона трех вторых и при больших напряжениях. При увеличении напряжения в конце концов наступает насыщение, и закон трех вторых утрачивает силу.

§ 102. Электронные лампы и их применения

1. Электроны в вакуумном диоде, испускаемые катодом, движутся к аноду. Если на анод подать положительный потенциал, то ток через диод пойдет. Если же подать отрицательный потенциал, то тока не будет. На этом основано применение вакуумных диодов для *выпрямления переменных токов*. Схема «однополупериодного» лампового выпрямителя представлена на рис. 225 *а*. Если входное напряжение представляется кривой рис. 225 *б*, то при отсутствии конденсатора C выходное напряжение изобразилось бы кривой рис. 225 *в*. Оно состояло бы из импульсов напряжения, следующих друг за другом через половину периода $T/2$. Такую же форму имела бы и кривая тока I, текущего через *нагрузочное сопротивление* R. Для того чтобы получить непрерывное течение тока I и сгладить его пульсации, вводится конденсатор C. Тогда за каждую половину периода, когда через лампу течет ток, последний не только поступает в сопротивление R, но и заряжает конденсатор C. В течение других полупериодов, когда лампа тока не пропускает, конденсатор разряжается через сопротивление R. Пульсации тока будут тем меньше, чем больше время $\tau = RC$, характеризующее быстроту разрядки конденсатора (см. § 48). Недостаток описанного выпрямителя состоит в том, что он пропускает ток только в течение одной половины периода T, а в течение другой половины не пропускает. На рис. 226 приведена схема «двухполупериодного» выпрямителя, в котором этот недостаток устранен.

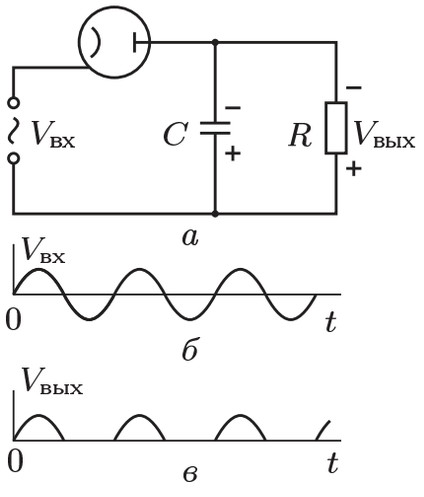


Рис. 225

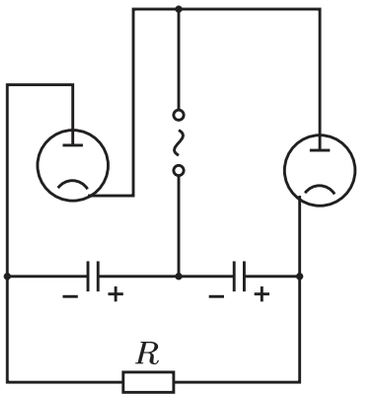


Рис. 226

2. Введением внутрь электронной лампы дополнительных электродов можно легко управлять электронным током. Эти дополнительные электроды называются *сетками*, так как им обычно придают форму металлических сеток или спиралей, окружающих катод. Простейшей является *трехэлектродная лампа*, или *триод*, имеющая всего одну сетку, называемую *управляющей сеткой*. Управляющая сетка располагается значительно ближе к като-

ду, чем анод. Поэтому изменение сеточного напряжения (т. е. разности потенциалов между сеткой и катодом) значительно сильнее влияет на напряженность электрического поля вблизи катода, а следовательно, и на электронный ток в лампе, чем равное по величине изменение анодного напряжения (т. е. разности потенциалов между анодом и катодом). Когда сеточное напряжение равно нулю, сетка практически не влияет на силу электронного тока в лампе. Если сетку зарядить положительно, то электронный ток увеличится. Если же ее зарядить отрицательно, то ток через лампу уменьшится и даже может совсем прекратиться. Изменениями потенциала сетки можно, следовательно, управлять силой электронного тока. Благодаря малости массы электрона инерционность триода также очень мала, и его управляющее действие сохраняется даже при очень быстрых изменениях сеточного напряжения.

Часть электронов, исходящих из катода, попадает на сетку, и в цепи последней возникает ток $I_\text{с}$, называемый *сеточным током.* Полный ток в лампе I слагается из анодного тока $I_\text{а}$ и тока на сетку $I_\text{с}$: $I = I_\text{а} + I_\text{с}$. Появление заметного сеточного тока нежелательно, так как оно приводит к бесполезному расходованию электрической энергии в цепи сетки. Однако в большинстве случаев основной поток электронов проходит сквозь редкую сетку, так что сеточный ток мал по сравнению с анодным током. Поэтому величиной $I_\text{с}$ мы, как правило, будем пренебрегать и считать, что $I = I_\text{а}$.

Анодный ток $I_\text{а}$ является функцией сеточного $V_\text{с}$ и анодного $V_\text{а}$ напряжений: $I_\text{а} = I_\text{а}(V_\text{с}, V_\text{а})$. На практике эту функцию принято характеризовать двумя семействами кривых, называемых *сеточными* и *анодными характеристиками.* Сеточная характеристика есть кривая $I_\text{а} = I_\text{а}(V_\text{с})$, изображающая зависимость анодного тока $I_\text{а}$ от сеточного напряжения $V_\text{с}$ при условии, что анодное напряжение и накал катода поддерживаются достоянными. Аналогично, анодная характеристика есть кривая, представляющая зависимость того же тока от анодного напряжения, если остаются постоянными сеточное напряжение и накал

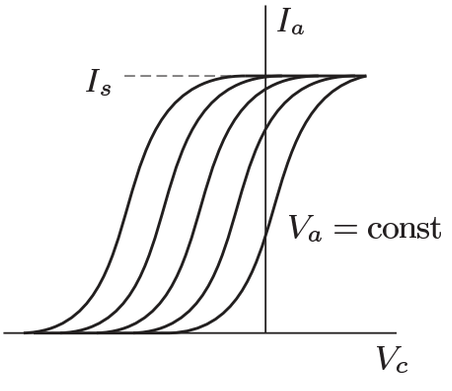


Рис. 227

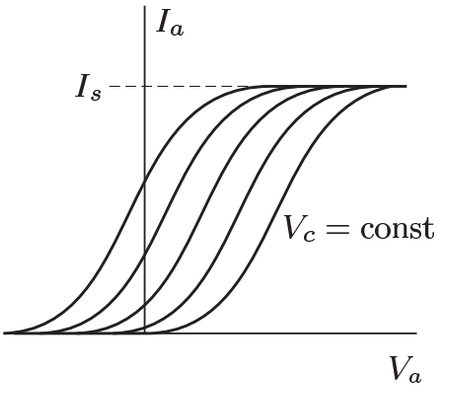


Рис. 228

катода. Схематически сеточные и анодные характеристики представлены на рис. 227 и 228. Сеточные характеристики поднимаются более

круто, чем анодные. При увеличении параметра V_a сеточные характеристики смещаются влево. То же происходит с анодными характеристиками при увеличении параметра V_c. При повышении температуры накала катода возрастает ток насыщения и оба семейства характеристик вытягиваются вверх.

Анодный ток I_a меняется нелинейно как с изменением V_c, так и с изменением V_a. Только на средних участках характеристики почти прямолинейны, так что здесь приближенно соблюдается закон Ома. На этих участках производные

$$S = \left(\frac{\partial I_\text{a}}{\partial V_\text{c}}\right)_{V_\text{a}} \qquad \text{и} \qquad R_i = \left(\frac{\partial V_\text{a}}{\partial I_\text{a}}\right)_{V_\text{c}} \tag{102.1}$$

практически постоянны. Производная S называется *крутизной* (*сеточной*) *характеристики*, а производная R_i — *дифференциальным* или *внутренним сопротивлением лампы.*

Как видно из рис. 227, при изменении параметра V_a форма сеточной характеристики практически остается неизменной, но вся характеристика смещается влево (при увеличении V_a) или вправо (при уменьшении V_a). При этом равным приращениям параметра V_a соответствуют равные смещения характеристики. Отсюда следует, что анодный ток I_a зависит не от напряжений V_c и V_a в отдельности, а только от их комбинации $V_\text{y} = V_\text{c} + DV_\text{a}$, где D — положительная постоянная, называемая *проницаемостью сетки.* Она меньше единицы, так как изменения сеточного напряжения значительно сильнее влияют на величину анодного тока, чем такие же изменения анодного напряжения. Таким образом, функция двух переменных $I_\text{a} = I_\text{a}(V_\text{c}, V_\text{a})$ фактически переходит в функцию одной переменной $V_\text{y} = V_\text{c} + DV_\text{a}$. По этой причине величина V_y называется *управляющим напряжением.* Дифференцируя формулу $I_\text{a} = I_\text{a}(V_\text{y}) = I_\text{a}(V_\text{c} + DV_\text{a})$, получаем

$$S = \left(\frac{\partial I_\text{a}}{\partial V_\text{c}}\right)_{V_\text{a}} = \frac{dI_\text{a}}{dV_\text{y}}, \qquad \frac{1}{R_i} = \left(\frac{\partial I_\text{a}}{\partial V_\text{a}}\right)_{V_\text{c}} = \frac{dI_\text{a}}{dV_\text{y}}\left(\frac{\partial V_\text{y}}{\partial V_\text{a}}\right)_{V_\text{c}} = SD.$$

Отсюда следует, что параметры лампы S, D, R_i связаны соотношением

$$SDR_i = 1. \tag{102.2}$$

3. Трехэлектродная лампа может служить для *усиления электрических сигналов и переменных токов.* Принципиальная схема усилителя на триоде приведена на рис. 229. В анодную цепь лампы введена батарея с электродвижущей силой $\mathscr{E}$ и нагрузочное сопротивление R_a. Усиливаемое напряжение прикладывается между сеткой и катодом. Рассматривая анодный ток как функцию управляющего напряжения V_y, напишем

$$\Delta I_\text{a} = S\Delta V_\text{y} = S(\Delta V_\text{c} + D\Delta V_\text{a}).$$

Применяя же закон Ома к внешнему участку цепи $K\mathscr{E}A$ и пренебрегая при этом внутренним сопротивлением батареи и подводящих проводов,

получим

$$V_{\text{a}} = \mathscr{E} - R_{\text{a}} I_{\text{a}}.$$

Так как величина $\mathscr{E}$ постоянна, то

$$\Delta V_{\text{a}} = -R_{\text{a}} \Delta I_{\text{a}} = -R_{\text{a}} S(\Delta V_{\text{c}} + D \Delta V_{\text{a}}),$$

и, следовательно,

$$\frac{\Delta V_{\text{a}}}{\Delta V_{\text{c}}} = -\frac{R_{\text{a}} S}{1 + SDR_{\text{a}}}.$$

Используя соотношение (102.2), приведем эту формулу к виду

$$\frac{\Delta V_{\text{a}}}{\Delta V_{\text{c}}} = -\frac{R_{\text{a}}}{R_{\text{a}} + R_i} \frac{1}{D}. \tag{102.3}$$

Полученная формула показывает, как связаны между собой амплитуды колебаний анодного и сеточного напряжений. В частном случае, когда $R_i \ll R_{\text{a}}$, она переходит в

$$\frac{\Delta V_{\text{a}}}{\Delta V_{\text{c}}} = -\frac{1}{D}. \tag{102.4}$$

Поэтому величина $K = 1/D$ называется *коэффициентом усиления лампы*. В другом предельном случае, когда $R_i \gg R_{\text{a}}$,

$$\frac{\Delta V_{\text{a}}}{\Delta V_{\text{c}}} = -\frac{R_{\text{a}}}{R_i D} = -SR_{\text{a}}, \tag{102.5}$$

и усиление тем больше, чем больше крутизна характеристики S. Знак минус в формулах (102.3)–(102.5) означает, что фазы колебаний сеточного и анодного напряжений противоположны.

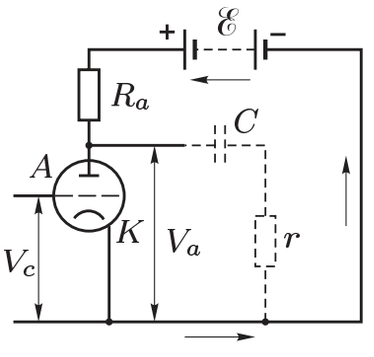


Рис. 229

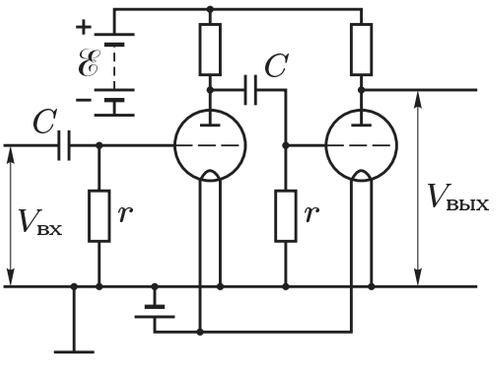


Рис. 230

Усиление электрического сигнала можно повторять многократно. Для этого напряжение, усиленное первой лампой, подается на сетку второй; напряжение, усиленное второй лампой, — на сетку третьей и т. д. Одна из схем такого *каскадного усилителя* представлена на рис. 230. Напряжение на сетку каждой лампы подается не непосредственно, а через *разделительный конденсатор* C. Последний, не препятствуя прохождению переменного усиливаемого сигнала, не допускает

на сетку высокое постоянное напряжение анодной батареи. Между сеткой и катодом каждой лампы включено сопротивление r (утечка сетки). Делается это с целью предотвращения *«запирания»* лампы. Дело в том, что в те периоды времени, когда потенциал сетки положителен, на нее попадают электроны. В отсутствие утечки на сетке накопились бы отрицательные заряды такой величины, что ток через лампу прекратился бы. Надлежащим подбором емкости C и сопротивления r можно добиться, чтобы колебания напряжения на сетке происходили около небольшого определенного значения, при котором еще сохраняется управляющее действие сетки, а сеточный ток пренебрежимо мал.

4. Для многих целей требуется, чтобы анодный ток зависел от *сеточного* и практически не зависел от анодного напряжения. Это достигается введением между управляющей сеткой и анодом *дополнительной сетки*, на которую подается положительный относительно катода потенциал, несколько меньший потенциала анода. Такая дополнительная сетка называется *экранной*, а *четырехэлектродная* лампа, в которую она входит, — *тетродом.* Введение экранной сетки ослабляет электрическое поле в окрестности анода, уменьшая тем самым влияние потенциала анода на анодный ток. В этом отношении экранная сетка действует так же, как уменьшение проницаемости D управляющей сетки. Поэтому коэффициент усиления у тетрода при прочих равных условиях гораздо больше, чем у триода. Однако введение экранной сетки легко может привести к нежелательным явлениям. Электроны, прошедшие через экранную сетку, попадают на анод и выбивают из него вторичные электроны. Это явление называется *динатронным эффектом.* Такое явление возникает и в триоде. Но в триоде при нормальной работе выбитые электроны обратно возвращаются к аноду под действием электрического поля последнего. В тетроде же экранная сетка всегда заряжена положительно, и в случае сильных колебаний анодного напряжения потенциал анода в отдельные периоды может стать ниже потенциала экранной сетки. Тогда выбитые электроны начнут попадать на экранную сетку и проходить через нее. Это ведет к уменьшению анодного тока, появлению провалов на характеристиках и ухудшению свойств тетрода. Нежелательные явления, вызванные динатронным эффектом, могут быть устранены введением третьей, так называемой *защитной* или *противодинатронной сетки.* Эта сетка помещается между экранной сеткой и анодом (рис. 231). Она соединяется с катодом (часто внутри лампы). Находясь под потенциалом катода, она тормозит вторичные электроны, препятствуя попаданию их на экранную сетку. В то же время она не оказывает существенного влияния на движение основного электронного потока. Лампы с пятью электродами, или *пентоды*, имеют высокий коэффи-

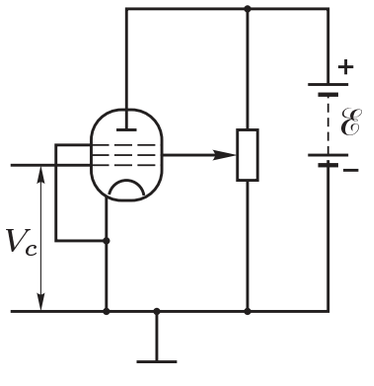


Рис. 231

циент усиления. Их характеристики гладкие, без провалов. Поэтому они получили более широкое распространение, чем тетроды.

Применение электронных ламп для генерации электрических колебаний будет рассмотрено в § 133.

§ 103. Вторичная и автоэлектронная эмиссия

1. При бомбардировке поверхностей металлов, полупроводников или диэлектриков пучком электронов наблюдается испускание *вторичных электронов*. Это явление называется *вторичной электронной эмиссией* (в электронных лампах его чаще называют *динатронным эффектом*, см. предыдущий параграф). В пучке эмиттируемых электронов наблюдаются три группы электронов: 1) электроны, упруго отраженные поверхностью эмиттера; 2) неупруго отраженные электроны; 3) вторичные электроны, т. е. такие электроны, которые выбиваются из эмиттера первичными электронами. Для количественного описания явления принято вводить *коэффициент вторичной эмиссии* σ. Так называют отношение σ полного количества электронов N, испущенных эмиттирующей поверхностью, к числу первичных электронов N_0:

$$\sigma = \frac{N}{N_0}.$$

Коэффициент вторичной электронной эмиссии зависит от природы, изготовления и от состояния поверхности облучаемого тела, а также от скорости электронов в падающем пучке и от угла падения последнего. Коэффициент σ не зависит от интенсивности пучка первичных электронов, если эта интенсивность не настолько велика, чтобы приводить к испарению и разрушению поверхности тела. Начальная энергия основной доли вторичных электронов, вышедших из эмиттера, составляет несколько электронвольт и в широком интервале не зависит от энергии первичных электронов. Первичные электроны, проникая внутрь бомбардируемого вещества, возбуждают на своем пути электроны эмиттера. Число электронов, возбуждаемых на единице пути первичного электрона, возрастает к концу пробега последнего. Именно здесь зарождается основная доля вторичных электронов, которые затем при благоприятных условиях могут выйти наружу. Чем больше скорость v первичных электронов, тем глубже они проникают в облучаемое тело и тем больше возбуждают вторичных электронов. Но, поскольку освобождение электронов происходит в более глубоких слоях тела, уменьшается вероятность выхода их наружу. Этим объясняется, почему коэффициент вторичной эмиссии сначала возрастает с увеличением скорости v первичных электронов, затем достигает размытого максимума, а при дальнейшем возрастании скорости v снова уменьшается. Аналогично объясняется возрастание коэффициента σ с увеличением угла падения первичных электронов: при скользящем падении первичного пучка вторичные электроны возникают в среднем ближе к поверхности эмиттера, чем при нормальном падении. Для

отчетливого наблюдения этого явления необходимо, конечно, чтобы поверхность эмиттера была достаточно гладкой.

Вторичная электронная эмиссия из толстых слоев металла всегда больше, чем из тонких металлических пленок, нанесенных на поверхность стекла. С увеличением толщины пленки до определенного предела эмиссия возрастает. Однако начиная с этого предела она становится такой же, как и у толстых металлических тел. Поэтому, экспериментируя с пленками, можно оценить толщину поверхностного слоя металла, из которого происходит эмиссия вторичных электронов. Таким путем найдено, что в металлах эта толщина не превосходит $5 \cdot 10^{-6}$ см.

Выход вторичных электронов, образовавшихся внутри эмиттера, в сильной степени зависит от природы последнего. В металлах, где велика концентрация электронов проводимости, вторичные электроны часто сталкиваются с ними и растрачивают свою энергию. В этих условиях вероятность выхода вторичных электронов наружу мала. Напротив, в полупроводниках и диэлектриках концентрация электронов проводимости мала, столкновения с ними происходят реже, а вероятность выхода электронов из эмиттера возрастает в несколько раз. Поэтому не существует металлов с большими коэффициентами σ и эффективные эмиттеры встречаются только среди полупроводников и диэлектриков. Значение σ в максимуме для всех чистых металлов не превышает 2, а для некоторых металлов даже меньше единицы. Так, для меди $\sigma_{\text{макс}} = 1{,}29$, никеля 1,25, серебра 1,47, платины 1,78, бериллия 0,53, алюминия 0,97. У полупроводников $\sigma_{\text{макс}}$ может достигать 10 и больше. Для получения сильной вторичной эмиссии на практике применяются *сложные катоды* (эмиттеры), состоящие из металлической подложки, на которой химическим путем и специальной обработкой создан тонкий слой полупроводника. Таковы, например, *сурьмяно-цезиевые эмиттеры*, получаемые обработкой сурьмы в парах цезия, *серебряно-цезиевые эмиттеры*, изготовляемые путем окисления серебра с последующей обработкой цезием, и другие аналогичные эмиттеры. Для таких эмиттеров $\sigma_{\text{макс}} \sim 10$.

2. Вторичная электронная эмиссия используется в *фотоэлектронных умножителях* (ФЭУ), предназначенных для усиления слабых электрических токов. Этот прибор представляет собой вакуумную трубку с катодом K и анодом A, между которыми расположено несколько электродов, эмиттирующих вторичные электроны (рис. 232). На эти электроды подается электрическое напряжение посредством делителей $Д$. Падающее электромагнитное излучение вырывает электроны с поверхности катода. Под действием электрического поля слабый электронный пучок ускоряется и направляется к эмиттеру $Э_1$, на котором происходит вторичная электронная эмиссия. Электроны с первого эмиттера направляются на второй эмиттер $Э_2$, где происходит вторичное усиление, и т. д. В результате получается усиленный во много раз (от 10^5 до 10^9) фототок, который и снимается с анода A.

3. Вторичная электронная эмиссия возникает не только при бомбардировке эмиттера электронами, но и при бомбардировке тяжелыми частицами — положительными и отрицательными ионами. Коэффициент

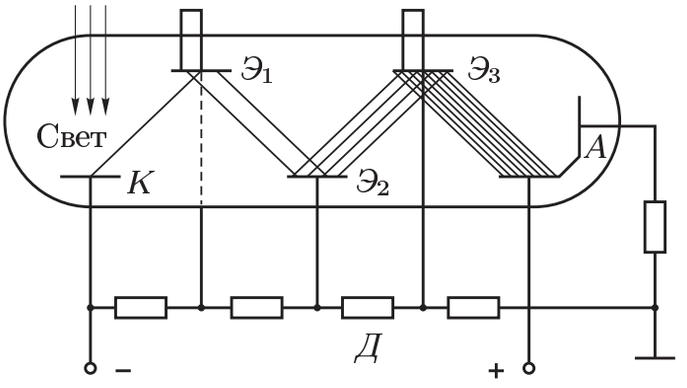


Рис. 232

вторичной эмиссии, т. е. среднее число электронов, вырываемых одной первичной частицей, здесь меньше, чем при бомбардировке электронами. Эмиссия электронов возникает также при облучении поверхности тела светом и другими короткими электромагнитными волнами (*фотоэлектрический эффект*). Это явление будет рассмотрено в т. V, ч. I нашего курса.

4. Эмиссия электронов из поверхности металлов может происходить под действием *сильного* электрического поля, вырывающего электроны из металла. Это явление называется *автоэлектронной*, или *холодной*, *эмиссией*. Для его наблюдения может служить хорошо откачанная вакуумная трубка, катодом которой служит металлическое острие, а анодом — обычный электрод с плоской или мало изогнутой поверхностью. При наложении напряжения на трубку у поверхности катода возникает сильное электрическое поле. Для оценки его числового значения катод можно считать маленьким шариком, а анод — удаленным в бесконечность. Если a — радиус шарика, а V — напряжение на трубке, то заряд шарика будет $q = Va$, а напряженность электрического поля на расстоянии r от центра шарика $E = q/r^2 = Va/r^2$. На самой поверхности шарика $E = V/a$. Пусть, например, радиус кривизны кончика острия $a = 10^{-2}$ мм. Тогда при напряжении $V = 1000$ В у поверхности катода возникнет поле $E \sim 10^6$ В/см. Если постепенно повышать напряжение на трубке, то при напряженности поля у поверхности катода $E \sim 10^5$–10^6 В/см в трубке возникает слабый ток, обусловленный автоэлектронной эмиссией с поверхности катода. Сила этого тока быстро увеличивается с повышением напряжения на трубке. Ток возникает и при холодном катоде По этой причине эмиссия и называется холодной. При дальнейшем повышении напряжения катод начинает сильно нагреваться и испаряться, а в трубке возникает газовый разряд. Объяснение механизма автоэлектронной эмиссии возможно только на основе волновой (квантовой) механики.

Глава VIII

ЭЛЕКТРИЧЕСКИЕ ЯВЛЕНИЯ В КОНТАКТАХ

§ 104. Контактная разность потенциалов

1. При соприкосновении двух различных металлов между ними возникает разность потенциалов, называемая *контактной разностью потенциалов.* Это явление было открыто Вольтой в 1797 г. Исследуя различные металлы, Вольта расположил их в следующий ряд: Al, Zn, Sn, Pb, Sb, Bi, Hg, Fe, Cu, Ag, Au, Pt, Pd. Он нашел, что если металлы в указанной последовательности привести в контакт друг с другом, то при этом каждый предыдущий металл получит более высокий потенциал, чем последующий. Далее, Вольта установил, что если несколько металлов 1, 2, ..., n привести в контакт друг с другом (рис. 233), то разность потенциалов $\varphi_n - \varphi_1$ между крайними металлами не будет зависеть от того, какими промежуточными металлами они разделены. Это положение называется *законом последовательных контактов Вольты.* Для справедливости его необходимо, чтобы все контактирующие металлы находились *при одной и той же температуре.* Если крайние металлы ряда привести в контакт между собой таким образом, чтобы образовалось замкнутое кольцо (рис. 234), то из закона последовательных контактов Вольты следует, что электродвижущая сила в кольце будет равна нулю и, следовательно, никакой электрический ток через цепь не пойдет, если только все металлы имеют одну и ту же температуру. В противном случае мы имели бы дело с нарушением закона сохранения энергии. Таким образом, закон последовательных контактов Вольты может рассматриваться как *следствие закона сохранения энергии.* Это рассуждение неприменимо в тех случаях, когда в цепи есть электролиты и, следовательно, могут происходить химические реакции, за счет энергии которых и возбуждается электрический ток. Закон последовательных контактов в таких случаях не имеет места. Контактная разность потенциалов в вольтах колеблется для различных пар металлов от нескольких десятых до нескольких единиц. Как и работа выхода электронов, она сильно изменяется даже при ничтожных загрязнениях поверхностей соприкасающихся металлов, их окислении и т п. Поэтому для получения верных значений контактных разностей

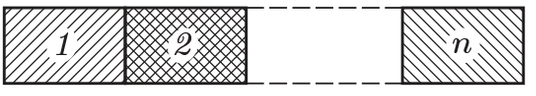


Рис. 233

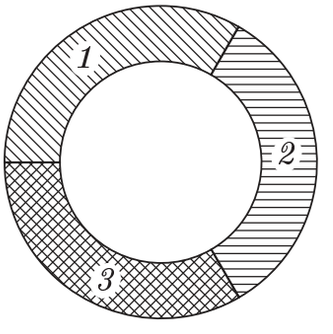


Рис. 234

потенциалов исследуемые вещества и их поверхности необходимо тщательно очищать, а измерения вести в вакууме.

2. Для объяснения возникновения контактной разности потенциалов будем пользоваться *моделью свободных электронов.* Допустим, что температура металла равна абсолютному нулю. Тогда все уровни энергии вплоть до границы Ферми μ будут заполнены электронами. Величина μ связана с концентрацией электронов проводимости n соотношением (99.6). Приведем два металла *1* и *2* в непосредственный контакт друг с другом (рис. 235). Так как энергии Ферми μ_1 и μ_2 у разных металлов разные, то будут разными и концентрации электронов проводимости n_1 и n_2. Пусть для определенности $\mu_2 > \mu_1$ и, следовательно, $n_2 > n_1$. В этом случае начнется диффузия электронов, в процессе которой электроны будут переходить от металла *2* к металлу *1*. Действительно, во втором металле имеются заполненные уровни, расположенные выше уровня Ферми μ_1 первого металла. Электроны с этих уровней могут переходить на соответствующие свободные уровни металла *1*, лежащие выше того же уровня μ_1. Обратный переход электронов из металла *1* в металл *2* невозможен, так как во втором металле все уровни с энергией μ_1 и ниже заполнены электронами. В результате диффузии металл *2* будет заряжаться положительно, а металл *1* — отрицательно. Потенциал металла *2* начнет повышаться, а металла *1* — понижаться. Уровень Ферми у первого металла поднимется, а у второго — опустится. Если потенциалы металлов обозначить через φ_{i1} и φ_{i2}, то энергии Ферми будут соответственно $\mu_1 + e\varphi_{i1}$ и $\mu_2 + e\varphi_{i2}$. На самой границе возникнет *скачок потенциала* или, что то же, электрическое поле, препятствующее процессу диффузии. Когда разность потенциалов $\varphi_{i2} - \varphi_{i1}$ достигнет определенной величины, диффузия прекратится. Это произойдет тогда, когда уровни Ферми обоих металлов сравняются, так как в этом случае в металле *1* уже не останется свободных уровней, на которые могли бы переходить электроны из металла *2*, а в металле *2* еще не освободятся уровни, на которые могли бы переходить электроны из металла *1*. Итак, в состоянии равновесия

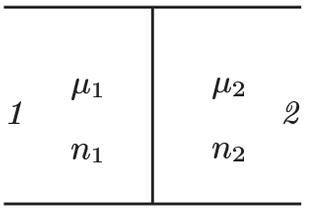


Рис. 235

$$\mu_1 + e\varphi_{i1} = \mu_2 + e\varphi_{i2}. \tag{104.1}$$

Отсюда

$$\varphi_{i2} - \varphi_{i1} = \frac{\mu_1 - \mu_2}{e}. \tag{104.2}$$

Так как $e < 0$, то при $\mu_2 > \mu_1$ получится $\varphi_{i2} > \varphi_{i1}$, как и должно быть.

При выводе предполагалось, что металл находится при абсолютном нуле температуры. Однако результат (104.2) остается верным и при других температурах. В самом деле, μ, а при наличии электрического поля $\mu + e\varphi$, есть не что иное, как *химический потенциал* электронного газа, отнесенный к одному электрону, а в состоянии равновесия химические потенциалы электронных газов обоих металлов должны

быть одинаковы (см т. II, § 82, 83). Нужно лишь иметь в виду, что при $T \neq 0$ химический потенциал μ зависит не только от концентрации электронов, но и от температуры металла.

В формуле (104.2) предполагается, что точка *1* лежит внутри металла *1*, а точка *2* — внутри металла *2*, что и отмечено индексом i. Поэтому разность потенциалов $\varphi_{i2} - \varphi_{i1}$ называется *внутренней контактной разностью потенциалов*. Из формулы (104.2) непосредственно следует, что внутренняя контактная разность потенциалов удовлетворяет закону последовательных контактов Вольты. Для оценки этой разности выразим с помощью формулы (99.6) энергии Ферми μ_1 и μ_2 через концентрации электронов проводимости n_1 и n_2. Тогда получим

$$\varphi_{i2} - \varphi_{i1} = \frac{h^2}{8me}\left(\frac{3}{\pi}\right)^{2/3}(n_1^{2/3} - n_2^{2/3}). \qquad (104.3)$$

После подстановки числовых значений получим

$$\varphi_{i2} - \varphi_{i1} = -3{,}66 \cdot 10^{-15}(n_1^{2/3} - n_2^{2/3}) \quad \text{(в вольтах)}.$$

Если принять, что на каждый атом металла приходится один электрон проводимости, то для меди, например, $\varphi_{i2} - \varphi_{i1} = 3{,}66 \cdot 10^{-15} n^{2/3} \sim$ $\sim$ 1 В. Поэтому внутренняя контактная разность потенциалов, по нашей оценке, будет порядка нескольких десятых или сотых долей вольта, а в некоторых случаях и порядка вольта. Понятно, что такая оценка не может считаться точной, поскольку в ее основе лежит модель свободных электронов.

3. Чтобы выйти из металла, электрон, как известно, должен совершить некоторую работу выхода A. Она совершается в тонком поверхностном слое, в котором на электрон действуют силы, стремящиеся втянуть его в металл. Толщина этого поверхностного слоя δ порядка нескольких атомных слоев. Заметив это, рассмотрим два металла, находящихся в непосредственном контакте между собой вдоль поверхности AB (рис. 236). При наличии такого контакта разность потенциалов между внутренними точками обоих металлов $\varphi_{i2} - \varphi_{i1}$ имеет вполне определенное значение. Пусть, далее, между теми же металлами имеется зазор CD макроскопической ширины. Возьмем внутри зазора две точки *1′* и *2′*, находящиеся в непосредственной близи от границ металлов CC' и DD'. Слова «в непосредственной близи» надо понимать в том смысле, что расстояние точки от границы металла больше толщины поверхностного слоя δ, но одинакового с ней порядка. Потенциалы во внешних точках *1′* и *2′* обозначим соответственно через φ_{e1} и φ_{e2}. Их разность $\varphi_{e2} - \varphi_{e1}$ называется *внешней контактной разностью потенциалов* между металлами. Ее величину можно связать с работой выхода электрона из металла. Сделаем это в предположении, что металл находится при абсолютном

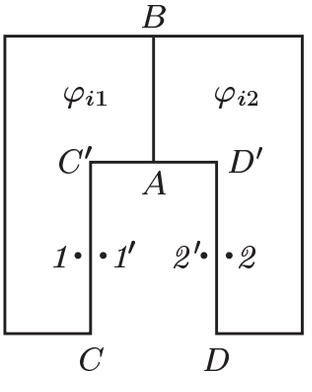


Рис. 236

нуле температуры. Тогда на границе CC' будем иметь

$$e\varphi_{e1} - (e\varphi_{i1} + \mu_1) = A_1,$$

а на границе DD'

$$e\varphi_{e2} - (e\varphi_{i2} + \mu_2) = A_2,$$

где A_1 и A_2 — работы выхода на этих границах. Вычитая почленно и воспользовавшись соотношением (104.1), получим

$$\varphi_{e2} - \varphi_{e1} = \frac{1}{e}(A_2 - A_1). \tag{104.4}$$

Наличие внешней контактной разности потенциалов означает, что в зазоре между CC' и DD', а также во внешнем пространстве между поверхностями контактирующих металлов существует электрическое поле, а сами поверхности заряжены электричеством.

Внешняя контактная разность потенциалов также подчиняется закону последовательных контактов Вольты. Действительно, образуем замкнутое кольцо из нескольких металлов (рис. 234), чтобы между соседними металлами были небольшие зазоры. Электродвижущая сила в таком кольце равна нулю, как это следует из закона сохранения энергии. Но ее можно представить в виде $\sum \Delta\varphi_i + \sum \Delta\varphi_e$, где суммирование ведется по всем контактам. Поскольку выше было доказано, что $\sum \Delta\varphi_i = 0$, то отсюда следует, что и $\sum \Delta\varphi_e = 0$.

4. Вольта доказал существование контактной разности потенциалов с помощью следующего классического опыта. На стержень электроскопа (рис. 237 *а*) был навинчен медный диск M, покрытый сверху

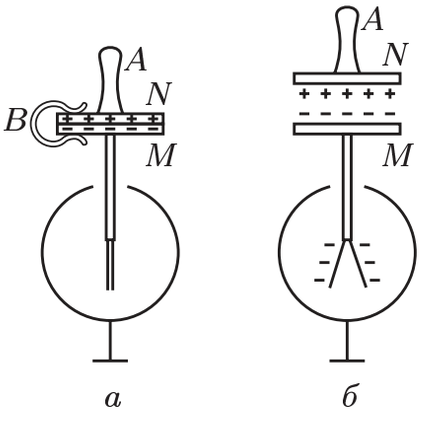


Рис. 237

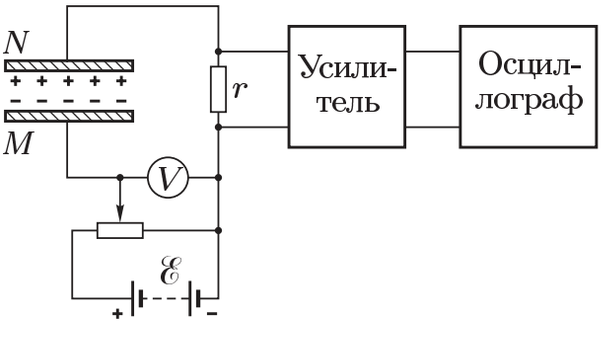


Рис. 238

тонким изолирующим слоем шеллака. На него был положен второй такой же цинковый диск N, снабженный изолирующей ручкой A. Образовавшийся конденсатор обладал довольно большой емкостью, так как слой шеллака был чрезвычайно тонок. Диски на короткое время соединялись медной проволокой B. Тогда между M и N возникала контактная разность потенциалов порядка одного вольта, причем цинк

заряжался положительно, а медь — отрицательно. Такая разность потенциалов, однако, слишком мала, чтобы листочки электроскопа заметно разошлись. Для ее обнаружения удаляют соединяющую проволоку B и за изолирующую ручку A поднимают верхний диск N. Так как при этом заряд конденсатора не изменяется, а емкость уменьшается во много раз, то разность потенциалов увеличивается во столько же раз, и листочки электроскопа заметно расходятся (рис. 237 *б*).

5. Для измерения внешней контактной разности потенциалов применяется *компенсационный метод.* Пластинки из исследуемых материалов M и N (рис. 238) располагают параллельно на небольшом расстоянии друг от друга. Одна из них закрепляется неподвижно, а другую при помощи соответствующего механического устройства заставляют колебаться с частотой в несколько десятков герц и амплитудой порядка долей миллиметра. Поскольку пластинки соединены между собой проводами измерительной схемы, устанавливаются контактная разность потенциалов φ и электрическое поле между ними. На внутренней поверхности каждой пластинки появляется электрический заряд $q = C\varphi$, где C — емкость между пластинками, в цепи — переменный электрический ток $\dot{q} = \varphi\, dC/dt$, а на нагрузочном сопротивлении r — переменное напряжение $r\dot{q} = r\varphi\, dC/dt$. Контактную разность потенциалов φ можно скомпенсировать напряжением противоположного знака от батареи. Тогда ток через сопротивление r прекратится. Для констатации этого напряжение с нагрузочного сопротивления, усиленное усилителем, подают на осциллограф. Пока компенсации нет, осциллограф регистрирует колебания напряжения на нагрузочном сопротивлении. Меняя с помощью делителя напряжение от батареи, добиваются того, чтобы колебания прекратились. Тогда вольтметр укажет искомую контактную разность потенциалов.

§ 105. Термоэлектрический ток

1. Согласно закону последовательных контактов Вольты в замкнутой цепи, состоящей из нескольких (например, трех) металлов или полупроводников (см. рис. 234), не происходит возбуждения электрического тока, если все тела цепи находятся *при одной и той же температуре.* Если, однако, температуры в местах контактов разные, то в цепи появляется электрический ток. Этот ток называется *термоэлектрическим.* Само явление возбуждения термоэлектрического тока, а также тесно связанные с ним *явления Пельтье и Томсона*, о которых будет сказано ниже, получили собирательное название *термоэлектричества.* Явление термоэлектрического тока было открыто в 1821 г. немецким физиком Зеебеком (1770–1831), который его подробно исследовал, хотя и дал неправильное истолкование существа своего открытия. (Зеебек думал, что под влиянием разности температур в разнородных проводниках, соединенных между собой, происходит выделение свободного магнетизма.)

Возбуждение термоэлектрического тока можно наблюдать на следующем приборе (рис. 239). К пластинке из сурьмы (Sb) припаяна пластинка из меди (Cu). Между пластинками располагается магнитная стрелка NS. Если один из спаев нагревать, то появляется электрический ток, и магнитная стрелка отклоняется. По направлению отклонения можно судить о направлении электрического тока. Оказывается, что через нагретый спай ток идет от меди к сурьме. Если охладить тот же спай, то направление термоэлектрического тока меняется на противоположное.

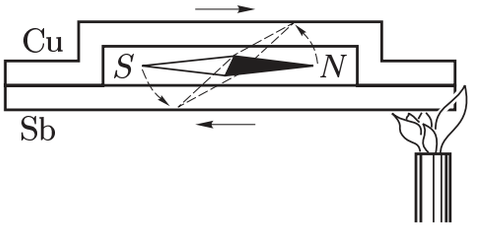


Рис. 239

Более эффектная демонстрация термоэлектрического тока осуществлена Полем. Толстая U-образная медная дуга перекрыта впаянным в нее в двух местах коротким мостиком *12* из константана (рис. 240 *а*). Место спая *1* подогревается пламенем газовой горелки через выступающий конец медной дуги. Для предотвращения нагревания спая *2* другой конец той же дуги загнут вниз и опущен в холодную воду. При разности температур между горячим и холодным спаями в 500 °C электродвижущая сила пары медь–константан составляет всего 0,027 В. Однако, ввиду малости сопротивления медной дуги, термоэлектрический ток может быть порядка сотен ампер. Для обнаружения такого большого тока берется кусок железа прямоугольной формы с двумя отверстиями, разрезанный на две половины *1* и *2*, как указано на рис. 240 *б* (поперечный разрез). Одна половина куска располагается сверху, а другая снизу медной дуги. Кусок в целом служит железным сердечником, а медная дуга — единственным витком, обвивающимся вокруг него. Нижний кусок 2 выполняет роль якоря. С помощью такого устройства на лекционных демонстрациях удается удерживать груз в несколько килограммов (вместе с якорем). Сам Поль доводил удерживаемый груз до 50 кг.

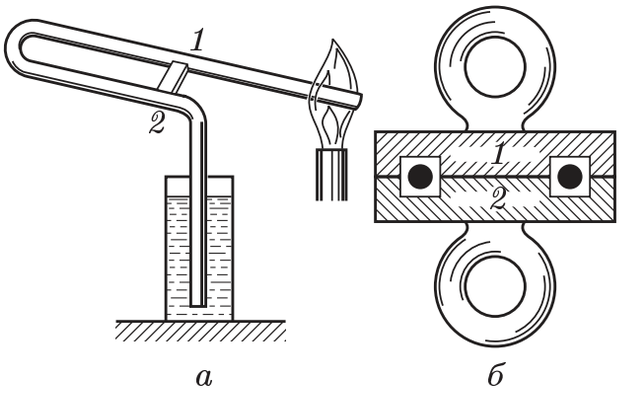


Рис. 240

Металл или полупроводник, к которому идет ток через более нагретый спай термоэлектрической пары, называется *положительным*, а другой — *отрицательным*. Первый играет роль анода, а второй — катода. Так, в термоэлектрической паре медь–сурьма (см. рис. 239) сурьма будет положительной, а медь — отрицательной. По этому признаку Зеебек, а затем и другие исследователи расположили металлы в так называемый *термоэлектрический ряд*, аналогичный ряду Вольты.

2. Электродвижущая сила всякой термопары складывается из электродвижущих сил обоих ее спаев. Электродвижущая сила одного спая зависит от рода контактирующих металлов и от температуры. Если обозначить ее через $f(t)$, то электродвижущая сила термопары представится разностью

$$\mathscr{E} = f(t_1) - f(t_2), \tag{105.1}$$

где t_1 — температура более, а t_2 — менее нагретого спаев. Производная $\alpha = df/dt$ называется *коэффициентом термоэлектродвижущей силы*. Как и функция f, коэффициент α является характеристикой обоих металлов термопары. На практике это создает определенные неудобства. Для избежания их условились величину α измерять по отношению к одному и тому же металлу, за который принимается свинец. Это означает, что α измеряется для термопары, у которой одна ветвь составлена из исследуемого материала, а другая — из свинца. Коэффициент термоэлектрической силы α_{12} металла 1 по отношению к металлу 2 определится формулой

$$\alpha_{12} = \alpha_1 - \alpha_2, \tag{105.2}$$

где α_1 и α_2 — значения коэффициентов термоэлектродвижущей силы металлов 1 и 2 соответственно по отношению к свинцу. Все эти величины зависят от чистоты материалов и сильно меняются с добавлением примесей.

Простейшей является линейная функция $f(t) = f_0 + \alpha t$. В этом случае

$$\mathscr{E} = \alpha(t_1 - t_2). \tag{105.3}$$

Существуют пары металлов, для которых такая простая формула оправдывается в довольно широких пределах. К ним относятся, например, термопары: (Cu, Bi), (Ag, Cu), (Au, Cu), (Pt, Fe). Для других термопар зависимость f от t более сложная, например: $f(t) = f_0 + \alpha t + \beta t^2$, где α и β — постоянные. В этом случае

$$\mathscr{E} = \alpha(t_1 - t_2)\left[1 + \frac{\beta}{\alpha}(t_1 + t_2)\right]. \tag{105.4}$$

К формуле такого вида пришел М. П. Авенариус (1835–1895) на основе анализа экспериментальных данных и Тэт (1831–1901) на основе полутеоретических соображений. Согласно формуле (105.4) величина $\mathscr{E}$ обращается в нуль при $t_1 = t_2$ и при $t_1 + t_2 = -\alpha/\beta$. Величина $\tau = (t_1 + t_2)/2$ называется *температурой нейтральной точки*. Если при неизменном значении t_2 повышать температуру t_1, то $\mathscr{E}$ будет возрастать по параболическому закону, достигнет максимума при $t_1 = \tau$, а затем обратится в нуль и изменит знак при температуре $t_1 = 2\tau - t_2$. Температура $t_1 = 2\tau - t_2$, при которой величина $\mathscr{E}$ проходит через нуль, называется *точкой инверсии*.

3. Необходимость возникновения термоэлектрического тока видна уже из формулы (104.2), определяющей внутреннюю контактную разность потенциалов на границе двух металлов. Дело в том, что химические потенциалы μ_1 и μ_2 зависят от температуры. Поэтому если

температуры спаев термопары неодинаковы, то будут неодинаковыми и внутренние контактные разности потенциалов. Это ведет к нарушению электрического равновесия и возникновению термоэлектрического тока.

Рассмотрим этот вопрос подробнее. Рассуждение проведем применительно к полупроводникам, так как таким путем будет достигнута большая общность, поскольку в полупроводниках ток переносят не только электроны, но и дырки. Предположим сначала, что носителями тока являются только электроны. Возьмем единственный полупроводник, имеющий форму стержня. Будем нагревать один из его концов, оставляя температуру второго неизменной. Концентрация электронов проводимости в полупроводниках возрастает с повышением температуры. Поэтому у нагретого конца она будет больше, чем у холодного. Электроны от нагретого конца начнут диффундировать к холодному. Горячий конец, теряя электроны, будет заряжаться положительно, а холодный — отрицательно. Внутри полупроводника возникнет электрическое поле $\mathbf{E}$, направленное от горячего конца к холодному. Тормозя диффузию электронов, оно будет нарастать до тех пор, пока диффузия не прекратится. Тогда установится *стационарное состояние*, в котором горячий конец полупроводника окажется заряженным положительно, а холодный — отрицательно. Градиент концентрации электронов, имеющийся в полупроводнике, в отношении возбуждения электрического тока формально эквивалентен полю сторонних сил $\mathbf{E}^{\text{стор}}$, уравновешивающему в установившемся состоянии электрическое поле $\mathbf{E}$: $\mathbf{E}^{\text{стор}} = -\mathbf{E}$. Таким образом, в электронном полупроводнике поле сторонних сил $\mathbf{E}^{\text{стор}}$ направлено от холодного к горячему концу (рис. 241 *а*).

Приведенное рассуждение применимо и к полупроводнику с дырочной проводимостью. Только теперь следует говорить о *диффузии дырок*, а не электронов. В дырочном полупроводнике нагретый конец зарядится отрицательно, а холодный — положительно. В установившемся состоянии поле $\mathbf{E}$ будет направлено от холодного конца к горячему, а поле сторонних сил $\mathbf{E}^{\text{стор}}$ — от горячего к холодному (рис. 241 *б*).

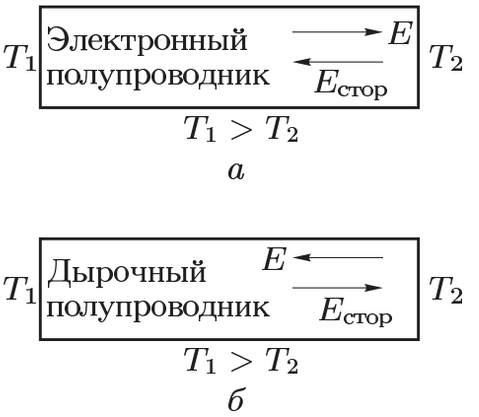
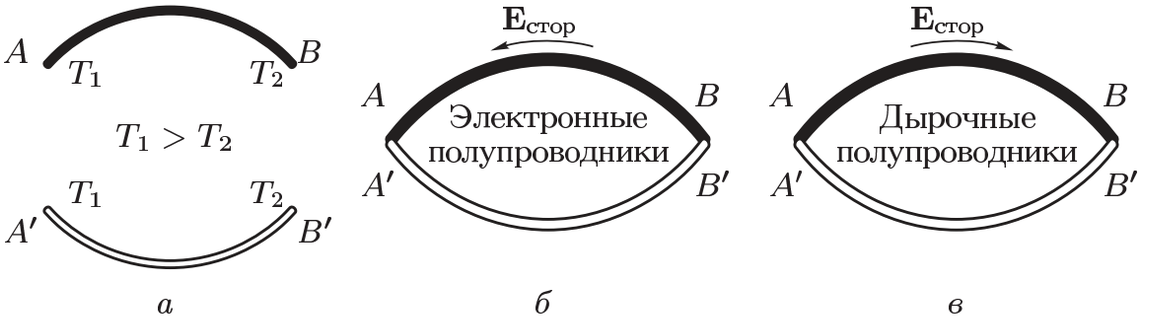


Рис. 241

В проводниках смешанного типа от горячего конца к холодному диффундируют одновременно и электроны, и дырки, возбуждая электрические поля в противоположных направлениях. В таких полупроводниках, в зависимости от величины концентраций и подвижностей электронов и дырок, результирующее электрическое поле $\mathbf{E}$ и поле сторонних сил $\mathbf{E}^{\text{стор}} = -\mathbf{E}$ могут быть направлены как к горячему, так и к холодному концам стержня. В некоторых случаях электрические поля, возникающие из-за диффузии электронов и дырок, компенсируют друг

друга, т. е. $\mathbf{E} = -\mathbf{E}^{\text{стор}} = 0$. Тогда никакой разности потенциалов между концами полупроводника не возникает. Такой случай имеет место в свинце. Вот почему коэффициент термоэлектродвижущей силы α всех материалов принято выражать относительно свинца.

Разность потенциалов между нагретым и холодным концами полупроводника зависит от материала последнего. Поэтому если взять две проволоки AB и $A'B'$ из разных материалов, концы которых A и A' поддерживаются при одной, а концы B и B' при другой температуре (рис. 242 *а*), то равновесные разности потенциалов на концах каждой



а *б* *в*

Рис. 242

проволоки будут разными. Соединим в термопару конец A с A', а конец B с B' (рис. 242 *б*, *в*). Тогда электрическое равновесие нарушится и в цепи $ABB'A'A$ потечет электрический ток. Это и есть термоэлектрический ток. Допустим, что проволока $A'B'$ сделана из свинца. Тогда в ней $\mathbf{E}^{\text{стор}} = 0$ и весь ток будет обусловлен сторонними силами, действующими в проволоке AB. Если AB изготовлена из электронного полупроводника, то ток через горячий спай потечет от A к A' (рис. 242 *б*). Если же она изготовлена из дырочного полупроводника, то он потечет в противоположном направлении (рис. 242 *в*). Мы видим, что по знаку коэффициента термоэлектродвижущей силы α можно судить о природе носителей тока в полупроводниках: коэффициент α положителен для дырочных и отрицателен для электронных полупроводников.

4. На основании изложенных соображений следует ожидать, что *термоэлектрические свойства должны быть выражены значительно сильнее у полупроводников, чем у металлов.* Действительно, электроны в металлах находятся в состоянии вырождения. Их энергия весьма слабо зависит от температуры, а концентрация имеет одинаковые значения как при низких, так и при высоких температурах. Слабо зависит от температуры у металлов и положение уровня химического потенциала. Поэтому коэффициент термоэлектродвижущей силы α для металлов и сплавов не превосходит нескольких микровольт на кельвин (табл. 6). В полупроводниках, напротив, концентрации дырок и электронов проводимости, а также все параметры, перечисленные выше, сильно меняются с температурой. Поэтому для них коэффициент термоэлектродвижущей силы α значительно больше, чем у металлов, и может превышать 1000 мкВ/К (табл. 7).

Таблица 6. Значения коэффициента термоэлектродвижущей силы для некоторых металлов и сплавов

Материал	α, мкВ/К	Материал	α, мкВ/К	Материал	α, мкВ/К
Висмут	−68,0	Натрий	−6,5	Золото	+2,9
Константан	−38,0	Платина	−4,4	Цинк	+3,1
Копель	−38,0	Ртуть	−4,4	Вольфрам	+3,6
Никель	−20,8	Алюминий	−0,4	Кадмий	+4,6
Нихром	−18,0	Олово	−0,2	Молибден	+7,6
Алюмель	−17,3	Магний	0,0	Железо	+15,0
Калий	−13,6	Свинец	0,0	Хромель	+24
Палладий	−8,9	Серебро	+2,7	Сурьма	+43

Таблица 7. Значения коэффициента термоэлектродвижущей силы для некоторых полупроводников

Материал	α, мкВ/К	Материал	α, мкВ/К	Материал	α, мкВ/К
Tl_2S	−780	Pb–Te–Se	−160	Bi_2Te_3	+170
MoS	−770	PbS	−160	Bi_2Se_3	+200
V_2O_5	−750	PbSe	−160	SbZn	+200
WO_3	−740	PbTe	−160	NiO	+240
CuO	−700	SnO	−140	MoS_2	+300
Fe_2O_3	−613	CdO	−40	Mn_2O_3	+385
FeO	−500	CuS	−7	CoO	+450
Fe_3O_4	−430	FeS	+26	SiC (черный)	+800
SiC (зеленый)	−300	CdO	+30	Tl_2S	+800
FeS_2	−200	Bi_2S_3	+80	Se	+1000
$CoSb_3$	−200	Sb_2Te_3	+100	CuO	+1120
MgO_3H_2	−200	$FeTiO_3$	+140	Cu_2O	+1200
Bi–Te–Se	−170	Bi–Sb–Te	+160		

5. Термоэлектричество может быть использовано для генераций электрического тока. Отдельная термопара (термоэлемент) имеет слишком малую электродвижущую силу. Для получения значительных напряжений термоэлементы соединяют последовательно в батареи, как указано на схематическом рис. 243. Все нечетные спаи поддерживаются при одной, а четные — при другой температуре. В результате электродвижущие силы отдельных термоэлементов складываются. Термобатарея подобна тепловой машине, включенной между нагревателем и холодильником. В такой машине большая часть теплоты, полученной от нагревателя, бесполезно теряется в результате теплопроводности и джоулева нагрева. Термобатареи из металлических термопар обладают слишком малыми КПД (около 0,1 %). Поэтому металлические тер-

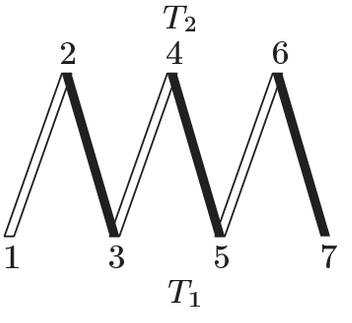


Рис. 243

мопары используются только для измерения температур и потоков лучистой энергии (см т. II, § 5). Значительно выгоднее термобатареи из полупроводников. Одна ветвь термопары в них изготовляется из проводника с электронной, а другая — с дырочной проводимостью. КПД полупроводниковых термоэлементов достигает 15% и в дальнейшем, несомненно, будет повышен еще больше. В нашей стране и других промышленно развитых странах ведется разработка полупроводниковых термоэлектрогенераторов для прямого преобразования в электрическую тепловой энергии Солнца, ядерных реакторов и пр.

§ 106. Явление Пельтье

1. В 1834 г. французский часовщик Пельтье (1785–1845) опубликовал статью о температурных аномалиях, наблюдаемых на границе двух различных проводников при прохождении через них электрического тока. Сам Пельтье не совсем уяснил сущность открытого им явления. Истинный смысл явления Пельтье был установлен в 1838 г. Ленцем (1804–1865). В углубление на стыке стержней из висмута и сурьмы Ленц поместил каплю воды. При пропускании электрического тока в одном направлении вода замерзала, а при пропускании в противоположном направлении образовавшийся лед таял. Тем самым было установлено, что при прохождении через контакт двух проводников электрического тока, в зависимости от направления последнего, помимо джоулевой теплоты выделяется или поглощается дополнительная теплота, которая получила название *теплоты Пельтье.* В этом и состоит *явление Пельтье.* Таким образом, оно является обратным по отношению к явлению Зеебека. В отличие от теплоты Джоуля–Ленца, которая пропорциональна квадрату силы тока, *теплота Пельтье пропорциональна первой степени силы тока и меняет знак при изменении направления последнего.* Теплоту Пельтье, как показали экспериментальные исследования, можно выразить формулой

$$Q_{П} = Пq, \tag{106.1}$$

где q — количество прошедшего электричества, а П — так называемый *коэффициент Пельтье*, значение которого зависит от природы контактирующих материалов и от их температуры. Теплота Пельтье $Q_{П}$ считается положительной, если она выделяется, и отрицательной, если она поглощается.

Сам Пельтье продемонстрировал открытое им явление на следующем изящном опыте. Две полоски из сурьмы AB и висмута CD (рис. 244) были спаяны в виде креста (*креста Пельтье*). К концам A и C можно было присоединять батарею, а к концам B и D — гальванометр. При пропускании электрического тока от батареи в направлении от сурьмы к висмуту спай нагревался. После этого батарея отключалась и присоединялся гальванометр. Последний обнаруживал ток в направлении от висмута к сурьме, т. е. в направлении, противоположном исходному току от батареи. Именно этого и следовало ожидать согласно принципу Ле Шателье–Брауна. Если повторить опыт, пропуская ток от

батареи в противоположном направлении, то спай охлаждается, а ток через гальванометр также меняет направление на противоположное.

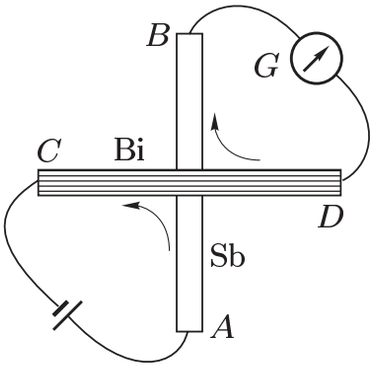


Рис. 244

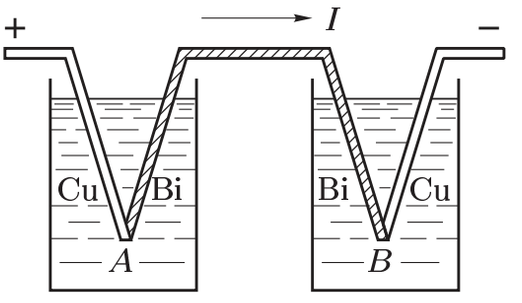


Рис. 245

2. Для количественного исследования явления Пельтье Леру (1832–1907) припаял к концам висмутовой проволоки AB медные провода и опустил спаи в два калориметра (рис. 245). Пропуская через оба спая один и тот же ток I в течение определенного времени t, Леру измерял теплоту, выделившуюся в каждом калориметре за это время. Если сопротивления проводов R, опущенных в калориметры, одинаковы, то в калориметрах выделится одна и та же джоулева теплота, а именно RI^2t. Теплота Пельтье, напротив, в одном калориметре будет положительна, а в другом отрицательна: если ток I идет от меди к висмуту, то теплота Пельтье выделяется, а если от висмута к меди, то поглощается. Таким образом, можно написать

$$Q_1 = RI^2t + \Pi It, \quad Q_2 = RI^2t - \Pi It,$$

где Q_1 — полное количество теплоты, выделившееся в первом, а Q_2 — во втором калориметрах. Исключая почленным вычитанием джоулеву теплоту, получим

$$2\Pi It = Q_1 - Q_2.$$

Из этого соотношения и можно найти коэффициент Пельтье. Таким путем найдено, что для металлов коэффициент Пельтье порядка 10^{-2}–10^{-3} В, а для полупроводников порядка $3 \cdot 10^{-1}$–10^{-3} В.

3. Классическая теория объясняла явление Пельтье тем, что электроны, переносимые током из одного металла в другой, ускоряются или замедляются под действием внутренней контактной разности потенциалов между металлами. В первом случае кинетическая энергия электронов увеличивается, а затем выделяется в виде тепла. Во втором случае она уменьшается и убыль энергии пополняется за счет тепловых колебаний атомов второго проводника. В результате происходит охлаждение. С этой точки зрения следовало бы ожидать, что коэффициент Пельтье будет совпадать с контактной разностью потенциалов. На самом деле это неверно. Дело в том, что по классической теории средняя кинетическая энергия теплового движения электронов в обоих контактирующих металлах одинакова. А это неверно вследствие различного

положения уровней Ферми в обоих металлах. Классическое объяснение учитывает только различие *потенциальных энергий* по разные стороны границы раздела металлов, считая средние кинетические энергии их одинаковыми. Для того чтобы исправить объяснение, надо изменение потенциальной энергии при переносе электрона из одного металла в другой заменить изменением *полной энергии.* Исправленное таким образом объяснение, разумеется, справедливо не только для металлов, но и для полупроводников с электронной проводимостью.

Совершенно аналогичное объяснение можно привести и для того случая, когда в контакте находятся два полупроводника с дырочной проводимостью. Через границу раздела переходят, конечно, электроны. По одну сторону границы происходит рождение пар электрон–дырка, по другую — рекомбинация электронов с дырками. Один из этих процессов сопровождается выделением, другой — поглощением энергии. От соотношения между выделяющейся и поглощающейся энергией зависит знак коэффициента Пельтье.

Эффект Пельтье, как и все термоэлектрические явления, выражен особенно сильно в цепях, составленных из электронных и дырочных полупроводников. Рассмотрим контакт таких полупроводников. Допустим, что электрическое поле имеет такое направление, что ток идет от дырочного полупроводника к электронному. Тогда электроны в электронном полупроводнике и дырки в дырочном будут двигаться навстречу друг другу. Электрон из свободной зоны электронного проводника после прохождения через границу раздела попадает в заполненную зону дырочного полупроводника и там занимает место дырки. В результате такой рекомбинации освобождается энергия, которая и выделяется в контакте в виде тепла. Рассмотрим теперь случай, когда ток проходит через границу раздела от электронного полупроводника к дырочному. Тогда электроны в электронном и дырки в дырочном полупроводниках будут двигаться в противоположные стороны. Дырки, уходящие от границы раздела, будут пополняться в результате образования новых пар при переходах электронов из заполненной зоны дырочного полупроводника в свободную. На образование таких пар требуется энергия, которая поставляется тепловыми колебаниями атомов решетки. Электроны и дырки, образующиеся при рождении таких пар, увлекаются в противоположные стороны электрическим полем. Поэтому пока через контакт идет ток, непрерывно происходит рождение новых пар. В результате в контакте теплота будет поглощаться. Таким образом, если ток идет от дырочного полупроводника к электронному, то теплота Пельтье выделяется. При обратном направлении тока она поглощается.

А.Ф. Иоффе предложил использовать явление Пельтье в полупроводниках для создания охлаждающих устройств. Отличаясь простотой, такие устройства успешно прошли испытания в лабораторных условиях. Термоэлектрический метод охлаждения обладает рядом преимуществ по сравнению с другими методами охлаждения.

§ 107. Термодинамика термоэлектрических явлений. Явление Томсона

1. Клаузиус в 1853 г. применил к явлениям термоэлектричества принципы термодинамики. Рассмотрим термопару, горячий спай которой поддерживается при постоянной температуре T_1, а холодный — при постоянной температуре T_2 (рис. 246). При прохождении тока I в спае *1* в единицу времени выделяется теплота Пельтье $\Pi_1 I$, а в спае *2* поглощается теплота $\Pi_2 I$. (Их следует рассматривать как величины алгебраические — они могут быть и положительными, и отрицательными.) Происходит также выделение джоулевой теплоты. Однако последним можно пренебречь, если разность температур $T_1 - T_2$ взять бесконечно малой. Действительно, теплота Пельтье пропорциональна первой степени силы тока I, тогда как джоулева теплота — второй. Когда $T_1 - T_2 \to 0$, ток I стремится к нулю, и джоулева теплота становится исчезающе малой по сравнению с теплотой Пельтье. Если отвлечься также от передачи теплоты посредством теплопроводности, то прохождение термоэлектрического тока можно рассматривать как обратимый круговой процесс и применить к нему равенство Клаузиуса

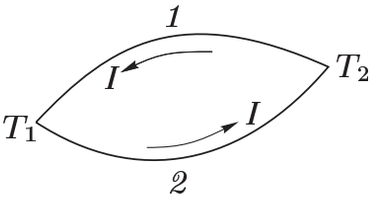


Рис. 246

$$\frac{\Pi_1}{T_1} - \frac{\Pi_2}{T_2} = 0. \tag{107.1}$$

Это соотношение получено для бесконечно малых разностей температур $T_1 - T_2$. Но оно в тех же предположениях остается верным и при конечных значениях $T_1 - T_2$. Чтобы убедиться в этом, запишем его в дифференциальной форме:

$$\frac{d}{dT}\left(\frac{\Pi}{T}\right) = 0, \tag{107.2}$$

а затем проинтегрируем. Тогда получим

$$\Pi/T = \text{const}. \tag{107.3}$$

Применим теперь к рассматриваемому процессу первое начало термодинамики. Термоэлектродвижущая сила $\mathscr{E} = \alpha(T_1 - T_2)$ совершает в единицу времени работу $\mathscr{E}I$. Приравнивая ее теплоте Пельтье и сокращая на I, придем к соотношению

$$\alpha(T_1 - T_2) = \Pi_1 - \Pi_2 = \frac{d\Pi}{dT}\,(T_1 - T_2),$$

откуда

$$d\Pi/dT = \alpha. \tag{107.4}$$

С учетом (107.3) отсюда получаем

$$\alpha = \Pi/T = \text{const}. \tag{107.5}$$

Таким образом, по теории Клаузиуса термоэлектродвижущая сила при всех температурах T_1 и T_2 должна быть пропорциональна $T_1 - T_2$, т. е. выражаться формулой $\mathscr{E} = \alpha(T_1 - T_2)$. Этот результат, как правило, не согласуется с опытом (см. § 105, п. 2).

2. Расхождение теории Клаузиуса с опытом было устранено Вильямом Томсоном, который независимо от Клаузиуса и почти одновременно с ним (1854 г.) развил *термодинамическую теорию термоэлектричества.* Томсон обратил внимание на то, что различные участки термопары нагреты неодинаково, а потому их физические состояния также неодинаковы. Неравномерно нагретый проводник должен вести себя как система находящихся в контакте физически разнородных участков. На этом основании Томсон пришел к заключению и подтвердил его экспериментально, что на границах таких участков должно происходить выделение или поглощение теплоты Пельтье. Такая теплота получила название *теплоты Томсона*, а само явление — *явления Томсона.*

С точки зрения электронной теории явление Томсона объясняется очень просто. Рассмотрим полупроводник с электронной проводимостью. Пусть $T_1 > T_2$, т. е. градиент температуры направлен от точки *2* к точке *1* (рис. 247 *а*). Из-за диффузии концентрация электронов в точ-

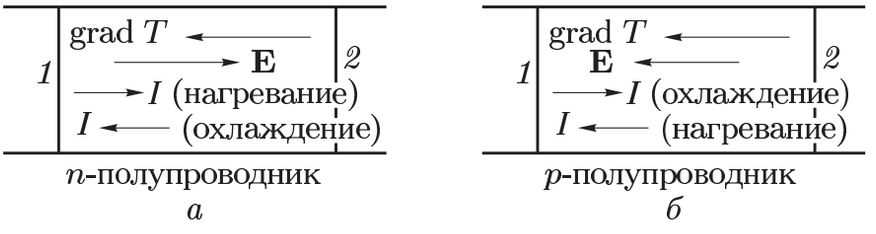


Рис. 247

ке *1* сделается меньше, чем в точке *2*. Возникнет электрическое поле $\mathbf{E}$, направленное от *1* к *2*, т. е. против градиента температуры. Если по проводнику течет ток в направлении $\operatorname{grad} T$ (т. е. электроны движутся в направлении поля $\mathbf{E}$), то поле $\mathbf{E}$ будет замедлять электроны, а участок полупроводника *12* станет охлаждаться. Если же ток течет в обратном направлении, то произойдет нагревание участка *12*. В дырочном полупроводнике соотношения будут обратными (рис. 247 *б*). Явление выглядит так, как если бы на обычный поток теплоты, вызванный теплопроводностью, накладывался дополнительный поток теплоты, связанный с прохождением электрического тока. В дырочных полупроводниках дополнительный поток теплоты направлен в ту же сторону, куда течет электрический ток, в электронных направления электрического тока и теплоты противоположны. Эффект Томсона считается положительным, если электрический ток, текущий в направлении градиента температуры, вызывает нагревание проводника, и отрицательным, если при том же направлении он охлаждает проводник.

Для количественного исследования явления Томсона Леру (1832–1907) в 1867 г. брал два одинаковых стержня AB и CD (рис. 248) из испытуемого материала. Концы A и C были соединены вместе

и поддерживались при температуре 100 °C. Температура свободных концов B и D была 0 °C. Пока электрический ток не был замкнут, термопары в точках a и b показывали одинаковые температуры. При пропускании электрического тока в одном стержне дополнительный поток теплоты проходил слева направо, а в другом — справа налево. В результате между точками a и b возникала разность температур, которая и регистрировалась термопарами. При изменении направления тока знак разности температур изменялся на противоположный.

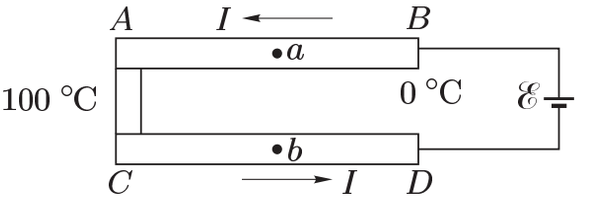


Рис. 248

Теплота Томсона, выделяющаяся в единицу времени на участке провода длиною dx, определяется выражением

$$dQ = \sigma I \frac{dT}{dx}\, dx, \tag{107.6}$$

где σ — так называемый *коэффициент Томсона.* Он зависит от материала провода и от температуры T. При этом за положительное принято направление градиента температуры, т. е. направление в сторону возрастания температуры.

3. Теперь нетрудно исправить теорию Клаузиуса, учтя в ней теплоту Томсона. Если разность температур $T_1 - T_2$ бесконечно мала, а ветви термопары бесконечно короткие (рис. 248), то в первой ветви в единицу времени выделится теплота Томсона $\sigma_1 I(T_1 - T_2)$, а во второй поглотится $\sigma_2 I(T_1 - T_2)$. С учетом этого равенство Клаузиуса примет вид

$$\frac{\Pi_1}{T_1} - \frac{\Pi_2}{T_2} + \left(\frac{\sigma_1}{T_1} - \frac{\sigma_2}{T_2}\right)(T_1 - T_2) = 0.$$

Переходя к дифференциальной форме и принимая во внимание, что разность $T_1 - T_2$ бесконечно мала, отсюда получим

$$\frac{d}{dT}\frac{\Pi}{T} = \frac{\sigma_2 - \sigma_1}{T}, \tag{107.7}$$

или

$$\frac{\Pi}{T} - \frac{d\Pi}{dT} = \sigma_1 - \sigma_2. \tag{107.8}$$

Первое начало теперь дает

$$\Pi_1 - \Pi_2 + (\sigma_1 - \sigma_2)(T_1 - T_2) = \alpha(T_1 - T_2),$$

или после дифференцирования и сокращения на $T_1 - T_2$

$$\frac{d\Pi}{dT} + (\sigma_1 - \sigma_2) = \alpha. \tag{107.9}$$

Отсюда с учетом (107.8) получаем

$$\Pi = \alpha T, \tag{107.10}$$

т. е. такое же соотношение (107.5), как и в теории Клаузиуса. Однако теперь коэффициент термоэлектродвижущей силы α не постоянен, а зависит от температуры.

4. Слабая сторона термодинамической теории Клаузиуса–Томсона состоит в том, что она принимает во внимание только *обратимые процессы*, происходящие в термоэлектрической цепи. Между тем в цепи происходят и *необратимые процессы*: теплопроводность и выделение джоулевой теплоты. От последнего можно освободиться, перейдя к бесконечно малым циклам, как сделано выше. Но теплота, переносимая теплопроводностью, того же порядка, а иногда и значительно больше, чем теплота Пельтье. Если все температуры поддерживаются постоянными, то теплопроводность не влияет на баланс энергии, так как она только переносит теплоту в неизменном количестве из одних участков цепи в другие. Однако наличие необратимых переносов теплоты делает сомнительным применение второго начала термодинамики в обратимой форме. Томсон обошел эту трудность, заметив, что теплопроводность и джоулево тепло являются побочными эффектами, органически не связанными с явлениями Зеебека, Пельтье и Томсона. Теплопроводность и джоулево тепло не влияют на термоэлектрические явления, и по этой причине от них можно совсем отвлечься. Эти соображения, конечно, мало убедительны и лишены доказательной силы. Онзагер позднее указал условия, когда допустимо раздельное рассмотрение обратимых и необратимых процессов. По-видимому, в металлах и полупроводниках эти условия выполняются, по крайней мере приближенно. В пользу этого говорит то обстоятельство, что выводы термодинамической теории Клаузиуса–Томсона в пределах точности измерений согласуются с опытом.

ЗАДАЧА

Тэт (1831–1901) ввел предположение, что коэффициент Томсона σ пропорционален термодинамической температуре. Показать, что это предположение приводит к формуле Авенариуса для электродвижущей силы $\mathscr{E}$.

Р е ш е н и е. Из уравнений (107.9) и (107.10) получаем

$$\frac{d}{dT}\left(T\frac{d\mathscr{E}}{dT}-\mathscr{E}\right)+\sigma_1-\sigma_2=0.$$

Интегрируя это уравнение с учетом справедливости предположения Тэта, приходим к требуемому результату.

§ 108. Выпрямляющее действие контактов полупроводников

1. Контакт двух металлов, полупроводников или металла с полупроводником обладает *выпрямляющим действием.* Это значит, что сопротивление такого контакта зависит от *направления* проходящего через него тока: в одном направлении (*запорном*) оно велико, в противоположном (*пропускном*) — мало. *Особенно резко выпрямляющее действие*

выражено на границе дырочного (p) *и электронного* (n) *полупроводников, когда работа выхода электрона из электронного полупроводника меньше, чем из дырочного.* О таком контакте говорят как об *электронно-дырочном p–n-контакте* или *переходе.* Хорошие p–n-переходы не удается получить, прижимая один полупроводник к другому, так как из-за шероховатости поверхностей тел соприкосновение всегда будет происходить лишь в нескольких точках, в воздушных зазорах между телами будут образовываться пленки окислов и т. п. Поэтому для получения хороших p–n-переходов в пластинку чистого полупроводника вводят две примеси — донорную и акцепторную (см. § 100). Первая сообщает полупроводнику *электронную*, а вторая — *дырочную проводимость.* Например, если пластинка сделана из германия или кремния, то в качестве донора можно взять элемент пятой группы периодической системы (фосфор, мышьяк и пр.), а в качестве акцептора — третьей (бор, индий и пр.). В результате в одной половине пластинки возникает электронная, в другой — дырочная проводимость, а между обеими половинами — тонкий переходный слой. Это и есть p–n-переход.

2. Рассмотрим теперь физические явления в электронно-дырочных контактах, с которыми связана односторонняя проводимость последних. Допустим сначала, что контакта между полупроводниками нет. Предположим, что оба полупроводника сделаны из одного и того же материала. Однако к одному из них добавлено небольшое количество донорной, а к другому — акцепторной примесей. В соответствии с этим границы энергетических зон в обоих полупроводниках совпадают (рис. 249 *а*). Но примесные (промежуточные) уровни в запре-

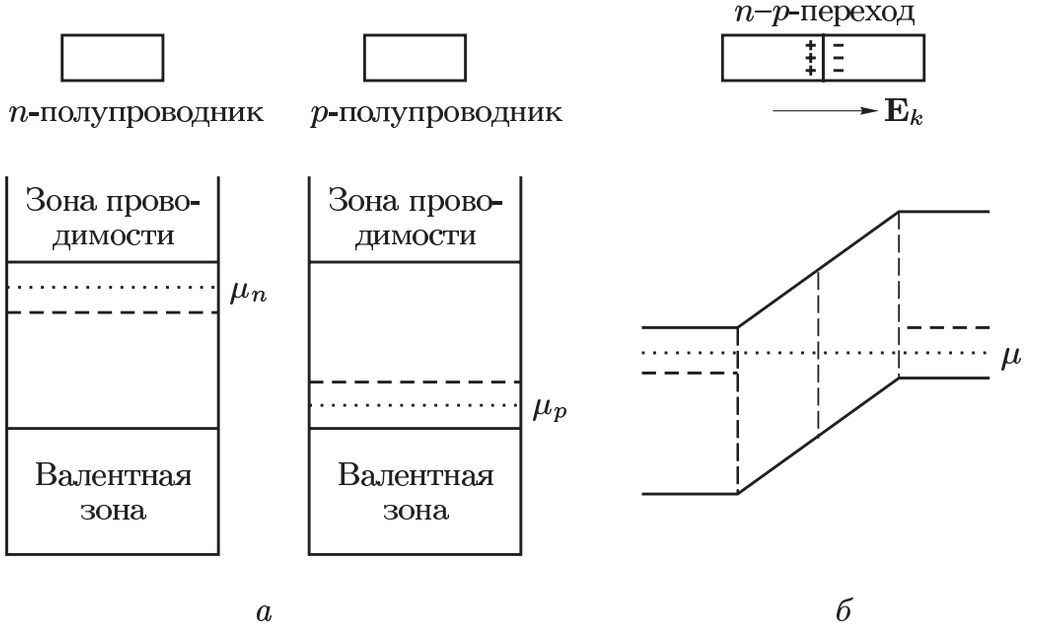


Рис. 249

щенной зоне расположены в электронном полупроводнике *вблизи зоны проводимости*, а в дырочном — *вблизи валентной зоны.* Благодаря этому средняя энергия электрона проводимости и уровень химического

потенциала μ в первом полупроводнике будут выше, а работа выхода — меньше, чем во втором полупроводнике.

Допустим теперь, что полупроводники приведены в контакт друг с другом (рис. 249 *б*). Так как работа выхода электрона из электронного полупроводника меньше, чем из дырочного, то электроны в большем количестве будут переходить из первого во второй. Электронный полупроводник начнет заряжаться положительно, а дырочный — отрицательно. В тонком слое между ними появится контактное электрическое поле, направленное от электронного к дырочному полупроводнику. В результате этого энергетические уровни электронного полупроводника начнут опускаться, а дырочного — подниматься. Контактное электрическое поле будет тормозить переход электронов из электронного в дырочный полупроводник. Процесс перехода электронов прекратится, когда уровни химического потенциала в обоих полупроводниках сделаются одинаковыми. Слева от переходной области энергетические уровни донорных примесей будут на тех же расстояниях от зоны проводимости, что и до контакта, а справа эти расстояния возрастут. Вследствие этого переходная область, в особенности ее правая граница, будет сильно обеднена электронами проводимости. Аналогично, переходная область, и в особенности ее левая граница, будет обеднена и дырками. Можно сказать, что контактное электрическое поле $\mathbf{E}_к$ выталкивает электроны проводимости в глубь электронного, а дырки — в глубь дырочного полупроводника. В результате переходный слой на границе обоих полупроводников оказывается сильно обедненным обоими носителями тока: и электронами, и дырками. Поэтому, несмотря на малую толщину (порядка 10^{-4}–10^{-6} см), электрическое сопротивление переходного слоя оказывается во много раз больше суммарного сопротивления обоих контактирующих полупроводников.

3. Допустим теперь, что наложено внешнее электрическое поле $\mathbf{E}$, направленное от электронного полупроводника к дырочному, т. е. одинаково с контактным полем $\mathbf{E}_к$. Ввиду большого сопротивления переходного слоя практически вся разность потенциалов, соответствующая наложенному полю, сосредоточится внутри переходного слоя, так что поле $\mathbf{E}_к$ внутри слоя может быть очень сильным. Такое поле усиливает контактное поле $\mathbf{E}_к$ и тем самым еще больше уменьшает концентрацию дырок и электронов проводимости в переходном слое. Сопротивление последнего еще больше возрастет. Практически ток через контакт не пойдет. Не то будет, когда внешнее поле $\mathbf{E}$ направлено против контактного поля $\mathbf{E}_к$. Достаточно уже небольшого поля $\mathbf{E}$, чтобы оно компенсировало поле $\mathbf{E}_к$. Тогда электроны проводимости и дырки будут беспрепятственно проникать в переходный слой, и сопротивление последнего практически исчезнет. Ток через контакт будет проходить. Если ток переменный, то в зависимости от его направления и силы сопротивление контакта становится пульсирующим, изменяясь от нуля практически до бесконечности. В соответствии с этим ток через контакт будет проходить только тогда, когда он направлен от

дырочного проводника к электронному. На этом принципе работают полупроводниковые выпрямители.

К этому надо добавить, что в любом полупроводнике, помимо *основных носителей тока*, имеется относительно малое количество *неосновных* (см. конец § 100). В электронных полупроводниках, помимо электронов проводимости, есть и дырки, а в дырочных — электроны. Очевидно, если внешнее поле **E** направлено от электронного к дырочному полупроводнику, то оно делает беспрепятственным проникновение в переходный слой неосновных носителей. В этом случае через контакт не идет ток основных носителей, но идет ток неосновных носителей. Однако, ввиду относительно малой концентрации последних, этот ток также мал.

4. Полупроводниковые выпрямители с успехом заменяют выпрямительные устройства старых систем. Эти выпрямители отличаются высокими КПД, малыми габаритами и невысокой стоимостью. Существует много различных систем полупроводниковых выпрямителей. Не останавливаясь на этом вопросе, опишем кратко устройство одного из типов германиевого выпрямителя. Он состоит из пластинки германия с электронной проводимостью, в которую с одной стороны вварен шарик индия, а с другой — шарик олова. Оловянный шарик служит только электродом для включения выпрямителя в цепь. Индий же играет принципиальную роль, сообщая германию дырочную проводимость. При нагревании в процессе сварки индий диффундирует в германий, вследствие чего вблизи индиевого электрода возникает дырочная проводимость, а на некоторой глубине — выпрямляющий p–n-переход. Такие выпрямители при площади контакта 1 мм^2 и напряжении 1 В пропускают токи больше 1 А, причем пропускаемые токи обратного направления обычно не превышают нескольких микроампер. При площади контакта в несколько квадратных сантиметров германиевые и кремниевые выпрямители могут пропускать токи в несколько сотен ампер, хотя они свободно умещаются на ладони руки. Их пробойные напряжения достигают многих сотен и даже нескольких тысяч вольт.

Полупроводниковые выпрямители, как и все полупроводниковые приборы, применяются не только для выпрямления обычных технических токов. В частности, полупроводниковые приборы нашли широкие применения в радиотехнике, для выпрямления и преобразования электрических колебаний высокой частоты, для усиления и генерации электрических колебаний, в счетно-решающих электронных устройствах и т. п. Они в значительной степени вытеснили из радиотехники электронные лампы. Без них было бы невозможно решение множества научно-технических задач. Применение полупроводников — необъятная область, которой посвящены специальные курсы. Однако в задачу настоящей книги не входит изложение практических применений полупроводников.

Глава IX

ЭЛЕКТРИЧЕСКИЕ ТОКИ В ГАЗАХ

§ 109. Ионизация и рекомбинация

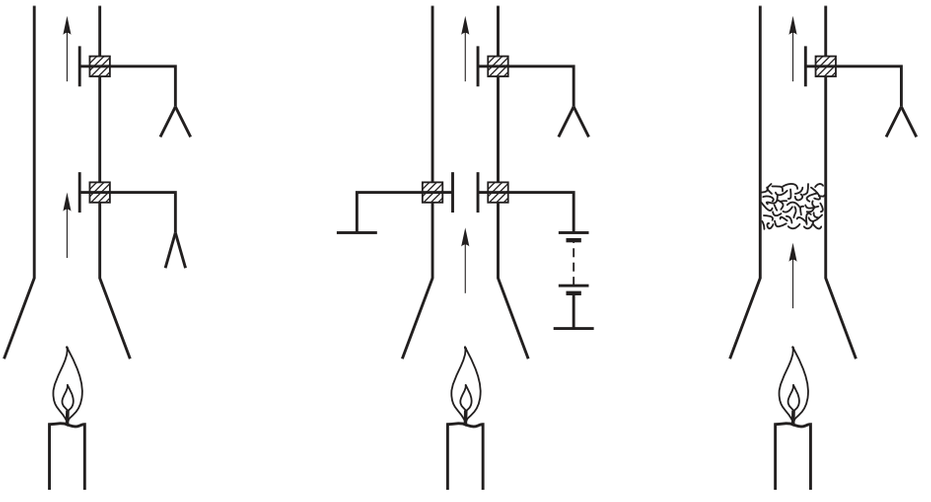
1. Газы в нормальном состоянии, в том числе и пары металлов, состоят из *электрически нейтральных* атомов и молекул и по этой причине не проводят электричества. Проводниками электричества могут быть только *ионизованные* газы. Помимо нейтральных молекул и атомов в них содержатся электроны, положительные и отрицательные ионы. Ионы в газах могут возникать под действием высоких температур, рентгеновских и ультрафиолетовых лучей, лучей радиоактивных элементов, космических лучей, в результате столкновений атомов газа с электронами и другими быстрыми элементарными и атомными частицами и т. д. Во всех этих случаях происходит вырывание из электронной оболочки атома или молекулы одного или нескольких электронов. Этот процесс называется *ионизацией.* Он приводит к освобождению электронов и образованию положительных ионов. Освободившиеся электроны могут присоединяться к нейтральным молекулам и атомам, превращая их в отрицательные ионы. Ионы и свободные электроны делают газ проводником электричества. Впрочем, и в нормальных условиях газы, например воздух, обладают электрической проводимостью, хотя и весьма ничтожной. Эта проводимость вызвана излучением радиоактивных веществ, имеющихся на поверхности Земли, а также космическими лучами, приходящими на Землю из мирового пространства. Электрическая проводимость воздуха была обнаружена еще Кулоном (1785). Наблюдая потери электричества с изолированных проводников, он пришел к заключению, что часть электричества теряется не через изоляторы, а прямо в воздух. Это заключение с особой убедительностью было подтверждено Чарлзом Бойсом (1855–1944) в 1889 г. Подвесив золотые листочки электроскопов к двум цилиндрам из кварца, из которых один был коротким и толстым, а другой — длинным и тонким, он наблюдал, что в обоих случаях потеря электричества была одинакова. Это было бы не так, если бы электричество уходило с листочков только через кварцевые цилиндры. Систематическое изучение электрических токов и разрядов в газах было начато лишь в конце XIX века. Была установлена природа газовых разрядов в различных условиях. Однако, ввиду сложности этих явлений, точной количественной теории их не существует до настоящего времени.

Ионизация газа, возникающая в результате вырывания электронов из молекул и атомов самого газа, называется *объемной ионизацией*, так как источники ионов здесь распределены в объеме, занимаемом газом. Помимо объемной ионизации существует *поверхностная ионизация.* При такой ионизации ионы или электроны поступают в газ со стенок

сосуда, в котором он заключен, или с поверхности тел, вносимых в газ. Например, источниками электронов могут служить раскаленные тела (*термоэлектронная эмиссия*) или поверхности металлов, освещаемые ультрафиолетовыми и прочими коротковолновыми электромагнитными излучениями (*фотоэлектрический эффект*).

2. После прекращения действия ионизатора положительные и отрицательные ионы газа соединяются между собой с образованием нейтральных молекул и атомов. Этот процесс называется *рекомбинацией.* В результате рекомбинации проводимость газа пропадает или, во всяком случае, возвращается к своему исходному значению. Проводимость спадает не сразу, а постепенно, так как для полной рекомбинации ионов газа требуется конечное время.

Следующий опыт может служить демонстрацией этого явления, а также явления ионизации газа. Внутри вертикальной металлической трубы на различной высоте укреплены металлические электроды, соединенные с наружными электроскопами проводами, изолированными от стенок трубы (рис. 250 *а*). Под нижнее отверстие трубы подносят



а *а* *в*

Рис. 250

зажженную свечу. В пламени последней образуются положительные и отрицательные ионы, уносимые вверх горячими газами. Воздух в трубе начинает проводить электричество, и листочки электроскопов спадают. Опыт показывает, что чем выше расположен электрод, тем медленнее спадают листочки соединенного с ним электроскопа. Дело в том, что за время поднятия вверх заметная доля ионов рекомбинирует и проводимость поднявшегося воздуха уменьшается. Поэтому скорость спада листочков электроскопа вверху должна быть меньше, чем внизу. Процесс спадания проводимости газа после прекращения действия ионизатора можно ускорить путем создания в газе электрического поля. Электрическое поле может практически мгновенно уничтожить повышенную проводимость газа. Изменим постановку предыдущего

опыта, поместив в трубе пластины плоского воздушного конденсатора, как указано на рис. 250 *б*. Если повторить опыт с незаряженным конденсатором, то после зажжения свечи листочки электроскопа спадают, как и раньше. Если же предварительно зарядить конденсатор, то после зажжения свечи спадания листочков не наблюдается. Ионы, поднимающиеся вверх вместе с потоком нагретого воздуха, увлекаются электрическим полем к пластинам конденсатора и выбывают из потока. Поэтому воздух после прохождения между пластинами конденсатора перестает проводить электричество, чем и объясняется результат опыта.

Проводимость ионизованного воздуха в предыдущих опытах можно также уничтожить, пропустив струю восходящего воздуха через слой ваты, как указано на рис. 250 *в*. При прохождении через вату происходит нейтрализация ионов.

3. Допустим, что источник создает в единицу времени в единице объема газа q пар ионов разных знаков. Допустим, далее, что в газе нет электрического тока и можно пренебречь уходом ионов из-за диффузии. Тогда единственным процессом потери ионов в газе останется рекомбинация. Обозначим через n число пар ионов разных знаков в единице объема газа. Рекомбинация происходит при встрече положительного иона с отрицательным. Число таких встреч пропорционально как числу положительных, так и числу отрицательных ионов, т. е. n^2. Поэтому убыль числа пар ионов в единицу времени в единице объема газа пропорциональна n^2 и может быть записана в виде αn^2, где α — постоянная, называемая *коэффициентом рекомбинации* ионов разных знаков. При справедливости введенных допущений уравнение баланса ионов в газе запишется в виде

$$dn/dt = q - \alpha n^2. \tag{109.1}$$

В стационарном состоянии $dn/dt = 0$,

$$n = \sqrt{q/\alpha}. \tag{109.2}$$

После выключения ионизатора

$$dn/dt = -\alpha n^2,$$

откуда

$$\frac{1}{n} - \frac{1}{n_0} = \alpha t, \tag{109.3}$$

где n_0 — концентрация пар положительных и отрицательных ионов при $t = 0$, т. е. в момент выключения ионизатора. Спустя время

$$\tau = 1/(n_0\alpha) \tag{109.4}$$

концентрация n убывает в 2 раза. Таким образом, убыль концентрации происходит не в геометрической прогрессии, а более медленно, так как время, в течение которого концентрация уменьшается вдвое, по мере уменьшения последней становится все больше и больше и в пределе при $n \to 0$ обращается в бесконечность.

Если включен ионизатор, то надо интегрировать уравнение (109.1). Допустим, что в момент включения ионизатора $n = 0$. Отсчитывая время t от этого момента и учитывая, что $q - \alpha n^2 > 0$, получим в результате интегрирования

$$n = \sqrt{q/\alpha}\,\mathrm{th}\,(t/\tau), \tag{109.5}$$

где время τ получится из (109.4), если подставить $n_0 = \sqrt{q/\alpha}$, т. е. $\tau = \sqrt{1/(q\alpha)}$.

ЗАДАЧА

В момент времени $t = 0$ начинает действовать ионизатор, создающий в единице объема газа q пар положительных и отрицательных ионов в единицу времени. Предполагая, что $q = \mathrm{const}$, найти выражение для концентрации пар ионов во все последующие моменты времени.

Ответ. $n = \sqrt{\dfrac{q}{\alpha}}\,\dfrac{(\sqrt{q}+n_0\sqrt{\alpha})-(\sqrt{q}-n_0\sqrt{\alpha})e^{-2t/\tau}}{(\sqrt{q}+n_0\sqrt{\alpha})+(\sqrt{q}-n_0\sqrt{\alpha})e^{-2t/\tau}}$, где $\tau = 1/\sqrt{q\alpha}$.

Заметим, что формула верна как при $n_0 < \sqrt{q/\alpha}$, так и при $n_0 > \sqrt{q/\alpha}$. В последнем случае концентрация убывает, несмотря на то что ионизатор непрерывно поставляет новые ионы.

§ 110. Измерение потенциала ионизации методом электронного удара

1. Чтобы из нейтрального атома или молекулы удалить электрон, требуется затрата энергии. Минимальное значение такой энергии называется *энергией ионизации* атома или молекулы. Энергию ионизации принято выражать в электронвольтах. Разность потенциалов, которую должен пройти электрон, чтобы приобрести энергию, равную энергии ионизации, называется *потенциалом ионизации* или *ионизационным потенциалом* атома или молекулы. Очевидно, ионизационный потенциал равен энергии ионизации, деленной на абсолютное значение заряда электрона. Можно говорить об энергии и потенциалах ионизации не только нейтральных атомов и молекул, но и ионов, у которых удаляется еще один электрон.

2. Джеймс Франк (1882–1964) и Густав Герц (1887–1975) в ряде классических работ, начатых в 1913 г., определяли ионизационные потенциалы атомов *методом электронного удара.* Схема их опыта с некоторыми изменениями изображена на рис. 251 *а*. Исследуемый газ при давлении 0,01–1 мм рт. ст. вводится в цилиндрический стеклянный сосуд, предварительно откачанный до высокого вакуума (для удаления посторонних газов). Источником электронов служит цилиндрический платиновый катод K, нагреваемый излучением помещенной внутри него вспомогательной спирали. При таком способе нагревания вся поверхность катода находится при одном и том же потенциале. Это было бы не так и вдоль катода возникло бы падение потенциала, если бы нагревание осуществлялось непосредственно пропусканием через катод

электрического тока (как было в первоначальных опытах Франка и Герца, в которых источником электронов служила платиновая проволока, нагреваемая электрическим током). Катод окружен цилиндрическими

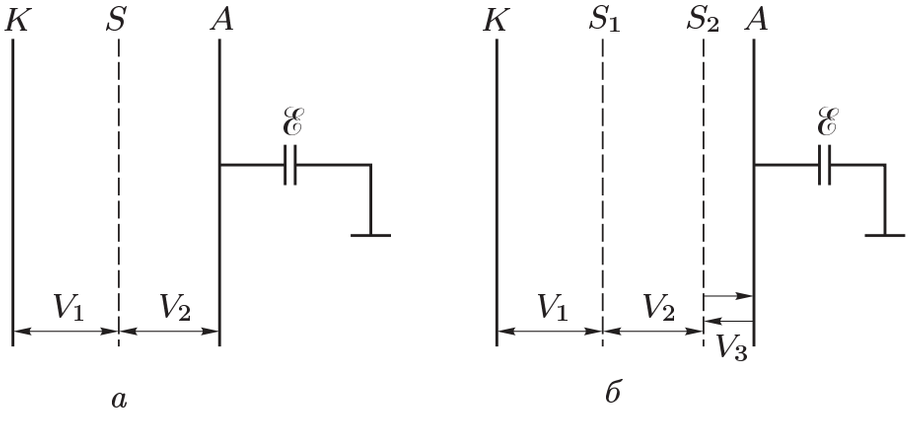


Рис. 251

сеткой S и коллектором A, изготовленными также из платины во избежание возникновения контактной разности потенциалов. Коллектор A соединен с чувствительным электрометром $\mathscr{E}$. Между катодом K и сеткой S прикладывается разность потенциалов V_1, ускоряющая электроны, а между S и коллектором A — несколько бóльшая (на полвольта или вольт) разность потенциалов V_2, замедляющая их. Таким образом, потенциал коллектора A оказывается ниже потенциала катода K на величину ΔV порядка 0,5–1 В. Допустим, что ускоряющая разность потенциалов V_1 меньше потенциала ионизации V_i. Тогда, помимо электронов, в газе не будет никаких заряженных частиц. Но кинетической энергии электронов, ускоренных разностью потенциалов V_1, недостаточно, чтобы преодолеть задерживающий потенциал V_2. Коллектор A заряжаться не будет, и электрометр $\mathscr{E}$ не даст отклонения. Повысим разность потенциалов V_1, чтобы она стала равной потенциалу ионизации. Тогда в газе начнутся такие столкновения электронов с атомами, в результате которых образуются положительные и отрицательные ионы. Отрицательные ионы на коллектор A не попадут, так как этому препятствует задерживающий потенциал V_2. Для положительных ионов тот же потенциал способствует их попаданию на коллектор. В результате коллектор A начнет заряжаться положительно, что и укажет соединенный с ним электрометр $\mathscr{E}$. Франк и Герц думали, что разность потенциалов V_1 между сеткой S и катодом K в этот момент и будет равна ионизационному потенциалу V_i. На самом деле это не так. Франк и Герц начали свои опыты в то время, когда Бор разрабатывал свою теорию атома. Однако в то время эта теория Франком и Герцем не была принята во внимание.

Бор указал на возможность другого механизма попадания ионов на коллектор A. Атом может находиться в определенных стационарных состояниях, характеризующихся уровнями энергии $\mathscr{E}_0, \mathscr{E}_1, \mathscr{E}_2, \ldots$ Если разность потенциалов V_1 достаточна для возбуждения атома, хотя и меньше потенциала ионизации V_i, то в атомах начнутся квантовые

переходы с нормального уровня $\mathscr{E}_0$ на вышележащие уровни $\mathscr{E}_1, \mathscr{E}_2, \ldots$ При обратных переходах на нормальный уровень атомы излучают свет, частота которого ν определяется соотношением $h\nu = \Delta\mathscr{E}$, где h — постоянная Планка. Если эта частота лежит в ультрафиолетовой области спектра, то может оказаться, что свет начнет вырывать электроны с поверхности коллектора A и сетки S (фотоэффект). Эти электроны под действием задерживающей разности потенциалов V_2 будут отбрасываться от коллектора A, и последний станет заряжаться положительно еще до того, как начнется ионизация газа.

3. Чтобы отделить эффект, вызванный ионизацией атомов, от эффекта их возбуждения, Дэвис и Гухер изменили первоначальную схему опытов Франка и Герца. Между катодом K и коллектором A они поместили две сетки S_1 и S_2 (рис. 251 *б*). Ускоряющее напряжение V_1 по-прежнему включалось между катодом K и сеткой S_1, а замедляющее напряжение V_2 — между сетками S_1 и S_2. (На рис. 251 *б* стрелки указывают направления, в которых электроны ускоряются.) Между сеткой S_2 и коллектором A создается разность потенциалов V_3, малая по сравнению с V_2. Знак этой разности потенциалов можно менять во время опыта. Два возможных направления поля V_3 указаны на рис. 251 *б* противоположно направленными стрелками. Допустим сначала, что энергия электрона, ускоренного разностью потенциалов V_1, недостаточна, чтобы ионизовать газ, но достаточна, чтобы вызвать возбуждение атомов и связанное с ним испускание ультрафиолетовых лучей. Последние, падая на сетку S_2 и коллектор A, вырывают с поверхности этих тел медленные электроны. Если электрическое поле между A и S_2 направлено вправо, то эти электроны будут отброшены коллектором A, и на нем останутся положительные заряды. Если же электрическое поле направлено влево, то электроны, вырванные из A, возвратятся обратно, а вырванные из S_2 направятся к A, так что на A образуется отрицательный заряд. Таким образом, при изменении знака разности потенциалов V_3 будет меняться знак заряда электрометра $\mathscr{E}$. Допустим теперь, что между сетками S_1 и S_2 происходит ионизация частиц газа. Тогда появятся отрицательные и положительные ионы. Отрицательные ионы будут увлекаться электрическим полем V_2 влево, а положительные ионы — вправо.

Пройдя через сетку S_2, положительные ионы попадут на коллектор A, каков бы ни был знак разности потенциалов V_3. Действительно, так как разность потенциалов V_3 мала по сравнению с V_2, она не сможет задержать положительные ионы, когда электрическое поле направлено влево. Когда же электрическое поле направлено вправо, то оно еще больше ускорит положительные ионы и тем самым облегчит попадание их на коллектор. Таким образом, возможно отличить возбуждение частиц газа от их ионизации. Если между первым возбуждением (переходом на первый возбужденный уровень) и ионизацией замечаются еще резкие возрастания электризации, то это указывает на дальнейшие возбуждения, т. е. переходы атома на более высокие энергетические уровни.

4. Существуют и другие способы определения ионизационных потенциалов, например спектроскопический. Об этом последнем способе будет сказано в т. V. В пределах отдельных периодов естественной системы элементов Менделеева (1834–1907) наибольшими потенциалами ионизации обладают атомы инертных газов, наименьшими — атомы щелочных металлов. Так, ионизационный потенциал атома гелия равен 24,58 В, неона 21,56 В, лития 5,390 В, натрия 5,138 В. Вот почему пламя особенно эффективно ионизует воздух, когда в него введено небольшое количество поваренной соли. Внутри групп естественной системы Менделеева ионизационный потенциал, как правило, уменьшается с увеличением атомного номера элемента.

При рекомбинации положительного и отрицательного ионов потенциальная энергия уменьшается. Частично она идет на излучение электромагнитных волн. Последнее называется *рекомбинационным излучением.* Рекомбинационное излучение проявляется, например, в виде свечения газа в рекламных трубках.

§ 111. Измерение слабых токов

1. Токи, возникающие в газах под действием внешнего ионизатора, обычно очень слабы, порядка 10^{-6}–10^{-12} А. Предел чувствительности обычных зеркальных гальванометров порядка 10^{-12} А. Дальнейшее повышение чувствительности невозможно из-за броуновского движения, вызывающего такие же по порядку величины отклонения зеркальца гальванометра, какие вызываются измеряемыми токами. Поэтому для измерения таких токов в большинстве случаев пользуются электрометрами. Обычно применяются *струнные* или *квадрантные* электрометры. Ниже изложение ведется применительно к струнному электрометру, хотя все методы и схемы включения остаются неизменными и в случае квадрантного электрометра.

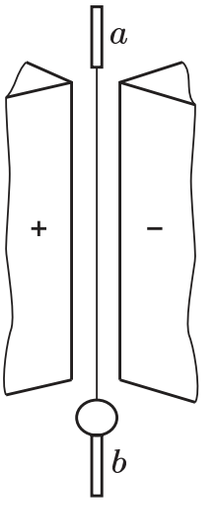


Рис. 252

В струнном электрометре подвижной частью прибора служит тончайшая платиновая нить диаметром 2–10 мкм, натянутая между двумя вертикальными металлическими «ножами», имеющими форму трехгранных призм (рис. 252). Призмы укреплены на янтарных изоляторах и могут быть заряжены до некоторой разности потенциалов (около 100 В), а нить с помощью стержней a и b, между которыми она натянута, соединяется с исследуемым источником электричества. Обычно укрепление платиновой нити на держателе b осуществляется посредством колечка или дужки из кварцевой нити. При зарядке нить электрометра изгибается по направлению к заряженным ножам в ту или в другую сторону, в зависимости от знака заряда. Изгиб получается тем больше, чем меньше натяжение нити. Для регулирования этого натяжения и получения желаемой чувствительности стержни a и b, к которым прикреплена нить, можно передвигать в направлении их длины с помощью микрометрического винта. Отклонение нити измеряется при помощи микроскопа с окулярной шкалой. Для градуировки прибора на нить подаются известные потенциалы и измеряются ее отклонения.

2. Для измерения слабых токов в газах применяются два метода, описываемые ниже.

Метод натекания. Схема этого метода изображена на рис. 253. Ножи электрометра M и N соединены с полюсами батареи B_1, середина которой заземлена. С помощью ключа K нить электрометра может либо заземляться, либо соединяться с электродом A ионизационной камеры, наполненной исследуемым газом. Между электродами A и B камеры создается напряжение с помощью батареи B_2. Допустим, что ионизация в камере создается рентгеновскими лучами от рентгеновской трубки. Сначала ключ K должен

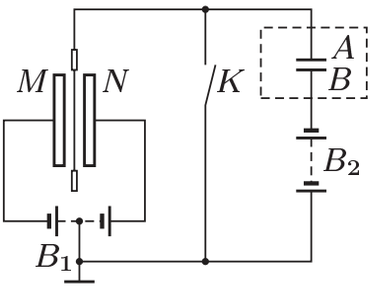


Рис. 253

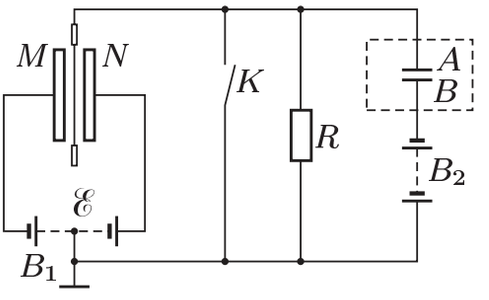


Рис. 254

быть замкнут. Тогда нить электрометра, находясь при нулевом потенциале, останется неподвижной, несмотря на то что в цепи B_2BAK циркулирует ток. Затем ключ K размыкают, отключая тем самым электрод A и нить электрометра от земли. Теперь ионы будут заряжать пластину A, и отклонение нити начнет увеличиваться. Пусть V — потенциал нити, который она приобретет за время t с момента размыкания ключа K. Тогда количество электричества, натекшего в систему за это время, будет $Q = CV$, где C — емкость конденсатора AB, нити и подводящих проводов. Средний ток за время t определится выражением

$$I = \frac{Q}{t} = \frac{CV}{t}.$$

Допустим, например, что $C = 25$ см и электрометр зарядился до 0,1 В $=$ $= 1/3000$ СГСЭ-ед. за 25 с. Тогда $I = 1/3000$ СГСЭ-ед. $\approx 10^{-13}$ А. Основной недостаток рассматриваемого метода состоит в том, что он дает лишь среднюю силу тока за время наблюдения t.

Метод постоянного отклонения (рис. 254). Этот метод применяется, когда ионизационный ток не слишком мал. От предыдущего метода он отличается тем, что между проводом, соединяющим электрод A с нитью, и землей включено высокоомное известное сопротивление R. Через это сопротивление во все время ионизации течет ток I. Вследствие этого на сопротивлении R существует падение напряжения $V = IR$. Если с помощью ключа K нить отключить от земли, то электрометр покажет напряжение V. После этого ионизационный ток найдется по формуле $I = V/R$. Преимущество описанного метода состоит в том, что благодаря практической безынерционности нити он позволяет определить не только среднее, но и мгновенное значение тока I. Недостаток же заключается в трудности точного измерения высокоомного сопротивления R.

§ 112. Несамостоятельная проводимость газов

1. Переходя к рассмотрению токов в газах, будем считать для простоты, что ток течет между двумя плоскими электродами, заряженными противоположно. Направление от положительного электрода к отрицательному примем за ось X. Как и в электролитах, плотность

электрического тока определяется выражением

$$j = n^+e^+u^+ + n^-e^-u^- - e^+D^+\frac{\partial n^+}{\partial x} - e^-D^-\frac{\partial n^-}{\partial x}, \qquad (112.1)$$

где сохранены все прежние обозначения. Первые два члена обусловлены движением ионов под действием электрического поля $\mathbf{E}$, последние два — диффузией ионов. Введем подвижности b^+ и b^- газовых ионов и предположим, что заряды положительных и отрицательных ионов по абсолютной величине одинаковы ($e^+ = -e^- = e$). Кроме того, будем считать одинаковыми концентрации ионов обоих знаков ($n^+ = n^- = n$). Если концентрация одна и та же по всему объему камеры, в которой течет ток, то диффузионного тока не будет, и можно написать

$$j = ne(b^+ + b^-)E. \qquad (112.2)$$

Предположим, что ионы образуются в камере исключительно под действием внешнего ионизатора. В таком случае газовый разряд и соответствующая ему проводимость газа называются *несамостоятельными.* Не имеет значения, является ли ионизация объемной или поверхностной. Пусть в единице объема газа ежесекундно образуется q пар новых ионов. Число ионов, рекомбинирующих за то же время во всем объеме Sl камеры, будет $Sl\alpha n^2$, где S — площадь электрода, а l — длина камеры. При наличии электрического тока убыль ионов будет происходить также за счет ухода их на электроды. Ежесекундно ток уносит на электроды Sj/e пар ионов. Поэтому вместо уравнения баланса (109.1) теперь следует писать

$$\frac{d(Sln)}{dt} = Slq - Sl\alpha n^2 - \frac{Sj}{e},$$

или после сокращения

$$\frac{dn}{dt} = q - \alpha n^2 - \frac{j}{el}. \qquad (112.3)$$

Для стационарных токов

$$q = \alpha n^2 + \frac{j}{el}. \qquad (112.4)$$

2. Рассмотрим решение последнего уравнения в двух предельных случаях.

Во-первых, предположим, что плотность тока j настолько мала, что членом $j/(el)$ можно пренебречь по сравнению с αn^2. Тогда $n = \sqrt{q/\alpha} = \text{const}$ и формула (112.2) дает

$$j = e\sqrt{q/\alpha}\,(b^+ + b^-)E,$$

т. е. плотность тока j пропорциональна электрическому полю E. Такой случай имеет место при малых E. Таким образом, *в слабых электрических полях выполняется закон Ома.*

Во-вторых, предположим, что концентрация ионов n мала. Тогда можно пренебречь рекомбинацией, поскольку рекомбинационный член αn^2 квадратичен по n. В этом приближении $j = qle$, т. е. ток не зависит

от приложенного напряжения. Этот результат справедлив в случае сильных электрических полей. Он объясняется тем, что за время, требующееся иону, чтобы пролететь в сильном поле **E** от одного электрода к другому, ионы не успевают сколько-нибудь заметно рекомбинировать. Поэтому все ионы, производимые ионизатором, уходят на электроды. Ежесекундно на электрод поступает заряд $Slqe$. Это и есть сила тока, текущего через газ. Величина $I_s = Slqe$ называется *током насыщения*, а величина $j_s = qle$ — *плотностью тока насыщения.* Обе эти величины пропорциональны длине ионизационной камеры l, поскольку полное число ионов, производимое ионизатором, также пропорционально l.

При промежуточных значениях напряженности электрического поля зависимость тока от напряжения приобретает сложный характер. Плотность тока j связана с полем E нелинейно, т. е. закон Ома не выполняется.

Эти выводы теории подтверждаются экспериментально. На рис. 255 представлена зависимость тока I в камере от приложенного напряжения V между электродами. Ионизацию в камере можно создать рентгеновскими лучами, излучением радиоактивных веществ или ультрафиолетовыми лучами, падающими на отрицательный электрод ионизационной камеры. Существенно только, чтобы внешний источник, поставляющий ионы, оставался неизменным ($q =$ $=$ const). Участок OA соответствует области применимости закона Ома. На участке AB ток меняется с напряжением нелинейно. Начиная с точки B, ток I достигает насыщения I_s и остается постоянным на некотором участке BC. Однако в точке C снова начинается возрастание ионизационного тока, сначала медленное, а затем очень резкое. Это означает, что появился новый, внутренний источник ионов. Если в некоторой точке D убрать внешний источник ионов, то разряд не прекращается. Из несамостоятельного разряд перешел в *самостоятельный*, когда образование новых ионов происходит в результате *внутренних процессов*, происходящих в самом газе.

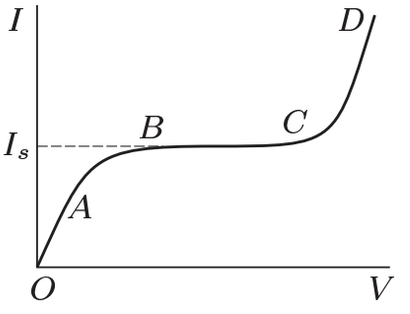


Рис. 255

3. Рассмотрим еще вопрос о распределении потенциала между электродами. Если между электродами не течет ток, то газ ведет себя как всякий диэлектрик. В газе возникнет однородная электрическая поляризация, но не появится никаких объемных зарядов. Поле **E** между электродами будет также однородным, а потенциал V будет линейно убывать с расстоянием x от значения $V_\text{а}$ на аноде до значения $V_\text{к}$ на катоде:

$$V = V_\text{а} - \frac{V_\text{а} - V_\text{к}}{l}\, x.$$

Не то будет, когда через газ проходит электрический ток. В этом случае у анода появятся избыточные отрицательные заряды, а у катода —

положительные. Иначе говоря, возникнут объемные заряды с объемной плотностью $\rho(x)$. В соответствии с уравнением $d^2V/dx^2 = -4\pi\rho$ потенциал V перестанет быть линейной функцией координаты x, а представится некоторой кривой $V = V(x)$. Эта кривая будет обращена выпуклостью вниз, если вторая производная d^2V/dx^2 положительна ($\rho < 0$), и вверх, если эта производная отрицательна ($\rho > 0$). В области, где нет пространственных зарядов, кривая переходит в прямолинейный участок. Изменение потенциала между электродами изображено на схематическом рис. 256 сплошной линией. Штриховая прямая соответствует распределению потенциала, когда между электродами пространственных зарядов нет. Пространственные заряды заставляют эту прямую изгибаться, так как отрицательные заряды, появляющиеся у анода, понижают потенциал вблизи этого электрода, а положительные заряды, накапливающиеся у другого электрода, повышают потенциал в окрестности катода. Мы видим, что прохождение тока через ионизованный газ делает падение потенциала вблизи электродов более крутым, а в середине между ними — более пологим. Это означает, что электрическое поле $\mathbf{E}$ перестает быть однородным. Оно максимально вблизи электродов и минимально между ними.

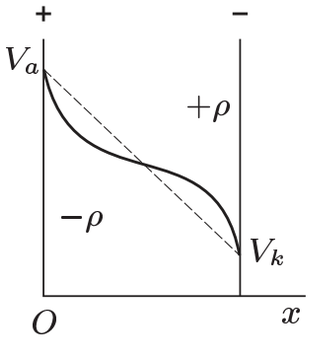


Рис. 256

Заметим в заключение, что требуется совсем ничтожное нарушение электрической нейтральности газа, чтобы вызвать в нем сильное электрическое поле. Поэтому даже при прохождении электрического тока в газе с высокой степенью точности выполняется равенство концентраций положительных и отрицательных ионов: $n^+ = n^-$. Если это имеет место, то говорят о *квазинейтральности* газа. Квазинейтральный — это такой идеализированный газ, в котором с большой точностью выполнено условие $n^+ = n^-$, хотя и могут существовать какие угодно неоднородные электрические поля. В расчетах пользуются такой идеализацией, несмотря на то что само наличие неоднородности электрического поля невозможно без нарушения равенства $n^+ = n^-$. Положение здесь аналогично тому, что имеет место при рассмотрении упругих деформаций тел. В достаточно твердых телах требуются ничтожные изменения плотности, чтобы возникли большие упругие напряжения. В теории упругости вводят представление об идеально недеформируемых упругих телах, в которых могут возникать сколь угодно большие напряжения, хотя реальной причиной возникновения последних и являются деформации.

§ 113. Измерение коэффициентов рекомбинации

Существует много методов измерения коэффициентов рекомбинации. Опишем три наиболее простых из них.

1. Метод Резерфорда (1897). Этот метод позднее (1900) был усовершенствован Таунсендом. Поток исследуемого газа продувается с постоянной скоростью v через латунную трубу, предварительно пройдя через слой ваты,

которой закрыт один из концов трубы (рис. 257). В начальном участке трубы имеется алюминиевое окошко AB, через которое могут проходить

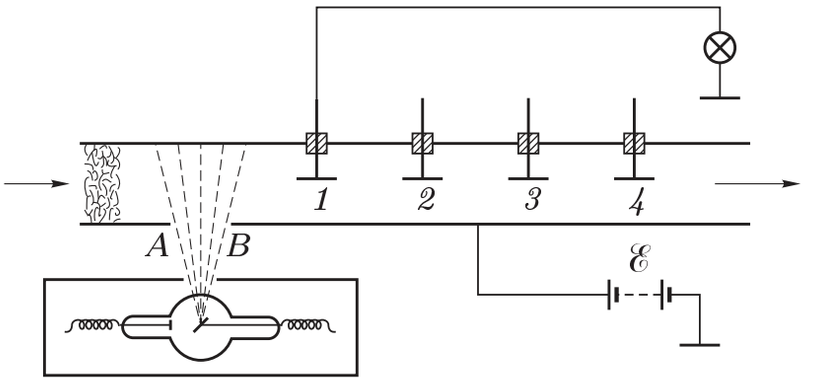


Рис. 257

рентгеновские лучи. Газ ионизуется рентгеновскими лучами или каким-либо другим способом. Образующиеся ионы увлекаются струей газа и по пути частично рекомбинируют. На некоторых расстояниях d_1, d_2, d_3, ... от места ионизации вдоль трубы располагаются тождественные электроды *1*, *2*, *3*, ... Каждый из них поочередно может соединяться с электрометром. При этом все остальные электроды соединяются с трубой, которая в свою очередь соединяется с одним из полюсов батареи, другой полюс которой заземлен. Сначала с электрометром соединяется электрод *1*. Тогда между этим электродом и стенками трубы возникает сильное электрическое поле, увлекающее на электрод *1* все проходящие мимо него ионы одного знака. Если n_1 — концентрация пар ионов у электрода *1*, то на него за время τ попадет $n_1 v S \tau$ ионов, которые передадут ему заряд $Q_1 = n_1 v S e \tau$, где S — площадь электрода, а e — заряд одного иона. После этого такие же измерения повторяют, соединив с электрометром электрод 2. Пусть в этом случае электрометр измерит заряд $Q_2 = n_2 v S e \tau$. Тогда на основании формулы (109.3)

$$\frac{1}{n_2} - \frac{1}{n_1} = \alpha t,$$

где t — время, затрачиваемое потоком газа на прохождение расстояния $d_2 - d_1$. Выражая его через скорость v по формуле $t = (d_2 - d_1)/v$, получим

$$\alpha = \left(\frac{1}{Q_2} - \frac{1}{Q_1}\right) \frac{S v^2 e \tau}{d_2 - d_1}. \tag{113.1}$$

Скорость v можно определить по расходу газа за определенный промежуток времени. Измерения можно повторить, соединив поочередно с электрометром электроды *1*, *1*, ... Резерфорд и его сотрудники убедились, что такие независимые измерения приводят к совпадающим результатам. Это может служить доказательством правильности исходных положений, лежащих в основе метода.

2. Метод переключения. Этот метод также был предложен Резерфордом (1897) и позднее усовершенствован Мак-Клюнгом (1902). Исследуемый газ, находящийся между двумя электродами ионизационной камеры, ионизуется рентгеновскими лучами. В определенный момент времени с помощью маятника выключается ток, питающий рентгеновскую трубку, а через некоторое время τ после этого тот же маятник замыкает цепь батареи, присоединенной к электродам камеры. В результате между электродами камеры создается сильное электрическое поле, увлекающее на один из электродов

ионы определенного знака. Поле выбирается сильным, чтобы при этом ионы в ионизационной камере не успели прорекомбинировать. Пусть Q — заряд, переходящий при этом на электрод камеры, соединенной с электрометром, с помощью которого он может быть измерен. Очевидно, $Q = Vne$, где V — объем камеры. Заряд Q измеряется для двух значений времени τ. Обозначим через Q_1 и Q_2 его значения для времени τ, равного соответственно τ_1 и τ_2. Тогда

$$\left(\frac{1}{Q_2} - \frac{1}{Q_1}\right) Ve = \alpha(\tau_2 - \tau_1), \tag{113.2}$$

откуда и можно определить коэффициент рекомбинации α.

3. Метод Мак-Клюнга. Исследуемый газ, помещаемый между пластинами плоского конденсатора, ионизуется рентгеновскими лучами. Пусть q — число пар ионов, создаваемых ионизатором в единице объема камеры в одну секунду. Для измерения q между обкладками конденсатора накладывается сильное электрическое поле, чтобы возник ток насыщения $I_s = Vqe$, где V — объем конденсатора. Из этого соотношения и можно найти q. Выключим электрическое поле. Тогда в установившемся состоянии, как следует из формулы (109.1), концентрация пар ионов n будет связана с q соотношением $\alpha n^2 = q$. Полный заряд ионов определенного знака в конденсаторе будет $Q = Vne = Ve\sqrt{q/\alpha}$. Его можно измерить, если в некоторый момент времени выключить ионизатор и тотчас же наложить на конденсатор сильное электрическое поле. Тогда заряд Q соберется на одной из пластин конденсатора и его можно измерить электрометром. После этого коэффициент рекомбинации найдется по формуле

$$\alpha = \frac{V^2 e^2 q}{Q^2} = \frac{Ve}{Q^2} I_s. \tag{113.3}$$

Значения коэффициента рекомбинации некоторых газов при давлении 1 атм и температуре 18 °С приведены в табл. 8.

Таблица 8. Значения коэффициента рекомбинации газов

Газ	α, 10^{-6} см^3/с	Газ	α, 10^{-6} см^3/с
Воздух	1,67	Углекислота	1,67
Кислород	1,61	Окись углерода	0,86
Водород	1,44	Водяные пары (100 °С)	0,86

От давления газа коэффициент α зависит слабо, немного уменьшаясь лишь при низких давлениях. С повышением температуры газа величина α уменьшается.

ЗАДАЧИ

1. Между плоскими электродами площадью $S = 100$ см^2 каждый, находящимися на расстоянии $l = 5$ см друг от друга, создана ионизация воздуха рентгеновскими лучами и наблюдается ток насыщения $I_s = 10^{-7}$ А. Определить число пар ионов q, создаваемых ионизатором в 1 см^3 в течение одной секунды, а также концентрацию этих пар n в установившемся состоянии. Ионы считать однозарядными.

Ответ. $q = \dfrac{I_s}{Sle} = 1{,}25 \cdot 10^9$ см$^{-3} \cdot$ с$^{-1}$; $n = \sqrt{q/\alpha} = 2{,}7 \cdot 10^7$ см$^{-3}$.

2. В атмосферном воздухе у поверхности Земли из-за радиоактивности почвы и ионизации космическими лучами в среднем образуется $q = 5$ ионов

в 1 см^3 в одну секунду. Определить ток насыщения, текущий благодаря этой естественной ионизации в плоском воздушном конденсаторе с площадью каждой обкладки $S = 100$ см^2 и расстоянием между обкладками $l = 5$ см.

Ответ. $I_s = qSle = 4 \cdot 10^{-16}$ А.

3. Оценить время разрядки конденсатора в условиях предыдущей задачи, если первоначально он был заряжен до разности потенциалов $V = 300$ В. Как будет меняться время разрядки с уменьшением давления воздуха в объеме конденсатора?

Ответ. $t \approx SV/(4\pi l I_s) \approx 1{,}3 \cdot 10^6$ с ≈ 15 сут. С уменьшением давления воздуха время разрядки будет увеличиваться. Круксу (1832–1919) удалось сохранить в вакууме заряженный электроскоп в течение нескольких месяцев.

4. Через какое время τ после выключения ионизатора число ионов в камере, наполненной воздухом, уменьшается 1) в 2 раза; 2) в 4 раза? Начальная концентрация пар ионов $n_0 = 10^7$ см$^{-3}$.

Ответ. 1) $\tau = 1/(n_0\alpha) = 0{,}06$ с; 2) $\tau = 3/(n_0\alpha) = 0{,}18$ с.

5. Определить эффективное сечение σ рекомбинации положительного молекулярного иона воздуха с отрицательным при комнатной температуре.

Решение. Массы положительных и отрицательных ионов и их концентрации одинаковы. Поэтому из формулы (86.15) второго тома находим $\nu = \sqrt{2}n^2\sigma\overline{v}$, где ν — среднее число столкновений положительных ионов с отрицательными в единице объема в единицу времени, сопровождающихся рекомбинацией. В стационарном состоянии это число должно равняться числу вновь образующихся пар ионов в том же объеме за то же время, т. е. $\sqrt{2}n^2\sigma\overline{v} = q$. Так как $q = \alpha n^2$, то $\sqrt{2}\sigma\overline{v} = \alpha$. Подставляя сюда

$$\overline{v} = \sqrt{\frac{8}{\pi}\frac{kT}{m}} = \sqrt{\frac{8}{\pi}\frac{RT}{M}},$$

где R — универсальная газовая постоянная, M — молекулярный вес воздуха ($M \approx 28{,}8$), получим

$$\delta = \frac{\alpha}{4}\sqrt{\frac{\pi M}{RT}} \approx 2{,}5 \cdot 10^{-11} \text{ см}^2.$$

Соответствующий эффективный диаметр d иона определяется по формуле $\sigma = \pi d^2$ и равен $d \approx 2{,}8 \cdot 10^{-6}$ см, т. е. примерно в сто раз больше газокинетических диаметров нейтральных молекул. Это объясняется электростатическим притяжением противоположно заряженных ионов, что ведет к увеличению числа столкновений между ними, а следовательно, и к возрастанию их эффективных диаметров.

6. Решить ту же задачу для рекомбинации положительного иона с электроном.

Решение. Если пренебречь массой электрона по сравнению с массой иона, то из формулы (86.15) второго тома получим $\nu_{ei} = n^2\sigma_{ei}\overline{v}_e$, где $\overline{v}_e$ — средняя тепловая скорость электрона. Рассуждая, как в предыдущей задаче, получим

$$\sigma_{ei} = \frac{\alpha_{ei}}{\overline{v}_e} = \frac{\alpha_{ei}}{4}\sqrt{\frac{2\pi m_e}{kT}},$$

где α_{ei} — коэффициент рекомбинации электрона с положительным ионом. Величины ν_{ei} и σ_{ei} мы также снабдили двумя индексами, чтобы явно отметить, что речь идет о столкновениях электронов с ионами. Так как радиус электрона можно считать бесконечно малым, то $\sigma = \pi r^2$, где r — радиус иона (см. т. II, § 86).

§ 114. Измерение подвижностей ионов

Из разнообразных методов измерения подвижностей газовых ионов опишем два простейших.

1. Метод переменного поля. Этот метод был предложен Резерфордом. В стеклянном сосуде S, наполненном исследуемым газом, имеются два плоских электрода: сплошной B и сетка A (рис. 258). Платиновая проволока C, нагреваемая электрическим током до высокой температуры, создает ионы в пространстве между C и A. Ионы определенного знака увлекаются к сетке A электрическим полем, создаваемым батареей B_2. Между электродами A и B с помощью трансформатора T создается переменная разность потенциалов $V = V_0 \sin \omega t$ и соответствующее ей электрическое поле $E = (V_0/l) \sin \omega t$, где l — расстояние между электродами A и B, а ω — круговая частота, с которой изменяется переменное напряжение V. Допустим, что ион, скажем положительный, проходит через сетку A в момент времени $t = 0$, когда электрическое поле E равно нулю. Далее он движется по направлению к электроду B со скоростью $v = bE = (V_0 b/l) \sin \omega t$, проходя за время t путь

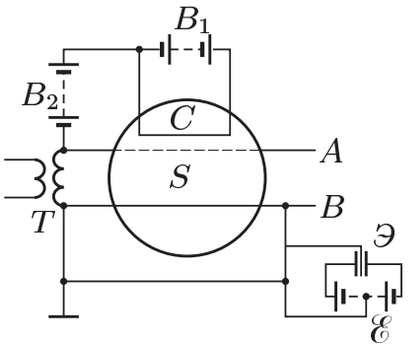


Рис. 258

$$x = \int v\, dt = \frac{V_0 b}{l\omega} (1 - \cos \omega t).$$

Наибольшее расстояние, на которое ион может удалиться от A, будет $x_{\text{макс}} = 2V_0 b/(l\omega)$. Если это расстояние меньше l, то заряд на электрод B и соединенный с ним электрометр Э не попадет. Повышая амплитуду напряжения V_0, можно добиться того, чтобы $x_{\text{макс}} = l$. Тогда электрометр начнет давать показания, и подвижность иона b можно вычислить по формуле

$$b = \frac{l^2 \omega}{2V_0} = \frac{\pi l^2}{V_0 T}, \tag{114.1}$$

где $T = 2\pi/\omega$ — период колебаний переменного напряжения.

2. Метод Зелени (1884–1954). Внутренний цилиндр цилиндрического конденсатора был разрезан на две части BB' и CC', изолированные друг от друга (рис. 259). Во внешнем цилиндре AA' имелась узкая щель mn, через которую перпендикулярно к оси прибора можно было впускать пучок рентгеновских лучей, ионизовавший газ в области mn, ограниченной на рис. 259 двумя вертикальными штриховыми плоскостями. Внутренний цилиндр CC' был соединен с электрометром, а наружный цилиндр AA' — с одним из полюсов батареи. Другой полюс батареи, электрометр и цилиндр BB' были заземлены. В конденсаторе создавалось радиальное электрическое поле

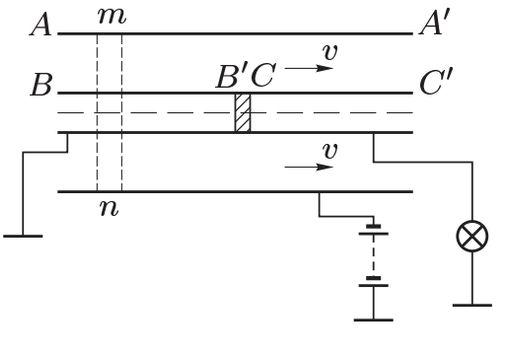


Рис. 259

$$E = \frac{V}{r \ln (r_2/r_1)},$$

где V — напряжение между обкладками конденсатора, r — расстояние от оси прибора, r_1 и r_2 — радиусы внутреннего и наружного цилиндров. Это электрическое поле собирает на цилиндрах BB' и AA' ионы, образованные рентгеновскими лучами. Если цилиндр BB' заряжен отрицательно, то к нему устремятся положительные ионы. Если же цилиндр BB' заряжен положительно, то он соберет отрицательные ионы. Существенно, что ни один ион не попадет на внутренний цилиндр CC', а следовательно, и на электрометр. Не то будет, если газ между цилиндрами продувается с определенной постоянной скоростью v. В этом случае ионы будут двигаться не только радиально к оси конденсатора, но и параллельно ей. Время, потребное иону для прохождения расстояния от наружного цилиндра до внутреннего, определяется выражением

$$t_1 = \int_{r_1}^{r_2} \frac{dr}{bE} = \frac{\ln(r_2/r_1)}{Vb} \int_{r_1}^{r_2} r\,dr = \frac{(r_2^2 - r_1^2)\ln(r_2/r_1)}{2Vb}.$$

Время же, которое требуется иону, чтобы от области ионизации mn дойти до цилиндра CC', равно $t_2 = l/v$, где l — расстояние от mn до цилиндра CC'. Если уменьшить напряжение V, то в некоторый момент ионы начнут попадать на цилиндр CC' и на электрометр. Это произойдет, когда $t_1 = t_2$, т. е. когда

$$b = \frac{v(r_2^2 - r_1^2)\ln(r_2/r_1)}{2Vl}. \tag{114.2}$$

По этой формуле и может быть вычислена подвижность b.

3. В табл. 9 приведены результаты измерений подвижностей различных ионов при давлении 1 атм и температуре 18 °С.

Таблица 9. Значения подвижностей ионов газов

Газ	Подвижность, $см^2/(с \cdot В)$		$\frac{b^-}{b^+}$
	положительных ионов b^+	отрицательных ионов b^-	
Водород	5,91	8,26	1,4
Кислород	1,29	1,79	1,4
Азот	1,27	1,84	1,4
Аргон	1,37	1,70	1,24
Окись углерода	1,10	1,14	1,04
Хлор	0,65	0,51	0,8
Водяной пар (100 °С)	0,62	0,51	0,78

Как показывает опыт, подвижность ионов в широких пределах обратно пропорциональна давлению $\mathscr{P}$. Это понятно, так как подвижность пропорциональна длине свободного пробега иона (см. § 42), а последняя в свою очередь обратно пропорциональна давлению $\mathscr{P}$.

Подвижность положительных ионов не зависит заметно от напряженности поля E. При не слишком больших E подвижность отрицательных ионов также не зависит от E, но при больших E она начинает возрастать с увеличением E. При дальнейшем увеличении E возрастание подвижности замедляется, и, наконец, в еще более сильных полях подвижность снова делается постоянной. Подвижность ионов,

в особенности отрицательных, очень сильно зависит от ничтожного количества некоторых примесей. Так, для гелия при давлении 1 атм, содержащего следы кислорода, $b^+ = 5{,}09$ см^2/(с·В), $b^- = 6{,}31$ см^2/(с·В). В совершенно чистом гелии подвижность положительных ионов b^+ остается почти той же, тогда как b^- достигает огромной величины 500 см^2/(с·В). Эти факты объясняются тем, что при ионизации газа происходит *вырывание* из атома или молекулы электрона, который вначале и является отрицательным ионом. Остаток же атома или молекулы становится положительным ионом. Поскольку масса электрона очень мала по сравнению с массой атома или молекулы, следовало бы ожидать, что его подвижность должна во много раз превосходить подвижность положительного иона. Однако, как показывает табл. 9, различие между b^- и b^+ не так уж велико. Это объясняется тем, что электрон очень скоро, столкнувшись с нейтральной частицей, «прилипает» к ней, образуя отрицательный ион, подвижность которого почти равна подвижности положительного иона. В особенности быстро процесс «прилипания» электронов должен происходить в газах, атомы которых обладают большим сродством к электрону. Таковым является, например, кислород, от ничтожных примесей которого вообще трудно избавиться. При рассмотрении влияния всех этих процессов на подвижность ионов необходимо иметь в виду, что измерение дает некоторую среднюю подвижность, которая, конечно, зависит от соотношения между числом быстрых и медленных ионов. Так же объясняется и влияние напряженности поля E на величину подвижности отрицательных ионов. Дело в том, что с увеличением напряженности электрического поля возрастает скорость электрона, вследствие чего уменьшается вероятность прилипания его к нейтральным частицам. Благодаря этому средняя подвижность отрицательного иона возрастает. Для положительного иона это произойти не может, так как он образуется уже в первый момент ионизации и в дальнейшем не меняет свою массу. В этом случае подвижность не должна зависеть от напряженности поля.

§ 115. Теория Таунсенда

1. До сих пор принималась во внимание только ионизация газа внешним ионизатором, но не учитывалась возможность ионизации при столкновениях ионов и электронов с нейтральными атомами и молекулами. Так можно поступать только в случае сравнительно слабых электрических полей, когда кинетическая энергия eEl, накопленная электроном (или ионом) на длине свободного пробега l, меньше энергии ионизации $\mathscr{E}_i$, и, следовательно, при столкновениях с нейтральными частицами электроны лишь изменяют *направление движения* (*упругое рассеяние*). Не так будет в сильных полях, когда $eEl > \mathscr{E}_i$. Тогда столкновения электронов с нейтральными частицами могут сопровождаться ионизацией последних. Ионизация может происходить и тогда, когда $eE\bar{l} < \mathscr{E}_i$, где $\bar{l}$ — средняя длина свободного пробега электрона.

Действительно, среди электронов найдутся и такие, у которых $l > \bar{l}$, так что для них условие $eEl > \mathscr{E}_i$ будет выполняться.

Допустим теперь, что под влиянием внешнего ионизатора или вследствие какой-либо другой причины у катода возник свободный электрон. Ускоренный электрическим полем, электрон ионизует атом при столкновении с ним. Вместо одного электрона становится два. После ускорения в электрическом поле они ионизуют два атома, а число электронов увеличивается до четырех и т. д. В результате по мере продвижения к аноду число электронов будет лавинообразно нарастать. Этот процесс называется *электронной лавиной*. Каждая ионизация атома сопровождается не только освобождением нового электрона, но и появлением положительного иона, а ионы также могут ионизовать газ.

2. Для количественной характеристики ионизующей способности электронов и ионов Таунсенд (1868–1957) ввел два *коэффициента ионизации* α и β. Первый из них определяется как среднее число ионов одного знака, производимое электроном на единице длины своего пути. Такой же смысл имеет коэффициент β, характеризующий ионизующую способность положительных ионов. Коэффициент ионизации электронами α значительно превосходит коэффициент ионизации положительными ионами β.

Следующий классический опыт Таунсенда доказывает это утверждение. Берется ионизационная камера в виде цилиндрического конденсатора, внутренним электродом которого служит тонкая металлическая нить (рис. 260). Между нитью и наружным цилиндром конденсатора прикладывается разность потенциалов V, достаточная для того, чтобы в объеме камеры происходила *ударная ионизация* газа. Последняя практически будет происходить лишь вблизи нити, где электрическое поле очень сильное. Допустим, что на нить подан положительный потенциал. Тогда к нити устремятся электроны и будут вблизи нее ионизовать газ. Положительные же ионы, устремляясь к наружному цилиндру, пройдут через область слабого поля и практически никакой ионизации не вызовут. Изменим теперь полярность напряжения V, не меняя его величину. Тогда роли положительных и отрицательных ионов поменяются местами. К нити устремятся положительные ионы, и ионизация в камере будет возбуждаться практически только ими. Опыт показывает, что в первом случае ионизационный ток больше и быстрее растет с напряжением V, чем во втором (на рис. 261 кривая I относится к случаю, когда внутренний электрод положителен, а кривая II — к случаю, когда он отрицателен).

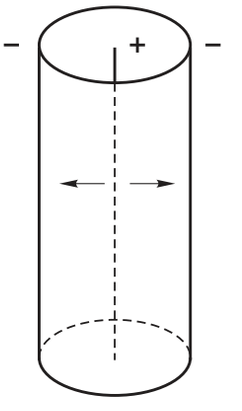


Рис. 260

Из изложенного следует, что *главную роль играет ионизация ударами электронов, по сравнению с которой ионизацией положительными ионами во многих случаях можно пренебречь.*

3. Перейдем теперь к изложению теории Таунсенда прохождения электрического тока через газ. В ней учитывается ударная ионизация атомов и молекул газа электронами и положительными ионами. Для простоты электроды разрядной трубки будем считать плоскими.

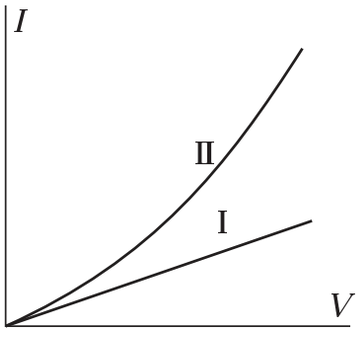


Рис. 261

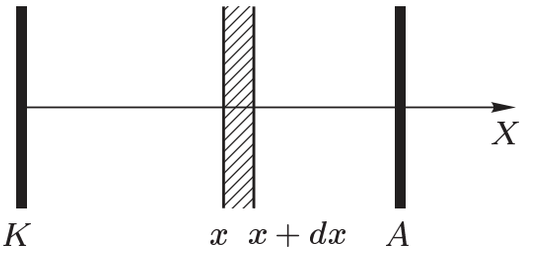


Рис. 262

Рекомбинацией ионов и электронов пренебрежем, предполагая, что за время прохождения между катодом и анодом эти частицы рекомбинировать не успевают. Кроме того, ограничимся стационарным режимом, когда все величины, характеризующие разряд, не зависят от времени. Поместим начало координат на поверхности катода K, направив ось X в сторону анода A (рис. 262). Пусть $n_e(x)$ и $n_p(x)$ — концентрации электронов и положительных ионов, а v_e и v_p — их средние дрейфовые скорости (по абсолютной величине). Возьмем в газе бесконечно тонкий плоский слой, параллельный плоскостям анода и катода, и выделим на нем единичную площадку. Через эту площадку слева в слой ежесекундно входит $n_e(x)v_e(x)$электронов, а справа выходит $n_e(x+dx)v_e(x+dx)$. В объеме dx слоя из-за ионизации электронами ежесекундно возникает $\alpha n_e v_e dx$ электронов и столько же положительных ионов. Аналогично из-за ионизации положительными ионами образуется $\beta n_p v_p dx$ электронов и столько же положительных ионов. Наконец, может существовать внешний источник ионизации, создающий ежесекундно q пар ионов в единице объема газа. А так как в случае стационарности процесса число электронов в слое не меняется, то должно выполняться соотношение

$$n_e(x)v_e(x) - n_e(x+dx)v_e(x+dx) + (\alpha n_e v_e + \beta n_p v_p)dx + q dx = 0.$$

Аналогично, для положительных ионов, движущихся от анода к катоду,

$$n_p(x+dx)v_p(x+dx) - n_p(x)v_p(x) + (\alpha n_e v_e + \beta n_p v_p)dx + q dx = 0.$$

Заменяя разности соответствующими дифференциалами и сократив на dx, получим

$$\begin{aligned} -\frac{d(n_e v_e)}{dx} + \alpha n_e v_e + \beta n_p v_p + q = 0, \\ \frac{d(n_p v_p)}{dx} + \alpha n_e v_e + \beta n_p v_p + q = 0. \end{aligned} \tag{115.1}$$

Введем плотности электрического тока электронов и положительных ионов:

$$j_e = n_e v_e e, \quad j_p = n_p v_p e, \tag{115.2}$$

где e — абсолютное значение заряда электрона. Тогда

$$\frac{dj_e}{dx} - \alpha j_e - \beta j_p - qe = 0, \quad \frac{dj_p}{dx} + \alpha j_e + \beta j_p + qe = 0. \tag{115.3}$$

Отсюда $\dfrac{d}{dx}(j_e + j_p) = 0$ и, следовательно,

$$j_e + j_p \equiv j = \text{const}. \tag{115.4}$$

Таким образом, полная плотность электрического тока j остается постоянной на всем протяжении от катода к аноду, как это и должно быть для квазистационарных процессов. Исключив теперь из уравнений (115.3) и (115.4) ток j_p, получим

$$\frac{dj_e}{dx} - (\alpha - \beta) j_e = \beta j + qe. \tag{115.5}$$

Оба коэффициента ионизации α и β зависят не только от рода газа, но также от давления и напряженности электрического поля. Давление газа $\mathscr{P}$ постоянно по всей длине разрядной трубки. Что касается поля E, то из-за наличия пространственных зарядов оно неоднородно, т. е. зависит от x. Благодаря этому коэффициенты α и β также зависят от x. Однако Таунсенд при интегрировании уравнений (115.3) ввел упрощающее предположение, приняв, что эти коэффициенты постоянны на протяжении всей длины трубки. Они могут измениться только при изменении напряжения на трубке (т. е. поля E). Такое упрощение означает, что поле E считается однородным во всем объеме разрядной камеры. Это может выполняться для сравнительно слабых токов, когда пространственные заряды, возникающие в объеме камеры, несущественны. Поэтому *теория Таунсенда применима лишь к начальной стадии разряда*, когда последнее условие может считаться выполненным. Разряд, для которого это условие выполняется, называется *таунсендовским.* Пренебрежение пространственными зарядами — существенный недостаток теории Таунсенда.

Предполагая α, β и q постоянными и интегрируя уравнение (115.5), находим j_e, а затем и j_p:

$$j_e = Ce^{(\alpha-\beta)x} - \frac{\beta j + qe}{\alpha - \beta}, \quad j_p = -Ce^{(\alpha-\beta)x} + \frac{\alpha j + qe}{\alpha - \beta}, \tag{115.6}$$

где C — постоянная интегрирования. Она найдется из граничных условий, которые должны выполняться на электродах.

4. Чтобы написать эти граничные условия, найдем сначала ток электронов и ток ионов, производимых объемным ионизатором. Число таких электронов, производимых во всем объеме камеры в одну секунду, будет Slq, а заряд, который они несут (по абсолютной величине), $Slqe$, где S — площадь поперечного сечения камеры, l — ее длина. Разделив на S, найдем для плотности электронного тока lqe. Такова же

будет плотность электрического тока, создаваемого положительными ионами. На аноде это и будет полная плотность тока положительных ионов. Таким образом, граничное условие на аноде имеет вид

$$j_p^{(a)} = qle, \tag{115.7}$$

где индекс a означает, что соответствующая величина берется на поверхности анода. Индекс k, вводимый ниже, означает то же самое для катода.

Напишем теперь граничное условие на поверхности катода. Предположим, что имеется внешний источник, создающий поверхностную ионизацию на катоде. Это могут быть рентгеновские или ультрафиолетовые лучи, освещающие поверхность катода. Это может быть высокая температура, при которой происходит термоэлектронная эмиссия. Обозначим через N число электронов, выходящих ежесекундно с единицы поверхности катода благодаря действию поверхностного ионизатора. Они вносят в плотность электронного тока слагаемое Ne. К нему надо добавить плотность электронного тока qle, создаваемого объемным ионизатором, а также плотность тока, создаваемого электронами, вырываемыми из катода положительными ионами (вторичная эмиссия). На единицу площади катода падает ежесекундно $j_p^{(k)}/e$ положительных ионов. Обозначим через γ среднее число электронов, вырываемых из катода одним положительным ионом. Тогда с единицы поверхности катода будет вырываться ежесекундно $\gamma j_p^{(k)}/e$ электронов, которые создадут электронный ток $\gamma j_p^{(k)}$. Заметим, что на границе катода с газом эмиссия электронов может происходить не только в результате ударов положительных ионов, но и в результате фотоэффекта и ударов других частиц. Но мы вместе с Таунсендом отвлечемся от этого обстоятельства и для простоты будем считать, что эмиссия электронов с катода вызывается только ударами положительных ионов. При таком упрощении полная плотность электронного тока у катода будет

$$j_e^{(k)} = Ne + qle + \gamma j_p^{(k)}. \tag{115.8}$$

Это и есть граничное условие на поверхности катода.

Из (115.6) находим

$$j_e^{(k)} = C - \frac{\beta j + qe}{\alpha - \beta}, \quad j_p^{(k)} = -C + \frac{\alpha j + qe}{\alpha - \beta},$$

$$j_p^{(a)} = -Ce^{(\alpha-\beta)l} + \frac{\alpha j + qe}{\alpha - \beta}.$$

После подстановки этих значений в (115.7) и (115.8) получим систему уравнений

$$\begin{gathered} Ce^{(\alpha-\beta)l} - \frac{\alpha}{\alpha-\beta}\, j = qe\left(\frac{1}{\alpha-\beta} - l\right), \\ (1+\gamma)C - \frac{\beta+\gamma\alpha}{\alpha-\beta}\, j = \left[N + ql + \frac{(1+\gamma)q}{\alpha-\beta}\right] e, \end{gathered} \tag{115.9}$$

из которой находим

$$j = \frac{A - Be^{(\alpha-\beta)l}}{\Delta}, \tag{115.10}$$

где введены обозначения

$$\begin{aligned} A &= (1+\gamma)[1+l(\alpha-\beta)]qe, \\ B &= (Ne+qle)(\alpha-\beta)+(1+\gamma)qe, \\ \Delta &= (\beta+\gamma\alpha)e^{(\alpha-\beta)l}-(1+\gamma)\alpha. \end{aligned} \tag{115.11}$$

Рассмотрим несколько частных случаев.

Допустим, что внешняя ионизация — чисто поверхностная ($q = 0$). Пренебрежем ионизацией положительными ионами ($\beta = 0$), а также вторичной эмиссией с катода ($\gamma = 0$). Тогда

$$j = Ne\exp(\alpha l). \tag{115.12}$$

В случае объемной ионизации ($N = 0$) в тех же предположениях ($\beta = \gamma = 0$)

$$j = qle\left\{\left(1+\frac{1}{\alpha l}\right)\exp(\alpha l)+\left(1-\frac{1}{\alpha l}\right)\right\}. \tag{115.13}$$

Учтем теперь ионизацию положительными ионами, но пренебрежем вторичной эмиссией на катоде ($\gamma = 0$). Тогда для случая поверхностной внешней ионизации ($N \neq 0$, $q = 0$) получим

$$j = \frac{(\alpha-\beta)Ne}{\alpha-\beta\exp\{(\alpha-\beta)l\}}\exp\{(\alpha-\beta)l\}. \tag{115.14}$$

Ионизации положительными ионами Таунсенд первоначально придавал большое значение, так как он убедился на опыте, что формула (115.14) правильно передает зависимость ионизационного тока от длины разрядной трубки l. Напротив, формула (115.12) согласовывалась с опытом лишь при малых значениях l, а при больших l давала большие ошибки.

5. Из формулы (115.10) видно, что при $\Delta \neq 0$ конечный ионизационный ток получается лишь при наличии внешних ионизаторов. В этом случае разряд остается *несамостоятельным.* Не то будет, когда знаменатель Δ обращается в нуль. С точностью до множителя $(\alpha - \beta)$ величина Δ представляет собой определитель, составленный из коэффициентов левых частей системы уравнений (115.9). Для того чтобы при $\Delta = 0$ указанная система имела нетривиальные решения ($C \neq 0$, $j \neq 0$), необходимо, чтобы обращались в нуль ее правые части. Это будет всегда, когда $q = N = 0$, т. е. когда нет внешних ионизаторов. Вспомним, что коэффициенты α, β и γ зависят от напряженности электрического поля E. С увеличением E ионизационный ток j растет. При некотором значении поля Δ обращается в нуль. Если убрать внешние ионизаторы, то в этом случае ток через газ все же будет идти. Разряд

становится *самостоятельным.* При выполнении условия

$$\Delta \equiv (\beta + \gamma\alpha)e^{(\alpha-\beta)l} - (1+\gamma)\alpha = 0, \tag{115.15}$$

согласно теории Таунсенда, происходит *пробой газа*, или *зажигание газового разряда.* Конечно, для зажигания разряда, т. е. развития электронных и ионных лавин, необходимо наличие в газе какого-то минимального начального количества электронов или ионов. Но таковые всегда имеются в газе, хотя бы и в ничтожных количествах. Однако изложенная теория не позволяет проследить процесс перехода несамостоятельного разряда в самостоятельный, так как она была развита только для процессов, не меняющихся во времени.

§ 116. Закон Пашена

1. Исследуем теперь зависимость коэффициентов ионизации α и β от напряженности электрического поля E и давления газа $\mathscr{P}$. Для определенности будем иметь в виду электроны. Примем вместе с Таунсендом, что при каждом столкновении электрон теряет скорость, которую он приобрел в электрическом поле. Чтобы электрон мог ионизовать газ, он должен на пути свободного пробега x приобрести энергию, не меньшую энергии ионизации, т. е. величина x должна удовлетворять условию $xE \geqslant V_i$, где V_i — потенциал ионизации. Примем, что все такие электроны ионизуют газ. Возьмем пучок из N_0 электронов, которые начали двигаться в электрическом поле E с нулевой начальной скоростью. Как известно из кинетической теории газов (см. т. II, § 88), среднее число электронов, проходящих путь x без столкновений, будет $N = N_0 e^{-x/\overline{l}}$ где $\overline{l}$ — средняя длина свободного пробега электрона. Если взять $x = V_i/E$, то, согласно нашему предположению, все такие электроны будут ионизовать газ. На пути 1 см электрон испытывает в среднем $1/\overline{l}$ столкновений, так что на этом пути все N электронов вызовут $N/\overline{l} = (N_0/\overline{l})e^{-x/\overline{l}}$ ионизаций. Среднее число ионизаций, производимое на том же пути одним электроном, будет

$$\alpha = (1/\overline{l})e^{-V_i/E\overline{l}}.$$

Это и есть коэффициент ионизации. Средняя длина свободного пробега $\overline{l}$ обратно пропорциональна давлению газа: $\overline{l} = 1/(A\mathscr{P})$, так что

$$\alpha = A\mathscr{P}e^{-B\mathscr{P}/E}, \tag{116.1}$$

где A и B — постоянные. Отсюда следует, что

$$\frac{\alpha}{\mathscr{P}} = f\left(\frac{E}{\mathscr{P}}\right), \tag{116.2}$$

т. е. отношение $\alpha/\mathscr{P}$ зависит не от E и $\mathscr{P}$ в отдельности, а только от их комбинации $E/\mathscr{P}$. Понятно, что этот результат, пока функция f не фиксирована, является более общим, чем результат (116.1). Таунсенд экспериментально подтвердил справедливость соотношения (116.2) для

ряда газов. Более поздние исследования показали, что при давлениях, больших атмосферного, это соотношение удовлетворяется значительно хуже, чем для низких давлений, а при высоких давлениях перестает быть верным.

Простым дифференцированием по $\mathscr{P}$ легко убедиться, что при заданном значении E выражение (116.1) достигает максимума при $\mathscr{P} = E/B$. Максимальное значение α при этом оказывается равным

$$\alpha_{\text{макс}} = \frac{A}{eB}\,E, \tag{116.3}$$

где e — основание натуральных логарифмов. Таким образом, максимальное значение α при заданной напряженности электрического поля E пропорционально этой напряженности. Опыт подтверждает и эту зависимость.

Разумеется, все выведенные соотношения справедливы и для положительных ионов.

2. Обратимся теперь к условию зажигания разряда (115.15). Для простоты предположим, что вторичная эмиссия электронов с катода роли не играет ($\gamma = 0$). Тогда (115.15) переходит в

$$\beta e^{(\alpha-\beta)l} - \alpha = 0.$$

Подставим сюда

$$\alpha = \mathscr{P} f\left(\frac{V}{l\mathscr{P}}\right), \quad \beta = \mathscr{P} f_1\left(\frac{V}{l\mathscr{P}}\right),$$

где V — напряжение на трубке. Тогда после сокращения на $\mathscr{P}$ мы придем к уравнению вида

$$F\left(\frac{V}{l\mathscr{P}}\right) = 0,$$

где F — некоторая функция аргумента $V/l\mathscr{P}$. Решая это уравнение, найдем «потенциал зажигания» $V_{\text{заж}} = V_{\text{заж}}(l\mathscr{P})$. Отсюда следует, что *разность потенциалов между электродами трубки, при которой начинается пробой газа, есть функция произведения давления газа $\mathscr{P}$ на расстояние между электродами.* Если в нескольких разрядных трубках с плоскими электродами создать условия, при которых произведения $\mathscr{P}l$ постоянны, то для всех трубок потребуется одна и та же разность потенциалов, чтобы вызвать газовый разряд. Этот закон был установлен экспериментально Пашеном (1865–1947) еще до создания Таунсендом теории пробоя газа.

Закон Пашена можно вывести и из более общего условия пробоя (115.15), если только γ, наряду с $\alpha/\mathscr{P}$ и $\beta/\mathscr{P}$, также является функцией $E/\mathscr{P}$. Такое предположение весьма правдоподобно, так как оно означает, что γ есть функция лишь кинетической энергии, приобретаемой в среднем положительным ионом на каждом свободном пробеге перед ударом о катод. При высоких давлениях газа (порядка сотен атмосфер) наблюдаются отступления от закона Пашена.

§ 117. Тлеющий разряд

1. *Тлеющим разрядом обычно называют самостоятельный разряд, в котором катод испускает электроны вследствие бомбардировки его положительными ионами и фотонами, образующимися в газе.* В отличие от таунсендовского разряда, где плотности электрического тока невелики, а влияние пространственного заряда несущественно, в тлеющем разряде плотности тока значительно больше, а пространственные заряды, возникающие из-за большого различия в массах электронов и положительных ионов (см. § 112, п. 3), делают электрическое поле в газе существенно неоднородным. Для тлеющего разряда характерна большая напряженность электрического поля и соответствующее ей большое падение потенциала вблизи катода (так называемое *катодное падение*).

2. Возьмем стеклянную трубку длиной 30–50 см, в которую впаяны два электрода (рис. 263). Приложим между ними постоянное напряжение в несколько сот вольт. При атмосферном давлении такое

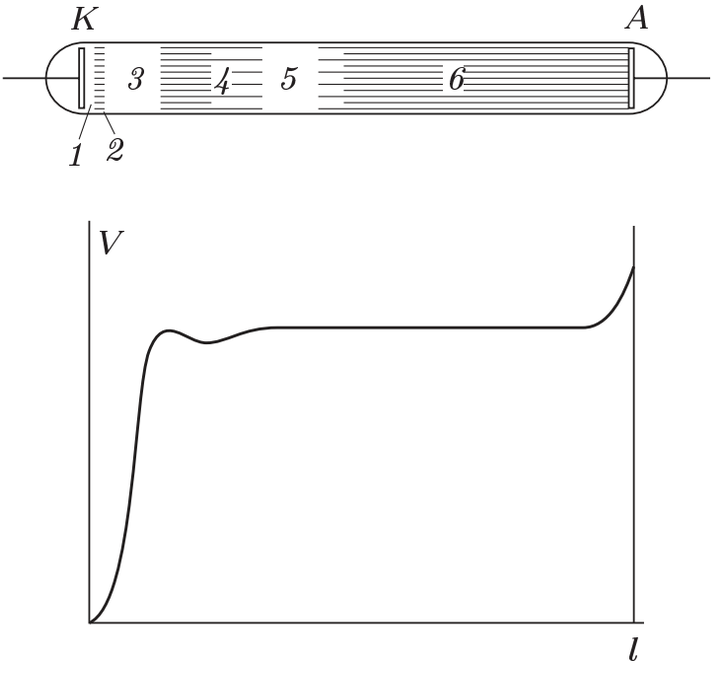


Рис. 263

напряжение недостаточно для пробоя газа, и трубка остается темной. Будем откачивать газ из трубки. При давлении 50 мм рт. ст. возникает самостоятельный разряд в виде светящегося шнура красноватого цвета, идущего от катода к аноду. При давлении 2–3 мм рт. ст. свечение заполняет всю трубку. При давлении газа порядка 0,1–0,01 мм рт. ст. разряд имеет вид, изображенный на схематическом рис. 263. На том же рисунке представлено примерное распределение потенциала V в трубке.

Непосредственно к катоду примыкает узкое так называемое *астоново темное пространство* (на рис. 263 обозначенное цифрой *1*), где электроны, исходящие из катода, еще не успели приобрести скорости, достаточной для возбуждения атомов и молекул газа. Оно впервые наблюдалось Фрэнсисом Астоном (1877–1945) в гелии, неоне и водороде, почему и получило его имя. Ширина астонова темного пространства (порядка десятых долей миллиметра) обратно пропорциональна давлению газа и уменьшается с увеличением плотности тока.

Затем идет тонкая светящаяся пленка, называемая *катодным слоем* (*2*), где происходит возбуждение атомов и молекул ударами электронов, но еще нет ионизации. Возвращаясь в нормальное состояние, возбужденные атомы излучают световые кванты, чем и объясняется свечение.

За катодным слоем следует *темное катодное пространство*, называемое также *темным круксовым* или *темным гитторфовым пространством* (*3*). На самом деле оно не совсем темное, но кажется таковым лишь на фоне примыкающих к нему более светлых областей разряда. В этой части пространства начинается *ионизация атомов и молекул* и *нарастание электронных лавин*. Из-за возможности ионизации уменьшается вероятность возбуждения атомов, с чем и связано ослабление свечения газа. Область темного катодного пространства наиболее важна для поддержания разряда, так как созданные здесь положительные ионы обеспечивают необходимую эмиссию электронов с катода.

Темное катодное пространство резко переходит в *отрицательное тлеющее свечение*, иногда называемое просто *тлеющим свечением* (*4*). Это свечение резко ограничено только со стороны катода. Свечение возникает из-за *рекомбинации* электронов с положительными ионами, а также вследствие *квантовых переходов* возбужденных атомов на более низкие энергетические уровни.

При продвижении к аноду яркость тлеющего свечения ослабевает, и оно постепенно переходит в так называемое *фарадеево темное пространство* (*5*), в которое уже не долетают быстрые электроны электронных лавин.

Перечисленные пять областей называются *катодными частями разряда. В них происходят все процессы, необходимые для поддержания разряда.*

3. За фарадеевым темным пространством следует так называемый *остов разряда*. В более или менее узких трубках он представляет собой столб ионизованного светящегося газа (*6*) и называется *положительным свечением* или *положительным столбом разряда*. Обычно положительный столб простирается до самой поверхности анода. При некоторых условиях между положительным столбом и анодом видно *темное анодное пространство*, а на самой поверхности анода — *анодное свечение*, или *анодная светящаяся пленка*. Положительный столб иногда разделяется на отдельные чередующиеся светлые и темные полосы — *страты*. Тогда разряд называется *сложным*. Наличие

положительного столба несущественно для поддержания разряда, хотя он и имеет большое значение в применениях разряда. В очень коротких трубках положительный столб отсутствует. В длинных трубках он служит проводящим мостом, соединяющим фарадеево темное пространство с анодом. Для положительного столба характерна относительно высокая степень ионизации и связанная с ней высокая проводимость газа. По той же причине плотность пространственного заряда в положительном столбе невелика (и даже равна нулю). Свечение в положительном столбе происходит в основном за счет рекомбинации электронов с положительными ионами. На последних нескольких свободных пробегах (в области так называемого анодного падения) электроны могут накопить достаточную кинетическую энергию, чтобы вызвать возбуждение атомов, в то время как положительные ионы оттягиваются от анода. Это приводит к анодному свечению.

4. Если анод трубки сделать подвижным и постепенно подвигать его к катоду, то катодные части разряда по своей форме и величине остаются неизменными вплоть до границы положительного столба, в то время как длина последнего уменьшается. При дальнейшем передвижении анода к катоду укорачивается длина фарадеева темного пространства, а затем и тлеющего свечения. При этом положение резкой границы тлеющего свечения со стороны катода остается неизменным. Когда расстояние от этой границы до анода уменьшается до долей миллиметра, тлеющий разряд прекращается. Аналогичные явления наблюдаются, если, оставляя анод неподвижным, передвигать к нему катод. При этом без изменения размеров и формы передвигаются и все катодные области разряда вместе с границей положительного столба. Затем положительный столб, фарадеево темное пространство и тлеющее свечение постепенно «съедаются» анодом. Разряд прекращается, когда толщина тлеющего свечения уменьшается до долей миллиметра. Если расстояние между катодом и анодом недостаточно, чтобы на нем поместить темное катодное пространство и начало тлеющего свечения, то разряд, когда это позволяет конструкция трубки, выбирает более длинный путь (рис. 264).

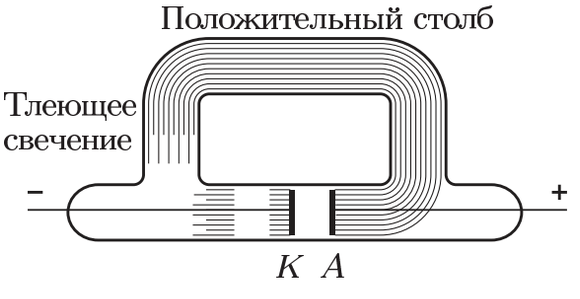


Рис. 264

5. При больших внешних сопротивлениях, когда сила тока в разрядной трубке невелика, поверхность катода, покрытая свечением и принимающая участие в разряде, пропорциональна силе тока в трубке (*закон Геля*). При изменении тока плотность его остается приблизительно постоянной. Вместе с ней остается постоянным и катодное падение потенциала. В этом случае оно называется *нормальным катодным падением.* До давлений в несколько десятков миллиметров ртутного столба нормальное катодное падение не зависит от давления газа,

а определяется только материалом катода и составом газа. В большинстве случаев оно лежит в пределах 100–300 В. Температура катода не оказывает влияния на величину нормального катодного падения, пока не возрастет заметно термоэлектронная эмиссия с поверхности катода. С хорошим приближением нормальное катодное падение пропорционально работе выхода электрона из катода. Это используется для устройства *газосветных трубок* с очень малым потенциалом зажигания. Такова, например, *неоновая лампочка*, в которой электродами служат два железных листочка, покрытых слоем бария для уменьшения работы выхода. Катодное падение составляет в этом случае всего 70 В, и тлеющий разряд зажигается в неоновой лампочке уже при включении в обычную осветительную сеть.

Когда с увеличением тока вся поверхность катода оказывается покрытой свечением, начинает возрастать и катодное падение. В этом случае оно называется *аномальным катодным падением*, а самый разряд — *аномальным тлеющим разрядом.*

6. Электроны, выбиваемые с поверхности катода положительными ионами, ускоряются в области катодного падения потенциала. Когда давление в трубке относительно велико (выше 0,1 мм рт. ст.), электроны теряют свои скорости в результате столкновений с частицами газа. При уменьшении давления газа увеличивается средняя длина свободного пробега электронов, а с ней и катодное темное пространство. При давлении 0,01–0,001 мм рт. ст. (в зависимости от размеров трубки) катодное темное пространство заполняет почти всю трубку и электронный пучок движется в трубке почти без столкновений. Такие электронные пучки получили название *катодных лучей.* Последние были открыты Круксом в конце 70-х годов XIX века, еще до установления их физической природы (и до открытия самого электрона). Если на пути катодных лучей поставить металлический экран, то за ним на противоположной стороне трубки наблюдается его тень. При поднесении магнита пучок лучей и образуемая им тень смещаются в сторону. Эти явления и наблюдались Круксом. Обычно в опытах с катодными лучами применяются короткие, но широкие стеклянные трубки, в которых анод помещается сбоку электронного пучка, исходящего из катода. Электроны, вышедшие с катода, ускоряются электрическим полем вблизи его поверхности и далее движутся перпендикулярно к ней по инерции. Попадая на стенки трубки, электроны сообщают им отрицательный заряд. Однако последний нейтрализуется положительными ионами, подтекающими из газа к стенкам трубки, а отрицательные ионы газа попадают на анод. Если поверхности катода придать вогнутую сферическую форму, то катодные лучи сфокусируются в центре этой сферы. Когда давление в трубке настолько мало, что область катодного темного пространства захватывает анод, тлеющий разряд в трубке прекращается. Вместе с ним прекращается также испускание катодных лучей и свечение стенок трубки.

7. Катодные лучи используются в так называемых *ионных рентгеновских трубках* для получения *рентгеновских лучей.* Устройство

такой трубки показано на схематическом рис. 265. Трубка имеет вогнутый катод K и два положительных электрода A и $Aк$. Один из них A, называемый *анодом*, более удален от катода и расположен в области тлеющего свечения. Он предназначен для поддержания в трубке тлеющего разряда. Второй электрод $Aк$ называется *антикатодом* и состоит из тугоплавкого металла (вольфрама, молибдена и пр.). На этом электроде фокусируются пучки электронов, испускаемые катодом. При их торможении из антикатода выходят рентгеновские лучи, представляющие собой очень короткие электромагнитные волны (длина волны порядка 10^{-8} см и меньше). Их свойства будут рассмотрены в четвертом и пятом томах нашего курса. Под действием бомбардирующих электронов антикатод сильно раскаляется. Поэтому в мощных рентгеновских трубках антикатод изнутри охлаждается проточной водой.

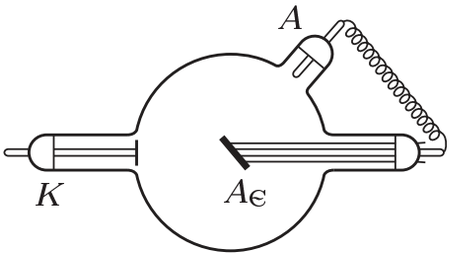


Рис. 265

Ионные рентгеновские трубки обладают тем недостатком, что в результате различных процессов количество газа в трубке уменьшается с течением времени. Когда давление газа в трубке делается меньше 0,001–0,0001 мм рт. ст., тлеющий разряд не зажигается, и трубка перестает работать. В настоящее время применяются почти исключительно *электронные рентгеновские трубки*, обладающие большей устойчивостью в работе, чем ионные трубки. В них тлеющий разряд не используется, а потому не нужен вспомогательный анод, служащий для поддержания такого разряда. Анод (в мощных трубках охлаждаемый) делается из таких же тугоплавких металлов, как и антикатод в ионных трубках. Катодом служит вольфрамовая спираль, накаливаемая электрическим током. Электроны, испускаемые катодом, образуются вследствие термоэлектронной эмиссии. Ускоренные электрическим полем, они бомбардируют анод и при торможении на нем испускают рентгеновские лучи.

8. Если в катоде просверлить малые отверстия, то положительные ионы, бомбардирующие катод, пройдя через эти отверстия, попадают в закатодное пространство и там распространяются в виде прямолинейных лучей (такой опыт впервые был произведен Гольдстейном (1850–1930) в 1882 г.). Эти лучи были названы *положительными*, или *каналовыми*, *лучами*, поскольку они выходят из отверстий катода, как из каналов. Каналовые лучи заметны в трубке в виде слабо светящихся пучков. Они, как и катодные лучи, вызывают свечение стекла трубки. Из-за наличия процессов *перезарядки* в пучке каналовых лучей имеются не только положительные, но и отрицательные ионы, а также быстрые, отчасти возбужденные нейтральные частицы. В магнитном поле такой пучок разделяется на три пучка: положительные ионы отклоняются в одну сторону, отрицательные — в противоположную, а нейтральные молекулы и атомы не испытывают никакого отклонения.

При повторном пропускании этих пучков через магнитное поле каждый из них снова распадается на три пучка. Отсюда следует, что процессы перезарядки происходят не только перед катодом, но и продолжаются в закатодном пространстве.

§ 118. Искровой разряд

1. *Искровой разряд* характеризуется прерывистой формой даже при пользовании источниками постоянного тока. Он возникает в газе обычно при давлениях порядка атмосферного. В естественных природных условиях искровой разряд наблюдается в виде *молнии*. По внешнему виду искровой разряд представляет собой пучок ярких зигзагообразных разветвляющихся тонких полосок, мгновенно пронизывающих разрядный промежуток, быстро гаснущих и постоянно сменяющих друг друга (рис. 266). Эти полоски называются *искровыми каналами*. Они начинаются как от положительного, так и от отрицательного электродов, а также от любой точки между ними. Каналы, развивающиеся от положительного электрода, имеют четкие нитевидные очертания, а развивающиеся от отрицательного, — диффузные края и более мелкое ветвление.

Так как искровой разряд возникает при больших давлениях газа, то потенциал зажигания очень высок. (Для сухого воздуха, например, при давлении 1 атм и расстоянии между электродами 10 мм пробивное напряжение 30 кВ.) Однако после того, как разрядный промежуток «пробит» искровым каналом, сопротивление этого промежутка становится очень малым, через канал проходит кратковременный импульс

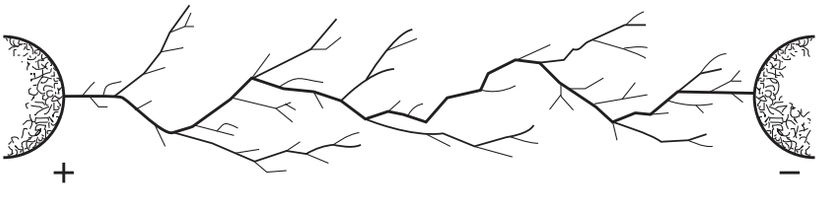


Рис. 266

тока большой силы, в течение которого на разрядный промежуток приходится лишь незначительное напряжение. Если мощность источника не очень велика, то после такого импульса тока разряд прекращается. Напряжение между электродами начинает повышаться до прежнего значения, и пробой газа повторяется с образованием нового искрового канала. Время τ нарастания напряжения тем больше, чем больше емкость C между электродами. Поэтому включение конденсатора параллельно разрядному промежутку увеличивает время между двумя последовательными искрами, а сами искры становятся более мощными. Через канал искры проходит большой электрический заряд, и поэтому увеличивается амплитуда и длительность импульса тока. При большой емкости C канал искры ярко светится и имеет вид широких полос. То же самое происходит при увеличении мощности источника тока.

Тогда говорят о *конденсированном искровом разряде*, или *конденсированной искре.* Максимальная сила тока в импульсе при искровом разряде меняется в широких пределах в зависимости от параметров цепи разряда и условий в разрядном промежутке, достигая нескольких сотен килоампер. При дальнейшем увеличении мощности источника искровой разряд переходит в *дуговой разряд.*

В результате прохождения импульса тока через канал искры в канале выделяется большое количество энергии (порядка 0,1–1 Дж на каждый сантиметр длины канала). С выделением энергии связано скачкообразное увеличение давления в окружающем газе — образование цилиндрической ударной волны, температура на фронте которой около 10^4 К. Происходит быстрое расширение канала искры, со скоростью порядка тепловой скорости атомов газа. По мере продвижения ударной волны температура на ее фронте начинает падать, а сам фронт отходит от границы канала. Возникновением ударных волн объясняются звуковые эффекты, сопровождающие искровой разряд: характерное потрескивание в слабых разрядах и мощные раскаты *грома* в случае молнии.

В момент существования канала, особенно при высоких давлениях, наблюдается наиболее яркое свечение искрового разряда. Яркость свечения неоднородна по сечению канала и имеет максимум в его центре.

2. Остановимся теперь на механизме искрового разряда. Вначале казалось, что искровой разряд обусловлен теми же основными процессами, что и тлеющий разряд. Согласно теории Таунсенда такими процессами являются объемная ионизация газа электронами и положительными ионами, а также вторичная эмиссия электронов с катода, обусловленная бомбардировкой его положительными ионами и фотонами, образующимися во время разряда. Однако такое предположение не согласуется с опытами. Если бы оно было верным, то время развития искрового разряда, например, было бы порядка времени, которое требуется ионам для прохождения расстояния между анодом и катодом. Это время, как нетрудно оценить, порядка 10^{-4}–10^{-5} с. Между тем осциллографические исследования показали, что фактическое время развития искрового разряда порядка 10^{-7} с и меньше.

В настоящее время общепринятой считается так называемая *стримерная теория искрового пробоя*, подтвержденная прямыми опытами. Качественно она объясняет основные особенности искрового разряда, хотя в количественном отношении и не может считаться завершенной. Если вблизи катода зародилась электронная лавина, то на ее пути происходит ионизация и возбуждение молекул и атомов газа. Существенно, что световые кванты, испускаемые возбужденными атомами и молекулами, распространяясь к аноду со скоростью света, сами производят ионизацию газа и дают начало новым электронным лавинам. Таким путем во всем объеме газа появляются слабо светящиеся скопления ионизованного газа, называемые *стримерами.* В процессе своего развития отдельные электронные лавины догоняют друг друга и, сливаясь вместе, образуют хорошо проводящий мостик из стримеров.

По этому мостику в последующий момент времени и устремляется мощный поток электронов, образующий канал искрового разряда. Поскольку проводящий мостик образуется в результате слияния практически одновременно возникающих стримеров, время его образования много меньше времени, которое требуется отдельной электронной лавине для прохождения расстояния от катода к аноду. Наряду с *отрицательными стримерами*, т. е. стримерами, распространяющимися от катода к аноду, существуют также *положительные стримеры*, которые распространяются в противоположном направлении.

§ 119. Коронный разряд

1. *Коронный разряд* возникает при сравнительно высоких давлениях газа (порядка атмосферного) в сильно неоднородном электрическом поле. Такое поле можно получить между двумя электродами, поверхность одного из которых обладает большой кривизной (тонкая проволочка, острие). Наличие второго электрода не обязательно, его роль могут играть окружающие заземленные электроды. Когда электрическое поле вблизи электрода с большей кривизной достигает примерно $3 \cdot 10^4$ В/м, вокруг этого электрода возникает свечение, имеющее вид оболочки или *короны*, откуда и произошло название разряда. Если корона возникает вокруг отрицательного электрода, то она называется *отрицательной.* В противоположном случае корона называется *положительной.* Механизм возникновения разряда в этих двух случаях разный.

В случае отрицательной короны положительные ионы, образуемые электронными лавинами, ускоряются в сильно неоднородном электрическом поле вблизи катода. Попадая на катод, они выбивают из него электроны (вторичная электронная эмиссия). Выбитые электроны, отталкиваясь от катода, на своем пути порождают новые электронные лавины. Так как электрическое поле убывает при удалении от проволоки, то на некотором расстоянии электронные лавины обрываются, электроны попадают в «темную» область и там прилипают к нейтральным молекулам газа. Образовавшиеся отрицательные ионы и являются основными носителями тока в «темной» области. Пространственный отрицательный заряд этих ионов вблизи анода ограничивает общий разрядный ток. В случае чистых электроположительных газов отрицательные ионы не образуются, а носителями тока в «темной» области являются *сами электроны.* В «темной» области разряд носит *несамостоятельный характер.*

В положительной короне, когда катодом служит электрод с большим радиусом кривизны, электрическое поле у катода слабое. Поэтому электронные лавины не могут порождаться электронами, выбиваемыми из катода вследствие вторичной эмиссии. Электронные лавины порождаются электронами, возникающими вблизи анода при *объемной ионизации газа* фотонами, излучаемыми коронирующим слоем. Они

зарождаются на внешней границе коронирующего слоя и распространяются к положительному электроду (обладающему большой кривизной). Положительные ионы, двигаясь через «темную» область к катоду, образуют *пространственный заряд*, который снова ограничивает силу разрядного тока.

При увеличении напряжения между электродами «темная» область коронного разряда исчезает и возникает *искровой разряд* с полным пробоем разрядного промежутка.

2. Корона иногда возникает в естественных условиях под влиянием *атмосферного электричества* на верхушках деревьев, корабельных мачт и пр. Это явление в старину получило название *огней святого Эльма.* С возникновением коронного разряда приходится считаться в технике высоких напряжений. Образуясь вокруг проводов высоковольтных линий передач электроэнергии, корона ионизует окружающий воздух, вследствие чего возникают вредные *токи утечки.* Для уменьшения этих вредных токов провода высоковольтных линий, а также подводящие провода к лабораторным высоковольтным установкам должны быть достаточно толстыми. Коронные разряды, поскольку они носят прерывистый характер, являются источниками значительных *радиопомех.*

Коронный разряд используется в *электрофильтрах*, предназначенных для очистки промышленных газов от примесей твердых и жидких частиц (дыма в производстве серной кислоты, в литейных цехах заводов цветных металлов и пр.). Принцип действия такого электрофильтра пояснен на рис. 267. Вдоль оси вертикальной трубы натянута заряженная (например, отрицательно) проволока AB, вокруг которой зажигается коронный разряд. Вследствие этого воздух внутри трубы становится сильно ионизованным. Ионы оседают на частицах дыма, поднимающегося вверх по трубе. Получив заряды, одноименные с зарядом коронирующей проволоки, частицы дыма отталкиваются от нее и оседают на стенках трубы. Отсюда они могут быть извлечены механическим способом и использованы для получения содержащихся в них ценных продуктов. Соответствующую демонстрацию легко осуществить в лекционной аудитории. Темным дымом заполняется прозрачная труба в отсутствие электрического поля между проволокой и стенками трубы. При включении напряжения воздух в трубе моментально становится светлым.

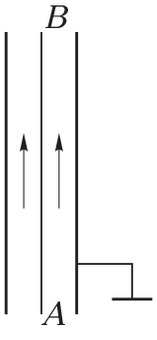


Рис. 267

§ 120. Дуговой разряд

1. Если после получения искрового разряда от мощного источника постепенно уменьшать расстояние между электродами (или сопротивление внешней цепи), то разряд из прерывистого становится *непрерывным* — возникает новая форма газового разряда, называемая *дуговым разрядом.* При этом ток резко увеличивается, достигая десятков и сотен

ампер, а напряжение на разрядном промежутке падает до нескольких десятков вольт.

Дуговой разряд можно получить от источника низкого напряжения, минуя стадию искры. Для этого электроды сближают до соприкосновения. В месте соприкосновения они сильно раскаляются электрическим током, после чего их разводят, получая при этом яркую *электрическую дугу.* Именно таким путем электрическая дуга была впервые получена в 1802 г. русским физиком В. В. Петровым (1761–1834). Электродами в его опытах служили два куска древесного угля, включенные в цепь мощной гальванической батареи. Угли раскалялись до ослепительного свечения, а между ними возникал яркий столб светящегося газа. Работа Петрова была опубликована на русском языке и осталась неизвестной заграничным ученым. В России на нее не обратили внимания, и она была основательно забыта. Явление электрической дуги было вновь открыто в 1810 г. английским химиком Дэви и без достаточных на то оснований названо *вольтовой дугой.*

В настоящее время электрическая дуга, горящая при атмосферном давлении, чаще всего получается между специальными угольными электродами, изготовляемыми из прессованного графита со связывающими веществами. Во время горения расстояние между электродами дуги бывает порядка 5 мм при токе 10–20 А и напряжении между ними 40–50 В. По мере горения дуги угольный катод заостряется, а на аноде образуется углубление, называемое *кратером.* Кратер — наиболее горячее место дуги. Его температура при атмосферном давлении достигает 4000 °С, а при давлении в 20 атм превышает 7000 °С (выше температуры границы фотосферы Солнца). В электрических дугах с металлическими электродами температура меньше (2000–2500 °С) из-за высокой теплопроводности электродов и быстрого испарения металлов, требующего большого количества тепла.

2. Согласно В. Ф. Миткевичу (1872–1951), дуговой разряд поддерживается главным образом за счет *термоэлектронной эмиссии с поверхности катода.* Подтверждением этой точки зрения может служить установленный на опыте факт, что во многих случаях устойчивая дуга получается только при условии, что температура катода достаточно высока. При охлаждении катода дуга горит неустойчиво, периодически гаснет и снова зажигается. Охлаждение же анода не вызывает нарушения устойчивого режима горения дуги.

С этим согласуется также поведение тлеющего разряда при увеличении тока в нем. Когда в результате бомбардировки положительными ионами катод газоразрядной трубки разогревается, появляется термоэлектронный ток, усиливающий электронный ток вторичной эмиссии. Вследствие этого пространственный положительный заряд вблизи катода, а с ним и катодное падение потенциала начинают уменьшаться. При дальнейшем возрастании разрядного тока последнее становится порядка потенциалов ионизации или возбуждения газа, т. е. около 10 В. Проводимость газового промежутка увеличивается, и тлеющий разряд переходит в дуговой. При охлаждении катода вследствие уменьшения

термоэлектронной эмиссии дуга не возникает или работает неустойчиво. Наоборот, тлеющий разряд можно перевести в дугу, нагревая проволочный катод газоразрядной трубки с помощью специальной батареи.

С возрастанием разрядного тока сопротивление дуги R сильно уменьшается из-за увеличения термоэлектронной эмиссии с катода и ионизации газа в разрядном промежутке. При этом сопротивление R убывает сильнее, чем возрастает ток I. Вследствие этого с увеличением тока I напряжение на разрядном промежутке $V = RI$ *не возрастает, а убывает.* Говорят, что дуга имеет *падающую вольт-амперную характеристику*, т. е. такую характеристику, когда напряжение на разрядном промежутке уменьшается с возрастанием тока. Поэтому для поддержания устойчивого горения дуги при случайных изменениях тока, например вследствие охлаждения катода, напряжение на электродах дуги должно быть повышено. Иначе дуга погаснет. С этой целью в цепь дуги включают последовательно *балластное сопротивление.* При случайном уменьшении тока напряжение на балластном сопротивлении уменьшается. Поэтому при неизменном подводимом общем напряжении напряжение на газоразрядном промежутке должно увеличиться, чем и обеспечивается стабильное горение дуги.

3. Наряду с дуговыми разрядами, обусловленными термоэлектронной эмиссией, существуют и разряды другого типа. Примером могут служить дуговые разряды в ртутных лампах. *Ртутная лампа* представляет собой предварительно откачанный кварцевый или стеклянный баллон, наполненный парами ртути. Дуговой разряд зажигается электрической искрой между двумя столбиками ртути, служащими электродами лампы. Ртутная дуга является мощным источником, ультрафиолетовых лучей. Поэтому такие лампы применяют в медицине и в научных исследованиях. Баллоны ламп делают из кварца или специальных сортов стекла, пропускающих ультрафиолетовые лучи.

Экспериментальные исследования показали, что источником мощной эмиссии электронов в ртутной лампе является небольшое, ярко светящееся пятно, возникающее на катоде и непрерывно бегающее по его поверхности (так называемое *катодное пятно*). Плотность тока в катодном пятне огромна и может достигать 10^6–10^7 А/см^2. Катодное пятно может возникнуть не только у поверхности ртутного, но и любого другого металлического электрода.

Ртутные дуги и аналогичные дуги с металлическими электродами получили название *электрических дуг с холодным катодом.* Дело в том, что раньше считалось, что катод действительно является холодным по всей его поверхности. Поэтому термоэлектронная эмиссия с катода не происходит или практически не играет никакой роли. Ленгмюр высказал предположение, что в случае холодного катода дуговой разряд поддерживается *автоэлектронной эмиссией с катода.* Действительно, катодное падение потенциала (10 В) происходит на протяжении порядка длины свободного пробега электрона. Поэтому вблизи

катода возникает сильное электрическое поле, достаточное, чтобы вызвать заметную автоэлектронную эмиссию. Несомненно, автоэлектронная эмиссия в дугах с *холодным* катодом играет существенную роль. Позднее появились указания на возможность нагрева таких катодов в отдельных точках до температур, при которых происходит большая термоэлектронная эмиссия, которая вместе с автоэлектронной эмиссией и поддерживает дуговой разряд. Этот вопрос еще недостаточно исследован.

§ 121. Плазма

1. *Плазмой называется ионизованный квазинейтральный газ, занимающий настолько большой объем, что в нем не происходит сколько-нибудь заметного нарушения квазинейтральности из-за тепловых флуктуаций.* Квазинейтральность газа означает, что количества положительных и отрицательных зарядов в нем *почти одинаковы.*

Оценим размеры области, в которой могут происходить заметные нарушения квазинейтральности, предполагая для простоты, что заряды положительных и отрицательных частиц одинаковы и равны элементарному заряду e. Пусть такая плазма заполняет пространство между плоскостями AB и MN (рис. 268). Допустим, что из-за тепловых флуктуаций отрицательные заряды сместились вверх на расстояние l. Тогда на границах плазмы возникнут макроскопические заряды противоположных знаков с поверхностной плотностью $\sigma = nle$, где n — концентрация частиц одного знака заряда. Напряженность электрического поля в плазме будет $E = 4\pi\sigma = 4\pi nle$, а плотность электрической энергии $E^2/8\pi = 2\pi(nle)^2$. Поскольку энергия электрического поля черпается из кинетической энергии теплового движения частиц газа, величина $E^2/8\pi$ не может превосходить $3nkT$. (На долю отрицательных частиц единицы объема приходится в среднем кинетическая энергия $(3/2)kT$, и такая же энергия — на долю положительных.) Следовательно, если опустить числовой коэффициент 3, то должно быть $2\pi(nle)^2 < nkT$, или $l < D$, где

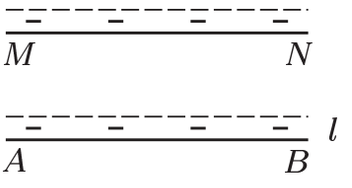


Рис. 268

$$D = \sqrt{\frac{kT}{2\pi ne^2}}. \tag{121.1}$$

Величина D называется *дебаевской длиной* или *дебаевским радиусом.* (Более точное определение дается в задаче 1.)

Таким образом, *чтобы плазма сохраняла квазинейтральность, ее линейные размеры должны намного превосходить дебаевский радиус* D. Только при соблюдении условия квазинейтральности плазма ведет себя как *связанный коллектив заряженных частиц*, а не как простая совокупность невзаимодействующих частиц. Например, если в плазме есть градиент концентрации, то, несмотря на различие коэффициентов

диффузии, положительные и отрицательные ионы не диффундируют с различными скоростями, а благодаря квазинейтральности плазмы перемещаются с одной и той же средней скоростью в одном и том же направлении. Такая диффузия называется *амбиполярной.* Физическая причина амбиполярного характера диффузии плазмы состоит в том, что из-за различия в коэффициентах диффузии и подвижности ионов в плазме возникает некоторое разделение электрических зарядов и связанное с ним электрическое поле, которое замедляет диффузию ионов одного знака и ускоряет диффузию ионов противоположного знака. В результате скорости диффузии положительных и отрицательных ионов выравниваются. Восстановление нарушенной квазинейтральности плазмы аналогично появлению восстанавливающих сил в упругих телах при их деформациях. С этим связана возможность разнообразных *коллективных колебаний* плазмы, значительно более разнообразных, чем в газах, состоящих из нейтральных частиц.

2. В зависимости от степени ионизации α различают *слабо ионизованную* (при α порядка долей процента), *умеренно ионизованную* (α порядка нескольких процентов) и *полностью ионизованную плазму.* В земных природных условиях плазма встречается довольно редко (например, в канале молнии). В верхних слоях атмосферы, в большей степени подверженных воздействию ионизующих факторов (ультрафиолетовые и космические лучи), постоянно присутствует слабо ионизованная плазма — *ионосфера*, отражающая радиоволны и делающая возможной радиосвязь на больших расстояниях (порядка расстояния между диаметрально противоположными точками земного шара). В космическом пространстве плазма представляет собой наиболее распространенное состояние вещества. Солнце и горячие звезды, имеющие высокие температуры, состоят из полностью ионизованной плазмы. Поэтому многие проблемы астрофизики связаны с изучением физических свойств плазмы. На почве астрофизики возникла *магнитная гидродинамика*, в которой плазма, движущаяся в магнитных полях, рассматривается как *сплошная жидкая среда*, обладающая *высокой проводимостью.* Плазма образуется в различных формах газового разряда, например в положительном столбе тлеющего разряда, а также в главном канале искрового разряда. Физика плазмы — сравнительно новый, быстро развивающийся раздел физики, которому посвящены специальные курсы. В курсе общей физики можно сообщить только некоторые отрывочные сведения о плазме.

3. Оценим удельную проводимость λ полностью ионизованной плазмы, состоящей из электронов и положительно заряженных ионов с зарядом Ze каждый. Движение ионов, ввиду их больших масс, можно не учитывать и считать, что весь ток создается движением легких электронов. Величина λ определяется столкновением электронов с ионами. Столкновения электронов между собой на величину тока не влияют, поскольку при таких столкновениях полный импульс электронов не изменяется. От этих столкновений можно отвлечься. Между

ионами и электронами плазмы действуют *кулоновские силы притяжения.* Это — *дальнодействующие силы.* Электрон сравнительно редко подходит к иону на такие малые расстояния, чтобы направление его движения изменилось резко и имело характер скачка. Гораздо большее значение имеют взаимодействия электрона сразу *с очень большим количеством ионов*, при которых направление траектории электрона меняется *плавно и непрерывно.* Отклонения электрона на большие углы от первоначального направления движения происходят в результате накопления *малых отклонений* при взаимодействии его с «далекими» ионами. Поэтому о столкновениях, длине и времени свободного пробега можно говорить лишь в условном смысле. Условимся называть временем свободного пробега электрона промежуток времени τ, в течение которою направление движения электрона меняется на угол порядка $90°$.

Для оценки величины τ допустим сначала, что единственный электрон движется в поле положительного иона с зарядом Ze. Если v — скорость электрона на бесконечности, а b — прицельный параметр, то при прохождении мимо иона траектория электрона отклоняется на угол, определяемый формулой

$$\operatorname{ctg}\frac{\vartheta}{2}=\frac{mv^2b}{Ze^2},$$

где m — масса электрона (см. т. I, § 58, где аналогичная формула получена для движений под действием гравитационных сил). Прицельный параметр b, для которого $\vartheta = 90°$, определяется выражением $b = Ze^2/(mv^2)$. Ему соответствует «эффективное поперечное сечение»

$$\sigma=\pi b^2=\pi\left(\frac{Ze^2}{mv^2}\right)^2.$$

Учет далеких взаимодействий приводит к тому же результату, но увеличенному в L раз:

$$\sigma=\pi L\left(\frac{Ze^2}{mv^2}\right)^2.$$

Коэффициент L называется *кулоновским логарифмом.* Он почти не зависит от температуры и плотности плазмы. Для плазмы, состоящей из полностью ионизованного дейтерия, при $kT \sim 10$ кэВ и концентрации электронов $n \sim 10^{12}$–10^{15} см$^{-3}$ $L \approx 15$. Так как каждый положительный ион содержит Z элементарных зарядов, то концентрация таких ионов будет n/Z, а средняя длина и время «свободного пробега»

$$\overline{l}=\frac{1}{\sigma n/Z}=\frac{Z}{\sigma n},\quad \overline{\tau}\approx\frac{\overline{l}}{v}=\frac{m^2\overline{v}^3}{\pi Zne^4L}.$$

Подставив сюда $m\overline{v}^2 \approx 3kT$, получим

$$\overline{\tau}\approx\frac{(3kT)^{3/2}\sqrt{m}}{\pi Zne^4L}. \tag{121.2}$$

Отсюда для удельной проводимости плазмы находим

$$\lambda \approx \frac{ne^2\overline{\tau}}{m} \approx \frac{(3kT)^{3/2}}{\pi Ze^2 L\sqrt{m}}. \tag{121.3}$$

Конечно, приведенный вывод надо рассматривать как грубую оценку, а не как доказательство формулы (121.3). Однако можно дать и более обоснованный вывод этой формулы. Подставляя числовые значения и выражая энергетическую температуру $\Theta = kT$ в электронвольтах, приходим к оценочной формуле для λ (в обратных секундах):

$$\lambda \approx \frac{10^{13}}{Z}\,\Theta^{3/2},$$

причем мы приняли $L \approx 15$. Проводимость плазмы растет пропорционально абсолютной температуре в степени $3/2$. В горячей плазме проводимость становится очень высокой. Так, при $\Theta = 10$ кэВ для дейтериевой плазмы $\lambda \approx 10^{19}$ с$^{-1}$, т. е. больше, чем у меди ($5\cdot 10^{17}$ с$^{-1}$). Еще быстрее растет с температурой теплопроводность плазмы, а именно пропорционально $\Theta^{5/2}$, так как для плазмы, очевидно, должен быть справедлив закон Видемана–Франца.

4. Большое различие в массах электронов и ионов плазмы делает возможным в плазме существование таких квазиравновесных состояний, которые в известном приближении могут быть характеризованы *двумя температурами.* Действительно, пусть начальное распределение скоростей электронов и ионов плазмы изотропно, но не максвелловское. При столкновении электрона с другим электроном они обмениваются энергией, величина которой порядка начальной энергии самих электронов. Поэтому время установления распределения электронов по энергиям (т. е. максвелловского распределения) из-за столкновений между ними можно оценить по формуле (121.2). (Это станет ясно, если в ней массу электрона m заменить приведенной массой $m/2$.) Это время, называемое *электронным временем релаксации* $\overline{\tau}_e$, пропорционально квадратному корню из массы электрона: $\overline{\tau}_e \sim \sqrt{m_e}$. Точно так же определяется *ионное время релаксации*, за которое успевает устанавливаться распределение по энергиям между ионами из-за столкновений между ними: $\overline{\tau}_i \sim \sqrt{m_i}$. Не то будет при взаимодействии электронов с ионами. Здесь по формуле (121.2) получается время порядка $\sqrt{m_e}$. Однако при каждом столкновении быстрая частица передает медленной лишь незначительную долю своей энергии. В среднем эта доля порядка m_e/m_i от первоначальной энергии быстрой частицы (см. т. I, § 28, задача 9). Для выравнивания энергий потребуется релаксационное время $\overline{\tau}_{ie}$, в m_i/m_e раз большее, чем $\overline{\tau}_e$. Таким образом,

$$\overline{\tau}_e : \overline{\tau}_i : \overline{\tau}_{ie} \sim 1 : \sqrt{\frac{m_i}{m_e}} : \frac{m_i}{m_e}. \tag{121.4}$$

Отсюда следует, что $\overline{\tau}_e \ll \overline{\tau}_i \ll \overline{\tau}_{ie}$. Если плазму предоставить самой себе, то сначала установится максвелловское распределение скоростей электронов, затем — ионов. Возникнет квазиравновесное состояние,

в котором электроны будут иметь температуру T_e, а ионы — температуру T_i. Вообще говоря, $T_e \neq T_i$. В этом случае плазму называют *неизотермической* или *двухтемпературной.* Затем в результате обмена энергиями между электронами и ионами установится максвелловское распределение для всей плазмы, характеризующееся общей температурой электронов и ионов (*изотермическая плазма*).

Когда плазма находится в электрическом поле, то в ней начинает течь электрический ток и выделяться джоулева теплота. При этом энергию от поля получают почти исключительно электроны, как наиболее подвижные частицы. Ионы нагреваются главным образом за счет энергии, которую они получают от «горячих» электронов при кулоновских взаимодействиях с ними. Так как последний процесс происходит сравнительно медленно, то *температура электронов в плазме оказывается выше температуры ионов.* Различие между ними может быть весьма значительным. Так, в положительном столбе тлеющего разряда при давлениях порядка 0,1 миллиметров ртутного столба температура электронов может достигать 50 000 °С и выше, тогда как температура ионов не превышает нескольких сотен градусов.

5. Основной практический интерес, который представляет физика плазмы, связан с решением проблемы *управляемого термоядерного синтеза.* Для того чтобы в веществе начались достаточно интенсивные термоядерные реакции, его необходимо нагреть до температуры в несколько килоэлектронвольт или десятков килоэлектронвольт, а при таких температурах всякое вещество находится в состоянии плазмы. Наиболее перспективными «рабочими веществами» для термоядерного реактора являются *изотопы водорода*: *дейтерий* (D) и *тритий* (T). Термоядерную реакцию синтеза легче получить не в чистом дейтерии, а в его смеси с тритием. Полное количество дейтерия в океанах $4 \cdot 10^{13}$ тонн, что эквивалентно энергии 10^{20} кВт · лет (полная потребляемая на всем земном шаре мощность составляет 10^{10} кВт). Тритий, как сильно радиоактивный элемент, в природных условиях не встречается, а получается искусственно. В будущих термоядерных реакторах расход трития должен с избытком пополняться воспроизводством (регенерацией) его в результате облучения ^{6}Li нейтронами, получающимися в самих термоядерных реакторах.

Так как термоядерные реакции должны происходить сравнительно плавно и медленно, то возникает необходимость достаточно длительного удержания горячей плазмы в ограниченном объеме рабочей камеры и изоляции ее от стенок этой камеры. Для этого предлагается использовать *магнитную термоизоляцию*, т. е. помещать плазму в сильное магнитное поле, препятствующее ионам и электронам перемещаться в поперечном направлении и уходить на стенки камеры.

Необходимое требование, которому должен удовлетворять всякий термоядерный реактор, состоит в том, чтобы энергия, выделяющаяся в ядерных реакциях, с избытком компенсировала затраты энергии от

внешних источников. Основными источниками потерь энергии является *тормозное излучение* электронов при кулоновских столкновениях последних, а также *магнитотормозное* (*циклотронное* или *бетатронное*) излучение, возникающее вследствие ускоренного движения электронов в магнитном поле. Для самоподдерживающихся термоядерных реакций требуется нагреть плазму до некоторой «критической» температуры (50 кэВ для реактора на чистом дейтерии и 10 кэВ для реактора на равнокомпонентной смеси дейтерия с тритием). При этом, как показывает расчет, должен выполняться так называемый критерий Лоусона: $n\tau > 10^{16}$ с/см^3 для реактора D–D и $n\tau > 10^{14}$ с/см^3 для реактора D–T. Здесь n — концентрация ионов плазмы (одного знака), а τ — среднее время удержания плазмы.

Основная трудность, стоящая на пути создания управляемого термоядерного синтеза, связана с получением *спокойной*, или *устойчивой*, *плазмы*. Дело в том, что из-за дальнодействующего характера кулоновских сил в плазме происходят разные коллективные процессы, например самопроизвольно возникающие *шумы* и *колебания*, делающие плазму *неустойчивой*. Основные усилия при решении проблемы управляемого термоядерного синтеза и направлены на подавление этих неустойчивостей.

ЗАДАЧИ

1. Полупространство справа от бесконечной заряженной плоскости AB (рис. 269) заполнено полностью ионизованной плазмой, состоящей из однозарядных положительных ионов и электронов. Найти электрическое поле и распределение зарядов в плазме на любых расстояниях от плоскости AB, если поверхностная плотность внешних зарядов на плоскости AB равна σ.

Р е ш е н и е. В плазме вблизи плоскости AB образуется избыток ионов, заряженных разноименно с σ, и недостаток ионов, заряженных одноименно. В плазме возникает электрическое поле $\mathbf{E}$, перпендикулярное к плоскости AB. Направим координатную ось X внутрь плазмы перпендикулярно к этой плоскости. Из соображений симметрии ясно, что все величины будут функциями только координаты x. Следовательно,

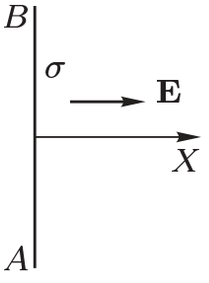


Рис. 269

$$\operatorname{div}\mathbf{E} = \frac{dE}{dx} = 4\pi\rho. \tag{121.5}$$

Плотность свободных электрических зарядов ρ можно представить в виде $\rho = (n^+ - n^-)e$, где n^+ — концентрация положительных, а n^- — отрицательных ионов. Вдали от границы AB (в бесконечности) плазма *электрически нейтральна*, там $n^+ = n^- \equiv n$. Вблизи границы концентрации n^+ и n^- различны. Если φ — электрический потенциал поля $\mathbf{E}$, то согласно формуле Больцмана

$$n^+ = n\exp\left(-\frac{e\varphi}{kT}\right), \quad n^- = n\exp\left(+\frac{e\varphi}{kT}\right). \tag{121.6}$$

При $x = +\infty$ потенциал φ равен нулю, а потому $n^+ = n^- = n$. Итак,

$$n^+ - n^- = n\left[\exp\left(-\frac{e\varphi}{kT}\right) - \exp\left(+\frac{e\varphi}{kT}\right)\right] = -2n\operatorname{sh}\frac{e\varphi}{kT}.$$

Следовательно,

$$\frac{dE}{dx} = -8\pi ne \operatorname{sh} \frac{e\varphi}{kT}.$$

Умножая обе части последнего уравнения на $2E dx = -2d\varphi$, получим

$$dE^2 = 16\pi ne \operatorname{sh} \frac{e\varphi}{kT}\, d\varphi.$$

Простое интегрирование дает

$$E^2 = 16\pi nkT \operatorname{ch} \frac{e\varphi}{kT} + C.$$

Постоянную C надо определить из условия, чтобы в бесконечности (где $\varphi = 0$) поле E обращалось в нуль. Это дает

$$E^2 = 16\pi nkT \left(\operatorname{ch} \frac{e\varphi}{kT} - 1 \right) = 32\pi nkT \operatorname{sh}^2 \frac{e\varphi}{kT}.$$

При извлечении квадратного корня перед ним надо взять знак плюс. Это следует из того, что когда поле E положительно, т. е. направлено вправо, то и потенциал φ также положителен. Если же поле отрицательно, т. е. направлено влево, то и потенциал φ отрицателен. Итак,

$$E = -\frac{d\varphi}{dx} = 4\sqrt{2\pi nkT} \operatorname{sh} \frac{e\varphi}{kT}. \tag{121.7}$$

После интегрирования получим

$$\operatorname{th} \frac{e\varphi}{kT} = C_1 e^{-x/D},$$

где D — постоянная:

$$D = \sqrt{\frac{kT}{8\pi ne^2}}. \tag{121.8}$$

Эта формула и дает точное выражение дебаевского радиуса, введенного выше в тексте настоящего параграфа.

Постоянную интегрирования C_1 можно выразить через потенциал φ_0 на границе плазмы. Таким путем получим

$$\operatorname{th} \frac{e\varphi}{kT} = \operatorname{th} \frac{e\varphi_0}{kT}\, e^{-x/D}. \tag{121.9}$$

Самый потенциал φ легко связать с напряженностью внешнего электрического поля $E_0 = 2\pi\sigma$. При $x = 0$ должно быть $E = E_0$, а потому

$$E_0 = 4\sqrt{2\pi nkT} \operatorname{sh} \frac{e\varphi_0}{kT}. \tag{121.10}$$

Таким образом, при удалении от границы плазмы величина $\operatorname{th}(e\varphi/kT)$ экспоненциально убывает. На протяжении дебаевского радиуса она убывает в e раз. Этим определяется и закон убывания поля E с координатой x. Можно сказать, что по порядку величины дебаевский радиус D определяет глубину, на которую электрическое поле проникает в плазму.

2. Решить предыдущую задачу для точечного заряда q, окруженного со всех сторон плазмой. Ограничиться расстояниями от заряда r, на которых выполняется условие $|e\varphi| \ll kT$.

Решение. При выполнении условия $|e\varphi| \ll kT$ выражения (121.6) можно разложить в ряды и оборвать их на членах первой степени. Подставляя полученные выражения в уравнение Пуассона $\Delta\varphi = -4\pi e(n^+ - n^-)$,

придадим ему вид

$$\Delta\varphi + \frac{\varphi}{D^2} = 0, \tag{121.11}$$

где D определяется прежним выражением (121.8). Ввиду сферической симметрии

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d\varphi}{dr}\right) + \frac{\varphi}{D^2} = 0.$$

Общее решение этого уравнения

$$\varphi = \frac{1}{r}\left(Ce^{-r/D} + C'e^{r/D}\right). \tag{121.12}$$

Постоянная интегрирования C' должна обращаться в нуль, чтобы удовлетворить условию $\varphi = 0$ при $r = \infty$. Для определения постоянной C допустим, что условие $|e\varphi| \ll kT$ выполняется уже при $r \ll D$. Тогда на таких расстояниях еще применимо решение (121.12). Но тогда оно переходит в кулоновский потенциал $\varphi = q/r$. Следовательно, должно быть $C = q$, т. е.

$$\varphi = \frac{q}{r}\,e^{-r/D}. \tag{121.13}$$

Этот потенциал называется *дебаевским потенциалом.* Формула (121.13) показывает, что дебаевский радиус по порядку величины определяет расстояние от заряда q, на котором кулоновское поле этого заряда экранируется противоположно заряженными ионами плазмы Можно также сказать, что действие кулоновского поля заряда q простирается на расстояние порядка дебаевского радиуса D, а на больших расстояниях практически не имеет места.

3. В электростатике считается, что в состоянии равновесия электричество распределяется по поверхности проводника. Фактически оно распределяется не по математической поверхности, а внутри поверхностного слоя конечной толщины l. Оценить эту толщину.

Ответ. Величина l порядка дебаевского радиуса D, если выражение для D изменить с учетом распределения Ферми для электронов металла. Для этого величину kT надо заменить энергией Ферми (99.6). Сделав это и опуская числовые множители порядка единицы, получим

$$l = \sqrt{a_1 n^{-1/3}},$$

где $a_1 = h^2/(4\pi^2 me) = 0{,}53\cdot 10^{-8}$ см (радиус первой боровской орбиты в атоме водорода), а $n^{-1/3}$ — среднее расстояние между свободными электронами металла.

Глава X
КОЛЕБАНИЯ И ВОЛНЫ

§ 122. Уравнение колебательного контура

1. Колебаниями в физике называют не только периодические или почти периодические движения тел, когда колеблющееся тело многократно повторяет одно и то же движение туда и обратно около определенного положения равновесия, а придают этому понятию более широкий смысл. *Под колебанием понимают всякий периодический или приблизительно периодический процесс, в котором значения той или иной физической величины повторяются точно или приближенно через равные или приблизительно равные промежутки времени.*

Допустим, например, что прямоугольная рамка (рис. 270) равномерно вращается в постоянном однородном магнитном поле с постоянной угловой скоростью ω. Если S — площадь рамки, то пронизывающий ее магнитный поток будет $\Phi = BS\cos\alpha$, где α — угол между направлением вектора магнитной индукции $\mathbf{B}$ и нормалью к плоскости рамки $\mathbf{N}$. При равномерном вращении $\alpha = \omega t$, $\Phi = \Phi_0\cos\omega t$, где $\Phi_0 = BS$. В рамке возбуждается электродвижущая сила $\mathscr{E}_{\text{инд}} = -d\Phi/dt = \mathscr{E}_0\sin\omega t$ и электрический ток $I = I_0\sin\omega t$, где $\mathscr{E}_0$ и I_0 — постоянные. Все эти выражения описывают колебательный процесс, в котором колеблющимися величинами являются магнитный поток Φ, электродвижущая сила $\mathscr{E}$ и электрический ток I.

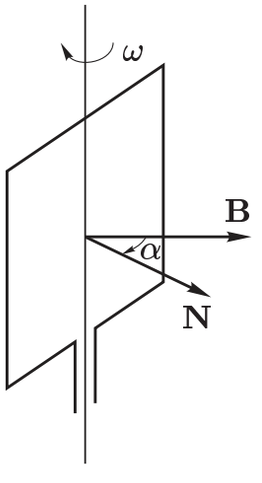


Рис. 270

Специфические закономерности колебательных явлений, определяющие не мгновенные значения колеблющихся величин, а характеризующие *колебательный процесс в целом*, не зависят от того, какова физическая природа величин, совершающих колебания. Такие закономерности изучает *теория колебаний*, характеризующаяся *единым подходом* к колебаниям различной физической природы. В дальнейшем эта особенность колебательных закономерностей будет проиллюстрирована на конкретных примерах.

В настоящей главе будут изучаться преимущественно электрические колебания. Однако для лучшего понимания этих явлений мы будем сопоставлять их с аналогичными колебаниями механических систем.

2. Изучение электрических колебаний мы начнем с вывода *уравнения колебательного контура.* Так называется система, состоящая из последовательно соединенных конденсатора, катушки самоиндукции

L и проводника с омическим сопротивлением R (рис. 271). Внешняя электродвижущая сила создает между полюсами *1* и *2* определенное напряжение $\mathscr{E}$, вообще говоря, меняющееся с течением времени. Одно из направлений при обходе контура тока примем за положительное. Оно обозначено на рис. 271 стрелками. Ток считается положительным, если он течет по контуру в положительном направлении, и отрицательным в противоположном случае. Обозначим через q заряд той из обкладок конденсатора, направление от которой к другой обкладке совпадает с положительным направлением контура *13421*. Применим к этому контуру уравнение Максвелла

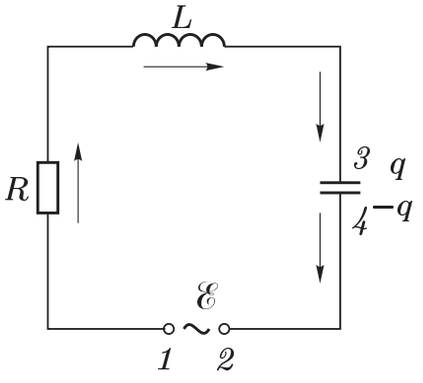


Рис. 271

$$\int E_l\,dl = -\frac{d\Phi}{dt}. \tag{122.1}$$

Мы пользуемся *практическими единицами*, которые при рассмотрении электрических колебаний более удобны. Пусть выполнено условие квазистационарности. Тогда, применяя к участку *13* закон Ома, найдем

$$\int\limits_{13} E_l\,dl = \int \frac{j}{\lambda}\,dl = I\int \frac{dl}{S\lambda} = RI,$$

где R — омическое сопротивление этого участка. Если сопротивление участка *42* пренебрежимо мало, то интеграл по пути *32* равен напряжению V между обкладками конденсатора. Для квазистационарных процессов

$$\int\limits_{32} E_l\,dl = V = \frac{q}{C}.$$

Наконец, интеграл $\int\limits_{21} E_l\,dl = -\int\limits_{12} E_l\,dl$ есть подводимое напряжение между полюсами *1* и *2*, взятое с противоположным знаком:

$$\int\limits_{21} E_l\,dl = -\mathscr{E}.$$

В результате уравнение (122.1) примет вид

$$\frac{d\Phi}{dt} + RI + \frac{q}{C} = \mathscr{E}. \tag{122.2}$$

Для квазистационарных токов $\Phi = LI$. Кроме того,

$$I = \frac{dq}{dt}, \tag{122.3}$$

так что

$$\frac{d}{dt}\left(L\frac{dq}{dt}\right) + R\frac{dq}{dt} + \frac{q}{C} = \mathscr{E}. \tag{122.4}$$

Это и есть уравнение колебательного контура. Если катушка самоиндукции не деформируется ($L = \text{const}$), оно переходит в

$$L\frac{d^2q}{dt^2} + R\frac{dq}{dt} + \frac{q}{C} = \mathscr{E}. \tag{122.5}$$

Механическим аналогом (122.5) может служить уравнение движения груза на пружине (рис. 272). Если справедлив закон Гука (1635–1703), а при движении груза возникает тормозящая сила $-\alpha\dot{x}$, пропорциональная скорости $\dot{x}$, то уравнение движения имеет вид

$$m\frac{d^2x}{dt^2} = -kx - \alpha\dot{x} + F,$$

или

$$m\frac{d^2x}{dt^2} + \alpha\dot{x} + kx = F, \tag{122.6}$$

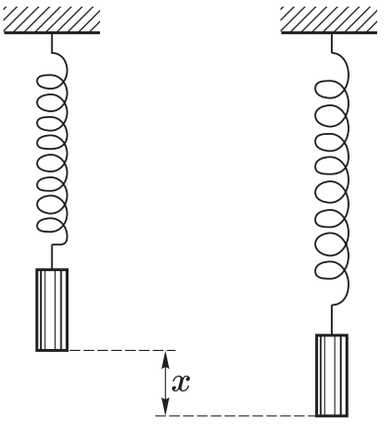
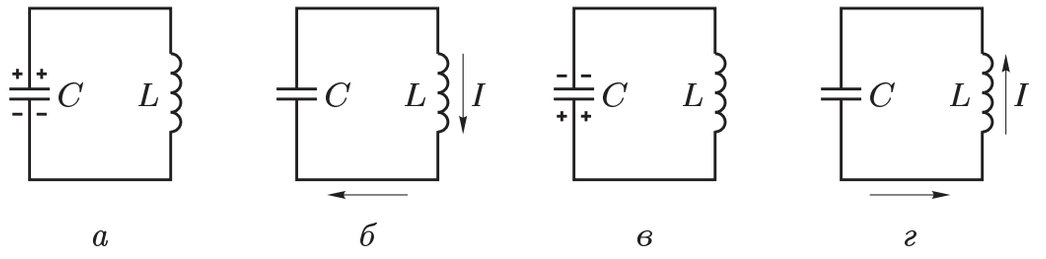


Рис. 272

где x — отклонение груза из положения равновесия, считаемое положительным, если оно направлено вниз. Величина $-kx$ есть *восстанавливающая сила*, равная сумме веса тела и силы натяжения пружины, F — результирующая всех прочих сил, действующих на груз. Это уравнение отличается от (122.5) только обозначениями и физическим смыслом входящих в него величин. Математически оба уравнения тождественны. В уравнении (122.5) роль отклонения x играет заряд конденсатора q; массы m — самоиндукция L; коэффициента сопротивления α — электрическое сопротивление R; коэффициента упругости пружины k — величина, обратная емкости, $1/C$; внешней силы F — внешняя электродвижущая сила $\mathscr{E}$. Одинаковые уравнения должны иметь и одинаковые решения. Заметив это, допустим, что в уравнении (122.6) $F = 0$, а коэффициент сопротивления α мал. Тогда, как хорошо известно из повседневного опыта, при отклонении груза из положения равновесия или сообщении ему толчка в вертикальном направлении возникнут колебания, слабо затухающие во времени. При $\alpha = 0$ затухания совсем не будет. Из математической тождественности уравнений (122.5) и (122.6) следует, что возникнут *электрические колебания*, если заряженный конденсатор замкнуть через катушку самоиндукции. При этом заряд конденсатора q будет меняться во времени по тому же закону, что и отклонение груза из положения равновесия. Если нет омического сопротивления, то электрические колебания в колебательном контуре будут *незатухающими.* При наличии сопротивления R колебания затухают.

3. И без обращения к уравнениям и механической аналогии нетрудно понять, почему в колебательном контуре возникают и поддерживаются электрические колебания. Для простоты будем считать, что электрическое сопротивление колебательного контура равно нулю. Пусть

в начальный момент верхняя пластинка конденсатора заряжена положительным электричеством, нижняя — отрицательным, а ток в колебательном контуре равен нулю (рис. 273 *а*). В этот момент вся энергия колебательного контура сосредоточена в конденсаторе. При отсутствии



а *б* *в* *г*

Рис. 273

внешних электродвижущих сил конденсатор начнет разряжаться, через катушку самоиндукции потечет электрический ток. Электрическая энергия конденсатора начнет превращаться в магнитную энергию катушки самоиндукции. Этот процесс закончится, когда заряд конденсатора обратится в нуль, а ток в контуре достигнет максимума (рис. 273 *б*). Начиная с этого момента ток, не меняя направления, начнет убывать. Однако он не сразу упадет до нуля, так как этому препятствует электродвижущая сила индукции. Ток будет заряжать нижнюю пластинку конденсатора положительно, а верхнюю — отрицательно. Возникнет электрическое поле, стремящееся ослабить ток. В конце концов ток обратится в нуль, а заряд на конденсаторе достигнет максимума (рис. 273 *в*). Тогда заряды на пластинках конденсатора по абсолютной величине станут такими же, что и в исходном положении *а*, только знаки их будут противоположными. С этого момента конденсатор начнет разряжаться вновь — по проводам потечет ток в направлении, противоположном направлению тока в положении *б*. В момент максимума тока (рис. 273 *г*) конденсатор разрядится, а затем колебательный контур вернется в исходное состояние *а*. После этого описанный цикл разрядки и зарядки конденсатора повторится снова. И если бы не было потерь энергии, то такое повторение происходило бы неограниченно долго — в контуре совершались бы строго периодические незатухающие электрические колебания.

4. Уравнения (122.5) и (122.6) — *дифференциальные уравнения второго порядка.* Если «внешних сил» $\mathscr{E}$ или F нет, то уравнения *линейны и однородны* относительно неизвестных q или x и их производных по времени. Они описывают так называемые *свободные колебания.* Колебательные системы, свободные колебания которых описываются линейными уравнениями, называются *линейными колебательными системами.* Введем обозначения:

$$\omega_0^2 = \frac{1}{LC} \quad \text{или} \quad \omega_0^2 = \frac{k}{m}, \tag{122.7}$$

$$2\gamma = \frac{R}{L} \quad \text{или} \quad 2\gamma = \frac{\alpha}{m}, \tag{122.8}$$

$$X = \frac{\mathscr{E}}{C} \quad \text{или} \quad X = \frac{F}{m}. \tag{122.9}$$

Тогда

$$\ddot{q} + 2\gamma\dot{q} + \omega_0^2 q = X, \tag{122.10}$$

$$\ddot{x} + 2\gamma\dot{x} + \omega_0^2 x = X. \tag{122.11}$$

Величина ω_0 называется *собственной частотой* колебательной системы, а γ — *коэффициентом затухания.* Смысл этих названий выяснится в дальнейшем.

5. Уравнение (122.5) можно переписать в виде

$$V_L + V_R + V_C = V_{\text{вх}},$$

где V_L, V_R, V_C — напряжения на катушке самоиндукции, омическом сопротивлении и конденсаторе, а $V_{\text{вх}}$ — входное напряжение, подводимое к колебательному контуру. Как видно из (122.5), эти величины связаны соотношениями

$$V_R = R\frac{dq}{dt} = CR\frac{dV_C}{dt}, \quad V_L = L\frac{d^2q}{dt^2} = \frac{L}{R}\frac{dV_R}{dt}.$$

На этих соотношениях основано применение электрических схем для *автоматического дифференцирования* и *интегрирования.* Соответствующие устройства называются *дифференцирующими* и *интегрирующими ячейками.* Допустим, например, что конденсатора в цепи нет. Тогда (122.5) запишется в виде

$$V_L + V_R = V_{\text{вх}},$$

причем $V_L = \dfrac{L}{R}\dfrac{dV_R}{dt}$, т. е. напряжение на самоиндукции с точностью до числового множителя равно производной от напряжения на сопротивлении R. Если параметры L и R подобрать так, чтобы было выполнено условие $|V_L| \ll |V_{\text{вх}}|$, то $V_R \approx V_{\text{вх}}$ и, следовательно, $V_L = \dfrac{L}{R}\dfrac{dV_{\text{вх}}}{dt}$. Подадим на осциллограф напряжение с катушки самоиндукции. Тогда получится осциллограмма, представляющая производную входного напряжения. Аналогичная RL-ячейка может быть использована и для интегрирования. Действительно, $V_R = \dfrac{R}{L}\int V_L\,dt$. Подберем параметры R и L так, чтобы было $|V_R| \ll |V_L|$. Тогда приближенно $V_L = V_{\text{вх}}$ и, следовательно, $V_R = \dfrac{R}{L}\int V_{\text{вх}}\,dt$. Подавая на осциллограф напряжение с сопротивления R, мы получим осциллограмму, представляющую интеграл от входного напряжения. Аналогично действуют дифференцирующие и интегрирующие RC-ячейки.

ЗАДАЧА

Колебательный контур содержит конденсатор с утечкой. Это значит, что небольшая часть тока, поступающего на одну из обкладок конденсатора, проходит через диэлектрик на другую обкладку. Емкость конденсатора равна C, его сопротивление R. Пренебрегая сопротивлением катушки самоиндукции и прочих проводов и предполагая, что выполнено условие квазистационарности, вывести уравнение собственных колебаний колебательного контура.

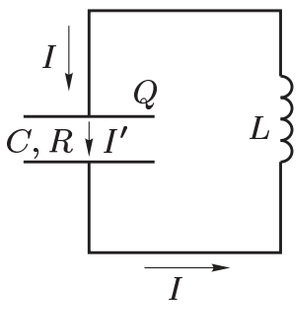


Рис. 274

Найти собственную частоту ω_0 и коэффициент затухания γ.

Р е ш е н и е. При выполнении условия квазистационарности (рис. 274)

$$L\frac{dI}{dt} + V = 0, \quad Q = CV,$$

$$\dot{Q} = I - I', \quad V = RI'.$$

Исключая Q, I и I', отсюда находим

$$\ddot{V} + 2\gamma\dot{V} + \omega_0^2 V = 0, \text{ где } \omega_0^2 = \frac{1}{LC}, \ \gamma = \frac{1}{2RC}.$$

§ 123. Свободные колебания гармонического осциллятора

1. Если нет омического сопротивления, то свободные колебания в колебательном контуре описываются уравнением

$$\ddot{q} + \omega_0^2 q = 0. \tag{123.1}$$

Уравнением такого же типа описываются свободные незатухающие колебания груза на пружине. Всякая система — механическая, электрическая или какая-либо другая, свободные колебания которой подчиняются уравнению типа (123.1), называется *гармоническим осциллятором.* При наличии силы сопротивления $2\gamma\dot{q}$ система называется *гармоническим осциллятором с затуханием.*

Для решения уравнения (123.1) умножим обе части его на $\dot{q}$. Тогда после небольших преобразований получим

$$\frac{d}{dt}\left(\dot{q}^2 + \omega_0^2 q^2\right) = 0.$$

Отсюда следует, что величина $\dot{q}^2 + \omega_0^2 q^2$ не меняется во времени. Так как эта величина есть сумма двух квадратов, то она существенно положительна и может быть представлена в виде

$$\dot{q}^2 + \omega_0^2 q^2 = \omega_0^2 q_0^2,$$

где q_0 — постоянная. Это равенство выражает сохранение энергии, так как его можно записать в виде

$$\frac{1}{2}LI^2 + \frac{q^2}{2C} = \text{const}.$$

Чтобы выполнить второе интегрирование, разделим переменные:

$$\frac{dq}{\sqrt{q_0^2 - q^2}} = \pm\omega_0\, dt.$$

Отсюда

$$\arccos\frac{q}{q_0} = \pm\omega_0 t + \text{const},$$

или

$$q = q_0 \cos(\omega_0 t + \delta). \tag{123.2}$$

Постоянные интегрирования q_0 и δ определяются начальными условиями. В качестве таковых можно, например, взять значения заряда q и тока $I = \dot{q}$ в момент времени $t = 0$.

2. Формулой вида (123.2) описываются также свободные колебания груза, подвешенного на пружине, физического или математического маятника при малых отклонениях, ножки звучащего камертона, а также колебания напряжения в цепи городского тока. Если какая-либо величина меняется во времени по закону (123.2), то говорят, что она совершает *гармоническое колебание.* С гармоническими колебаниями механических систем мы подробно ознакомились уже в первом томе, где нас больше всего интересовали периоды свободных колебаний таких систем. Величина ω_0 называется *круговой* или *циклической частотой* гармонического колебания. Она совпадает с собственной круговой частотой колебательной системы, определяемой формулой (122.7). Промежуток времени

$$T_0 = \frac{2\pi}{\omega_0}, \tag{123.3}$$

через который значения колеблющейся величины периодически повторяются, называется *периодом колебания.* Число колебаний в единицу времени

$$\nu_0 = \frac{1}{T_0} = \frac{\omega_0}{2\pi} \tag{123.4}$$

называется *частотой колебаний.* За единицу частоты принимают *герц.* Герц есть такая частота, когда в одну секунду совершается одно колебание. В дальнейшем прилагательное «круговая» будет часто опускаться. О какой частоте идет речь, будет видно из обозначений. Круговая частота всегда обозначается ω или Ω, просто частота — ν. Величина q_0 называется *амплитудой*, а величина $\omega_0 t + \delta$ — *фазой колебания.* Величину δ называют *начальной фазой.* Собственные частоты ω_0 и ν_0, а также период собственных колебаний T_0 зависят только от *устройства* колебательной системы. Напротив, амплитуда q_0 и начальная фаза δ определяются не самой колебательной системой, а *начальными условиями.*

Для электрических колебаний собственная частота определяется формулой (122.7). Поэтому

$$T_0 = 2\pi\sqrt{LC}. \tag{123.5}$$

Эта формула называется *формулой Вильяма Томсона.*

Если по оси абсцисс откладывать время, а по оси ординат — значение колеблющейся величины q, то получится *синусоида* (рис. 275). Это — периодическая кривая, значения ее ординат периодически повторяются через период T_0. Амплитуда q_0 есть максимальное отклонение величины q от ее нулевого значения.

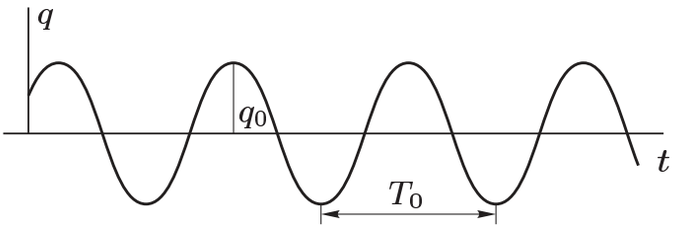


Рис. 275

Ток при электрических колебаниях найдется дифференцированием выражения (123.2):

$$I = \dot{q} = -\omega_0 q_0 \sin(\omega_0 t + \delta),$$

или

$$I = \omega_0 q_0 \cos(\omega_0 t + \delta + \pi/2).$$

Отсюда видно, что колебания тока I опережают по фазе колебания заряда q на $\pi/2$. Электрическая и магнитная энергии определяются выражениями

$$W_e = \frac{q^2}{2C} = \frac{q_0^2}{2C}\cos^2(\omega_0 t + \delta),$$

$$W_m = \frac{1}{2}LI^2 = \frac{1}{2}L\omega_0^2 q^2 \sin^2(\omega_0 t + \delta) = \frac{q_0^2}{2C}\sin^2(\omega_0 t + \delta).$$

Представим их в виде

$$W_e = \frac{q_0^2}{4C} + \frac{q_0^2}{4C}\cos(2\omega_0 t + \delta),$$

$$W_m = \frac{q_0^2}{4C} - \frac{q_0^2}{4C}\cos(2\omega_0 t + \delta).$$

Средние значения этих величин одинаковы и равны

$$\overline{W}_e = \overline{W}_m = \frac{q_0^2}{4C} = \frac{1}{4}LI_0^2.$$

Около этих средних значений величины W_e и W_m совершают гармонические колебания с круговой частотой $2\omega_0$. Непрерывно происходит переход электрической энергии в магнитную и обратно. Когда электрическая энергия достигает максимума, магнитная обращается в нуль, и наоборот. Полная энергия

$$W = W_e + W_m = \frac{q_0^2}{2C} \tag{123.6}$$

остается постоянной, как и должно быть по закону сохранения энергии. Она, как видно из формулы (123.6), пропорциональна *квадрату амплитуды*. Это справедливо и для механических гармонических колебаний.

3. В заключение точно сформулируем *условие квазистационарности*, выполнение которого предполагалось при рассмотрении всех колебаний в колебательном контуре. Квазистационарность означает,

что мгновенные значения тока I практически одинаковы на всех участках проводов, соединяющих обкладки конденсатора. Для этого все изменения во времени должны происходить настолько медленно, чтобы распространение электродинамических взаимодействий можно было считать *мгновенным.* Такие взаимодействия распространяются со скоростью, которая по порядку величины совпадает со скоростью света в вакууме c. Обозначим через l длину провода, соединяющего обкладки конденсатора (практически эта величина совпадает с длиной провода, из которого изготовлена обмотка катушки самоиндукции). На прохождение длины l электромагнитное возмущение затрачивает время порядка $\tau = l/c$. Условие квазистационарности будет выполнено, если $\tau \ll T_0$ или

$$l \ll \lambda, \tag{123.7}$$

где λ — длина электромагнитной волны в вакууме:

$$\lambda = cT_0. \tag{123.8}$$

ЗАДАЧА

Полностью ионизованная плазма, состоящая из электронов и положительных ионов, ограничена двумя параллельными плоскостями (рис. 268). Если все электроны сместить вдоль оси X(выбранной перпендикулярно к этим плоскостям), то возникнет сила, возвращающая их в положение равновесия. В плазме начнутся колебания. Они называются *плазменными колебаниями.* Определить их характер и собственную частоту. «Тяжелые» ионы можно считать неподвижными.

Р е ш е н и е. Если смещение электрона из положения равновесия равно x, то электрическое поле, возникающее в плазме, будет $E = 4\pi nex$, а сила, действующая на электрон, $F = -4\pi ne^2x$. Под действием этой силы электроны будут совершать гармонические колебания с круговой частотой

$$\omega_0 = \sqrt{\frac{4\pi ne^2}{m}}, \tag{123.9}$$

называемой *плазменной частотой.*

§ 124. Затухающие колебания

Учтем теперь тормозящие силы. Полагая в уравнении (122.10) $X = 0$, получим уравнение свободных колебаний

$$\ddot{q} + 2\gamma\dot{q} + \omega_0^2 q = 0. \tag{124.1}$$

Для его решения введем новую переменную ξ, полагая

$$q = \xi e^{-\gamma t}. \tag{124.2}$$

Тогда

$$\ddot{\xi} + (\omega_0^2 - \gamma^2)\xi = 0. \tag{124.3}$$

Формально это уравнение совпадает с дифференциальным уравнением свободных незатухающих колебаний (123.1). Однако коэффициент

$\omega_0^2 - \gamma^2$ может принимать как положительные, так и отрицательные значения. Надо различать три случая.

Случай 1. $\omega_0^2 - \gamma^2 > 0$. Введем обозначение

$$\omega^2 = \omega_0^2 - \gamma^2. \tag{124.4}$$

Тогда

$$\ddot{\xi} + \omega^2 \xi = 0.$$

Отсюда следует, что величина ξ должна совершать незатухающие гармонические колебания с круговой частотой ω:

$$\xi = a\cos(\omega t + \delta).$$

Следовательно,

$$q = ae^{-\gamma t}\cos(\omega t + \delta). \tag{124.5}$$

Кривая $q = q(t)$, представляемая этой формулой (рис. 276), *не периодична*. Однако величина q периодически проходит через нуль и бесконечное число раз достигает максимума и минимума. В этом смысле процессы, описываемые формулой (124.5), являются колебательными. Они называются *затухающими колебаниями*. Промежуток времени между двумя последовательными прохождениями величины q через нуль равен π/ω. Удвоенное его значение

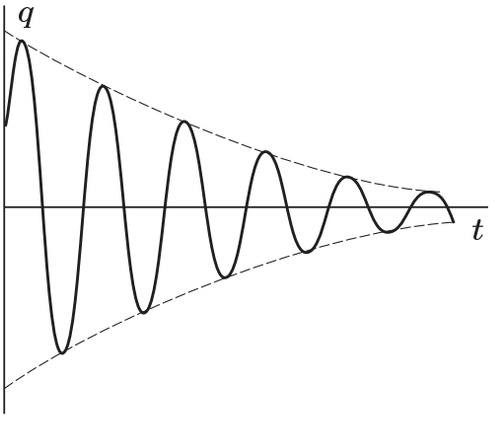


Рис. 276

$$T = \frac{2\pi}{\omega} = \frac{2\pi}{\sqrt{\omega_0^2 - \gamma^2}} = \frac{T_0}{\sqrt{1 - (\gamma/\omega_0)^2}} \tag{124.6}$$

называется *периодом колебаний*, хотя слово «период» здесь не совсем уместно, так как процесс не периодический. Из формулы (124.6) видно, что $T > T_0$, т. е. тормозящие силы понижают частоту колебаний и удлиняют их период. Приравнивая нулю производную $\dot{q}$, легко убедиться, что период T есть также время между двумя последовательными прохождениями величины q через максимум или минимум. Множитель

$$A = ae^{-\gamma t}, \tag{124.7}$$

стоящий перед периодической функцией $\cos(\omega t + \delta)$ в формуле (124.5), называется *амплитудой затухающих колебаний*. Она экспоненциально убывает во времени. Время

$$\tau = 1/\gamma, \tag{124.8}$$

по истечении которого амплитуда A убывает в e раз, называется *временем затухания*. Число полных колебаний, совершаемое за время τ, равно

$$N = \frac{\tau}{T} = \frac{1}{\gamma T}. \tag{124.9}$$

Отношение амплитуд в моменты последовательных прохождений колеблющейся величины через максимумы или минимумы равно $A_1/A_2 = e^{\gamma T}$. Логарифм этого отношения

$$d = \ln \frac{A_1}{A_2} = \gamma T \tag{124.10}$$

называется *логарифмическим декрементом колебания.* Он связан с числом колебаний N соотношением

$$N = 1/d. \tag{124.11}$$

Величина

$$Q = \pi N = \pi/d \tag{124.12}$$

называется *добротностью* колебательного контура. Физический смысл этой величины будет установлен ниже (см. § 127).

Постоянные интегрирования a и δ определяются начальными условиями. Допустим, например, что в начальный момент $q = 0$, $\dot{q} = \dot{q}_0$, где $\dot{q}_0$ — известная постоянная. Полагая в выражении (124.5) $t = 0$, получим $\cos\delta = 0$, и, следовательно, $\delta = \pm\pi/2 + 2n\pi$, $q = \pm ae^{-\gamma t}\sin\omega t$. Без потери общности двойной знак можно опустить и написать

$$q = ae^{-\gamma t}\sin\omega t,$$

так как постоянную интегрирования можно обозначить и через $+a$, и через $-a$. Отсюда

$$\dot{q} = (\omega\cos\omega t - \gamma\sin\omega t)ae^{-\gamma t}.$$

Подставляя сюда $t = 0$, получим $\dot{q}_0 = \omega a$,

$$q = \frac{\dot{q}_0}{\omega}\, e^{-\gamma t}\sin\omega t. \tag{124.13}$$

Случай 2. $\omega_0^2 - \gamma^2 = 0$. Это предельный случай предыдущего, когда период T обращается в бесконечность. Уравнение (124.3) переходит в $\ddot{\xi} = 0$, и, следовательно,

$$q = (a + bt)e^{-\gamma t}. \tag{124.14}$$

В зависимости от значений постоянных интегрирования a и b величина q будет или не будет проходить через максимум (один раз). На рис. 277 изображены два случая: 1) $a \neq 0$, $b = 0$ и 2) $a = 0$, $b \neq 0$. При любых a и b величина q асимптотически приближается к нулю, когда $t \to \infty$. Процесс не будет колебательным. Он называется *апериодическим.*

Если в начальный момент $q_0 = 0$, то $a = 0$,

$$q = bte^{-\gamma t}, \quad \dot{q} = (1 - \gamma t)be^{-\gamma t}.$$

Полагая $t = 0$, находим $b = \dot{q}_0$. Следовательно,

$$q = \dot{q}_0 te^{-\gamma t}. \tag{124.15}$$

Случай 3. $\omega_0^2 - \gamma^2 < 0$. Общее решение уравнения (124.3) будет

$$\xi = C_1 e^{-\sqrt{\gamma^2 - \omega_0^2}\,t} + C_2 e^{\sqrt{\gamma^2 - \omega_0^2}\,t},$$

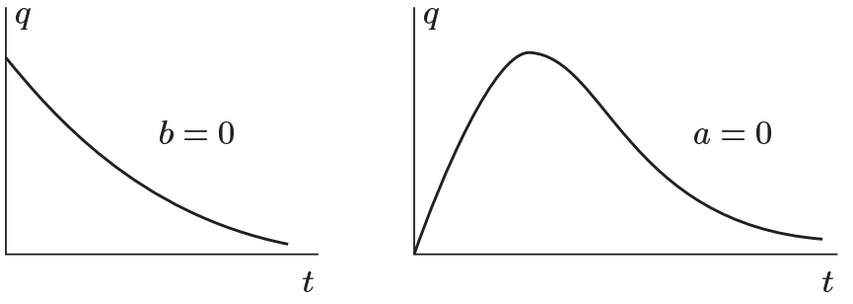


Рис. 277

а уравнения (124.1) —

$$q = C_1 e^{-\alpha_1 t} + C_2 e^{-\alpha_2 t}. \tag{124.16}$$

Через α_1 и α_2 обозначены положительные постоянные:

$$\alpha_1 = \gamma + \sqrt{\gamma^2 - \omega_0^2}, \quad \alpha_2 = \gamma - \sqrt{\gamma^2 - \omega_0^2}. \tag{124.17}$$

Если в начальный момент $q_0 = 0$, $\dot{q} = \dot{q}_0$, то

$$q = \frac{\dot{q}_0}{\alpha_1 - \alpha_2}\,(e^{-\alpha_2 t} - e^{-\alpha_1 t}) = \frac{\dot{q}_0}{2\sqrt{\gamma^2 - \omega_0^2}}\,(e^{-\alpha_2 t} - e^{-\alpha_1 t}). \tag{124.18}$$

§ 125. Баллистический гальванометр

Баллистический гальванометр предназначен для измерения количества электричества, проходящего через цепь при кратковременных импульсах тока. Подвижной частью прибора является прямоугольная рамка с обмоткой из тонкой изолированной проволоки, подвешенная на нити между полюсами магнита, где она может совершать крутильные колебания вокруг вертикальной оси. Возвращающий момент, стремящийся вернуть рамку в положение равновесия, создается закрученной нитью и пропорционален угловому отклонению φ. Момент инерции рамки Θ выбирают большим, чтобы период собственных колебаний ее был не меньше 10 с. При колебаниях в обмотке рамки возбуждается индукционный ток, тормозящий движение последней. Тормозящий момент должен быть пропорционален угловой скорости $\dot{\varphi}$. Для достижения этого полюсным наконечникам придают такую форму, чтобы они оканчивались цилиндрическими поверхностями, а между самими полюсами помещают цилиндр из мягкого железа (рис. 278, вид сверху, рамка не показана). При таком устройстве магнитные силовые линии в зазоре между полюсами магнита и цилиндром, где совершаются крутильные колебания рамки, практически радиальны. Вертикальные стороны рамки пересекают магнитные силовые линии, оставаясь все время перпендикулярными к ним. Вследствие этого возникающая электродвижущая сила и обусловленный ею тормозящий

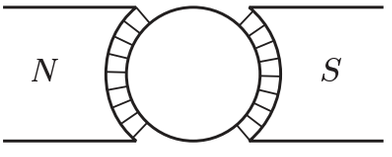


Рис. 278

момент оказываются пропорциональными угловой скорости $\dot{\varphi}$. Тормозящий момент, обусловленный сопротивлением воздуха, также пропорционален $\dot{\varphi}$. Поэтому крутильные колебания рамки гальванометра будут описываться дифференциальным уравнением того же типа, что и затухающие колебания, рассмотренные в предыдущем параграфе.

Допустим, что через гальванометр проходит кратковременный импульс тока I, длительность которого τ много меньше периода собственных колебаний рамки T. Во время прохождения такого импульса можно пренебречь всеми силами, за исключением амперовых сил, действующих на рамку со стороны внешнего магнитного поля. Эти силы создают вращающий момент, пропорциональный току I, так что уравнение движения рамки может быть записано в виде $\Theta\ddot{\varphi} = kI$, где k — постоянная. Интегрируя по времени прохождения импульса, найдем угловую скорость рамки, которую она приобретет за это время:

$$\dot{\varphi}_0 = \frac{k}{\Theta}\int I\,dt = \frac{k}{\Theta}\,q,$$

где q — количество прошедшего электричества. За то же время массивная рамка не успеет заметно сместиться. Импульс тока действует на рамку аналогично удару молотка по массивному шару. Пренебрегая временем τ, движение рамки гальванометра можно рассматривать как свободное затухающее колебание при начальных условиях $\varphi_0 = 0$, $\dot{\varphi}_0 = kq/\Theta$. Такая задача была решена в предыдущем параграфе. Было показано, что амплитуда, а с ней и максимальное отклонение $\varphi_{\text{макс}}$ (отброс гальванометра) пропорциональны $\dot{\varphi}_0$ независимо от того, будет ли режим колебательным или апериодическим. Таким образом, отброс гальванометра связан с количеством прошедшего электричества соотношением

$$q = B\varphi_{\text{макс}}, \tag{125.1}$$

где B — постоянная, называемая *баллистической постоянной гальванометра.* Она измеряется экспериментально. Для этого достаточно разрядить через гальванометр конденсатор известной емкости, заряженный до определенного потенциала, и измерить отброс гальванометра. После такой градуировки формулой (125.1) можно пользоваться для количественных измерений.

§ 126. Векторная диаграмма и комплексные обозначения

1. Существует наглядный геометрический способ представления гармонических колебаний, о котором уже говорилось в первом томе. Допустим, что геометрическая точка M равномерно вращается по окружности радиуса r (рис. 279) с угловой скоростью ω_0. Положение точки на окружности можно задать центральным углом φ между радиусом OM и положительным направлением оси X. Он равен $\varphi = \omega_0 t + \delta$, где δ — значение угла φ в начальный момент $t = 0$. При вращении точки M ее проекция N на ось X движется по диаметру AB

туда и обратно, совершая колебания между точками A и B с периодом $T_0 = 2\pi/\omega_0$. Абсцисса точки N

$$x = a\cos\varphi = a\cos(\omega_0 t + \delta), \tag{126.1}$$

как и сама точка N, совершает незатухающее гармоническое колебание. Этим способом можно представлять гармонические колебания любых

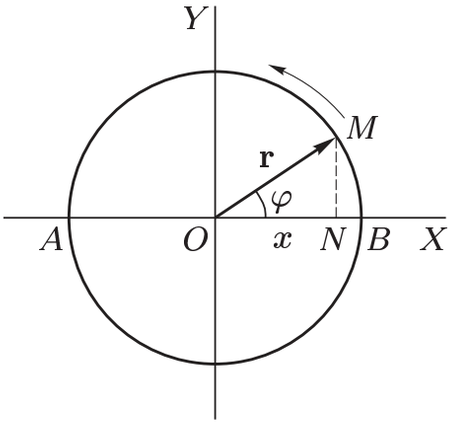


Рис. 279

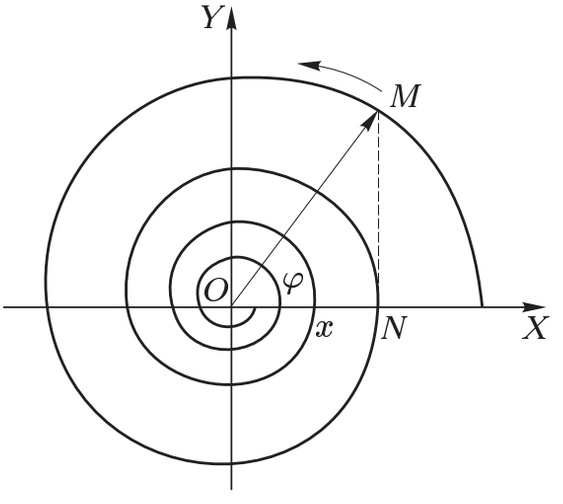


Рис. 280

величин. Надо только условиться изображать колеблющуюся величину абсциссой точки M, равномерно вращающейся по окружности. Вместо абсциссы можно, конечно, брать ординату $y = a\sin(\omega_0 t + \delta) = a\cos[\omega_0 t + (\delta - \pi/2)]$, но во избежание недоразумений условимся всюду пользоваться абсциссой.

Для представления затухающих колебаний вместо окружности надо взять логарифмическую спираль, асимптотически приближающуюся к фокусу O (рис. 280). Если точка M движется по спирали с постоянной угловой скоростью ω_0, приближаясь к фокусу, то ее проекция N на ось X будет совершать затухающее гармоническое колебание.

2. Вместо точки M можно взять радиус-вектор $\mathbf{r} = \overrightarrow{OM}$, равномерно вращающийся вокруг начала координат O. Гармонически колеблющаяся величина изобразится проекцией x этого радиуса-вектора на ось X. При этом во многих задачах оказывается удобным математические операции над величиной x заменить соответствующими операциями над самим радиусом-вектором $\mathbf{r}$. Например, если нужно вычислить сумму слагаемых

$$x_1 = a_1 e^{-\gamma_1 t}\cos(\omega_1 t + \delta_1), \quad x_2 = a_2 e^{-\gamma_2 t}\cos(\omega_2 t + \delta_2),$$

то можно сначала сложить по правилу параллелограмма векторы $\mathbf{r}_1$ и $\mathbf{r}_2$, проекциями которых являются эти слагаемые, а затем спроецировать полученный вектор $\mathbf{r} = \mathbf{r}_1 + \mathbf{r}_2$ на ось абсцисс. Результат этих операций, очевидно, будет равен $x = x_1 + x_2$. Операция проецирования производится в самом конце вычисления. При известном навыке можно совсем отвлечься от проекций, а изображать колеблющуюся величину

непосредственно самим вектором $\mathbf{r}$, равномерно вращающимся вокруг своего начала. Его проецирование на ось X *подразумевается, но не выполняется фактически.* Такой метод называется *методом векторных диаграмм.* Рисунок 279 можно поэтому назвать *векторной диаграммой незатухающего*, а рис. 280 — *затухающего гармонического колебания.* Метод векторных диаграмм широко применяется в электротехнике при изучении переменных токов.

3. В физике более широкое распространение получил другой метод, отличающийся от метода векторных диаграмм только по форме. В этом методе колеблющаяся величина представляется *комплексным числом.* Положение точки на плоскости (рис. 279) можно однозначно задать комплексным числом $z = x + iy$. Если точка M вращается, то

$$x = a\cos(\omega_0 t + \delta), \quad y = a\sin(\omega_0 t + \delta).$$

Поэтому, используя известную формулу Эйлера

$$e^{i\varphi} = \cos\varphi + i\sin\varphi,$$

величину z можно представить в виде

$$z = ae^{i(\omega_0 t+\delta)}.$$

Вещественная часть этого выражения

$$\mathrm{Re}\,(z) = a\cos(\omega_0 t + \delta) = x$$

представляет гармонические колебания величины x. Условимся опускать знак взятия вещественной части Re и писать просто

$$x = ae^{i(\omega_0 t+\delta)}. \tag{126.2}$$

Это символическое равенство не следует понимать буквально. Его надо понимать в том смысле, что физическая величина x равна *вещественной части* комплексного выражения, стоящего в этом равенстве справа. Модуль этого комплексного выражения a равен амплитуде колебания, а его аргумент $\omega_0 t + \delta$ — фазе. Можно формально упростить запись (126.2). Введем комплексную величину $A = ae^{i\delta}$, называемую *комплексной амплитудой колебания.* Тогда

$$x = Ae^{i\omega_0 t}. \tag{126.3}$$

Комплексность амплитуды A означает, следовательно, что колебание происходит с *начальной фазой*, отличной от нуля.

Наконец, можно формально рассматривать выражения типа (126.3) при комплексных значениях величины ω_0. Для раскрытия физического смысла таких выражений полагаем $\omega_0 = \omega_1 + i\omega_2$. Тогда

$$x = Ae^{i(\omega_1+i\omega_2)t} = ae^{-\omega_2 t}e^{i(\omega_1 t+\delta)} = ae^{-\omega_2 t}\cos(\omega_1 t + \delta).$$

Если $\omega_2 > 0$, то это выражение представляет *затухающее гармоническое колебание* с круговой частотой ω_1 и показателем затухания ω_2. Если $\omega_2 < 0$, то получится *колебание с неограниченно нарастающей амплитудой.* Таким образом, если частота комплексна, то это означает,

что амплитуда колебания *экспоненциально затухает или нарастает во времени.*

Очень важно научиться понимать физический смысл уравнений, записанных в комплексной форме, не переходя к вещественной форме записи. Комплексная форма позволяет часто избежать громоздкости формул и делает сами формулы более общими и легче обозримыми. Особенно широко комплексная форма применяется при изучении распространения волн.

4. Над комплексными величинами можно производить многие математические операции, как если бы эти величины были вещественными. Так можно поступать не всегда. Это можно делать только тогда, когда операции *вещественны и линейны.* К ним относятся, например, сложение, вычитание, умножение и деление на вещественное число, дифференцирование и интегрирование по вещественной переменной и пр. Вообще, операция L называется *линейной*, если результат действия ее на величину $\alpha_1 z_1 + \alpha_2 z_2$ представляется в виде

$$L(\alpha_1 z_1 + \alpha_2 z_2) = \alpha_1 L(z_1) + \alpha_2 L(z_2), \tag{126.4}$$

где z_1 и z_2 — какие угодно (вообще говоря, комплексные) величины, а α_1 и α_2 — любые постоянные. При этом в физике нас, в конце концов, интересуют лишь *вещественные операции*, т. е. такие, результаты действия которых на *вещественные величины* сами *вещественны.*

Комплексные выражения сами по себе не соответствуют никаким физическим величинам. Последние всегда вещественны и только из соображений удобства иногда представляются вещественными частями комплексных выражений. И математические операции в физике должны, в конце концов, определяться как операции над вещественными физическими величинами. Допустим, однако, что над вещественной величиной x надо выполнить линейную вещественную операцию L. Результат будет $L(x)$. Но тот же результат можно получить иначе. Возьмем комплексную величину $x + iy$ и применим к ней операцию L. Получим

$$L(x + iy) = L(x) + iL(y).$$

Отбросив мнимую часть, снова найдем $L(x)$. Если вычисление по второму методу окажется проще, то его применение оправдано.

Например, вместо того, чтобы дифференцировать по t функцию $x = a\cos(\omega_0 t + \delta)$, можно продифференцировать по тому же аргументу комплексное выражение (126.2) или (126.3), а затем взять вещественную часть результата. В обоих случаях получится одно и то же.

Но было бы грубой ошибкой переносить этот прием на *нелинейные операции.* Пусть, например, требуется возвести в квадрат вещественную величину x. Правильный результат будет x^2. Попробуем, однако, формально применить комплексный метод. Заменим x на $x + iy$ и возведем в квадрат, получим $(x^2 - y^2) + i2xy$. Вещественная часть этого выражения равна $x^2 - y^2$. Она зависит не только от вещественной части выражения $x + iy$, но и от мнимой. Ошибка получилась потому, что возведение в квадрат — нелинейная операция.

§ 127. Вынужденные колебания затухающего осциллятора под действием синусоидальной силы

1. Вынужденные колебания затухающего осциллятора описываются уравнением

$$\ddot{q} + 2\gamma\dot{q} + \omega_0^2 q = X(t), \tag{127.1}$$

где $X(t)$ — внешняя действующая сила, точнее, электродвижущая сила, деленная на индуктивность катушки самоиндукции, или (в случае механических колебаний) сила, деленная на массу колеблющегося тела. Уравнение (127.1) *линейно*, т. е. первой степени относительно неизвестного q и его производных по времени. Оно *неоднородно*, т. е. содержит правую часть $X(t)$. Для линейных однородных уравнений (т. е. уравнений без правой части) справедлив *принцип суперпозиции*, согласно которому сумма любых двух решений уравнения есть также решение того же уравнения. Для линейных неоднородных уравнений это несправедливо. Однако здесь имеет место суперпозиция решений *в другом смысле*. Пусть правая часть в уравнении (127.1) представляется в виде суммы $X = \sum X_i(t)$, а $q_i(t)$ — решение уравнения (127.1), в котором правая часть заменена на $X_i(t)$. Тогда сумма $q = \sum q_i(t)$ будет решением уравнения (127.1) с правой частью $X = \sum X_i(t)$. Для теории колебаний отмеченное свойство линейных уравнений имеет большое значение. Оно позволяет общую задачу о вынужденных колебаниях под действием *произвольно меняющейся силы* свести к частной задаче о вынужденных колебаниях под действием *синусоидальной силы*. Дело в том, что согласно известной математической *теореме Фурье* всякая функция $X(t)$ довольно общего вида может быть представлена в виде суммы синусоидальных функций (см. § 128).

Покажем теперь, как связаны между собой решения неоднородного и соответствующего ему однородного уравнений. Пусть $\overline{q}(t)$ — любое частное решение неоднородного уравнения (127.1). Тогда $\ddot{\overline{q}} + 2\gamma\dot{\overline{q}} + \omega_0^2\overline{q} = X(t)$. Вычитая это соотношение из (127.1) и вводя обозначение $Q = q - \overline{q}$, получим $\ddot{Q} + 2\gamma\dot{Q} + \omega_0^2 Q = 0$. Отсюда видно, что Q есть решение однородного уравнения. Таким образом, $q = \overline{q} + Q$, т. е. *общее решение неоднородного уравнения* (127.1) *может быть представлено в виде суммы частного решения того же уравнения и общего решения соответствующего однородного уравнения.*

2. После этих предварительных замечаний обратимся к задаче о колебаниях затухающего осциллятора при наличии внешней силы. Сначала исследуем частный случай, когда сила X меняется синусоидально, т. е. представляется выражением

$$X = X_0 \cos\omega t, \tag{127.2}$$

где X_0 и ω — постоянные. Задача сводится к решению уравнения

$$\ddot{q} + 2\gamma\dot{q} + \omega_0^2 q = X_0 \cos\omega t. \tag{127.3}$$

Среди частных решений этого уравнения есть такое, которое меняется во времени *синусоидально с частотой внешней возбуждающей силы* ω. Будем искать его в комплексной форме, что можно делать, так как все математические операции, с которыми придется иметь дело, линейны и вещественны. Заменим правую часть уравнения (127.3) на комплексную величину $X_0 e^{i\omega t}$, т. е. напишем

$$\ddot{q} + 2\gamma\dot{q} + \omega_0^2 q = X \equiv X_0 e^{i\omega t}. \tag{127.4}$$

Частное решение последнего уравнения ищем в виде $q = q_0 e^{i\omega t}$, откуда $\dot{q} = i\omega q$, $\ddot{q} = -\omega^2 q$. Подстановкой в уравнение (127.4) получаем

$$q = \frac{X}{\omega_0^2 - \omega^2 + 2i\omega\gamma}\, e^{i\omega t}. \tag{127.5}$$

Это частное решение описывает так называемые *вынужденные колебания осциллятора.* Они происходят с частотой внешней возбуждающей силы ω. Добавив к частному решению (127.5) общее решение соответствующего однородного уравнения, получим

$$q = \frac{X_0}{\omega_0^2 - \omega^2 + 2i\omega\gamma}\, e^{i\omega t} + e^{-\gamma t}(C_1 \cos\omega_0 t + C_2 \sin\omega_0 t). \tag{127.6}$$

Добавленное слагаемое описывает *свободные колебания* осциллятора. Выбором произвольных постоянных C_1 и C_2 можно удовлетворить любым начальным условиям. Однако, каковы бы ни были эти условия, свободные колебания всегда *экспоненциально затухают*, причем за время $\tau = 1/\gamma$ амплитуда свободных колебаний убывает в e раз. Процесс затухания свободных колебаний называется *установлением колебаний*, а время τ — *временем затухания* или *временем установления.* Если $t \gg \tau$, то свободные колебания практически совсем затухнут. Останутся одни только вынужденные колебания, *совершенно не зависящие от начальных условий.* Исследованием таких колебаний мы прежде всего и займемся.

3. Конечно, решение (127.5) лишь символически представляет вынужденное колебание. В нем должна быть оставлена *только вещественная часть.* Для нахождения последней введем обозначение

$$\omega_0^2 - \omega^2 + 2i\omega\gamma = \rho e^{i\delta}, \tag{127.7}$$

где ρ и δ — величины вещественные. Тогда $q = (X_0/\rho)e^{i(\omega t - \delta)}$, или в вещественной форме

$$q = a\cos(\omega t - \delta). \tag{127.8}$$

Величины ρ и δ найдем, приравнивая вещественные и мнимые части в соотношении (127.7). Таким путем получаем

$$\omega_0^2 - \omega^2 = \rho\cos\delta, \quad 2\omega\gamma = \rho\sin\delta,$$

откуда

$$a = \frac{X}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4\omega^2\gamma^2}}, \tag{127.9}$$

$$\operatorname{tg}\delta = \frac{2\omega\gamma}{\omega_0^2 - \omega^2}. \tag{127.10}$$

Таким образом, *вынужденное колебание будет гармоническим, амплитуда и фаза которого определяются формулами* (127.9) и (127.10).

Формулы (127.9) и (127.10) можно получить также с помощью векторной диаграммы. Это будет сделано в § 129 на примере переменных токов, где соотношения совершенно аналогичны.

4. Исследуем сначала поведение *амплитуды* вынужденных колебаний в зависимости от частоты ω. Электрические и механические колебания будем рассматривать совместно, называя величину q либо *зарядом конденсатора*, либо *смещением колеблющегося тела из положения равновесия.* Оставляя амплитуду силы X_0 неизменной, будем менять ее частоту ω. При $\omega = 0$ получаем *статическое отклонение* под действием постоянной силы X_0: $a_0 = X_0/\omega_0$. При возрастании частоты ω амплитуда смещения a сначала также возрастает, затем проходит через максимум и, наконец, асимптотически стремится к нулю (рис. 281). Приравнивая нулю производную $da/d\omega$, убеждаемся, что амплитуда смещения (заряда) a достигает максимума при $\omega = \sqrt{\omega_0^2 - 2\gamma^2}$. Максимумы амплитуды скорости (тока) ωa и амплитуды ускорения (напряжения $L\ddot{q}$) $\omega^2 a$ достигаются соответственно при $\omega = \omega_0$ и $\omega = \omega_0^2/\sqrt{\omega_0^2 - \gamma^2}$.

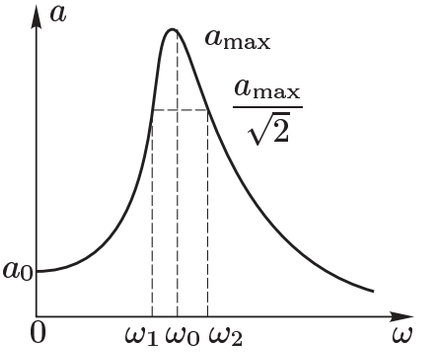


Рис. 281

Найдем еще частоту, при которой максимальна средняя мощность, развиваемая силой X. Мощность выражается произведением силы X на скорость $\dot{q}$, т. е. равна $P = X\dot{q}$. Так как произведение гармонически колеблющихся величин X и $\dot{q}$ — нелинейная операция, то надо пользоваться вещественной формой представления колебаний, т. е. формулами (127.2) и (127.8). Из них получаем

$$P = -\omega a X_0 \cos\omega t \sin(\omega t - \delta) = \frac{\omega a X_0}{2}[\sin\delta - \sin(2\omega t - \delta)].$$

Второе слагаемое в квадратных скобках синусоидально меняется во времени с частотой 2ω. Его среднее значение по времени равно нулю. На величину средней мощности оно не влияет. Последняя определяется только первым слагаемым и равна $^1/_2 X_0 \omega a \sin\delta$. Она достигает максимума при той же частоте, что и амплитуда скорости ωa, т. е. при частоте $\omega = \omega_0$.

В наиболее важном случае, когда *затухание невелико*, положения всех максимумов *почти не отличаются* друг от друга. Поэтому за максимум амплитуды смещения можно принять ее значение при $\omega = \omega_0$, т. е.

$$a_{\text{макс}} = \frac{X_0}{2\omega_0\gamma} = \frac{\omega_0}{2\gamma}\, a_0. \tag{127.11}$$

Отношение максимального значения амплитуды $a_{\text{макс}}$ к статическому отклонению a_0 называется *добротностью осциллятора* или *колебательного контура*. Обозначая добротность через Q, имеем

$$\frac{a_{\text{макс}}}{a_0} \equiv Q = \frac{\omega_0}{2\gamma} = \frac{\pi}{d}, \tag{127.12}$$

где d — логарифмический декремент.

Кривая, изображающая зависимость амплитуды колебаний a от частоты внешней возбуждающей силы ω, называется *резонансной кривой* (см. рис. 281). Одной из характеристик резонансной кривой может служить значение амплитуды в максимуме $a_{\text{макс}}$. Другой важной характеристикой является *ширина резонансной кривой*. Пусть ω_1 и ω_2 — значения частоты ω, при которых энергия колебаний вдвое меньше энергии в максимуме. Тогда

$$(\omega_1^2 - \omega_0^2)^2 = 4\omega_1^2\gamma^2, \quad (\omega_2^2 - \omega_0^2)^2 = 4\omega_2^2\gamma^2.$$

Если $|\omega_1 - \omega_0| \ll \omega_0$, $|\omega_2 - \omega_0| \ll \omega_0$, то отсюда получаем приближенно

$$\Delta\omega \equiv \omega_2 - \omega_1 = 2\gamma = \omega_0/Q. \tag{127.13}$$

Величина $\Delta\omega$ и называется *шириной* (или *полушириной*) резонансной кривой. Мы видим, что чем больше добротность осциллятора, тем у́же резонансная кривая. Далее, из формул (127.11) и (127.13) получаем

$$\Delta\omega \cdot a_{\text{макс}} = \omega_0 a_0. \tag{127.14}$$

Чем больше максимум резонансной кривой, тем он острее, т. е. тем у́же резонансная кривая.

5. Итак, наиболее интенсивные колебания будут наблюдаться при частоте $\omega = \omega_0$. Явление возбуждения сильных колебаний при частоте внешней возбуждающей силы, равной или близкой собственной частоте колебательной системы, называется *резонансом*. Физическая причина этого явления предельно проста. Ее лучше всего пояснить не на примере синусоидально меняющейся силы, а на примере силы, состоящей из одинаковых кратковременных *толчков*, следующих друг за другом через одинаковые промежутки времени. Между двумя последовательными толчками система движется *свободно*. При каждом толчке скорость $\dot{q}$ скачкообразно получает приращение $\Delta\dot{q}$. Если толчки следуют друг за другом через промежутки времени, равные периоду собственных колебаний системы, то каждый новый толчок вызывает приращение скорости *того же знака и величины*, что и предыдущий толчок: *толчки усиливают друг друга*. В этом случае энергия, вкладываемая в систему, *максимальна*. Амплитуда колебаний будет нарастать

до тех пор, пока возросшие силы трения в среднем за период колебаний не компенсируют действие каждого нового толчка. Тогда установятся наиболее интенсивные колебания, характеризующиеся *максимумом амплитуды скорости* и *максимальной энергией.* Это и есть *резонанс.* Синусоидально меняющаяся сила также может рассматриваться как последовательность одинаковых толчков. Только теперь толчки *не мгновенные, а длительные* и непрерывно примыкают друг к другу. Но это обстоятельство в приведенном объяснении не играет принципиальной роли. Естественно, что в случае толчков конечной длительности появляется зависимость формы резонансной кривой от *формы толчка.* Толчки синусоидальной формы характеризуются тем, что для них на резонансной кривой имеется *только один максимум* при $\omega = \omega_0$. Для толчков другой формы может быть и не так. Например, в разобранном выше случае последовательности мгновенных толчков максимум амплитуды появляется не только при основной частоте $\omega = \omega_0$, но и при меньших частотах $\omega_0/2$, $\omega_0/3$, ... При таких частотах толчки следуют один за другим реже. Однако они по-прежнему усиливают друг друга, а это есть основное условие резонанса. На резонансной кривой появляются новые максимумы, хотя и менее интенсивные, чем максимум при основной частоте.

6. Явление резонанса удобно демонстрировать на камертонах, так как они отличаются высокими добротностями. Возьмем два камертона с одинаковыми собственными частотами и установим их на резонансных ящиках (рис. 282). Заставим звучать один камертон. Излучаемые им звуковые волны будут периодически воздействовать на второй камертон. Так как частоты совпадают, то наступает резонанс — второй камертон начинает колебаться и звучать. В этом можно убедиться, заглушив рукой колебания первого камертона. Тогда отчетливо слышен звук второго камертона. Для демонстрации в большой аудитории рядом с ножкой второго камертона можно подвесить на нити маленький шарик. Он отскакивает в сторону под действием ударов ножки камертона. Наденем теперь на ножку одного из камертонов небольшой грузик. Это поведет к «расстройке» камертонов — их собственные частоты перестанут совпадать. Если повторить опыт, то второй камертон уже не будет отзываться на звучание первого камертона. Этот опыт показывает, что уже ничтожная расстройка нарушает условие резонанса. Вместе с тем он позволяет судить, насколько велика добротность камертона как колебательной системы.

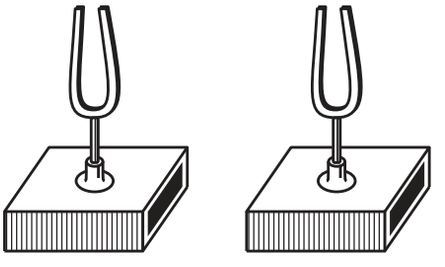


Рис. 282

7. Электрическим аналогом описанного опыта может служить опыт с высокочастотными колебательными контурами. Берется генератор высокочастотных колебаний с частотой в несколько мегагерц. Емкостью колебательного контура генератора служит конденсатор переменной емкости. Последнюю можно менять поворотом специальной ручки.

Возле генератора помещается второй такой же колебательный контур, в который включена электрическая лампочка от карманного фонарика. Роль резонансных ящиков выполняют металлические стержни, привинченные к пластинам конденсаторов и установленные параллельно друг другу. Этим между колебательными контурами устанавливается емкостная связь: электрические заряды одного стержня индуцируют заряды на другом стержне. Когда работает генератор, в его колебательном контуре текут токи высокой частоты, возбуждающие токи во втором колебательном контуре, содержащем лампочку накаливания. Если собственные частоты колебательных контуров различаются сильно, то лампочка не горит. Можно увеличить накал лампы генератора и тем самым увеличить амплитуду генерируемых им колебаний. Однако лампочка все равно не загорится. Но если плавно менять емкость конденсатора в одном из колебательных контуров, то в определенный момент лампочка накаливания ярко вспыхивает. Это происходит при резонансе, когда собственные частоты обоих контуров совпадают или почти совпадают между собой.

На этом примере видна польза явления резонанса. Вся техника радиоприема основана на резонансе. Для того чтобы радиоприемник «принимал радиостанцию», необходимо его «настроить», т. е. добиться совпадения собственной частоты колебательного контура радиоприемника с частотой электромагнитных волн, излучаемых работающей радиостанцией. Формула (127.12) показывает, что *чувствительность радиоприемника тем выше, чем выше добротность его колебательного контура. Чувствительность радиоприемника и полоса частот, которые он способен принимать, обратно пропорциональны друг другу и связаны между собой соотношением* (127.14). *При большой чувствительности радиоприемник способен принимать только очень узкую область частот. Чем выше чувствительность приемника, тем меньше мешают приему другие радиостанции, работающие на близких частотах.* Приемник бесконечно высокой чувствительности принимал бы только волны вполне определенной частоты, на которую он настроен. Все эти закономерности относятся не только к радиоприемнику. Они носят *общий характер* и справедливы для любых осцилляторов независимо от физической природы совершающихся колебаний.

8. Но явление резонанса может быть и вредным, и тогда с ним приходится бороться. Пятикратный и даже десятикратный запас прочности, предусматриваемый при статических расчетах механических систем, может оказаться недостаточным, если системы подвергаются действию периодических, хотя и относительно слабых сил. Так, во всех армиях мира принято правило подавать команду «вольно», когда отряд пехоты или кавалерии проходит по мосту. При прохождении «в ногу» с отбиванием шага периодические толчки, которым подвергается мост, могут попасть в такт с собственными колебаниями моста. Тогда наступает резонанс. Амплитуда колебаний моста может увеличиться настолько, что он рухнет. Такие случаи наблюдались неоднократно

в прошлом столетии, что и послужило поводом для введения в войсковые уставы упомянутого выше правила. Конечно, одного совпадения частот *недостаточно* для разрушения моста. Мальчик, стреляющий по мосту из рогатки, не может рассчитывать на его разрушение, как бы идеально он ни вошел в ритм стрельбы и как бы долго ни продолжалась последняя. Существуют силы трения и другие тормозящие силы, *ограничивающие возрастание амплитуды колебаний.* Помимо совпадения частот для разрушения необходимо еще, чтобы каждый толчок был *достаточно сильным.* Точно так же в случае электрических систем при расчетах предельно допустимых напряжений, подводимых к системе, *нельзя ограничиваться статическими полями.* Если система содержит емкости и индуктивности, то во избежание электрического пробоя при расчетах необходимо считаться с возможностью резонанса.

9. Резонансная кривая, приведенная на рис. 281 и представляемая формулой (127.9), называется *амплитудной резонансной кривой.* Обратимся теперь к рассмотрению *фазовой резонансной кривой* (127.10). Она определяет угол δ, на который отстает по фазе смещение (заряд) q относительно внешней силы X. Зависимость δ от ω изображена на рис. 283. Если $\omega \to 0$, то $\delta \to 0$: при низких частотах q и X колеблются

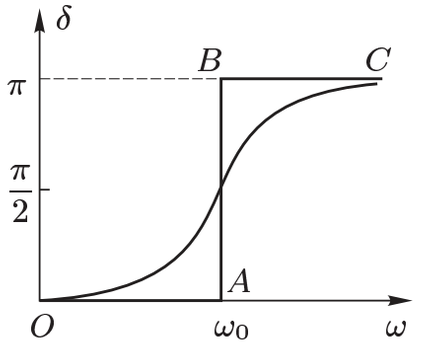


Рис. 283

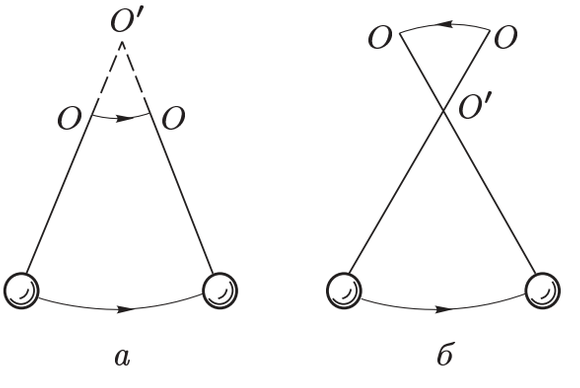


Рис. 284

в *одинаковых фазах.* При $\omega \to \infty$ сдвиг фаз δ стремится к π — колебания совершаются в *противоположных фазах.* При резонансе $\operatorname{tg}\delta = \infty$, $\delta = \pi/2$. Если $\gamma \to 0$, то фазовая резонансная кривая претерпевает разрыв непрерывности при $\omega = \omega_0$ и переходит в ломаную $OABC$: при $\omega < \omega_0$ фазы колебаний q и X одинаковы, а при $\omega > \omega_0$ — противоположны. Ломаная $OABC$ представляет *предельное положение*, к которому стремятся резонансные кривые при $\gamma \to 0$.

Следующая механическая иллюстрация помогает уяснить фазовые соотношения при вынужденных колебаниях. Возьмем рукой маятник за точку O (рис. 284 *а*) и будем медленно двигать эту точку туда и обратно. Возникнут вынужденные колебания, как если бы маятник стал длиннее и был закреплен в неподвижной точке подвеса O'. В этом случае колебания руки (внешней силы) совершаются с частотой, меньшей собственной частоты ω_0 маятника. Направления движения руки и шарика маятника все время совпадают, их колебания происходят

в одинаковых фазах. Не то будет, когда точка O колеблется с частотой, превышающей собственную частоту колебаний маятника. Теперь маятник колеблется так, как если бы его точка подвеса была неподвижна и смещена вниз относительно точки O: точка O и шарик маятника движутся противоположно, т. е. колеблются в противоположных фазах (рис. 284 *б*).

§ 128. Вынужденные колебания под действием несинусоидальной силы. Теорема Фурье

1. Вынужденные колебания под действием несинусоидальной силы можно исследовать с помощью математической *теоремы Фурье* (1768–1830). Согласно этой теореме всякая периодическая функция достаточно общего вида может быть разложена в *ряд Фурье*, т. е. представлена в виде суммы конечного или бесконечного числа *синусоидальных функций.* Мы не будем доказывать эту теорему, но приведем выражение для *коэффициентов* ряда Фурье.

Периодической функцией $f(t)$ называется всякая функция, при любом t удовлетворяющая условию $f(t) = f(t+T)$, где T — положительная постоянная, не равная нулю. Она называется *периодом функции* f. Если T — период, то, очевидно, $2T$, $3T$ и т. д. также будут периодами. Действительно, $f(t+2T) = f(t+T+T) = f(t+T) = f(t)$. Среди всех этих периодов можно указать *наименьший*, который называется *основным периодом*, а частота $\Omega = 2\pi/T$ — *основной частотой.* В дальнейшем под T и Ω мы будем понимать основной период и основную частоту.

Теорема Фурье утверждает, что всякую периодическую функцию с основным периодом T можно представить в виде суперпозиции синусоид с периодами T, $T/2$, $T/3$, ... или с частотами $\omega_k = k\Omega$, где $k = 0, 1, 2, \ldots$ Запишем этот ряд Фурье в комплексной форме:

$$f(t) = \sum_{k=0}^{\infty} c_k E^{i\omega_k t}, \tag{128.1}$$

подразумевая, что функция $f(t)$ равна *вещественной части* ряда, стоящего справа. Для вычисления коэффициентов c_k умножим обе части соотношения (128.1) на $e^{-i\omega_m t}$ и проинтегрируем от 0 до T. Получим

$$\int_0^T f(t) e^{-i\omega_m t}\, dt = \sum_{k=0}^{\infty} c_k \int_0^T e^{i(\omega_k - \omega_m)t}\, dt.$$

В силу периодичности показательной функции

$$e^{i(\omega_k - \omega_m)T} = e^{i(k-m)T\Omega} = e^{2\pi(k-m)i} = 1.$$

Отсюда следует, что интеграл в правой части равен нулю, если $k \neq m$. Если же $k = m$, то этот интеграл равен T. Таким образом,

$$c_m = \frac{1}{T}\int_0^T f(t)e^{-i\omega_m t}\,dt. \tag{128.2}$$

По этой формуле и можно вычислить коэффициенты ряда Фурье c_m. Интегрирование не обязательно проводить в пределах от 0 до T. Можно взять любые пределы, лишь бы длина промежутка интегрирования была равна T.

2. Применение теоремы Фурье к задаче о вынужденных колебаниях производится по следующей схеме. Если внешняя «сила» $X(t)$, действующая на затухающий осциллятор, периодична, то ее следует разложить в ряд Фурье и решить задачу о вынужденных колебаниях для каждого члена этого ряда в отдельности. Тогда сумма таких решений и даст решение задачи о вынужденных колебаниях осциллятора под действием силы X.

Рассмотрим, например, с изложенной точки зрения явление резонанса под действием периодической силы, имеющей характер толчков (рис. 285). Пусть T — основной период, через который толчки следуют друг за другом. В разложении силы $X(t)$ будут содержаться синусоидальные члены (гармоники) с периодами T, $T/2$, $T/3$, ..., T/m, ... Резонанс на m-й гармонике получится при условии $T/m = T_0$, или $T = mT_0$, где T_0 — собственный период колебательного контура. Отсюда видно, что резонанс возникает, когда толчки следуют друг за другом не только через время T_0, но и через время $2T_0$, $3T_0$, ..., т. е. вдвое реже, втрое реже и т. д. Этот результат понятен из простых физических соображений.

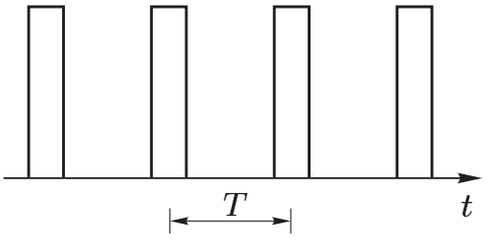


Рис. 285

3. Для всего изложенного не имеет значения, что сила $X(t)$ является *периодической* функцией. Существенно только, что она представляется *суммой синусоидальных членов.* Допустим, например, что

$$X(t) = A(t)\cos\omega t, \tag{128.3}$$

где функция $A(t)$ *медленно меняется во времени* (по сравнению с «быстро меняющейся» величиной $\cos\omega t$). Про колебание (128.3) говорят, что оно *модулировано по амплитуде.* Простейшей является *синусоидальная модуляция*, когда

$$A(t) = A_0(1 + \alpha\cos\Omega t) \tag{128.4}$$

с постоянными A_0, α, Ω. Величину ω называют *несущей частотой*, α — глубиной, а Ω — *частотой модуляции*. В этом случае

$$X = A_0 \cos \omega t + \frac{\alpha A_0}{2} [\cos (\omega + \Omega)t + \cos (\omega - \Omega)t], \qquad (128.5)$$

т. е. *синусоидально модулированное колебание эквивалентно суперпозиции трех синусоидальных колебаний с частотами* ω, $\omega + \Omega$, $\omega - \Omega$. Если эти частоты соизмеримы, то сила X периодична. Если же они несоизмеримы, то она не периодична. Хорошей иллюстрацией может служить следующий акустический опыт. Два одинаковых камертона устанавливаются возле друг друга на резонансных ящиках (рис. 282). Мы видели (см. с. 533), что если возбудить один камертон, то начинает звучать и другой (резонанс). Если же на ножку второго камертона надеть небольшой грузик, то он вызовет «расстройку», и второй камертон перестанет «отзываться» на колебания первого. Введем теперь между камертонами диск с отверстием и приведем его во вращение. Когда вращающееся отверстие проходит мимо отверстий резонансных ящиков, звук, воздействующий на второй камертон, усиливается, звуковая волна становится модулированной по амплитуде. Меняя частоту вращения диска, можно подобрать ее такой, чтобы второй камертон снова начал звучать, когда колеблется первый.

Модулированное колебание (128.3) часто называют *гармоническим* или *синусоидальным колебанием с переменной амплитудой*. С формально логической точки зрения такой термин внутренне противоречив, так как по самому определению амплитуда A и частота синусоиды ω — *величины постоянные*. Однако в некоторых случаях этот термин может быть оправдан. Все зависит от того, *на какую систему* (приемник) воздействует модулированная сила (128.3). Если приемник остронастроенный (т. е. его коэффициент затухания γ мал), то он будет отвечать практически на одну частоту, совпадающую с его собственной частотой ω_0. С помощью такого приемника из суперпозиции (128.5) можно выделить одновременно только колебания с частотой либо ω, либо $\omega + \Omega$, либо $\omega - \Omega$. В этом случае недопустимо функцию (128.3) отождествлять с одной синусоидой, а следует рассматривать ее как суперпозицию (128.5) трех синусоид с различными частотами. Положение меняется, когда приемник не остронастроенный. Такой приемник воспринимает одновременно *все три колебания* с частотами ω, $\omega + \Omega$, $\omega - \Omega$, и притом практически с *одним и тем же усилением*. В результате суперпозиции таких колебаний в приемнике возникает вынужденное колебание $A(t) \cos \omega t$. Явление протекает так, *как если бы на приемник действовала синусоидальная сила, амплитуда которой* $A(t)$ *медленно менялась во времени*. То же самое можно сказать и про частотную модуляцию, когда медленно меняется не амплитуда A, а частота колебания ω. Все это справедливо только тогда, когда время установления колебаний приемника $\tau = 1/\gamma$ мало по сравнению с временем, в течение которого заметно меняются амплитуда и фаза действующей силы. За время τ вынужденные колебания в приемнике успевают установиться,

тогда как A и ω практически остаются постоянными. Поэтому в течение времени установления τ сила $X = A(t)\cos[\omega(t)t]$ и вызывает такой же эффект, как если бы она оставалась чисто синусоидальной с постоянными $A(t)$ и $\omega(t)$.

4. Допустим теперь, что сила $X(t)$ *непериодична*, но воздействует на колебательную систему в течение *конечного времени*, например изображается функцией, показанной на рис. 286 сплошной линией. Этот случай сводится к случаю периодической силы. Для этого возьмем

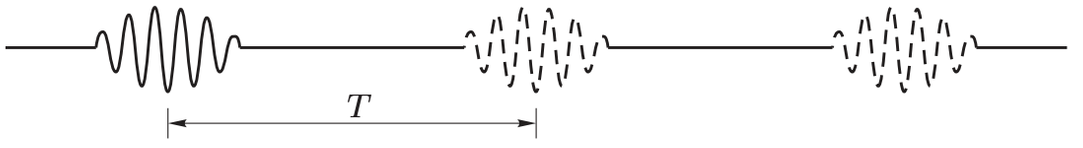


Рис. 286

промежуток времени T, очень большой по сравнению с временем затухания свободных колебаний системы, и «периодически продолжим» $X(t)$, чтобы получилась периодическая функция с периодом T, как это изображено на рис. 286. Поскольку время T ничем не ограничено, его можно выбрать настолько большим, чтобы замена исходной функции ее периодическим продолжением практически не отразилась на поведении колебательной системы в течение интересующего нас времени. А если выполнить предельный переход $T \to \infty$, то вспомогательная величина T вообще выпадет из результата, а самый результат сделается *вполне точным.* Разлагая периодически продолженную функцию в ряд Фурье и полагая $\Omega = 2\pi/T$, пишем

$$X(t) = \sum_{n=0}^{\infty} c_n e^{in\Omega t}, \text{ где } c_n = \frac{1}{T}\int_{-T/2}^{+T/2} X(t)e^{-in\Omega t}\,dt.$$

Для выполнения предельного перехода введем новые обозначения: $\omega = n\Omega$, $\Delta\omega = \Omega$. Тогда

$$X(t) = \sum \frac{c_n}{\Omega}\, e^{i\omega t}\,\Delta\omega.$$

Поскольку при $T \to \infty$ величины $\Delta\omega$ стремятся к нулю, заменим сумму интегралом:

$$X(t) = \int_0^{\infty} a(\omega)e^{i\omega t}\,d\omega, \tag{128.6}$$

где

$$a(\omega) = \lim_{n\to\infty}\frac{c_n}{\Omega} = \frac{1}{T\Omega}\int_{-\infty}^{+\infty} X(t)e^{-i\omega t}\,dt,$$

или

$$a(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} X(t) e^{-i\omega t}\, dt. \tag{128.7}$$

Формула (128.6) называется *интегралом Фурье.* Она представляет функцию $X(t)$ в виде *суперпозиции непрерывного множества синусоидальных колебаний,* частоты которых непрерывно заполняют определенный интервал. При этом подразумевается, что от правой части формулы (128.6) следует брать лишь *вещественную часть.* Разумеется, приведенные рассуждения не могут служить строгим доказательством формулы (128.6). Они устанавливают только связь между рядом и интегралом Фурье. Строгое доказательство дается в теории интеграла Фурье.

§ 129. Закон Ома для переменных токов (синусоидально меняющихся во времени)

1. Рассмотрим участок цепи, состоящий из последовательно соединенных *омического сопротивления* R, *катушки самоиндукции* L и *конденсатора* C, к концам которого приложена синусоидальная электродвижущая сила $\mathscr{E} = \mathscr{E}_0 \cos \omega t$ (см. рис. 271). Найдем ток I, который установится в цепи под действием этой электродвижущей силы. С этой целью перейдем к комплексной форме $\mathscr{E} = \mathscr{E}_0 e^{i\omega t}$. Тогда заряд конденсатора в установившемся режиме представится выражением (127.5). Дифференцируя его по времени с учетом соотношений (122.7)–(122.9), находим

$$I = \mathscr{E}/Z, \quad \mathscr{E} = IZ, \tag{129.1}$$

где введено обозначение

$$Z = R + i\left(\omega L - \frac{1}{\omega C}\right). \tag{129.2}$$

Формула (129.1) называется *законом Ома для переменных* (*точнее, синусоидальных*) *токов.* Роль сопротивления играет комплексная величина Z, называемая *комплексным сопротивлением*, или *импедансом.* Физическое содержание соотношения (129.1) раскроется полностью, если представить его в вещественной форме. Вопрос сводится к определению *амплитуды* и *фазы тока.* Начнем с рассмотрения частных случаев.

2. Случай 1. Цепь не содержит конденсатора и катушки самоиндукции. При отсутствии самоиндукции $L = 0$. Отсутствие конденсатора означает, что точки *3* и *4* на рис. 271 сливаются в одну точку, т. е. напряжение между этими точками все время равно нулю. Поэтому член q/C в уравнении (122.5) следует опустить, что можно сделать формально, полагая $C = \infty$, а не $C = 0$, как могло показаться на первый взгляд. В результате формула (129.1) переходит в обычный закон Ома $\mathscr{E} = RI$ — между током и напряжением *нет сдвига фаз.*

3. Случай 2. Цепь не содержит конденсатора и омического сопротивления ($R = 1/C = 0$). В этом случае

$$Z = i\omega L, \quad \mathscr{E} = i\omega LI. \tag{129.3}$$

Импеданс Z *чисто мнимый.* Это значит, что сдвиг фаз между током и напряжением равен 90°. Действительно,

$$\mathscr{E} = \omega LIe^{i\pi/2}, \text{ или } \mathscr{E} = \omega LI_0 e^{i(\omega t+\pi/2)}.$$

В вещественной форме

$$\mathscr{E} = \omega LI_0 \cos(\omega t + \pi/2), \quad I = I_0 \cos\omega t.$$

Отсюда видно, что *ток отстает от напряжения по фазе на* $\pi/2$.

Амплитуда тока связана с амплитудой напряжения соотношением $\mathscr{E}_0 = \omega LI_0$. Величина ωL (или $i\omega L$) называется *индуктивным сопротивлением.* Индуктивное сопротивление тем больше, чем больше частота ω и индуктивность катушки L. Для увеличения индуктивного сопротивления в катушку самоиндукции вводят железный сердечник, состоящий из железных полос или проволок, изолированных друг от друга, например, лаком. Такая катушка называется *дросселем.*

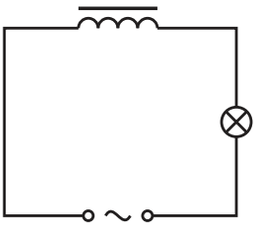


Рис. 287

Соберем схему, состоящую из последовательно соединенных лампового реостата и дросселя (рис. 287). Вынем из дросселя железный сердечник и подберем сопротивление лампового реостата таким, чтобы лампы горели нормальным накалом. Затем вдвинем железный сердечник. При этом увеличивается индуктивное сопротивление, а ток уменьшается. В результате лампы начинают гореть более тускло и даже могут совсем погаснуть. Если переключить цепь на постоянное напряжение, то вдвигание и выдвигание сердечника не будет влиять на силу тока и накал ламп.

При высоких частотах, даже когда величина L ничтожна, индуктивное сопротивление ωL может значительно превзойти омическое. Соединим концы толстого медного стержня (диаметром около 5 мм), согнутого в дугу, с источником тока высокой частоты (порядка мегагерца). Параллельно дуге включим обычную лампочку накаливания с сопротивлением около 100 Ом (рис. 288). По сравнению с ним омическое сопротивление дуги (для постоянного тока около 0,001 Ом) ничтожно. Несмотря на это, дуга не закорачивает лампочку. Последняя горит ярко, что объясняется большим индуктивным сопротивлением дуги.

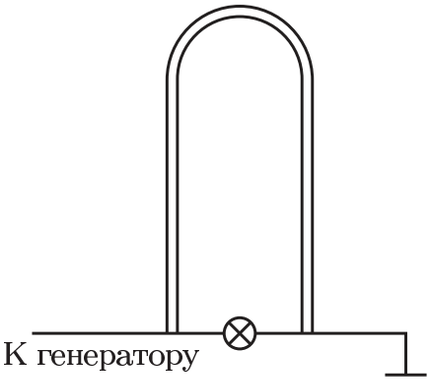


Рис. 288

4. Случай 3. Цепь не содержит катушки самоиндукции и омического сопротивления. В этом случае

$$Z = -\frac{i}{\omega C}, \quad \mathscr{E} = -\frac{i}{\omega C}\, I, \quad I = i\omega C \mathscr{E}. \tag{129.4}$$

Отсюда видно, что ток опережает по фазе напряжение на $\pi/2$. Амплитуда тока связана с амплитудой напряжения соотношением $I_0 = \omega C \mathscr{E}_0$. Величина $1/\omega C$ (или $-i/\omega C$) называется *емкостным сопротивлением.* Емкостное сопротивление тем меньше, чем больше емкость конденсатора. И без вычислений ясно, что конденсатор бесконечно большой емкости не будет оказывать никакого сопротивления переменному току. При прохождении переменного тока на обкладках конденсатора накапливались бы электрические заряды противоположных знаков. Однако при $C = \infty$ это не привело бы к возникновению напряжения между обкладками ($V = q/C = 0$). В этом отношении конденсатор бесконечно большой емкости ведет себя так же, как кусок проволоки, не обладающий омическим сопротивлением.

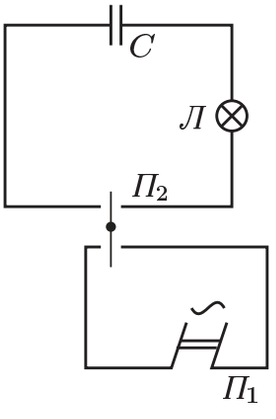


Рис. 289

Соберем схему, изображенную на рис. 289. Перекинем переключатель $П_1$ на постоянный ток. Лампа $Л$ гореть не будет, так как конденсатор C оказывает ей бесконечно большое сопротивление. Будем изменять направление тока с помощью переключателя $П_2$. При каждом переключении через лампу проходит зарядный или разрядный ток конденсатора, и она вспыхивает. При достаточно частом переключении получается переменный ток, и лампа горит почти полным накалом. При перекидывании переключателя $П_1$ на переменный ток лампа $Л$ горит ровно и ярко. Если увеличить емкость конденсатора C, то уменьшится емкостное сопротивление, и лампа будет гореть еще ярче. Уменьшение емкости конденсатора приводит к обратному эффекту.

5. Исследуем теперь общий случай, когда R, L, C произвольны. Импеданс Z представим в виде

$$Z = \rho e^{i\delta},$$

где ρ и δ — величины вещественные. Тогда

$$I = \frac{\mathscr{E}}{Z} = \frac{\mathscr{E}_0}{\rho}\, e^{i(\omega t - \delta)} = I_0 e^{i(\omega t - \delta)},$$

или в вещественной форме

$$I = \frac{\mathscr{E}_0}{\rho} \cos(\omega t - \delta).$$

Ток отстает по фазе от напряжения на δ. Для вычисления ρ и δ воспользуемся формулой (129.2). Она дает

$$\rho e^{i\delta} = R + i\left(\omega L - \frac{1}{\omega C}\right).$$

Приравнивая вещественные и мнимые части, получим

$$\rho\cos\delta = R, \quad \rho\sin\delta = \omega L - \frac{1}{\omega C}.$$

Отсюда

$$\rho^2 = R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2, \tag{129.5}$$

$$\operatorname{tg}\delta = \frac{\omega L - 1/\omega C}{R}.$$

В результате получаем

$$I = \frac{\mathscr{E}_0\cos(\omega t - \delta)}{\sqrt{R^2 + (\omega L - 1/\omega C)^2}}. \tag{129.6}$$

Эта формула совместно с формулой (129.5) выражает закон Ома для переменных токов *в вещественной форме.* По сравнению с вещественной комплексная форма того же закона

$$I = \frac{\mathscr{E}}{R + i(\omega L - 1/\omega C)} \tag{129.7}$$

более компактна и лучше приспособлена для вычислений.

6. Закон Ома для переменных токов можно представить также на векторной диаграмме. С этой целью перепишем предыдущую формулу в виде

$$RI + i\omega LI - \frac{i}{\omega C}\,I = \mathscr{E}. \tag{129.8}$$

Она означает, что сумма падений напряжения на омическом сопротивлении, катушке самоиндукции и конденсаторе равна внешнему приложенному напряжению. В методе векторных диаграмм все эти напряжения должны рассматриваться не как числа, а как *векторы.* Падение напряжения на омическом сопротивлении RI откладывается вдоль оси X, падения напряжения на катушке самоиндукции $I\omega L$ и конденсаторе $I/\omega C$ — вдоль оси Y: первое в положительном, второе — в отрицательном направлениях (рис. 290). Векторная диаграмма позволяет

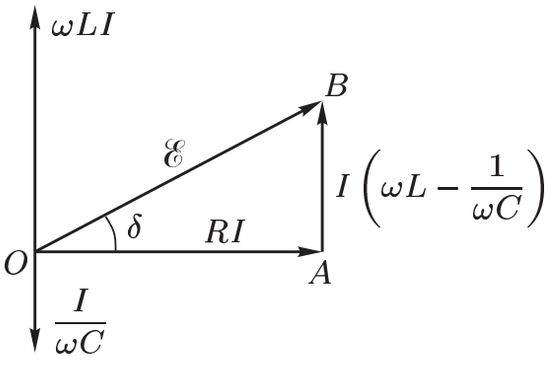


Рис. 290

определить амплитуду и фазу колебаний. Амплитуда колебаний тока найдется из прямоугольного треугольника OAB по теореме Пифагора:

$$(RI_0)^2 + \left(\omega L - \frac{1}{\omega C}\right)^2 I_0^2 = \mathscr{E}_0^2.$$

Из того же треугольника легко найти тангенс угла сдвига фаз δ. В результате получаются прежние формулы (129.5) и (129.6).

Квадратный корень, стоящий в знаменателе формулы (129.6), называется *полным сопротивлением цепи.* Омическое сопротивление R называют *активным сопротивлением*, а величину $\omega L - 1/\omega C$ — *реактивным сопротивлением.* Впрочем, реактивным сопротивлением иногда называют саму мнимую часть импеданса $i(\omega L - 1/\omega C)$.

7. Допустим, что при неизменном внешнем напряжении $\mathscr{E}_0 \cos(\omega t + \delta)$ меняются параметры L и C. Когда реактивное сопротивление $\omega L - 1/\omega C$ обратится в нуль, полное сопротивление достигнет минимума R. Это произойдет при резонансной частоте $\omega = \omega_0 \equiv 1/\sqrt{LC}$. Тогда амплитуда тока станет максимальной: $I_{0\text{макс}} = \mathscr{E}_0/R$. Общее падение напряжения в цепи будет равно падению напряжения на омическом сопротивлении $V_R = RI$. Суммарное падение напряжения на конденсаторе и катушке самоиндукции обратится в нуль: $V_L + V_C = 0$, т. е. напряжения V_L и V_C будут равны по величине и противоположны по фазе. Таким образом, при резонансе

$$\left|\frac{V_L}{V_R}\right| = \left|\frac{V_C}{V_R}\right| = \frac{\omega L}{R} = \frac{\omega}{2\gamma} = Q. \qquad (129.9)$$

При больших добротностях Q напряжения V_L и V_C могут во много раз превзойти напряжение на омическом сопротивлении, а следовательно, и общее напряжение в цепи $\mathscr{E}$. Это явление называется *резонансом напряжений.* При расчете изоляции электрических линий, содержащих емкости и самоиндукции, это явление необходимо учитывать. Иначе может получиться пробой линии.

Для демонстрации явления резонанса напряжений колебательный контур собирается из магазина емкостей C и катушки самоиндукции L с железным сердечником, который можно вдвигать в катушку и выдвигать и тем самым менять ее индуктивность L (рис. 291). В качестве сопротивления R служит электрическая лампочка или несколько лампочек, соединенных параллельно. С помощью четырех вольтметров, не показанных на рисунке, можно одновременно измерять напряжения V_L, V_C, V_R (точнее, их амплитуды), а также входное напряжение $\mathscr{E}$. Опыт показывает, что при включении переменного тока напряжение V_R на сопротивлении R меньше входного напряжения $\mathscr{E}$: часть напряжения падает на реактивном сопротивлении. Поэтому лампы горят сравнительно тускло, а амплитуды напряжений

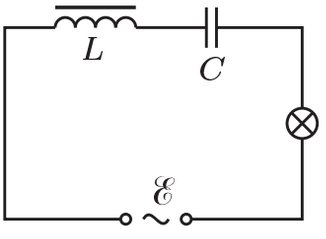


Рис. 291

V_L и V_C отличаются друг от друга. Изменяя самоиндукцию L или емкость C, можно добиться того, чтобы напряжения V_L и V_C сделались по амплитуде почти одинаковыми. Так как их фазы противоположны, то реактивное сопротивление обращается в нуль, сила тока достигает максимума, и лампа начинает гореть ярко. При этом напряжения V_L и V_C намного превосходят внешнее напряжение $\mathscr{E}$. Получился резонанс напряжений.

§ 130. Правила Кирхгофа для переменных токов

1. *К переменным токам без всяких изменений применимо первое правило Кирхгофа.* Действительно, точки схождения проводов обладают пренебрежимо малыми емкостями, в них не могут накапливаться электрические заряды. Поэтому в любой момент времени сумма сил токов, подходящих к точке разветвления, должна равняться сумме сил токов, отходящих от нее. *Второе правило Кирхгофа также применимо к синусоидальным переменным токам, если омические сопротивления R всюду заменить на соответствующие комплексные сопротивления* (*импедансы*) Z. Это правило непосредственно следует из уравнения Максвелла

$$\oint E_l\, dl = -\frac{d\Phi}{dt}.$$

Для доказательства выделим в разветвленной сети какой-либо замкнутый контур, например контур, изображенный на рис. 292. Предположим, что выполнено условие квазистационарности. Тогда предыдущее уравнение для выделенного контура запишется в виде

$$\sum_k \left(I_k R_k + \frac{q_k}{C_k} - \mathscr{E}_k \right) = -\sum L_k \frac{dI_k}{dt}.$$

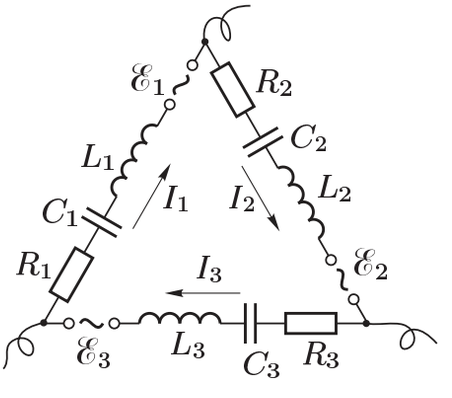


Рис. 292

Приложенные напряжения $\mathscr{E}_k$ считаются положительными, если при обходе контура источник тока проходится от отрицательного полюса к положительному. Написанное уравнение справедливо и для неустановившихся процессов, причем электродвижущие силы $\mathscr{E}_k$ могут меняться во времени как угодно. Допустим теперь, что все электродвижущие силы $\mathscr{E}_k$ меняются во времени синусоидально, т. е. в комплексной форме $\mathscr{E}_k \sim e^{i\omega t}$. Тогда в случае установившихся процессов заряды q_k и токи $I_k = \dot{q}_k$ будут меняться во времени так же, а предыдущее уравнение перейдет в

$$\sum_k I_k \left(R_k + i\omega L_k - \frac{i}{\omega C_k} \right) = \sum_k \mathscr{E}_k,$$

или

$$\sum Z_k I_k = \sum \mathscr{E}_k, \tag{130.1}$$

а это и есть второе правило Кирхгофа.

2. Все результаты, полученные формальным применением правил Кирхгофа к постоянным токам, в комплексной форме сохраняют силу и для установившихся синусоидальных токов. В частности, при параллельном соединении нескольких комплексных сопротивлений результирующее комплексное сопротивление определяется формулой

$$\frac{1}{Z} = \sum \frac{1}{Z_k}. \tag{130.2}$$

Рассмотрим частный случай, когда соединены параллельно катушка самоиндукции и конденсатор (рис. 293). По первому правилу Кирхгофа

$$I + I_2 = I_1.$$

По второму правилу

$$i\omega L I_1 - \frac{iI_2}{\omega C} = 0.$$

Исключим из этих уравнений, например, ток I_2. Получим

$$I = I_1(1 - \omega^2 LC),$$

или

$$I = I_1\left(1 - \frac{\omega^2}{\omega_0^2}\right). \tag{130.3}$$

Если $\omega = \omega_0$, то $I = 0$, и, следовательно, $I_1 = I_2$. Токи I_1 и I_2 могут быть отличны от нуля. Однако колебания в контуре становятся совершенно независимыми от внешнего тока. Причина всего этого в том, что при $\omega = \omega_0$ (т. е. при $\omega L = 1/\omega C$) комплексное сопротивление Z контура обращается в бесконечность, как в этом легко убедиться с помощью формулы (130.2). Контур ведет себя как непроницаемая пробка, через которую внешний ток пройти не может, а в катушке могут совершаться только свободные колебания. Амплитуды этих колебаний легко найти из условия, что напряжение на конденсаторе или на катушке самоиндукции в любой момент времени должно равняться внешнему приложенному напряжению $\mathscr{E}$. Это дает

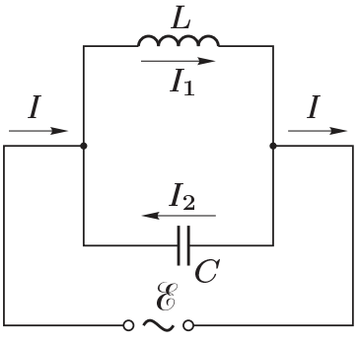


Рис. 293

$$I_1 = \frac{\mathscr{E}}{i\omega L}, \quad I_2 = -\frac{\mathscr{E}}{-i/\omega C} = \frac{\mathscr{E}}{i\omega L}.$$

Возникает вопрос, как в колебательном контуре могли появиться токи I_1 и I_2, если его сопротивление переменному току бесконечно велико. На этот вопрос законы и соотношения, которыми мы пользовались,

ответа дать не могут, так как они относятся *только к установившимся состояниям.* Последние устанавливаются в результате *переходных процессов.* Во время переходных процессов ток во внешней цепи не равен нулю, в колебательный контур поступают токи и заряды, идет накопление электромагнитной энергии. Это происходит до тех пор, пока в любой момент времени напряжения на конденсаторе и катушке самоиндукции не уравновесят внешнее приложенное напряжение. Когда это произойдет, дальнейшее поступление новых зарядов и токов в колебательный контур прекратится. Начнут совершаться свободные колебания, как если бы контур был *автономной колебательной системой.* Однако эта автономия сохраняется только до тех пор, пока внешняя электродвижущая сила $\mathscr{E} = \mathscr{E}_0 \cos \omega t$ остается неизменной. Если изменить амплитуду $\mathscr{E}_0$, то $\mathscr{E}$ перестанет уравновешивать напряжения на конденсаторе и катушке самоиндукции. Появится ток во внешней цепи, и начнется новый процесс установления колебаний.

Разумеется, без поступления энергии извне незатухающие колебания возможны только тогда, когда омическое сопротивление контура равно нулю. Учтем теперь омическое сопротивление. Обозначим буквой R омическое сопротивление катушки самоиндукции. Омическим сопротивлением всех остальных проводов пренебрежем. Тогда

$$I_1 = \frac{\mathscr{E}}{R + i\omega L}, \quad I_2 = -i\omega C \mathscr{E},$$

$$I = I_1 - I_2 = \frac{R - i\omega L + i\omega^2 L^2 \omega C + i\omega C R^2}{R^2 + \omega^2 L^2}\, \mathscr{E}.$$

При резонансе $\omega = \omega_0$ и, следовательно, $\omega C = 1/\omega L$. В этом случае

$$I = \frac{R + iR \cdot R/\omega L}{R^2 + \omega^2 L^2}\, \mathscr{E}.$$

Пусть сопротивление R пренебрежимо мало по сравнению с индуктивным сопротивлением ωL. Тогда

$$I \approx \frac{R}{\omega^2 L^2}\, \mathscr{E} = i \frac{2\gamma}{\omega}\, I_1 = i \frac{I_1}{Q}.$$

Следовательно,

$$|I_1 / I| = Q.$$

Чем больше добротность колебательного контура, тем меньше ток I в общей цепи. Таким образом, при $\omega = \omega_0$ малым током I в колебательном контуре большой добротности можно возбудить и поддерживать большие токи. Это явление называется *резонансом токов.* Его можно наблюдать, если в схему рис. 313 ввести амперметры для измерения токов I, I_1 и I_2. Вдали от резонанса токи I, I_1 и I_2 не очень сильно отличаются по величине. Изменением индуктивности катушки или емкости конденсатора настроим колебательный контур в резонанс. Токи I_1 и I_2 при этом сделаются почти одинаковыми, а ток I — близким к нулю.

§ 131. Эффективные напряжение и ток

1. Сдвиг фаз между током и напряжением влияет на *работу и мощность переменного тока.* Работа δA, совершаемая электродвижущей силой $\mathscr{E}$ за время dt, определяется выражением $\delta A = \mathscr{E}\, dq$, где dq — заряд, прошедший через поперечное сечение провода за это время. Мгновенная мощность будет

$$P = \mathscr{E}\frac{dq}{dt} = \mathscr{E} I. \tag{131.1}$$

Поскольку операция умножения нелинейна, пользоваться комплексными выражениями так, как это делалось до сих пор, нельзя. Надо перейти к вещественной форме, т. е. в случае синусоидальных токов положить

$$\mathscr{E} = \mathscr{E}_0 \cos \omega t, \quad I = I_0 \cos(\omega t - \delta). \tag{131.2}$$

Но обычно проще проводить вычисления в комплексной форме, применяя следующий прием. Обозначим через $\mathscr{E}^*$ и I^* величины, комплексно сопряженные с $\mathscr{E}$ и I. Тогда

$$\operatorname{Re}\mathscr{E} = \frac{1}{2}(\mathscr{E} + \mathscr{E}^*), \quad \operatorname{Re} I = \frac{1}{2}(I + I^*).$$

Подставляя эти вещественные величины в формулу (131.1), получим

$$P = \frac{1}{4}(\mathscr{E} I^* + \mathscr{E}^* I) + \frac{1}{4}(\mathscr{E} I + \mathscr{E}^* I^*). \tag{131.3}$$

Здесь уже можно положить $\mathscr{E} = \mathscr{E}_0 e^{i\omega t}$, $I = I_0 e^{i(\omega t - \delta)}$ и воспользоваться формулой $\cos\alpha = \frac{1}{2}(e^{i\alpha} + e^{-i\alpha})$. В результате получится

$$P = \frac{1}{2}\mathscr{E}_0 I_0 \cos\delta + \frac{1}{2}\mathscr{E}_0 I_0 \cos(2\omega t - \delta).$$

Это выражение, конечно, легко получить и без использования комплексных выражений. Второе слагаемое в последней формуле быстро колеблется во времени с удвоенной частотой 2ω. Среднее значение его по времени равно нулю. Первое слагаемое от времени не зависит и дает *среднюю мощность переменного тока*:

$$\overline{P} = \frac{1}{2}\mathscr{E}_0 I_0 \cos\delta. \tag{131.4}$$

Величины $\mathscr{E}_0$ и I_0 называются *амплитудными значениями напряжения и тока.* Вместо них в электротехнике чаще употребляют *эффективные значения*, определяемые выражениями

$$\mathscr{E}^2_{\text{эфф}} = \frac{1}{T}\int_0^T \mathscr{E}^2\, dt, \quad I^2_{\text{эфф}} = \frac{1}{T}\int_0^T I^2\, dt. \tag{131.5}$$

Для синусоидальных токов

$$\mathscr{E}_{\text{эфф}} = \mathscr{E}_0/\sqrt{2}, \quad I_{\text{эфф}} = I_0/\sqrt{2}. \tag{131.6}$$

С введением этих величин формула для средней мощности переменного тока принимает вид

$$\overline{P} = \mathscr{E}_{\text{эфф}} I_{\text{эфф}} \cos\delta. \tag{131.7}$$

2. Аналогичным образом разность фаз δ сказывается на *взаимодействии переменных токов.* Рассмотрим, например, переменные токи I_1 и I_2, текущие вдоль бесконечно длинных прямолинейных проводов, находящихся на расстоянии r друг от друга. Мгновенная сила, действующая на единицу длины каждого провода, определяется выражением

$$F = \frac{2I_1 I_2}{cr}\,\mu.$$

(Здесь применяется гауссова система единиц.) Если токи I_1 и I_2 синусоидальны, а сдвиг фаз между ними равен δ, то для средней силы отсюда находим

$$\overline{F} = \frac{2\mu}{cr}\, I_{1\,\text{эфф}} I_{2\,\text{эфф}} \cos\delta. \tag{131.8}$$

В зависимости от разности фаз δ средняя сила $\overline{F}$ может быть либо силой притяжения, либо силой отталкивания. Если $\delta = \pi/2$, то $\overline{F} = 0$.

Примером может служить опыт Элиу Томсона, описанный в § 65. Если I — ток в обмотке электромагнита, то магнитный поток, пронизывающий алюминиевое кольцо, будет $\Phi = L_{12} I$. Допустим, что $I = I_0 e^{i\omega t}$. Тогда электродвижущая сила индукции в кольце $\mathscr{E}^{\text{инд}} = -d\Phi/dt = -i\omega L_{12} I$, а ток

$$I' = -\frac{i\omega L_{12} I}{R + i\omega L}.$$

Здесь L — индуктивность, а R — омическое сопротивление кольца. Если R пренебрежимо мало по сравнению с индуктивным сопротивлением ωL, то

$$I' = -\frac{L_{12}}{L}\, I.$$

Так как коэффициенты L_{12} и L положительны, то фазы токов I и I' противоположны: $\delta = \pi$. Средняя сила взаимодействия кольца и электромагнита будет *отталкивательной.*

§ 132. Процессы установления колебаний

В общем случае колебания под действием внешней силы слагаются из вынужденных и свободных (см. § 127). Собственные колебания затухают, если время t, прошедшее с момента начала действия силы, велико по сравнению с временем затухания $\tau = 1/\gamma$. Исследуем теперь этот вопрос более подробно. Для простоты проведем вычисления в предположении, что коэффициент затухания γ равен нулю. Понятно, что в этом случае процесс установления колебаний никогда не закончится, так как $\tau = 1/\gamma = \infty$. Однако, получив решение для $\gamma = 0$, легко понять качественно, что будет при $\gamma \neq 0$.

Предположим, что на гармонический осциллятор в момент времени $t = 0$ начинает действовать периодическая «сила» $f = f_0 \cos \omega t$. Тогда при $t > 0$ колебания будут описываться уравнением

$$\ddot{x} + \omega_0^2 x = f_0 \cos \omega t.$$

Если $\omega \neq \omega_0$, то общее решение этого уравнения имеет вид

$$x = \frac{f_0}{\omega_0^2 - \omega^2} \cos \omega t + a \cos \omega_0 t + b \sin \omega_0 t.$$

Постоянные a и b определяются начальными условиями. Допустим, что в начальный момент $t = 0$ координата x и скорость $\dot{x}$ равны нулю. Чтобы удовлетворить первому условию, необходимо положить $a = -f_0/(\omega_0^2 - \omega^2)$. После этого второе условие будет удовлетворено, если $b = 0$. Решение, удовлетворяющее обоим условиям, запишется в виде

$$x = \frac{f_0}{\omega_0^2 - \omega^2} (\cos \omega t - \cos \omega_0 t). \tag{132.1}$$

Получилась суперпозиция двух гармонических колебаний: собственного с частотой ω_0 и вынужденного с частотой ω. С суперпозицией гармонических колебаний разных частот приходится встречаться в самых разнообразных явлениях. Примером могут служить два звучащих камертона с разными собственными частотами. Особое значение имеет случай, когда частоты ω и ω_0 отличаются друг от друга мало. В этом случае выражение (132.1) целесообразно преобразовать к виду

$$x = A(t) \sin \overline{\omega} t, \tag{132.2}$$

где

$$A(t) = \frac{\pm 2 f_0}{\omega_0^2 - \omega^2} \sin \omega_м t, \tag{132.3}$$

$$\overline{\omega} = (\omega + \omega_0)/2, \quad \omega_м = |\omega - \omega_0|/2. \tag{132.4}$$

Колебание, представляемое выражением (132.2), есть *амплитудно-модулированное колебание* с несущей частотой, равной *средней частоте* ω, и *частотой модуляции* $\omega_м$. Общий вид его представлен на рис. 294. В случае гармонических колебаний воздуха, возбуждаемых двумя камертонами, ухо обычно воспринимает результирующее колебание как

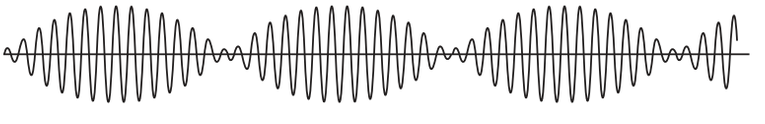


Рис. 294

гармоническое колебание с «переменной амплитудой» (см. § 126) Ухо слышит музыкальный тон, интенсивность которого периодически меняется с периодом $T_б = \pi/\omega_м$ и частотой $\omega_б = 2\pi/T_б = 2\omega_м = |\omega - \omega_0|$. Это явление называется *биениями*, а величины $T_б$ и $\omega_б$ — *периодом* и *частотой биений* соответственно. То обстоятельство, что ухо

воспринимает биения как музыкальный тон периодически меняющейся громкости, связано, конечно, с тем, что ухо как колебательная система обладает относительно малой добротностью; время установления колебаний этой системы мало по сравнению с периодом биений $T_б$.

В действительности собственные колебания затухают. Результирующее колебание представляется суперпозицией незатухающей и затухающей синусоид. Когда собственные колебания затухнут, останется только колебание, представляемое незатухающей синусоидой, т. е. процесс установления колебаний закончится. В зависимости от величины коэффициента затухания γ характер установления колебаний может быть разным, как это видно из рис. 295.

Рассмотрим теперь случай, когда $\omega = \omega_0$ (резонанс). В этом случае выражение (132.1) принимает неопределенный вид $x = 0/0$. Неопреде-

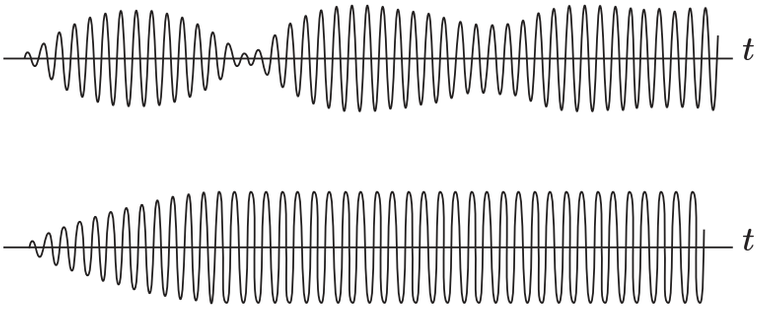


Рис. 295

ленность можно раскрыть, перейдя к пределу $\omega \to \omega_0$. Дифференцируя по ω, с помощью известного правила Лопиталя находим

$$\lim_{\omega\to\omega_0} \frac{\cos\omega t - \cos\omega_0 t}{\omega_0^2 - \omega^2} = \lim_{\omega\to\omega_0} \frac{-t\sin\omega t}{-2\omega} = \frac{t}{2\omega_0}\sin\omega_0 t.$$

Следовательно,

$$x = \frac{f_0 t}{2\omega_0}\sin\omega_0 t. \tag{132.5}$$

На это выражение можно смотреть как на колебание с частотой ω_0, амплитуда которого $A(t) = f_0 t/2\omega_0$ линейно растет во времени. Если бы не было трения и других тормозящих сил, то процесс нарастания амплитуды колебаний никогда не закончился бы (рис. 296). В действительности тормозящие силы всегда есть. Пока колебания малы, они не играют большой роли. Но по мере нарастания амплитуды колебаний роль тормозящих сил становится все более и более существенной. Они в конце концов приостанавливают дальнейший рост амплитуды колебаний, и последние переходят в колебания постоянной амплитуды.

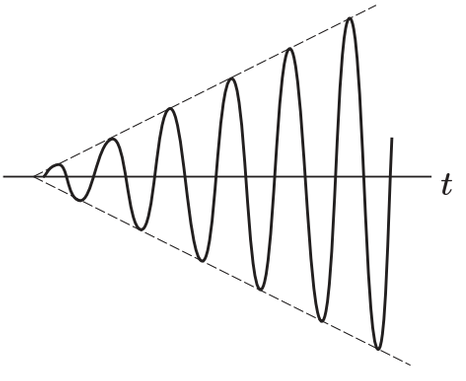


Рис. 296

ЗАДАЧИ

1. При рассмотрении квазистационарных процессов в колебательном контуре пренебрегают магнитной энергией, локализованной в катушке самоиндукции. На примере следующей задачи (и задачи 2) убедиться в допустимости такого приближения. Обкладки плоского конденсатора имеют форму дисков радиуса R. Толщина конденсатора d. Пространство между обкладками заполнено однородным диэлектриком с диэлектрической и магнитной проницаемостями ε и μ. Конденсатор включен в цепь переменного тока $I = I_0 \cos \omega t$. Пренебрегая краевыми эффектами, вычислить электрическую и магнитную энергии, локализованные в конденсаторе. Найти отношение максимальной магнитной к максимальной электрической энергии. Провести численный расчет для $R = 10$ см и частоты $\nu = 100$ Гц, $\varepsilon = \mu = 1$.

Р е ш е н и е. Электрическая энергия (в гауссовой системе)

$$W_e = \frac{q^2}{2C} = \frac{2d}{\varepsilon \omega^2 R^2} I_0^2 \sin^2 \omega t.$$

Магнитное поле внутри конденсатора создается током смещения. На расстоянии r от оси конденсатора оно найдется из соотношения

$$H \cdot 2\pi r = \frac{4\pi}{c} I_{\text{см}} = \frac{4\pi}{c} \left(\frac{r}{R}\right)^2 I,$$

откуда

$$H = \frac{2r}{cR^2} I = \frac{2r}{cR^2} I_0 \cos \omega t.$$

Магнитная энергия, локализованная в конденсаторе:

$$W_m = \frac{\mu}{8\pi} \int H^2 \, dV = \frac{\mu d}{4c^2} I_0^2 \cos^2 \omega t.$$

Отношение максимальных энергий:

$$\frac{W_m^{\text{макс}}}{W_e^{\text{макс}}} = \frac{\varepsilon \mu}{2} \left(\frac{\omega R}{2c}\right)^2 \approx 0{,}5 \cdot 10^{-14}.$$

2. Пространство внутри длинного соленоида, состоящего из N витков проволоки, заполнено однородным веществом с диэлектрической проницаемостью ε и магнитной проницаемостью μ. Длина соленоида равна l, радиус R. По обмотке соленоида течет переменный ток $I = I_0 \cos \omega t$. Пренебрегая краевыми эффектами, вычислить магнитную и электрическую энергии, локализованные внутри соленоида, и найти отношение максимальных значений этих энергий. Провести численный расчет для $R = 5$ см, $\varepsilon = \mu = 1$ и частоты $\nu = 100$ Гц.

О т в е т.

$$W_m = \frac{2\pi^2 \mu R^2 H^2}{c^2 l} I_0 \cos^2 \omega t, \quad W_e = \frac{\varepsilon \mu^2 \pi^2 \omega^2 R^4 N^2}{4c^4 l} I_0^2 \sin^2 \omega t,$$

$$\frac{W_e^{\text{макс}}}{W_m^{\text{макс}}} = \frac{\varepsilon \mu}{2} \left(\frac{\omega R}{2c}\right)^2 \approx 1{,}3 \cdot 10^{-15}.$$

3. Прямой однослойный соленоид с индуктивностью L совершает вынужденные крутильные гармонические колебания вокруг своей оси: $\varphi = \varphi_0 \cos \omega t$. Соленоид гибкими проводами подсоединен к конденсатору емкости C (опыт Мандельштама и Папалекси). Найти напряжение на конденсаторе при резонансе, когда частота ω равна собственной частоте колебательного

контура $\omega_0 = 1/\sqrt{LC}$. Радиус соленоида a, длина проволоки, из которой он изготовлен, l, сопротивление соленоида R (см. § 97).

Ответ. $V = \dfrac{m}{e}\,\dfrac{lLa\omega^3}{R}\,\varphi_0 \sin \omega t$, где m — масса, e — заряд электрона.

4. В § 78 был описан опыт Эйнштейна и де Гааза, в котором для определения гидромагнитного отношения атомов было использовано явление резонанса. Предполагая, что магнитное поле, в котором совершаются крутильные колебания железного цилиндрика, очень сильное, так что цилиндрик почти всегда намагничен до насыщения, а при прохождении магнитного поля через нуль перемагничивание происходит практически мгновенно, определить амплитуду φ_0 установившихся крутильных колебаний цилиндрика при резонансе.

Решение. Уравнение вынужденных крутильных колебаний цилиндрика записывается в виде (78.2). Во время перемагничивания можно пренебречь всеми вращающими моментами, действующими на цилиндрик, за исключением момента $-VI/\Theta\Gamma$. За это время цилиндрик не успеет повернуться на заметный угол, но его угловая скорость получит приращение

$$\Delta\dot{\varphi}_0 = -\frac{V}{\Theta\Gamma}\int \dot{I}\,dt = -2\frac{V}{\Theta\Gamma}\,I_{\text{нас}}.$$

Получив такую начальную угловую скорость, цилиндрик начнет совершать свободные затухающие колебания

$$\varphi = (\Delta\dot{\varphi}_0/\omega)e^{-\gamma t}\sin \omega t.$$

Через половину периода амплитуда колебаний уменьшится в $e^{\gamma T/2}$ раз. Но в это время произойдет новое перемагничивание, в результате которого амплитуда угловой скорости получит новое приращение $\Delta\dot{\varphi}_0$. Новая амплитуда отклонения станет равной $(\Delta\dot{\varphi}_0/\omega)(1+e^{-\gamma T/2})$ и т. д. Амплитуда установившихся колебаний представится геометрической прогрессией

$$\varphi_0 = \frac{\Delta\dot{\varphi}_0}{\omega}\,(1 + e^{-\gamma T/2} + e^{-2\gamma T/2} + \ldots) = \frac{\Delta\dot{\varphi}_0/\omega}{1 - e^{-\gamma T/2}}$$

или при малых значениях коэффициента затухания

$$\varphi_0 = \frac{2\Delta\dot{\varphi}_0}{\omega T\gamma} = \frac{\Delta\dot{\varphi}_0}{\pi\gamma} = -2\frac{V}{\pi\gamma\Theta\Gamma}\,I_{\text{нас}}.$$

5. Через баллистический гальванометр с баллистической постоянной B пропускается кратковременный импульс тока, в течение которого через него проходит количество электричества q. Спустя половину периода, когда рамка гальванометра вернется в исходное положение, через гальванометр пропускается такой же импульс тока, но в противоположном направлении; через следующую половину периода пропускается снова такой же импульс, но в первоначальном направлении и т. д. Таким образом, всякий раз, когда рамка гальванометра проходит через положение равновесия, она испытывает одинаковые толчки в направлении своего движения. Найти максимальный угол отклонения рамки при установившихся колебаниях. Период (затухающих) колебаний гальванометра T, логарифмический декремент d.

Ответ. $\varphi_{\text{макс}} = \dfrac{1}{1 - e^{-Td/2}}\,\dfrac{q}{B}$.

6. Катушка колебательного контура с параметрами $L, C, R = 0$ помещена в постоянное магнитное поле, создающее в ней постоянный магнитный поток

Φ_0. В момент времени $t = 0$ магнитное поле выключается. Время выключения τ пренебрежимо мало по сравнению с периодом собственных колебаний контура. Найти ток I в контуре в зависимости от времени после выключения поля.

Ответ. $I = \dfrac{\Phi_0}{L}\cos\omega_0 t$, $\omega_0 = \dfrac{1}{\sqrt{LC}}$.

7. Вблизи катушки колебательного контура с параметрами L_1, C, $R = 0$ расположена вторая катушка с индуктивностью L_2. Коэффициент взаимной индукции между катушками равен L_{12}. Какой будет резонансная частота контура, если выводы второй катушки замкнуты накоротко? Считать, что индуктивное сопротивление второй катушки на рассматриваемой частоте значительно больше ее активного сопротивления. При каком условии резонанс недостижим?

Ответ. $\omega_{\text{рез}} = 1/\sqrt{(L_1 - L_{12}^2/L_2)C}$. Резонанс недостижим, если $L_{l2}^2 = L_1 L_2$.

8. В колебательном контуре с индуктивностью L и емкостью C совершаются незатухающие колебания силы тока

$$I = I_0\cos\omega t, \quad \omega^2 = 1/(LC).$$

Катушкой самоиндукции служит прямая длинная проволочная спираль. Как изменится частота, амплитуда и энергия колебаний, если в момент времени $t = 0$ очень быстро (т. е. в течение времени, малого по сравнению с периодом колебаний $T = 2\pi/\omega$) растянуть спираль до удвоенной длины? Объяснить, почему при этом меняется энергия колебаний.

Ответ. Частота увеличится в $\sqrt{2}$ раз. Амплитуда колебаний и энергия возрастут вдвое.

9. Две одинаковые катушки, намотанные на общий каркас, включены последовательно в колебательный контур с емкостью C двумя способами, изображенными на рис. 297. Резонансные частоты колебательных контуров оказались равными ω_1 и ω_2 соответственно. Найти индуктивность L каждой из катушек и коэффициент их взаимной индукции L_{12}.

Ответ. $L = \dfrac{1}{4C}\left(\dfrac{1}{\omega_1^2} + \dfrac{1}{\omega_2^2}\right)$, $L_{12} = \dfrac{1}{4C}\left(\dfrac{1}{\omega_1^2} - \dfrac{1}{\omega_2^2}\right)$.

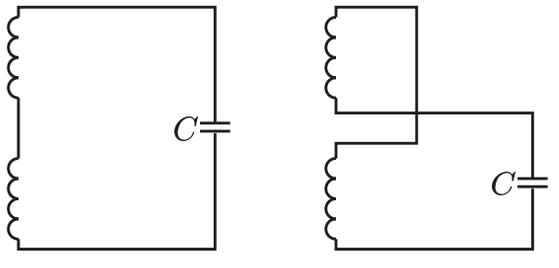


Рис. 297

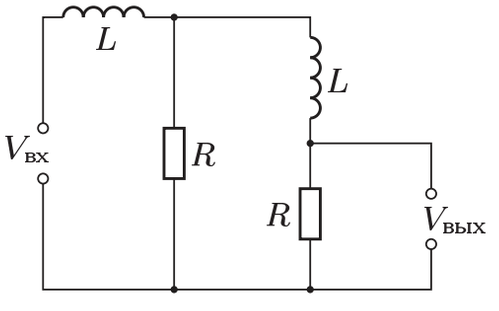


Рис. 298

10. С помощью схемы, показанной на рис. 298, требуется получить фазовый сдвиг на угол 90° между напряжением на входе $V_{\text{вх}}$ и напряжением на выходе $V_{\text{вых}}$. Какому условию должны удовлетворять параметры схемы R и L, если круговая частота входного напряжения равна ω? Чему при этом будет равно отношение амплитуд входного и выходного напряжений?

Ответ. $\omega L = R$, $|V_{\text{вх}}/V_{\text{вых}}| = 3$.

11. Найти ток I (в установившемся режиме) в цепи, изображенной на рис. 299. При какой частоте ω амплитуда установившихся колебаний будет максимальна и при какой минимальна? Чему равен максимум и минимум тока?

Ответ. $I = \dfrac{V\omega C(1-\omega^2 LC)}{2\omega^2 LC - 1}\sin\omega t$, $I_{\text{мин}} = 0$ при $\omega^2 = 1/LC$, $I_{\text{макс}} = \infty$ при $\omega^2 = 1/(2LC)$.

12. Генератор синусоидальной ЭДС замкнут на активное сопротивление R и реактивное X, соединенные параллельно. Убрав R и X, тот же генератор замыкают на активное сопротивление r и реактивное x, соединенные после-

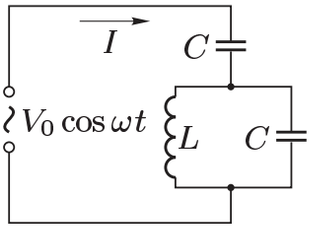


Рис. 299

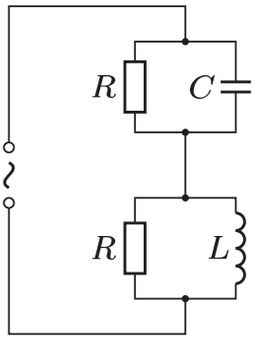


Рис. 300

довательно. При каком условии амплитуда и фаза тока при этом останутся неизменными, если X и x — величины вещественные?

Ответ. $r = \dfrac{X^2}{R^2+X^2}R$, $x = \dfrac{R^2}{R^2+X^2}X$.

13. При каком условии амплитуда тока I в цепи, изображенной на рис. 300, зависит только от амплитуды приложенного напряжения $V = V_0\cos\omega t$, но не от частоты? Найти при этом условии разность фаз между приложенными напряжениями на концах RC-пары.

Ответ. $L = R^2C$, $\operatorname{tg}\varphi = -\omega RC$.

14. Найти комплексный импеданс Z бесконечной цепочки, изображенной на рис. 301.

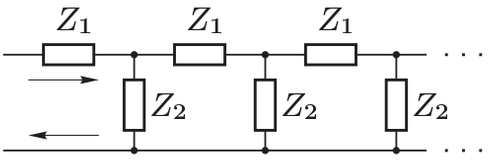


Рис. 301

Решение. Пусть параметры цепочки таковы, что при наложении синусоидального напряжения в ней устанавливается синусоидальный ток. В этом случае можно пользоваться понятием импеданса. Если удалить первые два звена цепочки Z_1 и Z_2, то останется такая же бесконечная цепочка. Ее можно заменить одним звеном с импедансом Z. Тогда получится схема, изображенная на рис. 302. Импедансы Z и Z_2 соединены параллельно, их

результирующий импеданс $ZZ_2/(Z+Z_2)$ соединен последовательно с импедансом Z_1. В результате должен получиться импеданс Z, т. е.

$$Z = Z_1 + \frac{ZZ_2}{Z+Z_2}, \text{ откуда } Z = \frac{Z_1}{2} + \sqrt{Z_1^2/4 + Z_1Z_2}.$$

Знак плюс перед квадратным корнем означает, что из двух значений корня следует брать то, которое имеет положительную вещественную часть. Действительно, квадратный корень есть не что иное, как импеданс Z' бесконечной цепочки, изображенной на рис. 303, а во всякой реальной системе активная часть комплексного сопротивления должна быть положительной.

15. Решить предыдущую задачу в предположении, что все импедансы, из которых составлена цепь, чисто мнимые (состоят из катушек самоиндукции и конденсаторов).

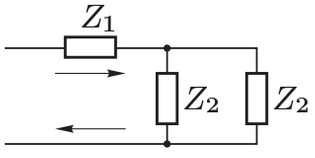


Рис. 302

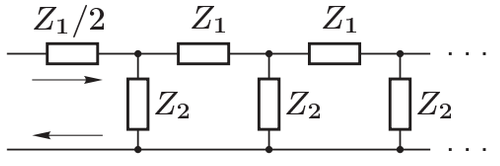


Рис. 303

Р е ш е н и е. Полагая в предыдущей задаче $Z_1 = iX_1$, $Z_2 = iX_2$, получим

$$Z = \frac{1}{2}(iX_1 \pm \sqrt{-(4X_1X_2 + X_1^2)}).$$

Если $4X_1X_2 + X_1^2 < 0$, т. е. подкоренное выражение положительно, то импеданс Z будет содержать вещественную часть, а потому цепь будет потреблять или отдавать энергию, в зависимости от знака этой вещественной части. Стационарное состояние в этом случае невозможно, а решение, полученное в предыдущей задаче, неприменимо. Для решения необходимо использовать начальные условия.

Синусоидальный ток в цепочке возможен только при выполнении условия $4X_1X_2 + X_1^2 \geqslant 0$. Только тогда можно пользоваться понятием импеданса и решением предыдущей задачи. Вопрос сводится к выбору знака перед квадратным корнем. Для этого предположим, что импеданс Z_1 содержит малое омическое сопротивление R_1 и перейдем к пределу $R_1 \to 0$. Полагая $Z_1 = iX_1 + R_1$, $Z_2 = iX_2$, из решения предыдущей задачи получим

$$Z' = \frac{1}{2}\sqrt{-(4X_1X_2 + X_1^2) + 2iR_1(X_1 + 2X_2)},$$

где $Z' = Z - Z_1/2$ — импеданс бесконечной цепочки, изображенной на рис. 303. При извлечении квадратного корня пренебрежем квадратами R_1. В этом приближении

$$Z' = \frac{i}{2}\left[\sqrt{4X_1X_2 + X_1^2} - i\frac{R_1(X_1 + 2X_2)}{\sqrt{4X_1X_2 + X_1^2}}\right].$$

Знак надо выбрать так, чтобы вещественная часть этого выражения была положительна. Так как $R_1 > 0$, то в случае $X_1 + 2X_2 > 0$ надо взять знак плюс, а в случае $X_1 + 2X_2 < 0$ — минус. Полагая $R_1 = 0$, окончательно

находим

$$Z = \begin{cases} \dfrac{i}{2}\,[X_1 + \sqrt{4X_1X_2 + X_1^2}], & \text{если } X_1 + 2X_2 > 0, \\ \dfrac{i}{2}\,[X_1 - \sqrt{4X_1X_2 + X_1^2}], & \text{если } X_1 + 2X_2 < 0. \end{cases}$$

Например, если цепочка составлена только из катушек самоиндукции, то $X_1 > 0$, $X_2 > 0$, и перед корнем надо взять знак плюс. В этом случае ток отстает по фазе от напряжения на $\pi/2$. Если же цепочка состоит только из конденсаторов, то $X_1 < 0$, $X_2 < 0$, так что годится только знак минус. В этом случае фаза тока опережает фазу напряжения на $\pi/2$.

16. Найти импеданс бесконечной цепочки, изображенной на рис. 304.

Ответ. $Z = Z_1 + \sqrt{Z_1^2 + Z_1Z_2}$.

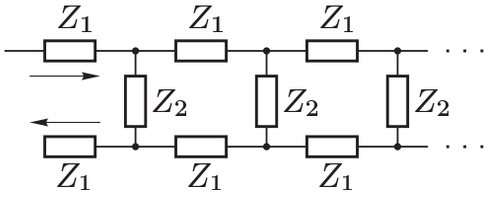


Рис. 304

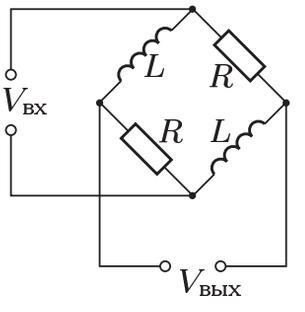


Рис. 305

17. На вход схемы, изображенной на рис. 305, подается синусоидальное напряжение частоты ω. Исследовать зависимость амплитуды и фазы выходного напряжения от величины сопротивления R.

Ответ. Амплитуды на входе и выходе одинаковы. Сдвиг фазы выходного напряжения относительно входного определяется формулой

$$\operatorname{tg}\delta = \frac{2\omega RL}{\omega^2L^2 - R^2}.$$

18. По длинному прямому проводу течет синусоидальный ток I высокой частоты $\nu = 10^8$ Гц. К проводу подносится квадратный проволочный контур со стороной $a = 17{,}2$ см, в который включена лампочка (рис. 306 *а*). Когда контур поднесен на расстояние $b = 10$ см, лампочка горит нормальным накалом. Определить эффективное значение силы тока в проводе $I_{эфф}$, если для нормального накала лампочки требуется постоянное напряжение $V = 6$ В. Уменьшится или увеличится напряжение на лампочке и во сколько раз, если квадрат заменить двойным квадратом, изображенным на рис. 306 *б*? Сопротивлением контуров пренебречь.

Ответ. $I_{эфф} = \dfrac{V}{4\pi\nu a \ln(a/b + 1)} = 0{,}028$ СГСМ-ед.$= 0{,}28$ А. Напряжение на лампочке уменьшится в

$$\ln\left(\frac{a}{b} + 1\right) \Big/ \ln\frac{(a+b)^2}{b(b+2a)} \approx 2 \text{ раза}.$$

19. Цепь, состоящая из последовательно соединенных сопротивления R и большой индуктивности L, присоединена к источнику постоянного тока,

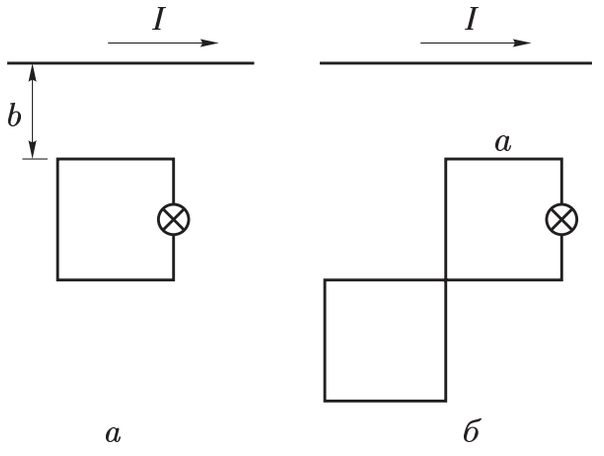


Рис. 306

поддерживающего на зажимах постоянное напряжение V_0. Для ограничения перенапряжений во время отключения источника параллельно с цепью включен конденсатор емкости C (рис. 307). Определить напряжение на конденсаторе $V(t)$ после отключения источника постоянного напряжения. Параметры контура удовлетворяют условию $4L > CR^2$.

Ответ. $V(t) = V_0 e^{-\gamma t}\left[\cos\omega t + \left(\dfrac{1}{\omega RC} + \dfrac{\gamma}{\omega}\right)\sin\omega t\right]$, где $\omega^2 = \omega_0^2 - \gamma^2$, $\omega_0^2 = 1/LC$, $\gamma = R/2L$.

20. К контуру L, C, R (рис. 308) с малым затуханием в момент $t = 0$ подключают источник постоянной ЭДС с ничтожно малым внутренним сопротивлением. Определить напряжение V на конденсаторе C в зависимости от времени t. На какое минимальное напряжение должен быть рассчитан конденсатор?

Ответ. $V = \mathscr{E}[1 - e^{-\gamma t}(\cos\omega t + (\gamma/\omega)\sin\omega t)]$. Минимальное напряжение, на которое должен быть рассчитан конденсатор, не меньше $2\mathscr{E}$.

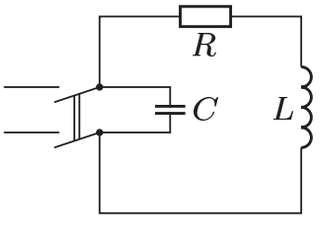


Рис. 307

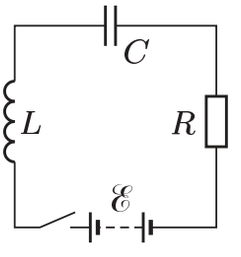


Рис. 308

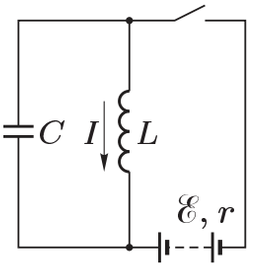


Рис. 309

21. Катушка с индуктивностью L, конденсатор с емкостью C и батарея с электродвижущей силой $\mathscr{E}$ и внутренним сопротивлением R соединены параллельно (рис. 309). Найти силу тока $I(t)$ в катушке после включения батареи. Параметры L, C, R удовлетворяют условию $L < 4CR^2$.

Ответ. $I = \dfrac{\mathscr{E}}{R}\left[1 - e^{-\gamma t}\left(\cos\omega t + \dfrac{\gamma}{\omega}\sin\omega t\right)\right]$.

22. К синусоидальному напряжению $\mathscr{E} = \mathscr{E}_0\cos(\omega t + \delta)$ в момент времени $t = 0$ подключаются последовательно соединенные сопротивление R и индуктивность L. Найти силу тока I в цепи в зависимости от времени. При

каком условии после замыкания цепи в ней сразу установятся синусоидальные колебания?

Ответ.

$$I = \frac{\mathscr{E}_0}{R^2+\omega^2L^2}\,[R\cos(\omega t+\delta)+\omega L\sin(\omega t+\delta)-e^{-Rt/L}(R\cos\delta+\omega L\sin\delta)].$$

При условии $\operatorname{tg}\delta = -R/\omega L$ установятся синусоидальные колебания.

23. В цепь переменного тока с напряжением $\mathscr{E} = 440$ В и частотой $\nu = 50$ Гц включены последовательно нормально горящая лампочка накаливания и конденсатор. Чему равна емкость конденсатора C, если лампочка рассчитана на напряжение 220 В и силу тока 1 А? Чему равен сдвиг фаз между током и полным напряжением в цепи?

Ответ. $C = 1/(2\sqrt{3}\pi\nu R) = 8{,}4$ мкФ. Ток опережает напряжение по фазе на $60°$.

24. В цепь переменного тока с напряжением $\mathscr{E}_{\text{эфф}} = 440$ В и частотой $\nu = 50$ Гц включены последовательно нормально горящая лампочка накаливания и катушка самоиндукции. Лампочка рассчитана на 110 В и 1 А. При замене лампочки другой, рассчитанной на 220 В и 0,8 А, оказалось, что новая лампочка горит также нормальным накалом. Найти сопротивление R и самоиндукцию L катушки.

Ответ. $R = \dfrac{3}{2}\,\dfrac{R_2^2-5R_1^2}{R_2-R_1} = 137$ Ом, $L = \dfrac{1}{2\pi\nu}\sqrt{16R_1^2-(R_1+R_2)^2} =$ $= 1{,}16$ Г, где $R_1 = 110$ Ом и $R_2 = 275$ Ом — сопротивления первой и второй лампочек соответственно.

25. Два одинаковых проволочных кольца радиуса r каждое расположены так, как указано на рис. 310. Расстояние l между центрами колец велико по сравнению с r. В кольце *1* поддерживается переменный ток $I = I_0\cos\omega t$. Найти величину и направление средней силы $\overline{F}$ взаимодействия между кольцами. Индуктивность одного кольца равна L, омическое сопротивление R. Исследовать два предельных случая: 1) $\omega L \gg R$; 2) $\omega L \ll R$.

Ответ. $\overline{F} = \dfrac{6\pi^4\omega^2Lr^8I_0^2}{R^2+\omega^2L^2}\,\dfrac{1}{l^7}$. Сила $\overline{F}$ — отталкивательная. Если $\omega L \gg R$, то

$$\overline{F} = \frac{6\pi^4r^8I_0^2}{L}\,\frac{1}{l^7}.$$

В другом предельном случае $\omega L \ll R$

$$\overline{F} = \frac{6\pi^4\omega^2Lr^8I_0^2}{R^2}\,\frac{1}{l^7}.$$

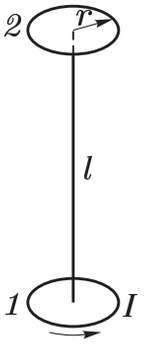


Рис. 310

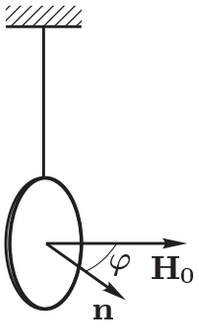


Рис. 311

26. Металлическое проволочное кольцо площади S с омическим сопротивлением R и индуктивностью L подвешено в горизонтальном однородном магнитном поле $\mathbf{H} = \mathbf{H}_0\cos\omega t$ и удерживается в нем таким образом, что угол между вектором $\mathbf{H}_0$ и нормалью $\mathbf{n}$ к плоскости кольца равен φ (рис. 311). Определить средний момент сил $\overline{\mathbf{M}}$, действующий на кольцо со стороны магнитного поля. Найти положения равновесия кольца и исследовать их устойчивость. Рассмотреть два предельных случая: 1) $\omega L \gg R$; 2) $\omega L \ll R$. В каком случае при одинаковых L вращающий момент меньше?

Ответ. $\overline{\mathbf{M}} = \dfrac{\omega^2 S^2 L H_0 \cos\varphi}{2(R^2+\omega^2 L^2)}\,[\mathbf{H}_0\mathbf{n}]$. Возможны два положения равновесия: а) плоскость кольца перпендикулярна к магнитному полю (неустойчивое равновесие); б) плоскость кольца параллельна магнитному полю (устойчивое равновесие): при $\omega L \gg R$

$$\overline{M} \equiv \overline{M}_1 = \frac{H_0^2 S^2}{2L}\sin\varphi\cos\varphi;$$

при $\omega L \ll R$,

$$\overline{M} \equiv \overline{M}_2 = \frac{\omega^2 H_0^2 S^2 L}{2R^2}\sin\varphi\cos\varphi = \overline{M}_1\left(\frac{\omega L}{R}\right)^2.$$

Во втором случае вращающий момент меньше.

§ 133. Автоколебания. Ламповый генератор

1. *Автоколебаниями называются вынужденные незатухающие колебания в реальных системах, период и амплитуда которых не зависят от характера внешнего воздействия, а определяются свойствами самой автоколебательной системы.* Автоколебания поддерживаются за счет поступления энергии из внешнего источника, причем количество поступающей энергии регулируется самой системой. В отличие от незатухающих собственных колебаний гармонического осциллятора и аналогичных систем, амплитуды которых определяются начальными условиями, *амплитуды автоколебаний от начальных условий не зависят.* Собственные незатухающие колебания относятся к идеализированному типу колебаний, который в реальных системах никогда не реализуется точно, — реальные собственные колебания всегда затухают. Напротив, *автоколебания в реальных системах могут продолжаться сколь угодно долго, пока не израсходуется энергия источника, поддерживающего эти колебания.*

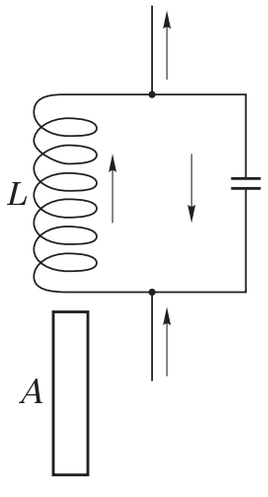


Рис. 312

Примерами автоколебательных систем могут служить часы, духовые и смычковые музыкальные инструменты и пр. В часах автоколебания возбуждаются и поддерживаются силой закрученной пружины или весом гири, в духовых музыкальных инструментах — воздушной струей, в скрипке — силами трения, возникающими при равномерном движении смычка, и т. д. Гудение телеграфных проводов под действием ветра, «пение» водопроводных труб, звучание человеческого голоса или трель соловья — все это примеры автоколебаний, вызываемых непериодическими силами.

Автоколебания могут возбуждаться и поддерживаться также и периодическими силами. Однако период последних не имеет никакого отношения к периоду возбуждаемых автоколебаний. Хорошим примером может служить демонстрационный опыт, предложенный и осуществленный М. И. Маклаковым,

лекционным ассистентом Московского физико-технического института. Катушка L колебательного контура помещается над столом в вертикальном положении (рис. 312). Снизу в нее входит железная трубка A, нижний конец которой стоит на столе. В трубке сделана вертикальная прорезь для уменьшения токов Фуко. Параметры колебательного контура подбираются так, чтобы его собственная частота совпадала с частотой переменного городского тока (50 герц). После включения тока через некоторое время наступает резонанс токов, и железная трубка втягивается в катушку. Индуктивность катушки увеличивается, колебательный контур выходит из резонанса, а амплитуда колебаний тока в катушке уменьшается. Благодаря этому трубка возвращается в исходное положение под действием силы тяжести. После этого колебания тока в колебательном контуре начинают нарастать, и снова наступает резонанс. Трубка опять втягивается в катушку. Трубка совершает автоколебания, т. е. периодические движения вверх и вниз, и при этом громко стучит по столу, подобно молотку. Период этих механических автоколебаний в десятки раз превосходит период переменного тока, поддерживающего их. Прибор получил название *молотка Маклакова.* Вместо параллельного включения катушку самоиндукции и конденсатор можно включить также последовательно.

Строгая теория автоколебаний весьма сложна. Это связано с тем, что автоколебания *нелинейны*, т. е. описываются нелинейными уравнениями. *Принцип суперпозиции в этих случаях не имеет места*, что затрудняет получение и исследование решений самих уравнений.

2. Важнейшей автоколебательной системой является *ламповый генератор*, нашедший широкое применение в радиотехнике. Простейшая схема такого генератора представлена на рис. 313. В цепь сетки лампы включен колебательный контур, в котором возбуждаются автоколебания. В цепь анода введена катушка, индуктивно связанная с катушкой колебательного контура. Она называется *катушкой обратной связи.* В использовании такой катушки заключена основная идея генератора. При наличии катушки обратной связи ток в колебательном контуре может индукционно воздействовать на ток в анодной цепи, и обратно. В реальных генераторах применяются *многосеточные лампы.* Однако для выяснения принципа действия генератора мы ограничимся простейшим случаем, когда в качестве лампы взят *триод*, имеющий всего одну сетку.

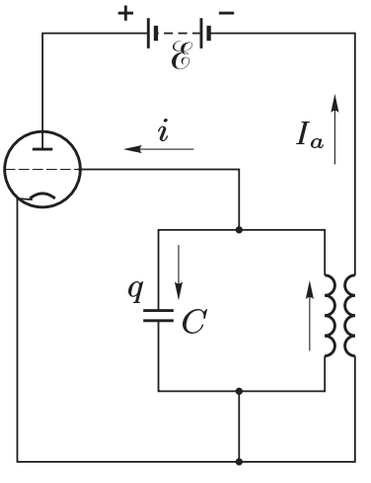


Рис. 313

В реальных условиях в колебательном контуре всегда совершаются колебания, возникающие в результате внешних влияний или тепловых флуктуаций. Обозначим через I ток в колебательном контуре, а через I_a — в анодной цепи. Некоторая часть тока i из колебательного контура ответвляется и идет на сетку.

Однако ток i мал и им можно пренебречь. Дело в том, что емкость конденсатора C очень велика по сравнению с емкостью между нитью и сеткой лампы. Поэтому емкостное сопротивление конденсатора много меньше сопротивления, оказываемого лампой, присоединенной параллельно конденсатору. В этих условиях основной ток устремляется через конденсатор. Однако, несмотря на свою малость, сеточный ток i играет принципиальную роль, так как он меняет заряд и потенциал сетки. Анодный ток I_a можно представить в виде функции сеточного V_g и анодного V_a напряжений. Так как сетка расположена значительно ближе к катоду, то изменение потенциала сетки значительно сильнее сказывается на силе анодного тока, чем равное изменение потенциала анода. Влиянием изменений анодного напряжения на анодный ток I_a можно пренебречь. В этом приближении

$$\frac{dI_a}{dt} = \left(\frac{dI_a}{dV_g}\right)_{V_a} \frac{dV_g}{dt} = S\frac{dV_g}{dt},$$

где S — крутизна сеточной характеристики. При изменении анодного тока в колебательном контуре индуцируется ЭДС

$$\mathscr{E}^{\text{инд}} = -M\frac{dI_a}{dt} = -MS\frac{dV_g}{dt}, \tag{133.1}$$

где M — коэффициент взаимной индукции между катушкой колебательного контура и катушкой обратной связи. Напряжение на сетке равно напряжению на конденсаторе, т. е. $V_g = q/C$, где q — заряд на верхней пластине конденсатора (при положительном обходе по контуру конденсатор проходится сверху вниз). Таким образом, $dV_g/dt = \dot{q}/C = I/C$, и, следовательно,

$$\mathscr{E}^{\text{инд}} = -\frac{MS}{C}\, I. \tag{133.2}$$

Если $M < 0$, то $\mathscr{E}^{\text{инд}}$ и I имеют одинаковые знаки. В этом случае при отсутствии сопротивления колебания в контуре будут нарастать — произойдет *самовозбуждение колебаний. Если же $M > 0$, то под действием электродвижущей силы $\mathscr{E}^{\text{инд}}$ колебания в контуре прекратятся.* Пересоединив концы катушки обратной связи или повернув на 180° саму катушку, можно изменить знак M и таким образом перейти от одного случая к другому.

Для получения более точного условия самовозбуждения колебательного контура надо учесть его омическое сопротивление R. При наличии электродвижущей силы $\mathscr{E}^{\text{инд}}$ уравнение колебаний в колебательном контуре имеет вид

$$L\ddot{q} + R\dot{q} + \frac{q}{C} = -\frac{MS}{C}\,\dot{q},$$

где L — индуктивность колебательного контура. Разделив на L и введя обозначения

$$\omega_0^2 = \frac{1}{LC}, \quad 2\delta = \frac{R}{L} + \frac{MS}{LC}, \tag{133.3}$$

получим

$$\ddot{q} + 2\delta\dot{q} + \omega_0^2 q = 0. \tag{133.4}$$

Допустим, что лампа работает на прямолинейном участке сеточной характеристики. Тогда S и δ можно считать величинами постоянными. В этом случае формально уравнение (133.4) описывает свободные колебания в контуре:

$$q = q_0 e^{-\delta t}(A\cos\omega t + B\sin\omega t), \tag{133.5}$$

где A и B — постоянные, а $\omega^2 = \omega_0^2 - \delta^2$. Однако эти колебания могут не только затухать ($\delta > 0$), но и нарастать ($\delta < 0$). Условие нарастания колебаний $\delta < 0$ можно привести к виду

$$M < -CR/S. \tag{133.6}$$

Это и есть *условие самовозбуждения колебаний.* Из него видно, что *для самовозбуждения колебаний необходимо выполнение неравенства* $M < 0$, которое было получено выше для частного случая $R = 0$.

3. Изложенная теория «линейна», т. е. она основана на линейных дифференциальных уравнениях. *Она не полна и годится только на начальной стадии процесса*, когда колебания малы и поэтому с достаточным приближением крутизна характеристики S может считаться постоянной. Линейная теория приводит к правильному условию самовозбуждения колебаний (133.6), но она неприменима на более поздних стадиях процесса, когда амплитуда колебаний сделается достаточно большой. В частности, она не может ответить на вопрос, каковы будут установившиеся колебания. Если $\delta < 0$, то по линейной теории амплитуда колебаний должна возрастать неограниченно, а при $\delta = 0$ колебания должны быть незатухающими с амплитудой и фазой, *определяемыми начальными условиями.* Эти выводы не согласуются с опытом. Причина расхождения состоит в том, что при сильных колебаниях крутизна характеристики S даже приближенно перестает быть постоянной, а потому уравнения, описывающие колебания, становятся *нелинейными. Полная теория автоколебаний вообще и колебаний в ламповом генераторе в частности должна исходить из нелинейных уравнений, справедливых на всех стадиях процесса.*

Запишем уравнение колебаний в колебательном контуре в виде

$$L\dot{I} + RI + \frac{q}{C} = -M\frac{dI_a}{dt}. \tag{133.7}$$

Для упрощения предположим, что анодный ток I_a зависит только от знака сеточного напряжения V_g. Когда $V_g > 0$, через лампу идет *максимально возможный ток*, равный *току насыщения* I_s. Когда $V_g < 0$, анодный ток I_a равен нулю. Соответствующая идеализированная сеточная характеристика представлена на рис. 314 жирной ломаной линией (ср. с рис. 227). В этом случае уравнение (133.7) становится нелинейным. Введенная схематизация, конечно, не годится в точных расчетах. Но она достаточна, чтобы *качественно передать* основные особенности автоколебаний. На рис. 315 изображена кривая сеточного

напряжения V_g. Под ней приведена ступенчатая кривая $I_a(t)$ для анодного тока I_a. Производная $\dot{I}_a$ всюду равна нулю, за исключением отдельных точек, в которых ток I_a испытывает скачки. Формально

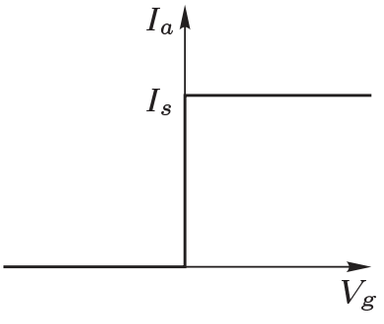


Рис. 314

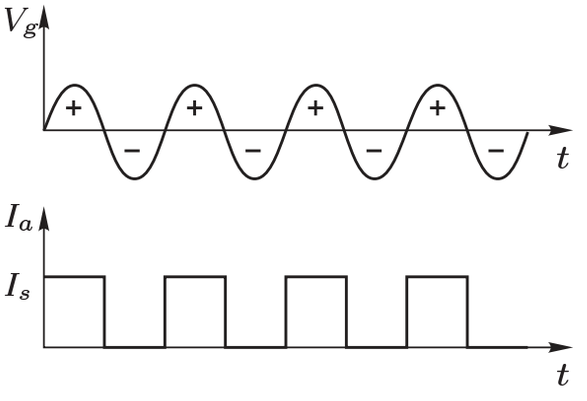


Рис. 315

в этих точках производная $\dot{I}_a$ обращается в бесконечность. В действительности функция $I_a(t)$ всюду непрерывна и дифференцируема. Точки разрыва должны быть заменены малыми интервалами времени τ, в течение которых происходят резкие изменения анодного тока на величину $\Delta I_a = \pm I_s$, а производная $\dot{I}_a$ по абсолютной величине очень велика. Такие изменения вызывают скачки тока ΔI и в колебательном контуре. Для нахождения ΔI проинтегрируем выражение (133.7) по одному из интервалов времени τ. Поскольку величины I и q остаются конечными, при таком интегрировании они в пределе не вносят никакого вклада, и мы получаем

$$L\Delta I = -M\Delta I_a. \tag{133.8}$$

Скачки ΔI аналогичны кратковременным толчкам, действующим на колебательную систему. Если $M < 0$, то такие толчки усиливают колебания; если $M > 0$, то ослабляют. Между двумя последовательными толчками система совершает затухающие свободные колебания. Допустим, что $M < 0$. Пусть перед некоторым толчком амплитуда колебаний тока в колебательном контуре равна I_0. После толчка она получает приращение $|\Delta I| = |M\Delta I_a/L| = |MI_s/L|$ и становится равной $I_1 = I_0 + |MI_s/L|$. К следующему толчку колебания приходят с амплитудой

$$I_2 = I_1 e^{-\gamma T/2} = \left(I_0 + \left|\frac{MI_s}{L}\right|\right) e^{-\gamma T/2},$$

где $\gamma = R/(2L)$, а T — период колебания. Колебания будут нарастать, если $I_2 > I_0$. Амплитуда колебаний установится, когда $I_2 = I_0$. Таким образом, амплитуда установившихся колебаний будет

$$I_0 = \frac{|MI_s/L|}{e^{\gamma T/2} - 1}. \tag{133.9}$$

Характер установившихся колебаний (т. е. автоколебаний) представлен на рис. 316. Благодаря наличию скачков и затуханию колебания *не*

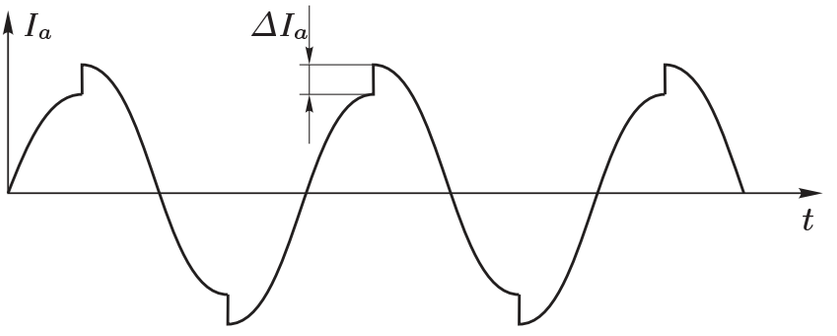


Рис. 316

гармонические. Допустим, однако, что добротность Q колебательного контура очень велика, т. е. $\gamma T \ll 1$. Тогда скачки тока I станут малозаметными, а автоколебания — *практически гармоническими* с амплитудой

$$I_0 = \frac{2|MI_s/L|}{\gamma T} = \frac{2Q}{\pi}\left|\frac{M}{L}\,I_s\right|. \qquad (133.10)$$

Конечно, в реальных ламповых генераторах, характеристики которых — плавные кривые, никаких резких скачков не возникает, *колебания непрерывны, хотя и не совсем синусоидальны.*

4. В настоящее время для генерации синусоидальных и несинусоидальных колебаний и их усиления, наряду с электронными лампами, успешно применяются *полупроводниковые приборы*, обладающие рядом преимуществ по сравнению с электронными лампами. Полупроводниковые приборы не имеют накаливаемого катода и потому потребляют относительно меньшую мощность. Они не требуют вакуума, который может портиться при работе лампы. По этой причине они более долговечны и надежны в работе, чем электронные лампы. Благодаря своим малым габаритам полупроводниковые приборы незаменимы во многих радиотехнических схемах, например в счетно-решающих устройствах. Однако полупроводниковые приборы обладают и рядом недостатков (изменение свойств полупроводников с температурой, усложнения схем из-за малости входных сопротивлений, шумы, вызываемые случайными изменениями путей токов внутри сопротивлений). По этой причине, когда габариты и вес не являются определяющими, в технике эксперимента пока отдается предпочтение схемам на электровакуумных и газонаполненных приборах. Не останавливаясь на всех этих вопросах, относящихся к области радиотехники, ограничимся краткими замечаниями.

Для усиления и генерации колебаний применяются полупроводниковые приборы не с двумя (как в выпрямителях), а *с тремя и более электродами.* Рассмотрим *полупроводниковый триод* (*транзистор*), содержащий три электрода. Он представляет собой кристалл германия или кремния, в котором введением донорной и акцепторной примесей созданы три области с чередующимися типами проводимости: элек-

тронной и дырочной. Существуют два типа транзисторов. В транзисторах n–p–n-типа (рис. 317 *а*) крайние области обладают *электронной*, а средняя (*база*, или *основание*) — *дырочной проводимостью.* В транзи-

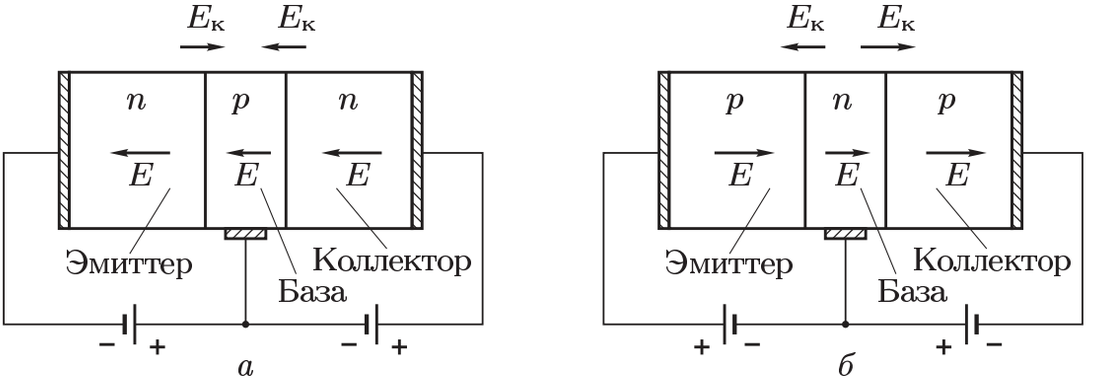


Рис. 317

сторах p–n–p-типа (рис. 317 *б*), наоборот, электронной проводимостью обладает *база*, а дырочной — *крайние области.* Кристалл снабжен соответствующими металлическими электродами, с помощью которых транзистор включается в цепь.

Контактное поле $\mathbf{E}_к$ в обоих n–p-переходах направлено *от электронного к дырочному полупроводнику* (см. § 108). Направление $\mathbf{E}_к$ — *запорное*, противоположное направление — *пропускное.* Включим транзистор в схему согласно рис. 317 *а.* Половина, включенная в проходном направлении, называется *эмиттером*, а включенная в запорном направлении — *коллектором.* Ширина базы, разделяющей эти половины, всегда мала и измеряется десятками (или даже единицами) микрометров. Электрический ток внутри эмиттера (рис. 317 *а*) создается главным образом движением *электронов*, являющихся *основными носителями заряда.* Эти электроны проходят через левый n–p-переход в область базы и там под действием электрического поля $\mathbf{E}$ движутся по направлению к коллектору. Толщина базы должна быть такова, чтобы значительная часть электронов прошла через нее. Пройдя через правый p–n-переход, электроны попадают в коллектор уже в качестве *основных носителей заряда.* Тем самым они меняют ток в коллекторе. Те же рассуждения относятся и к рис. 317 *б* (только роль электронов будут выполнять положительные дырки). Таким образом, всякое изменение тока в цепи эмиттера будет вызывать изменение тока и в цепи коллектора. В этом отношении транзистор действует аналогично электронной лампе. Роль катода играет эмиттер, анода — коллектор, сетки — база.

§ 134. Релаксационные колебания

На рис. 318 *а* представлена характеристика неоновой лампы (см. § 117). Это — *нелинейная характеристика.* Если повышать напряжение на лампе V, то при $V = V_2$ она вспыхивает и начинает светиться красным светом. При дальнейшем повышении напряжения ток в лампе возрастает вдоль кривой

AB. Если уменьшать напряжение на лампе в обратном порядке, то она гаснет при другом напряжении $V_1 < V_2$. Величины V_2 и V_1 называются *потенциалами зажигания* и *погасания* соответственно.

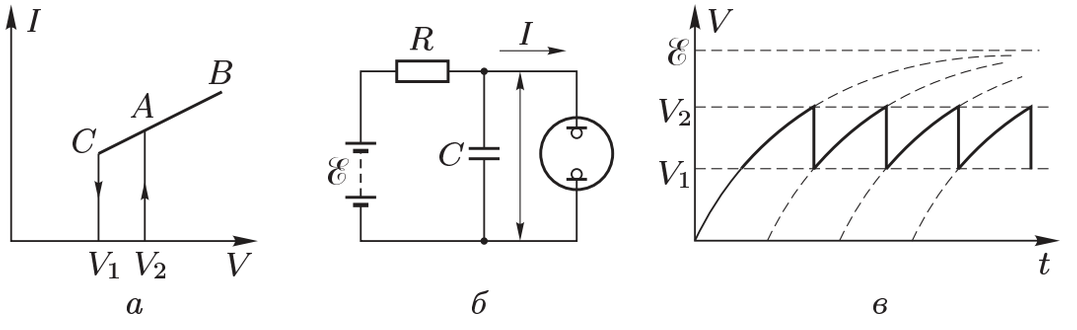


Рис. 318

Включим неоновую лампу в цепь, изображенную на рис. 318 *б*. Сопротивление R должно быть очень велико, а электродвижущая сила батареи $\mathscr{E}$ — больше V_2. Если замкнуть цепь, то конденсатор C начнет заряжаться. Напряжение на нем (равное напряжению на неоновой лампе V) будет возрастать по закону

$$V = \mathscr{E}(1 - e^{-t/\tau}),$$

где $\tau = RC$ (см. § 48). Когда напряжение V достигнет значения V_2, лампа загорится и начнет пропускать ток. Ее сопротивление практически обратится в нуль. Поэтому конденсатор очень быстро (почти мгновенно) начнет разряжаться через лампу. Однако когда напряжение на конденсаторе упадет до V_1, лампа погаснет и перестанет пропускать ток. С этого момента снова начнется зарядка конденсатора, пока потенциал V не достигнет значения V_2. Тогда лампа опять загорится и начнется новая разрядка конденсатора. И этот процесс будет продолжаться периодически с периодом

$$T = \tau \ln \frac{\mathscr{E} - V_1}{\mathscr{E} - V_2}.$$

Зависимость напряжения V от времени представлена на рис. 318 *в* жирной пилообразной кривой. Если период колебаний T порядка секунды или больше, то будут видны кратковременные вспышки света, разделенные более продолжительными паузами. Уменьшая R или C, можно получить период T гораздо меньше, и глаз уже не будет различать отдельных вспышек. В рассматриваемом случае автоколебания возникают потому, что существует

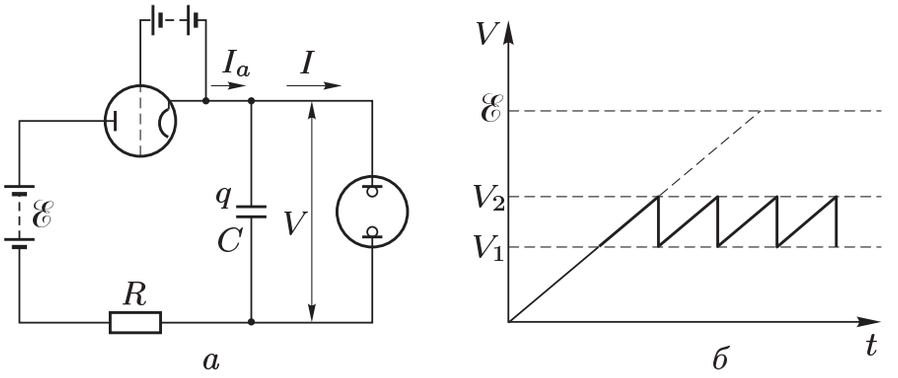


Рис. 319

определенное *время успокоения* (или *время релаксации*) контура $\tau = RC$. По этой причине колебания рассматриваемого типа получили название *релаксационных колебаний.*

Зубцы на пилообразной кривой рис. 318 *в* не прямые. Для многих целей, в особенности в электронных осциллографах, требуются пилообразные напряжения с *прямолинейными зубцами* — напряжение в пределах каждого зубца должно меняться во времени по линейному закону. Этого можно достигнуть, если ввести в цепь, помимо неоновой лампы, *второй нелинейный элемент* — электронную лампу (триод или лучше пентод), как указано на рис. 319 *а*. Через лампу потечет анодный ток $\dot{I}_a = \dot{q}$, практически не зависящий от анодного напряжения. Поэтому во время зарядки заряд на конденсаторе будет меняться во времени по линейному закону: $q = I_a t + \text{const}$. По линейному закону будет меняться и напряжение на конденсаторе C (равное напряжению на неоновой лампе). Поэтому вместо кривой рис. 318 *в* получится такая же кривая, но с прямыми зубцами (рис. 319 *б*).

§ 135. Параметрическое возбуждение колебаний

1. Допустим, что с помощью надлежащего приспособления (например, электрического мотора) индуктивность L или емкость C колебательного контура (или то и другое) периодически меняются во времени. Свободные колебания такой системы описываются уравнением

$$\frac{d\Phi}{dt} + RI + \frac{q}{C} = 0, \tag{135.1}$$

или

$$\frac{d}{dt}\left(L\frac{dq}{dt}\right) + R\frac{dq}{dt} + \frac{q}{C} = 0 \tag{135.2}$$

(см. § 122). При постоянном R это — *линейное дифференциальное уравнение с периодическими коэффициентами*, переходящее в нелинейное, когда сопротивление R зависит от тока I. Аналогичным уравнением описывается и движение механической системы — *качелей.* Качающийся на качелях, приседая и распрямляясь, периодически поднимает и опускает центр масс своего тела и тем самым меняет параметры системы. При определенных условиях все рассмотренные системы становятся *неустойчивыми* — случайно возникшее отклонение от состояния равновесия приводит в них к возникновению и нарастанию колебаний. Это явление, поскольку оно вызывается изменениями параметров системы, называется *параметрическим возбуждением колебаний*, а сами колебания — *параметрическими.*

Нахождение условий возбуждения параметрических колебаний сводится к исследованию решений линейных дифференциальных уравнений с периодическими коэффициентами. Решение таких уравнений представляет, вообще говоря, очень трудную математическую задачу. Найти решение в конечной аналитической форме обычно не удается. К тому же линейные уравнения позволяют получить *только условие возбуждения колебаний*, но не позволяют решить вопрос об *установлении их стационарной амплитуды*, так как при достаточно больших амплитудах дифференциальные уравнения, описывающие колебания,

становятся *существенно нелинейными.* Мы рассмотрим только возбуждение параметрических колебаний и ограничимся при этом простейшим случаем, когда параметры системы изменяются *скачкообразно*, а в промежутках между этими скачками остаются постоянными. Можно, например, через равные промежутки времени очень быстро раздвигать и сближать пластины плоского конденсатора или растягивать и сжимать спираль, служащую «катушкой самоиндукции» колебательного контура, меняя тем самым скачкообразно величины C и L. «Очень быстро» или «скачкообразно» означает, что за время изменения параметров Δt заряд конденсатора q практически не успевает измениться. Отсюда следует, что за то же время Δt практически не изменится и магнитный поток Φ через катушку самоиндукции. В самом деле, проинтегрировав почленно уравнение (135.1) по промежутку времени Δt, получим

$$\int\limits_t^{t+\Delta t} \frac{d\Phi}{dt}\, dt + \int\limits_t^{t+\Delta t} RI\, dt + \int\limits_t^{t+\Delta t} \frac{q}{C}\, dt = 0,$$

или при постоянном сопротивлении R

$$\Delta\Phi + R\Delta q + \int \frac{q}{C}\, dt = 0.$$

Так как по предположению изменение заряда Δq за время Δt пренебрежимо мало, а заряд q во время изменения остается конечным, то при $\Delta t \to 0$ из последнего соотношения следует $\Delta\Phi = \text{const}$. Результат остается верным и в том случае, когда R зависит от силы тока.

2. После этих замечаний возьмем колебательный контур и будем через определенные промежутки времени скачкообразно изменять его индуктивность L, оставляя емкость C неизменной. Таким образом, индуктивность будет принимать два значения, большее из которых обозначим через L_1, а меньшее через L_2. Соответствующие значения собственной частоты колебательного контура обозначим через $\omega_1 = 1/\sqrt{L_1 C}$, $\omega_2 = 1/\sqrt{L_2 C}$, а периоды собственных колебаний — через T_1 и T_2. Для простоты будем считать, что омическое сопротивление контура равно нулю. В контуре всегда текут токи, вызванные случайными внешними наводками или тепловыми флуктуациями. Пусть сначала $L = L_1$. В момент, когда ток в катушке максимален и равен I_{10}, а заряд конденсатора обращается в нуль, скачкообразно уменьшим индуктивность от L_1 до L_2. Так как магнитный поток при этом останется неизменным, то ток возрастет до $I_{20} = (L_1/L_2)I_{10}$. С этого момента начнутся свободные колебания тока $I = I_{20}\cos\omega_2 t$. Через время $t = T_2/4$, когда I обратится в нуль, увеличим L до прежнего значения L_1. Так как во время изменения индуктивности ток через катушку не течет, то амплитуда колебаний не изменится, а изменится только их частота. Колебания тока будут описываться уравнением $I = I_{20}\sin\omega_1 t$. (Условимся время t отсчитывать всякий раз от момента последнего скачкообразного изменения индуктивности.) В момент $t = T_1/4$, когда

ток достигнет максимального значения I_{10}, снова уменьшим индуктивность от L_1 до L_2. В результате амплитуда колебаний сделается равной $I_{30} = (L_1/L_2)I_{20} = (L_1/L_2)^2 I_{10}$. И так будем поступать дальше, уменьшая L всякий раз, когда ток максимален, и увеличивая, когда он проходит через нуль. В результате амплитуда колебаний тока будет неограниченно возрастать в геометрической прогрессии

$$A_0 = A_{10}\left[1 + \frac{L_1}{L_2} + \left(\frac{L_1}{L_2}\right)^2 + \ldots\right],$$

несмотря на то, что в системе отсутствуют какие бы то ни было источники тока или напряжения. Такая раскачка колебаний называется *параметрическим резонансом.* Мы видим, что параметрический резонанс возникает, когда параметры системы меняются с частотой, *вдвое большей* собственной частоты этой системы. Но легко видеть, что параметрический резонанс можно получить также, если частоту изменения параметров уменьшить в 2, 3, ... раз. Однако в этом случае он будет выражен слабее.

Совершенно аналогичное явление получится и в том случае, когда с теми же периодами скачкообразно изменять *емкость конденсатора.* При таких изменениях остается постоянным *заряд конденсатора* q, а меняется напряжение на его обкладках $V = q/C$. В момент прохождения заряда через максимум надо уменьшить емкость C, повысив тем самым напряжение V. В момент же, когда q обращается в нуль, надо возвратить емкость C к ее исходному значению. В результате снова возникнет усиление колебаний напряжения V с амплитудой, возрастающей в геометрической прогрессии.

3. Все это легко понять и *с энергетической точки зрения.* Энергия конденсатора $W_e = q^2/(2C)$, а катушки самоиндукции $W_m = \Phi^2/(2L)$. Для возбуждения параметрических колебаний надо уменьшить C, когда заряд конденсатора максимален, и уменьшить L, когда в контуре максимален ток. При таких изменениях электрическая и магнитная энергии увеличиваются соответственно на $\Delta W_e = q^2\Delta(1/2C)$ и $\Delta W_m = \Phi^2\Delta(1/2L)$. Возвращать же C и L к их исходным значениям надо в те моменты, когда обращаются в нуль q и L, так как в этом случае электрическая и магнитная энергии остаются неизменными. Таким образом, в колебательную систему будет периодически вкладываться энергия, что и приводит к раскачке колебаний. Увеличение электрической энергии конденсатора при уменьшении его емкости легко понять на примере плоского конденсатора. Между пластинами конденсатора действуют силы кулоновского притяжения. При раздвижении пластин емкость конденсатора уменьшается и одновременно совершается работа против этих сил. Она-то и идет на увеличение энергии конденсатора. Аналогичное явление имеет место и в случае проволочной спирали, по

которой течет ток. Растягивая спираль, мы уменьшаем ее индуктивность L и одновременно совершаем работу против амперовых сил притяжения между ее витками. В результате совершается положительная внешняя работа, и магнитная энергия тока увеличивается.

Аналогичные рассуждения применимы и к раскачке качелей. Для пояснения возьмем математический маятник, колеблющийся на нити, верхний конец которой пропущен через малое отверстие. Будем втягивать нить, когда маятник проходит через нижнее положение, и настолько же выпускать ее, когда он проходит через крайние положения. В первом случае мы совершаем над маятником положительную работу, сообщая маятнику энергию, а во втором — отрицательную работу, отбирая от него энергию обратно. Однако положительная работа по абсолютной величине превосходит отрицательную. Действительно, натяжение нити максимально, когда колеблющийся маятник проходит через среднее положение, так как это натяжение должно не только уравновесить вес маятника, но и сообщить ему ускорение. Напротив, в крайних положениях натяжение нити минимально, так как здесь оно должно уравновесить только составляющую силы веса вдоль направления нити. Поэтому при равных перемещениях вдоль нити сила натяжения в среднем положении совершит большую работу, чем в крайнем положении. В результате в среднем положении маятник будет получать больше энергии, чем возвращать в крайнем. Поэтому произойдет параметрическая раскачка колебаний. Человек, качающийся на качелях, подобен маятнику: он приседает в крайних положениях и выпрямляется в среднем.

4. Наличие омического сопротивления, пока оно остается постоянным, не вносит никаких затруднений. Надо только учесть, что теперь свободные колебания системы между моментами изменения ее параметров происходят с затуханием. Если выполнено условие, что потери энергии, связанные с этим затуханием, меньше энергии, вкладываемой в систему за тот же промежуток времени, то опять будет происходить раскачка колебаний с амплитудой, возрастающей в геометрической прогрессии. Таким образом, если бы система все время подчинялась линейному дифференциальному уравнению, то при соблюдении указанного условия амплитуда колебаний непрерывно возрастала бы до тех пор, пока не произошел бы «пробой» конденсатора или изоляции подводящих проводов. *Возникновение параметрических колебаний с установившейся амплитудой теория, основанная на линейных дифференциальных уравнениях с периодическими коэффициентами, объяснить не может.* Для получения колебаний со стационарной амплитудой в систему приходится вводить проводники с *нелинейной характеристикой*, например катушку с железным сердечником, лампы накаливания и пр. Но тогда дифференциальные уравнения, описывающие систему, становятся *нелинейными.* Из этих уравнений можно получить не только условие возникновения стационарных установившихся колебаний, но и найти их амплитуду.

5. Все изложенное подтверждается опытом. На рис. 320 изображена *емкостная параметрическая машина Л. И. Мандельштама и Н. Д. Папалекси.* Конденсатор машины состоял из двух систем обкладок — неподвижной (статор) и подвижной (ротор). Статор был изготовлен из 26 квадратных алюминиевых пластин с симметрично расположенными радиальными вырезами, а ротор — из 25 таких же пластин круглой формы с аналогичными вырезами. С помощью мотора ротор можно было приводить во вращение со скоростью до 4000 оборотов в минуту. При этом периодически менялась емкость конденсатора и возбуждались параметрические колебания тока. Для того чтобы сделать систему нелинейной, параллельно конденсатору включалась цепочка из 6 неоновых ламп. При наличии последних на конденсаторе получалось устойчивое напряжение, достигавшее 600–700 В. В отсутствие неоновых ламп напряжение не устанавливалось, а продолжало нарастать до 2000–3000 В, пока не проскакивали искры между обкладками конденсатора. Аналогичные опыты можно произвести и с «индукционной машиной», в которой периодически меняется индуктивность контура.

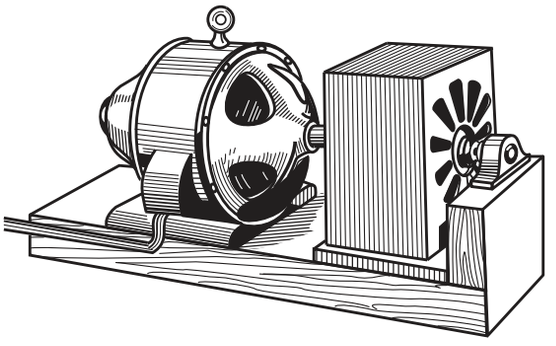


Рис. 320

§ 136. Трансформатор

1. Трансформатор состоит из двух обмоток — *первичной* и *вторичной*, навитых на общий железный сердечник (рис. 321). Уравнения колебаний в такой системе записываются в виде

$$R_1 I_1 = \mathscr{E} - \dot{\Phi}_1, \quad R_2 I_2 = -\dot{\Phi}_2, \tag{136.1}$$

где индексом 1 обозначены величины, относящиеся к первичной, а индексом 2 — к вторичной обмоткам. Для простоты пренебрежем рассеянием магнитного потока, проходящего через железный сердечник трансформатора. В этом предположении

$$\frac{\Phi_1}{\Phi_2} = \frac{n_1}{n_2}, \tag{136.2}$$

где n_1 и n_2 — числа витков в первичной и вторичной обмотках. Записав это соотношение в виде $n_1\Phi_2 = n_2\Phi_1$ и продифференцировав по времени, убеждаемся, что для производных магнитного потока справедливо такое же соотношение:

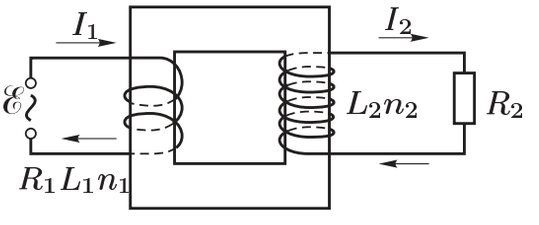


Рис. 321

$$\dot{\Phi}_1/\dot{\Phi}_2 = n_1/n_2.$$

Оно позволяет исключить из уравнений (136.1) магнитные потоки. Таким путем получаем

$$R_1 I_1 - \frac{n_1}{n_2} R_2 I_2 = \mathscr{E}. \tag{136.3}$$

Отсюда видно, что наличие вторичной обмотки меняет ток в первичной цепи. Однако уравнения (136.3) недостаточно для определения двух неизвестных

I_1 и I_2. Для получения недостающего уравнения введем упрощающее предположение, что трансформатор *идеальный*, т. е. не обладает ферромагнетизмом. (В реальных трансформаторах, конечно, это не так.) Тогда связь между магнитными потоками и токами будет линейной:

$$\Phi_1 = L_1 I_1 + L_{12} I_2, \quad \Phi_2 = L_{21} I_1 + L_2 I_2,$$

где L_1 — индуктивность первичной обмотки, L_2 — вторичной, а $L_{12} = L_{21}$ — коэффициент взаимной индукции этих обмоток. Ввиду (136.2) *при любых токах* I_1 и I_2 соблюдается соотношение

$$n_2(L_1 I_1 + L_{12} I_2) = n_1(L_{21} I_1 + L_2 I_2).$$

Приравнивая коэффициенты при I_1 и I_2, из него находим

$$n_1 L_{21} = n_2 L_1, \quad n_2 L_{12} = n_1 L_2, \tag{136.4}$$

а потому

$$L_{12} L_{21} = L_1 L_2. \tag{136.5}$$

Если воспользоваться еще теоремой взаимности ($L_{12} = L_{2l}$), то получится

$$L_{12} = L_{21} = \sqrt{L_1 L_2}. \tag{136.6}$$

Теперь система уравнений (136.1) принимает вид

$$\begin{aligned} R_1 I_1 &= \mathscr{E} - L_1 \dot{I}_1 - \sqrt{L_1 L_2}\, \dot{I}_2, \\ R_2 I_2 &= -\sqrt{L_1 L_2}\, \dot{I}_1 - L_2 \dot{I}_2. \end{aligned} \tag{136.7}$$

Предположим, далее, что электродвижущая сила $\mathscr{E}$ меняется во времени синусоидально: $\mathscr{E} \sim e^{i\omega t}$. Тогда для установившихся колебаний получится

$$\begin{aligned} (R_1 + i\omega L_1) I_1 + i\omega \sqrt{L_1 L_2}\, I_2 &= \mathscr{E}, \\ i\omega \sqrt{L_1 L_2}\, I_1 + (R_2 + i\omega L_2) I_2 &= 0. \end{aligned} \tag{136.8}$$

Отсюда

$$\begin{aligned} I_1 &= \frac{R_2 + i\omega L_2}{R_1 R_2 + i\omega(L_1 R_2 + L_2 R_1)}\, \mathscr{E}, \\ I_2 &= \frac{-i\omega \sqrt{L_1 L_2}}{R_1 R_2 + i\omega(L_1 R_2 + L_2 R_1)}\, \mathscr{E}. \end{aligned} \tag{136.9}$$

Эти формулы и решают задачу о трансформаторе. В практически важном случае омическое сопротивление первичной цепи R_1 мало по сравнению с индуктивным ωL_1. Пренебрегая им, получим

$$I_1 = \frac{R_2 + i\omega L_2}{i\omega L_1 R_2}\, \mathscr{E}, \quad I_2 = -\sqrt{\frac{L_2}{L_1}}\, \frac{\mathscr{E}}{R_2} = -\frac{n_2}{n_1}\, \frac{\mathscr{E}}{R_2}. \tag{136.10}$$

При выводе последнего соотношения было учтено, что индуктивность обмотки пропорциональна квадрату числа витков. Формула делает понятной основную идею трансформатора. Если бы сопротивление R_2 было непосредственно присоединено к источнику электродвижущей силы $\mathscr{E}$, то получился бы ток $\mathscr{E}/R_2$. Трансформатор увеличивает этот ток в n_2/n_1 раз или уменьшает в n_1/n_2 раз. Этот факт обычно выражают несколько иначе. Величина $V_2 = R_2 I_2$ дает падение напряжения на сопротивлении R_2. Ее называют напряжением во вторичной цепи. Из второй формулы (136.10) получаем

$$V_2 = -\frac{n_2}{n_1}\, \mathscr{E}. \tag{136.11}$$

Трансформатор повышает напряжение в n_2/n_1 раз или понижает в n_1/n_2 раз. С этим и связано соответствующее увеличение (или уменьшение) тока в сопротивлении R_2.

2. Что касается тока I_1, то при его рассмотрении удобнее обратиться к векторной диаграмме. Ограничимся случаем $R_1 = 0$. Тогда первое уравнение (136.8) запишется в виде

$$L_1 I_1 + \sqrt{L_1 L_2}\, I_2 = \mathscr{E}/i\omega.$$

Если $R_2 = \infty$, то $I_2 = 0$. В этом случае ток в первичной цепи называется *током холостого хода трансформатора.* Обозначим его через I_0. В нашем случае ($R_1 = 0$) $I_0 = \mathscr{E}/(i\omega L_1)$ и, следовательно,

$$L_1 I_1 + \sqrt{L_1 L_2}\, I_2 = L_1 I_0, \quad \text{или} \quad n_1 I_1 + n_2 I_2 = n_1 I_0. \tag{136.12}$$

Ток холостого хода $I_0 = -i\mathscr{E}/\omega L_1$ отстает по фазе от напряжения на $\pi/2$. Если на векторной диаграмме напряжение изобразить горизонтальным отрезком, направленным вправо, то ток холостого хода I_0 изобразится отрезком, направленным вниз (рис. 322). Ток I_2, как видно из (136.10), изобразится горизонтальным отрезком, направленным влево. Вместо самого тока удобнее откладывать ток, умноженный на соответствующее число витков (так называемые *ампер-витки*). Согласно (136.12) величина $n_1 I_1$ изобразится вектором, равным геометрической разности векторов $n_1 I_0$ и $n_2 I_2$. При увеличении нагрузки (т. е. уменьшении сопротивления R_2) ток I_2, как показывают формулы (136.10), растет. Вместе с ним растет и ток I_1. А так как ток холостого хода остается неизменным, то из рис. 322 следует, что должен уменьшаться сдвиг фаз φ между током I_1 и напряжением $\mathscr{E}$. Оба эти обстоятельства ведут к увеличению потребляемой мощности.

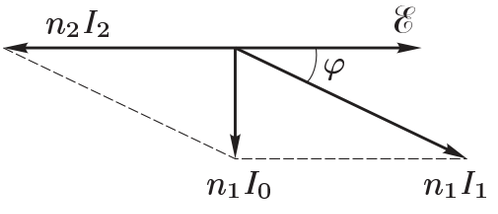


Рис. 322

3. Вычислим среднюю мощность электрической энергии P_0 в первичной цепи. Считая для общности сопротивление R_1 каким угодно, запишем ток I_1 в виде

$$I_1 = \frac{a + ib}{c + id}\,\mathscr{E}.$$

Значения постоянных a, b, c, d легко установить, сравнивая это выражение с первой формулой (136.9). Умножая числитель и знаменатель на $c - id$, получим

$$I_1 = \frac{(ac + bd) + i(bc - ad)}{c^2 + d^2}\,\mathscr{E}.$$

Отсюда для косинуса сдвига фаз между током I_1 и напряжением $\mathscr{E}$ получим

$$\cos\varphi = \frac{ac + bd}{|(ac + bd) + i(bc + ad)|},$$

а для средней мощности

$$P_0 = \frac{1}{2}\,|\mathscr{E} I_1| \cos\varphi = \frac{1}{2}\,\frac{ac + bd}{c^2 + d^2}\,|\mathscr{E}|^2.$$

Подставим сюда значения коэффициентов a, b, c, d. Кроме того, введем отношение индуктивного сопротивления цепи к соответствующему омическому сопротивлению:

$$\alpha_к = \omega L_к / R_к. \tag{136.13}$$

В результате найдем

$$P_0 = \frac{1}{2}\,\frac{1+\alpha_2(\alpha_1+\alpha_2)}{1+(\alpha_1+\alpha_2)^2}\,\frac{|\mathscr{E}|^2}{R_1}, \tag{136.14}$$

$$\operatorname{tg}\varphi = -\frac{\alpha_1}{1+\alpha_2(\alpha_1+\alpha_2)}. \tag{136.15}$$

Ток отстает по фазе от напряжения на угол $|\varphi|$.

Средняя мощность, *потребляемая в первичной цепи*, будет $P_1 = R_1|I_1|^2/2$, а во вторичной $P_2 = R_2|I_2|^2/2$. Вычисляя их, получим

$$P_1 = \frac{1+\alpha_2^2}{1+(\alpha_1+\alpha_2)^2}\,\frac{|\mathscr{E}|^2}{R_1}, \quad P_2 = \frac{1}{2}\,\frac{\alpha_1\alpha_2}{1+(\alpha_1+\alpha_2)^2}\,\frac{|\mathscr{E}|^2}{R_1}. \tag{136.16}$$

Таким образом, $P_0 = P_1 + P_2$, как это и должно быть. При неизменных параметрах первичной цепи ($\mathscr{E}$, R_1, α_1) мощность P_2, потребляемая во вторичной цепи, максимальна, когда $\alpha_2 = \sqrt{\alpha_1^2+1}$, или приближенно, когда $\alpha_2 = \alpha_1$, так как в практически важных случаях всегда $\alpha_1 \gg 1$. Максимальное значение P_2 дается приближенным выражением

$$P_{2\,\text{макс}} \approx \frac{1}{8}\,\frac{|\mathscr{E}|^2}{R_1}. \tag{136.17}$$

Мощность тока в первичной цепи при этом будет $P_0 = |\mathscr{E}|^2/(4R_1)$, т. е. вдвое больше. Таким образом, КПД трансформатора составляет всего 50 %. В общем случае КПД дается выражением

$$\eta = \frac{P_2}{P_0} = \frac{\alpha_1\alpha_2}{1+\alpha_2(\alpha_1+\alpha_2)}. \tag{136.18}$$

При заданном α_1 КПД максимален, когда $\alpha_2 = 1$, т. е. когда омическое сопротивление вторичной цепи равно ее индуктивному сопротивлению. Максимальное значение КПД будет $\eta_{\text{макс}} = \dfrac{\alpha_1}{\alpha_1+2}$. Если $\alpha_1 \gg 1$, то $\eta_{\text{макс}} \approx 1$. При этом для мощности, потребляемой во вторичной цепи, получаем

$$P_2 \approx \frac{1}{2}\,\frac{|\mathscr{E}|^2}{\alpha_1 R_1} = \frac{1}{2}\,\frac{|\mathscr{E}|^2}{\omega L_1} \approx P_0. \tag{136.19}$$

Мощность холостого хода трансформатора $P_0^{(0)}$ найдется из формулы (136.14), если в ней положить $\alpha_2 = 0$. Это дает

$$P_0^{(0)} = \frac{1}{2(\alpha_1^2+1)}\,\frac{|\mathscr{E}|^2}{R_1} \approx \frac{|\mathscr{E}|^2}{2\alpha_1^2 R_1}, \tag{136.20}$$

$$\frac{P_0}{P_0^{(0)}} = \frac{1+\alpha_2(\alpha_1+\alpha_2)}{1+(\alpha_1+\alpha_2)^2}\,(1+\alpha_1^2) \approx \alpha_1\alpha_2. \tag{136.21}$$

Приближенные выражения, как всегда, получены в предположении $\alpha_1 \gg 1$. Кроме того, последнее выражение предполагает также, что $\alpha_1 \gg \alpha_2$ (чтобы КПД был большим). При максимальной мощности $P_2\alpha_1 = \alpha_2$ и из (136.21) получаем

$$P_0/P_0^{(0)} \approx \alpha_1^2/2. \tag{136.22}$$

Таким образом, при нагрузке трансформатора до максимальной мощности мощность тока в первичной цепи возрастает в $\alpha_1^2/2$ раз по сравнению с мощностью холостого хода, а при нагрузке с максимальным КПД ($\alpha_2 = 1$) — в α_1 раз.

4. Трансформатор применяется не только для повышения или понижения напряжения переменного тока, но и для *передачи электрической энергии на расстояние.* Для исследования этого вопроса рассмотрим два связанных трансформатора, параметры которых обозначены на рис. 323. Запишем урав-

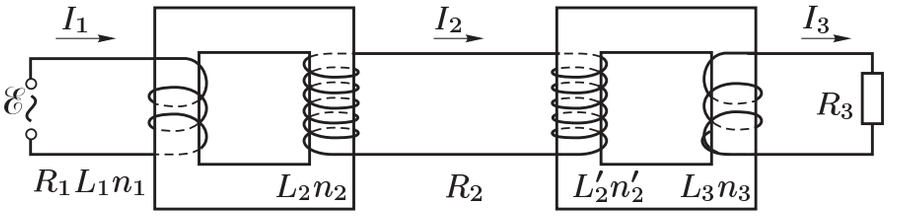


Рис. 323

нения колебаний в них:

$$\begin{aligned} R_1 I_1 &= \mathscr{E} - L_1 \dot{I}_1 - \sqrt{L_1 L_2}\, \dot{I}_2, \\ R_2 I_2 &= -\sqrt{L_1 L_2}\, \dot{I}_1 - L_2 \dot{I}_2 - L_2' \dot{I}_2 - \sqrt{L_2' L_3}\, \dot{I}_3, \\ R_3 I_3 &= -\sqrt{L_2' L_3}\, \dot{I}_2 - L_3 \dot{I}_3. \end{aligned} \tag{136.23}$$

Для уменьшения громоздкости формул предположим, что омическое сопротивление R_1 пренебрежимо мало, и положим $R_1 = 0$. Если, кроме того, электродвижущая сила $\mathscr{E}$ меняется синусоидально, то для установившегося режима получим

$$\begin{aligned} I_1 &= \frac{(\alpha_2\alpha_3 - 1) - i(\alpha_2 + \alpha_2' + \alpha_3)}{\alpha_2' + \alpha_3 - i}\, \frac{\mathscr{E}}{\omega L_1}, \\ I_2 &= -\sqrt{\frac{L_2}{L_1}}\, \frac{\alpha_3 - i}{\alpha_2' + \alpha_3 - i}\, \frac{\mathscr{E}}{R_2}, \\ I_3 &= \sqrt{\frac{L_2}{L_1}\, \frac{L_3}{L_2'}}\, \frac{\alpha_2'}{\alpha_2' + \alpha_3 - i}\, \frac{\mathscr{E}}{R_3}, \end{aligned} \tag{136.24}$$

где использованы следующие обозначения:

$$\alpha_2 = \frac{\omega L_2}{R_2}, \quad \alpha_2' = \frac{\omega L_2'}{R_2}, \quad \alpha_3 = \frac{\omega L_3}{R_3}. \tag{136.25}$$

Здесь R_2 означает сумму сопротивлений вторичной обмотки первого трансформатора, первичной обмотки второго и соединяющих их проводов. Если $R_2 = 0$, то $\alpha_2' = \infty$ и из последней формулы (136.24) получаем

$$I_3 = \sqrt{\frac{L_2}{L_1}\, \frac{L_3}{L_2'}}\, \frac{\mathscr{E}}{R_3} = \frac{n_2}{n_1}\, \frac{n_3}{n_2'}\, \frac{\mathscr{E}}{R_3}. \tag{136.26}$$

Смысл этой формулы ясен. Первый трансформатор повышает напряжение в n_2/n_1 раз, второй — в n_3/n_2' раз, так что напряжение на выходе второго трансформатора становится равным $\dfrac{n_2}{n_1}\, \dfrac{n_3}{n_2'}\, \mathscr{E}$. Если оба трансформатора одинаковы ($n_1 = n_3$, $n_2 = n_2'$), то $I_3 = \mathscr{E}/R_3$. Ток на выходе I_3 получается таким же, как если бы источник напряжения $\mathscr{E}$ был непосредственно замкнут на сопротивление R_3.

Мощность тока в первичной цепи

$$P_0 = \frac{1}{2}\,\frac{\alpha_2\alpha_3(\alpha_2'+\alpha_3)+\alpha_2}{(\alpha_2'+\alpha_3)^2+1}\,\frac{|\mathscr{E}|^2}{\omega L_1}. \tag{136.27}$$

Потребляемые мощности равны соответственно

$$P_1 = 0, \quad P_2 = \frac{1}{2}\,\frac{\alpha_2(\alpha_3^2+1)}{(\alpha_2'+\alpha_3)^2+1}\,\frac{|\mathscr{E}|^2}{\omega L_1}, \quad P_3 = \frac{1}{2}\,\frac{\alpha_2\alpha_2'\alpha_3}{(\alpha_2'+\alpha_3)^2+1}\,\frac{|\mathscr{E}|^2}{\omega L_1}. \tag{136.28}$$

Если $R_2 = 0$, то $\alpha_2 = \infty$, $\alpha_2' = \infty$, причем $\alpha_2/\alpha_2' = L_2/L_2'$, т. е. α_2 и α_2' — бесконечно большие одного порядка. Используя это, находим $P_2 = 0$, $P_3 = P_0$. Вся энергия, вырабатываемая в первичной цепи, передается без потерь потребителю.

При неизменных параметрах всех обмоток, за исключением последней, потребляемая мощность P_3 максимальна при $\alpha_3 = \sqrt{1+\alpha'^2_2}$, или приблизительно при $\alpha_3 = \alpha_2'$. Таким образом,

$$P_{3\,\text{макс}} \approx \frac{\alpha_2|\mathscr{E}|^2}{8\omega L_1}. \tag{136.29}$$

Такова же мощность, теряемая в линии передачи, так что КПД в этом случае составляет всего около 50 %. В общем случае КПД

$$\eta = \frac{P_3}{P_0} = \frac{\alpha_2'\alpha_3}{1+\alpha_3(\alpha_2'+\alpha_3)} \tag{136.30}$$

достигает максимума при $\alpha_3 = 1$. Если учесть, что $\alpha_2' \gg 1$, то максимальный КПД приблизительно равен $\eta_{\text{макс}} \approx 1$. При этом

$$\begin{aligned} P_0 &= \frac{\alpha_2}{2}\,\frac{\alpha_2'+\alpha_3+1}{(\alpha_2'+\alpha_3)^2+1}\,\frac{|\mathscr{E}|^2}{\omega L_1} \approx \frac{\alpha_2}{2\alpha_2'}\,\frac{|\mathscr{E}|^2}{\omega L_1}, \\ P_3 &= \frac{\alpha_2}{2}\,\frac{\alpha_2'}{(\alpha_2'+\alpha_3)^2+1}\,\frac{|\mathscr{E}|^2}{\omega L_1} \approx \frac{\alpha_2}{2\alpha_2'}\,\frac{|\mathscr{E}|^2}{\omega L_1} \approx P_0. \end{aligned} \tag{136.31}$$

Наконец, мощность холостого хода трансформатора ($\alpha_3 = 0$) равна

$$P_0^{(0)} = \frac{1}{2}\,\frac{\alpha_2}{1+\alpha'^2_2}\,\frac{|\mathscr{E}|^2}{\omega L_1} \approx \frac{1}{2}\,\frac{\alpha_2}{\alpha'^2_2}\,\frac{|\mathscr{E}|^2}{\omega L_1}. \tag{136.32}$$

Таким образом, при максимальном КПД

$$P_3/P_0^{(0)} \approx \alpha_2'. \tag{136.33}$$

§ 137. Колебания с двумя степенями свободы

1. Рассмотрим электрические колебания в двух колебательных контурах, индуктивно связанных между собой (рис. 324). Будем считать, что нет омических сопротивлений и внешних сил, действующих на систему (свободные колебания). Поскольку колебания в одном контуре влияют на колебания в другом, они называются *связанными колебаниями.* Такие колебания описываются дифференциальными уравнениями

$$\frac{Q_1}{C_1} + L_1\dot{I}_1 + L_{12}\dot{I}_2 = 0, \quad \frac{Q_2}{C_2} + L_2\dot{I}_2 + L_{21}\dot{I}_1 = 0, \tag{137.1}$$

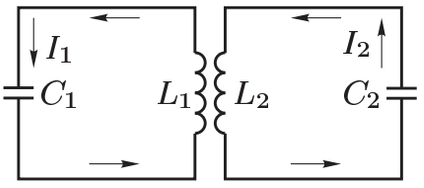


Рис. 324

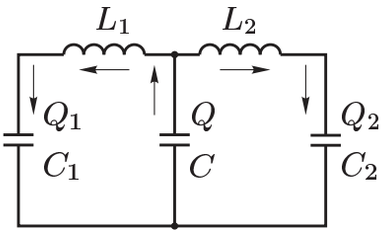


Рис. 325

или

$$L_1\ddot{Q}_1 + L_{12}\ddot{Q}_2 + \frac{Q_1}{C_1} = 0, \quad L_{21}\ddot{Q}_1 + L_2\ddot{Q}_2 + \frac{Q_2}{C_2} = 0. \tag{137.2}$$

Разрешив эти уравнения относительно производных, приведем их к виду

$$\ddot{Q}_1 + a_{11}Q_1 + a_{12}Q_2 = 0, \quad \ddot{Q}_2 + a_{21}Q_1 + a_{22}Q_2 = 0, \tag{137.3}$$

где введены обозначения:

$$\begin{aligned} a_{11} &= \frac{L_2}{C_1(L_1L_2 - L_{12}L_{21})}, \quad a_{12} = \frac{L_{12}}{C_2(L_1L_2 - L_{12}L_{21})}, \\ a_{21} &= -\frac{L_{21}}{C_1(L_1L_2 - L_{12}L_{21})}, \quad a_{22} = \frac{L_1}{C_2(L_1L_2 - L_{12}L_{21})}. \end{aligned} \tag{137.4}$$

Прежде чем идти дальше, рассмотрим такие же колебательные контуры, но с емкостной связью (рис. 325). В этом случае

$$\frac{Q_1}{C_1} + \frac{Q}{C} + L_1\dot{I}_1 = 0, \quad \frac{Q_2}{C_2} + \frac{Q}{C} + L_2\dot{I}_2 = 0. \tag{137.5}$$

Продифференцировав эти уравнения по времени и приняв во внимание, что $I = I_1 + I_2$, или $\dot{Q} = \dot{Q}_1 + \dot{Q}_2$, получим

$$\ddot{I}_1 + a_{11}I_1 + a_{12}I_2 = 0, \quad \ddot{I}_2 + a_{21}I_1 + a_{22}I_2 = 0, \tag{137.6}$$

где введены обозначения:

$$\begin{aligned} a_{11} &= \frac{1}{L_1}\left(\frac{1}{C_1} + \frac{1}{C}\right), \quad a_{12} = \frac{1}{L_1C}, \\ a_{21} &= \frac{1}{L_2C}, \quad a_{22} = \frac{1}{L_2}\left(\frac{1}{C_2} + \frac{1}{C}\right). \end{aligned} \tag{137.7}$$

Мы видим, что в обоих случаях колебания описываются *однотипными системами* линейных дифференциальных уравнений второго порядка с постоянными коэффициентами. К уравнениям такого же типа приводит задача о малых колебаниях механических систем с двумя степенями свободы, например двух *связанных маятников.* Целесообразно рассмотреть все эти колебания совместно. Поэтому мы не будем конкретизировать колебательную систему, а предположим только, что

ее конфигурация определяется какими-то координатами x_1 и x_2 подчиняющимися системе дифференциальных уравнений

$$\ddot{x}_1 + a_{11}x_1 + a_{12}x_2 = 0, \quad \ddot{x}_2 + a_{21}x_1 + a_{22}x_2 = 0, \tag{137.8}$$

где a_{ik} — постоянные коэффициенты. Значения этих коэффициентов, как и смысл координат x_1 и x_2, устанавливаются в каждом конкретном случае в отдельности.

2. Попытаемся сначала найти частное решение системы уравнений (137.8):

$$x_1 = A_1 e^{i\omega t}, \quad x_2 = A_2 e^{i\omega t}, \tag{137.9}$$

где A_1 и A_2 — постоянные. После подстановки в (137.8) получаем

$$(a_{11} - \omega^2)A_1 + a_{12}A_2 = 0, \quad a_{21}A_1 + (a_{22} - \omega^2)A_2 = 0. \tag{137.10}$$

Такая система линейных однородных уравнений имеет отличное от нуля решение только при выполнении условия

$$\begin{vmatrix} a_{11} - \omega^2 & a_{12} \\ a_{21} & a_{22} - \omega^2 \end{vmatrix} = 0. \tag{137.11}$$

Это — квадратное уравнение относительно ω^2. Обозначим его корни через ω_1^2 и ω_2^2. Не представляет труда показать, что в обоих примерах, приведенных выше, оба эти корня вещественны и притом положительны. В общем случае, когда система не конкретизирована, условие вещественности и положительности корней должно быть введено в качестве независимого требования, которому должны удовлетворять коэффициенты a_{ik}. Для ω получаются четыре значения: $\pm\omega_1$ и $\pm\omega_2$. Однако введение отрицательных ω не дает ничего нового. Действительно, коэффициенты уравнений (137.8) *вещественны* и нас должны интересовать лишь *вещественные решения* этих уравнений. Но вещественные части выражений $Ae^{i\omega t}$ и $Ae^{-i\omega t}$ имеют один и тот же вид, а именно $C\cos(\omega t + \delta)$, где C и δ — произвольные постоянные. Поэтому, не теряя общности, можно ограничиться лишь положительными корнями ω_1 и ω_2.

Как видно из (137.10), коэффициенты A_1 и A_2 не независимы. Их отношение однозначно определяется значениями коэффициентов a_{ik} и частоты ω:

$$h \equiv \frac{A_2}{A_1} = \frac{\omega^2 - a_{11}}{a_{12}} = \frac{a_{21}}{\omega^2 - a_{22}}. \tag{137.12}$$

Соответственно двум значениям частоты ω получаются и два значения отношения h, обозначаемые в дальнейшем через h_1 и h_2. Таким образом, мы нашли два частных решения уравнений (137.8).

Первое решение: $x_1 = e^{i\omega_1 t}$, $x_2 = h_1 e^{i\omega_1 t}$.

Второе решение: $x_1 = e^{i\omega_2 t}$, $x_2 = h_2 e^{i\omega_2 t}$.

Общее решение выражается *линейной комбинацией* этих двух частных решений с постоянными коэффициентами, т. е.

$$x_1 = C_1 e^{i\omega_1 t} + C_2 e^{i\omega_2 t}, \quad x_2 = h_1 C_1 e^{i\omega_1 t} + h_2 C_2 e^{i\omega_2 t}, \tag{137.13}$$

где C_1 и C_2 — произвольные комплексные постоянные. Они определяют *амплитуды* и *фазы колебаний* и могут быть найдены из *начальных условий*: по двум значениям x_1 и x_2 и их производных в момент времени $t = 0$.

Введем обозначения:

$$\xi_1 = C_1 e^{i\omega_1 t}, \quad \xi_2 = C_2 e^{i\omega_2 t}. \tag{137.14}$$

Тогда

$$x_1 = \xi_1 + \xi_2, \quad x_2 = h_1\xi_1 + h_2\xi_2, \tag{137.15}$$

$$\xi_1 = \frac{h_2 x_1 - x_2}{h_2 - h_1}, \quad \xi_2 = \frac{-h_1 x_1 + x_2}{h_2 - h_1}. \tag{137.16}$$

Отсюда видно, что ξ_1 и ξ_2 могут быть приняты за *новые координаты*, определяющие *конфигурацию системы*. Координата ξ_1 совершает гармонические колебания с частотой ω_1 а координата ξ_2 — с частотой ω_2. Эти координаты называются *нормальными координатами*, а совершаемые ими колебания — *нормальными колебаниями* или *модами*. Таким образом, в общем случае *колебание системы представляет собой суперпозицию двух нормальных колебаний с частотами* ω_1 и ω_2.

3. Рассмотрим симметричный случай, когда $a_{11} = a_{22}$, $a_{12} = a_{21}$. В примерах, приведенных выше (см. рис. 324 и 325), такой случай реализуется, когда параметры обоих колебательных контуров одинаковы. Уравнение (137.11) в этом случае переходит в $(\omega^2 - a_{11})^2 = a_{12}^2$, и, следовательно,

$$\omega_1^2 = a_{11} + a_{12}, \quad \omega_2^2 = a_{11} - a_{12}, \quad h_1 = 1, \quad h_2 = -1,$$

$$x_1 = \xi_1 + \xi_2, \quad x_2 = \xi_1 - \xi_2, \quad \xi_1 = \frac{x_1 + x_2}{2}, \quad \xi_2 = \frac{x_1 - x_2}{2}.$$

Например, в случае рис. 325 получаем

$$\omega_1^2 = \frac{1}{L_1}\left(\frac{1}{C_1} + \frac{2}{C}\right), \quad \omega_2^2 = \frac{1}{L_1 C_1}.$$

Последнюю формулу легко понять. Действительно, пусть в системе совершается только нормальное колебание с частотой ω_2, а колебание с частотой ω_1 не возбуждено. Это значит, что $2\xi_1 = I_1 + I_2 = I = 0$, т. е ток через конденсатор C не течет. Поэтому на колебания с частотой ω_2 конденсатор C не оказывает никакого влияния. Система ведет себя как один колебательный контур, в котором катушки самоиндукции и конденсаторы соединены последовательно. Результирующая индуктивность такого контура $2L_1$, емкость $C_1/2$, а собственная частота $\omega = 1/\sqrt{L_1 C_1} = \omega_2$.

4. Интересное явление наблюдается в симметричном случае, когда связь между обеими подсистемами, из которых состоит сложная система, *слабая*. В этом случае, ввиду малости коэффициента a_{12}, частоты нормальных колебаний ω_1 и ω_2 мало отличаются друг от друга. Одна из них немного больше, а другая немного меньше собственной частоты подсистемы $\omega_0 = \sqrt{a_{11}}$. Колебания координат x_1 и x_2 представляются

суперпозицией нормальных колебаний с мало отличающимися частотами. На протяжении нескольких колебаний обе координаты x_1 и x_2 колеблются почти так, как если бы связи не было. Наличие слабой связи приводит к возникновению *биений*, схематически изображенных на рис. 294. Когда амплитуда координаты x_1 проходит через максимум, амплитуда координаты x_2 обращается в нуль, и наоборот. Явление легко демонстрируется с помощью двух одинаковых математических маятников, между которыми установлена слабая связь. Отклонив первый маятник, наблюдают, что амплитуда его колебаний медленно убывает и второй маятник также начинает колебаться. На протяжении нескольких десятков периодов колебания первого маятника полностью затухнут, а колебания второго станут максимальными. После этого начнется затухание колебаний второго маятника. Первый маятник, наоборот, начнет раскачиваться, и по истечении такого же числа десятков периодов его амплитуда вернется к исходному значению. Затем процесс передачи колебаний от одного маятника к другому будет повторяться, пока в результате действия сил трения колебания не прекратятся.

5. Затухающие собственные колебания в связанных системах рассматриваются так же, как и незатухающие. Надо только ввести в уравнения (137.6) линейные члены, содержащие первые производные $\dot{x}_1$ и $\dot{x}_2$. Можно также рассмотреть вынужденные колебания, введя внешние силы, действующие на систему. Это делается так же, как и в случае системы с одной степенью свободы. Наконец, когда число степеней свободы системы n больше двух, то для задания ее конфигурации требуется n координат. Если эти координаты описываются линейными уравнениями типа (137.8), то всякое колебание системы также представится суперпозицией *нормальных колебаний с n частотами*. Некоторые из этих частот могут совпадать. Тогда говорят о *вырождении*. Вырожденный случай можно свести к невырожденному. Для этого надо «снять вырождение», т. е. слегка изменить коэффициенты уравнений, а затем совершить предельный переход к первоначальным значениям этих коэффициентов.

§ 138. Волновое уравнение

1. Об *электромагнитных возмущениях*, или *волнах*, уже говорилось в § 83. Там на примере плоского возмущения было выяснено, как возбуждаются электромагнитные волны, и вычислена скорость их распространения. Вернемся снова к этому вопросу, чтобы придать изложению математически более простой и систематический характер. Кроме того, мы рассмотрим некоторые новые вопросы. Многие вопросы, относящиеся к учению о волнах (отражение, преломление, интерференция, дифракция, дисперсия и пр.), мы сознательно опускаем, так как они будут подробно рассмотрены в следующем томе.

Начнем с простой механической аналогии. Если ударить по какому-либо месту натянутого шнура, то от места удара в противоположных направлениях побегут два поперечных возмущения. Рассмотрим одно

из них, например возмущение, распространяющееся вправо. Положение невозмущенного натянутого шнура примем за ось X. Тогда каждую материальную точку шнура можно характеризовать абсциссой x, которую она имела на невозмущенном шнуре, а само возмущение — смещением s этой точки из положения равновесия, как функции координаты x и времени t: $s = s(x, t)$. Однако эта функция зависит не от x и t в отдельности, а от определенной комбинации их, которая будет найдена ниже. На рис. 326 вверху изображено положение возмущенного шнура в момент времени $t = 0$. Эта начальная форма шнура может быть представлена уравнением $s(\xi, 0) = f(\xi)$, где ξ — абсцисса какой-то произвольной материальной точки шнура $A(\xi)$. Через время t возмущение на шнуре переместится вправо на расстояние $OO' = vt$, где v — *скорость распространения возмущения.* Это значит, что смещение $s(x, t)$ точки $A(x)$ с координатой x в момент t будет таким же, каким было смещение точки $A(\xi)$ с координатой ξ в момент $t = 0$, если только $x - \xi = vt$, т. е. $s(x, t) = s(\xi, 0) = f(\xi) = f(x - vt)$. Таким образом, опуская аргументы x и t в функции $s(x, t)$, находим для смещения s следующее выражение:

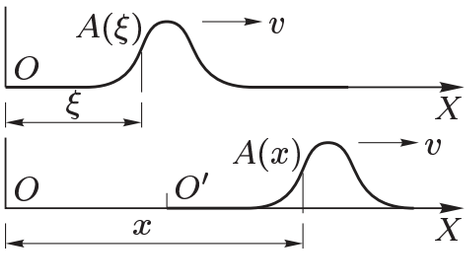


Рис. 326

$$s = f(x - vt). \tag{138.1}$$

Следовательно, если возмущение распространяется вправо, то величина смещения s зависит только от комбинации аргументов $x - vt$. Если эта комбинация остается постоянной, то будет оставаться постоянным и смещение s. Это значит, что уравнение $x - vt = \text{const}$ есть *уравнение фронта распространяющегося возмущения.* Дифференцируя его по t, находим $dx/dt = +v$, т. е. v, как и должно быть, есть скорость распространения волнового фронта.

Таким же путем убеждаемся, что возмущение, распространяющееся влево, описывается уравнением

$$s = f(x + vt). \tag{138.2}$$

Если же возмущение идет и вправо, и влево, то

$$s = f_1(x - vt) + f_2(x + vt). \tag{138.3}$$

Вид функций f_1 и f_2 определяется *начальными условиями*, т. е. заданием начальной формы шнура и начального распределения скоростей, а потому может быть весьма разнообразным.

2. Можно получить уравнение, *совсем не содержащее начальных условий*, а потому пригодное для описания распространения любых волновых возмущений в шнуре. В этом отношении оно аналогично уравнениям Ньютона в механике, которые также не содержат начальных условий. Независимость от начальных условий связана с тем, что это уравнение (как и уравнение, выражающее второй закон Ньютона) *дифференциальное.* Для его получения дифференцируем выражение

(138.1) сначала по x, а затем по t и находим

$$\frac{\partial s}{\partial x} = f', \quad \frac{\partial s}{\partial t} = f' \frac{\partial (x - vt)}{\partial t} = -vf',$$

где f' означает производную по аргументу $x - vt$, от которого зависит функция f. Исключая f', находим

$$\frac{\partial s}{\partial x} = -\frac{1}{v} \frac{\partial s}{\partial t}.$$

Это уравнение не содержит начальных условий. Однако оно описывает не все возмущения, а только *возмущения, распространяющиеся вправо.* Возмущения, распространяющиеся влево, описываются таким же уравнением, но со знаком плюс в правой части. Для нахождения уравнения, справедливого для обоих возмущений, а также их суперпозиции (138.3), дифференцируем вторично и находим

$$\frac{\partial^2 s}{\partial x^2} = f'', \quad \frac{\partial^2 s}{\partial t^2} = -vf'' \frac{\partial (x - vt)}{\partial t} = v^2 f'',$$

или после исключения f''

$$\frac{\partial^2 s}{\partial x^2} - \frac{1}{v^2} \frac{\partial^2 s}{\partial t^2} = 0. \tag{138.4}$$

Легко убедиться, что такому же уравнению удовлетворяет и возмущение (138.2), а также более общее возмущение (138.3). Дифференциальное уравнение (138.4) называется *волновым уравнением.* Оно справедливо для любых возмущений, распространяющихся в шнуре.

Выражение (138.3), в котором f_1 и f_2 — произвольные функции, есть *общее решение* волнового уравнения (138.4). Чтобы убедиться в этом, введем новые независимые переменные $\xi = x - vt$ и $\eta = x + vt$. В этих переменных уравнение (138.4) принимает вид

$$\frac{\partial^2 s}{\partial \xi \partial \eta} = 0.$$

Общее решение его $s = f_1(\xi) + f_2(\eta)$, и наше утверждение доказано. Следовательно, всякий процесс, описываемый волновым уравнением (138.4), в общем случае представляет собой два возмущения, распространяющихся со скоростью v в противоположных направлениях оси X.

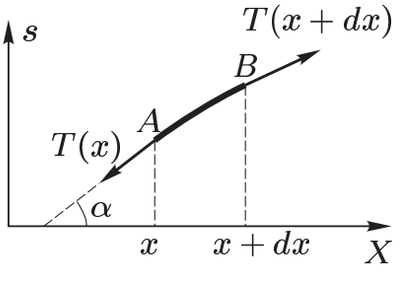


Рис. 327

3. Приведем два примера на применение уравнения (138.4). Выведем уравнение малых поперечных колебаний гибкого натянутого шнура, исходя из уравнений механики. Будем считать, что вся упругость шнура вызвана его *натяжением* T, упругостью формы пренебрегаем. На элемент AB шнура (рис. 327) слева действует сила натяжения $T(x)$. Ее вертикальная составляющая будет $-T(x) \operatorname{tg} \alpha = T(x) \partial s / \partial x$ (положительное направление выбрано вверх). Аналогичная сила действует на правый конец элемента AB. Результирующая

этих двух сил будет

$$\left(T\frac{\partial s}{\partial x}\right)_B - \left(T\frac{\partial s}{\partial x}\right)_A = \frac{\partial}{\partial x}\left(T\frac{\partial s}{\partial x}\right)dx.$$

Для малых колебаний угол α будет мал, и можно пренебречь изменением натяжения T вдоль шнура. В этом приближении предыдущее выражение переходит в $T\dfrac{\partial^2 s}{\partial x^2}\,dx$. Кроме того, можно также пренебречь изменениями линейной плоскости шнура δ при его удлинениях и сжатиях. Приравняв массу элемента $\delta\,dx$, умноженную на его ускорение $\partial^2 s/\partial t^2$, действующей силе, находим

$$\delta\frac{\partial^2 s}{\partial t^2} = T\frac{\partial^2 s}{\partial x^2}.$$

Это уравнение совпадает с (138.4), а потому для скорости распространения v в шнуре получаем

$$v = \sqrt{T/\delta}. \tag{138.5}$$

Такая формула уже была выведена нами другим способом (см. т. I, § 84). Совершенно так же могут быть выведены формулы для скорости распространения упругих возмущений в стержнях, в неограниченных упругих средах, а также в жидкостях и газах (см. т. I, § 81, 83, 85).

В качестве второго примера рассмотрим плазму в постоянном магнитном поле B, обладающую достаточно высокой проводимостью. Выделим в ней какую-либо тонкую силовую трубку. Если плазма сместится поперек магнитного поля, то благодаря высокой проводимости ее магнитный поток через поперечное сечение трубки сохранится (см. § 71). Магнитные силовые линии как бы *вморожены в вещество* и движутся вместе с ним. Но вдоль магнитной трубки действует максвелловское натяжение $\tau = B^2/8\pi$. Поэтому в плазме вдоль магнитных силовых линий могут распространяться *поперечные возмущения*, аналогичные возмущениям в натянутом шнуре. Такие возмущения называются *магнитогидродинамическими* или *альфвеновскими волнами*. Они были теоретически предсказаны Альфвеном (р. 1908). Есть, однако, отличие гидродинамических волн от волн в натянутом шнуре. Оно состоит в том, что магнитная силовая трубка подвергается не только продольному натяжению, но и равному ему боковому давлению (рис. 328 *а*). Однако боковое давление легко исключить. Для этого к основаниям элемента трубки надо приложить натяжение τ и равное ему давление. Эти напряжения ничего не меняют, так как они взаимно компенсируют друг друга. Но тогда система максвелловских натяжений сведется к продольному натяжению $2\tau = B^2/4\pi$ и всестороннему давлению τ (рис. 328 *б*). Всесторон-

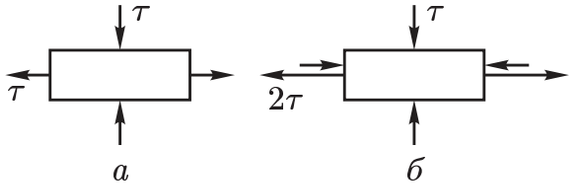


Рис. 328

нее давление на распространение поперечных возмущений не влияет. Поэтому можно воспользоваться формулой (138.5), полагая в ней $T = 2S\tau$, $\delta = S\rho$, где S — площадь поперечного сечения магнитной трубки, а ρ — плотность плазмы. Таким путем для скорости распространения магнитогидродинамических волн найдем

$$v = \frac{B}{\sqrt{4\pi\rho}}. \tag{138.6}$$

4. Уравнение (138.4) есть «одномерное» волновое уравнение, поскольку оно относится к распространению процессов только *вдоль одного направления*, а величина s, характеризующая описываемый процесс, зависит только от одной пространственной координаты x и времени t. Волновые процессы, распространяющиеся в пространстве, описываются *«трехмерным»* волновым уравнением

$$\frac{\partial^2 s}{\partial x^2} + \frac{\partial^2 s}{\partial y^2} + \frac{\partial^2 s}{\partial z^2} - \frac{1}{v^2}\frac{\partial^2 s}{\partial t^2} = 0. \tag{138.7}$$

Однако, как правило, мы этим уравнением пользоваться не будем.

§ 139. Плоские электромагнитные волны

1. Пусть в неограниченной однородной среде распространяется *без поглощения* какое-то электромагнитное возмущение. Отсутствие поглощения означает, что при любом возмущении в среде не выделяется джоулево тепло. Следовательно, величина $\mathbf{jE} = \lambda\mathbf{E}^2$ должна обращаться в нуль, каково бы ни было поле $\mathbf{E}$. Это возможно тогда и только тогда, когда $\lambda = 0$, т. е. когда среда является диэлектриком. Допустим, кроме того, что объемных электрических зарядов в среде нет. Тогда уравнения Максвелла примут вид

$$\begin{gathered}\operatorname{rot}\mathbf{H} = \frac{1}{c}\dot{\mathbf{D}}, \quad \operatorname{rot}\mathbf{E} = -\frac{1}{c}\dot{\mathbf{B}}, \\ \operatorname{div}\mathbf{D} = 0, \quad \operatorname{div}\mathbf{B} = 0.\end{gathered} \tag{139.1}$$

Рассмотрим частное решение их, когда все величины зависят только от x и t. Переходя к координатной форме, получим в этом случае

$$\begin{gathered}\frac{\partial H_z}{\partial x} = -\frac{1}{c}\frac{\partial D_y}{\partial t}, \quad \frac{\partial H_y}{\partial x} = \frac{1}{c}\frac{\partial D_z}{\partial t}, \\ \frac{\partial E_y}{\partial x} = -\frac{1}{c}\frac{\partial B_z}{\partial t}, \quad \frac{\partial E_z}{\partial x} = \frac{1}{c}\frac{\partial B_y}{\partial t}, \\ \frac{\partial D_x}{\partial x} = \frac{\partial D_x}{\partial t} = \frac{\partial B_x}{\partial x} = \frac{\partial B_x}{\partial t} = 0.\end{gathered} \tag{139.2}$$

Из последних четырех уравнений следует, что D_x и B_x не зависят от x и t, т. е. величины постоянные. Это — *статические поля*, накладывающиеся на переменное поле электромагнитного возмущения. Они

не влияют на распространение возмущения и могут быть отброшены без ущерба для общности. Оставшиеся четыре уравнения распадаются на две группы *независимых уравнений.* В одну из них входят y-составляющие электрического поля и z-составляющие магнитного поля, в другую — z-составляющие электрического и y-составляющие магнитного поля. Обе эти группы *однотипны*, а потому можно ограничиться рассмотрением одной из них. В качестве таковой возьмем систему уравнений, содержащую E_y и H_z. С помощью соотношений $\mathbf{D} = \varepsilon\mathbf{E}$ и $\mathbf{B} = \mu\mathbf{H}$ преобразуем ее к виду

$$\frac{\partial H}{\partial x} = -\frac{\varepsilon}{c}\frac{\partial E}{\partial t}, \quad \frac{\partial E}{\partial x} = -\frac{\mu}{c}\frac{\partial H}{\partial t}, \tag{139.3}$$

причем мы опустили у полей индексы y и z, предполагая, что вектор $\mathbf{E}$ направлен параллельно оси Y, а вектор $\mathbf{H}$ — параллельно оси Z. Дифференцируя первое уравнение по t, а второе по x, исключаем H и находим

$$\frac{\partial^2 E}{\partial x^2} - \frac{1}{v^2}\frac{\partial^2 E}{\partial t^2} = 0. \tag{139.4}$$

Аналогично,

$$\frac{\partial^2 H}{\partial x^2} - \frac{1}{v^2}\frac{\partial^2 H}{\partial t^2} = 0, \tag{139.5}$$

где введено обозначение

$$v = c/\sqrt{\varepsilon\mu}. \tag{139.6}$$

Таким образом, векторы $\mathbf{E}$ и $\mathbf{H}$ удовлетворяют *одному и тому же волновому уравнению.* Это доказывает, что рассматриваемое возмущение состоит из плоских волн, распространяющихся со скоростью v параллельно оси X. Возмущение поперечно, т. е. векторы $\mathbf{E}$ и $\mathbf{H}$ перпендикулярны к оси X, вдоль которой происходит распространение. Возьмем волну, распространяющуюся в положительном направлении оси X:

$$E = f(x - vt), \quad H = g(x - vt). \tag{139.7}$$

Тогда $\partial E/\partial t = -vf'$, $\partial H/\partial x = g'$ и после подстановки в первое уравнение (139.3) получится $g' = (\varepsilon v/c)f'$. Интегрируя это соотношение и опуская постоянные интегрирования (имеющие смысл статических полей, не представляющих интереса в рассматриваемом вопросе), придем к соотношению $g = (\varepsilon v/c)f$, или $H = (v/c)D$. Аналогично поступаем со вторым уравнением (139.3). Таким образом,

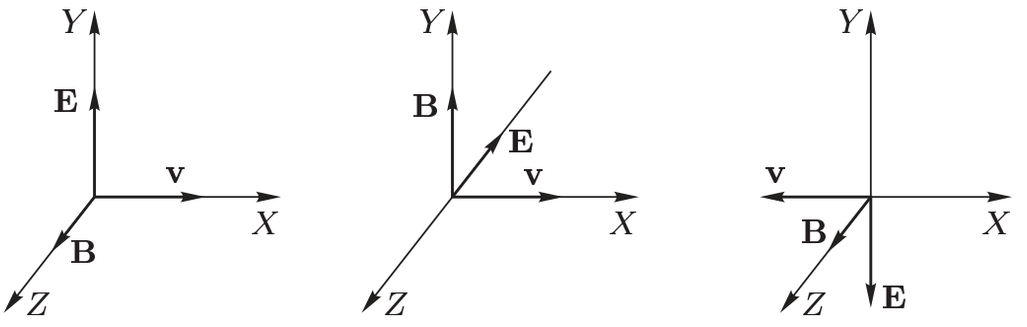
$$H = \frac{v}{c}D, \quad E = \frac{v}{c}B, \tag{139.8}$$

или в векторной форме

$$\mathbf{H} = \frac{1}{c}[\mathbf{vD}], \quad \mathbf{E} = -\frac{1}{c}[\mathbf{vB}], \tag{139.9}$$

где $\mathbf{v}$ — вектор скорости, с которой распространяется электромагнитное возмущение.

Векторы **E**, **B**, **v** (а также векторы **D**, **H**, **v**) *взаимно перпендикулярны и образуют правовинтовую систему.* Взаимное расположение их показано на рис. 329 *а*. Это правовинтовое соотношение указанных векторов есть *внутреннее свойство* бегущей электромагнитной волны,



а *б* *в*

Рис. 329

не зависящее ни от какой координатной системы. Повернем на рис. 329 *а* тройку векторов **E**, **B**, **v** вокруг оси X на 90°. Получится расположение, представленное на рис. 329 *б*. Теперь электрический вектор будет направлен по оси Z, а магнитный — по оси Y. Такому расположению соответствует вторая группа уравнений, входящая в систему (139.2). Мы видим, что действительно нет необходимости в особом исследовании этой группы. Повернем, далее, тройку векторов **E**, **B**, **v** на рис. 329 *а* вокруг оси Z на 180°. Получится расположение, приведенное на рис. 329 *в*, которому соответствует волна, распространяющаяся влево. Таким образом, нет необходимости особо рассматривать и эту волну.

2. Вид функции f (или g) в плоской бегущей электромагнитной волне зависит от *начальных условий* и может быть каким угодно. Особо важное значение имеют *синусоидальные*, или *монохроматические*, *волны.* Они могут быть представлены в виде

$$\mathbf{E} = \mathbf{E}_0 \cos \omega(t - x/v), \quad \mathbf{H} = \mathbf{H}_0 \cos \omega(t - x/v), \tag{139.10}$$

где $\mathbf{E}_0$ и $\mathbf{H}_0$ — постоянные, называемые *амплитудами волны.* Если ввести обозначение

$$k = \omega/v, \tag{139.11}$$

то

$$\mathbf{E} = \mathbf{E}_0 \cos(\omega t - kx), \quad \mathbf{H} = \mathbf{H}_0 \cos(\omega t - kx). \tag{139.12}$$

Величина k называется *волновым числом.* Если фиксировать координату x, то получатся синусоидальные функции времени, описывающие гармонические колебания с круговой частотой ω. Напротив, если фиксировать время t, то получится синусоидальное распределение поля **E**,

H в пространстве в рассматриваемый момент времени. Пространственный период поля **E**, **H** называется *длиной волны*. Он равен

$$\lambda = \frac{2\pi}{k} = \frac{2\pi v}{\omega} = \frac{v}{\nu}. \tag{139.13}$$

Можно записать бегущую монохроматическую электромагнитную волну *в векторной форме*, не содержащей никаких координат. Для этого введем единичный вектор **N** нормали к фронту волны AB, т. е. к плоскости постоянной фазы $\omega t - kx = \text{const}$. Тогда, как видно из рис. 330, $x = (\mathbf{Nr})$ и, следовательно, $kx = (\mathbf{kr})$, где $\mathbf{k} = k\mathbf{N}$ — так

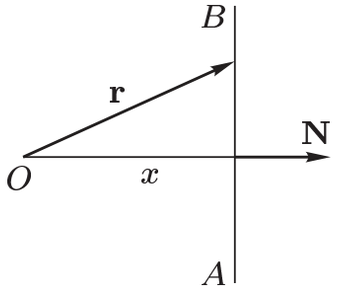


Рис. 330

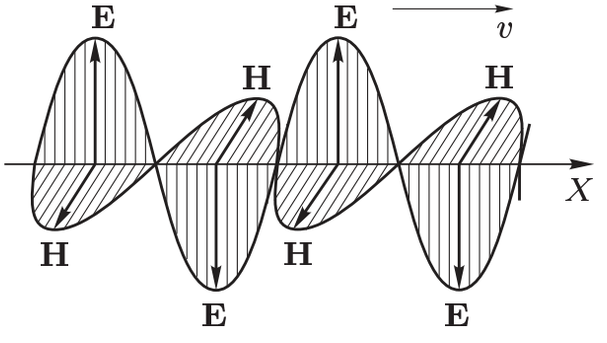


Рис. 331

называемый *волновой вектор*. В результате получится

$$\mathbf{E} = \mathbf{E}_0 \cos(\omega t - \mathbf{kr}), \quad \mathbf{H} = \mathbf{H}_0 \cos(\omega t - \mathbf{kr}), \tag{139.14}$$

или в комплексной форме

$$\mathbf{E} = \mathbf{E}_0 e^{i(\omega t - \mathbf{kr})}, \quad \mathbf{H} = \mathbf{H}_0 e^{i(\omega t - \mathbf{kr})}. \tag{139.15}$$

Здесь $\mathbf{E}_0$ и $\mathbf{H}_0$ уже могут быть *комплексными*, что означает введение начальных фаз. Однако, ввиду соотношений (139.8) или (139.9), в бегущей монохроматической волне электрический и магнитный векторы всегда колеблются *в одинаковых фазах*. Вид такой волны в пространстве в какой-либо момент времени изображен на рис. 331. Чтобы составить представление об изменении поля во времени, надо вообразить, что весь рисунок равномерно движется вправо со скоростью v. Чтобы получить волну, распространяющуюся влево, надо изменить на противоположное направление одного из векторов: **E** или **H**.

§ 140. Стоячие волны

1. Пусть в натянутом шнуре слева направо распространяется поперечная синусоидальная волна $s_1 = a\cos(\omega t - kx)$. Если изменить знак у kx, то получится волна $s_2 = a\cos(\omega t + kx)$, распространяющаяся справа налево. Такую волну можно получить, если отразить от конца шнура первую волну. Поэтому волну s_1 можно назвать *падающей*, а волну s_2 — *отраженной*. Никакой добавочной фазы в выражение для отраженной волны можно не вводить, если условиться помещать начало координат в точке шнура, в которой падающая и отраженная

волны находятся в одинаковых фазах. Это и предполагается в дальнейшем. Предположим, что отражение полное, т. е. амплитуды падающей и отраженной волн одинаковы. От наложения таких волн возникает возмущение

$$s = s_1 + s_2 = 2a \cos kx \cos \omega t, \tag{140.1}$$

называемое *стоячей волной.* В этом возмущении каждая точка шнура, характеризуемая координатой x, совершает гармоническое колебание с частотой ω и амплитудой $2a \cos kx$. Амплитуда таких колебаний обращается в нуль в тех точках, где $\cos kx = 0$. Такие точки называются *узлами смещения.* Посередине между двумя соседними узлами амплитуда колебаний $2a \cos kx$ максимальна, соответствующие точки называются *пучностями смещения.* Расстояние Δx между двумя соседними узлами или пучностями определится из условия $k\Delta x = \pi$, откуда $\Delta x = \pi/k = \lambda/2$. Все точки между двумя соседними узлами колеблются в одинаковых фазах. Они одновременно проходят через положение равновесия и одновременно достигают максимума. При переходе через узел знак s меняется на противоположный. Это значит, что при этом фаза колебания скачкообразно изменяется на π. Однако такой скачок не ведет к нарушению непрерывности колебательного процесса, так как он совершается при нулевой амплитуде. Картина колебаний в стоячей волне представлена на рис. 332. Две синусоиды на этом рисунке изображают крайние положения, которых достигает шнур при своих колебаниях, стрелками указано направление движения, которое возникает из этих крайних положений. Узлы смещения как бы разделяют шнур на автономные области, в которых совершаются *независимые гармонические колебания.* Никакой передачи движения от одной области к другой, а следовательно, и перетекания энергии через узлы не происходит. Иначе говоря, нет никакого распространения возмущения вдоль шнура. Вот почему возмущение, представляемое выражением (140.1), называется *стоячей волной.* Заметим еще, что в узлах смещения максимальны производные $\partial s/\partial x$, т. е. *деформации шнура*, а в пучностях смещения $\partial s/\partial x = 0$. Поэтому узлы смещения являются *пучностями деформации*, а пучности смещения — *узлами деформации.*

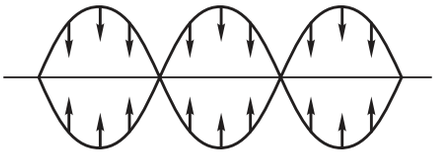


Рис. 332

2. Выше мы рассуждали так, как если бы длина шнура была не ограничена. В этом случае частота ω, а следовательно, и длина волны $\lambda = vT$ могут быть какими угодно. Не то будет, когда оба конца шнура закреплены. Если в шнуре с закрепленными концами возбудить какое-то произвольное возмущение и затем предоставить его самому себе, то это возмущение побежит в обе стороны и начнет отражаться от концов шнура. В шнуре возникнет довольно сложное *нестационарное движение. Стационарное движение в виде стоячей волны возможно лишь при вполне определенных частотах.* Дело в том, что на закрепленных концах шнура должны выполняться определенные *граничные*

условия: в них смещение s все время должно равняться нулю. Значит, если в шнуре возбуждена стоячая волна, то концы шнура должны быть ее узлами. Отсюда следует, что на длине шнура l должно укладываться целое число полуволн: $l = n\lambda/2$, откуда

$$\lambda = \frac{2l}{n}, \quad \omega = \frac{2\pi v}{\lambda} = \frac{\partial v}{l}\, n. \tag{140.2}$$

Целое число n может быть каким угодно. Получается бесконечный набор возможных типов стационарных колебаний, которым соответствует дискретный ряд частот. Эти колебания называются *собственными* или *нормальными* колебаниями шнура. Они имеют такой же смысл, что и нормальные колебания, или моды, введенные в § 137 для дискретных систем. В шнуре возможных типов нормальных колебаний получилось бесконечно много. Это связано с тем, что шнур рассматривается как непрерывная система, обладающая *бесконечным числом степеней свободы*. Собственное колебание с наинизшей частотой $\omega = \pi v/l$ называется *основным колебанием*, все остальные собственные колебания — *обертонами* или *гармониками*.

Все изложенное справедливо и для колебаний упругих стержней, как продольных, так и поперечных. Только здесь спектр возможных собственных колебаний богаче. Дело в том, что концы стержня могут быть либо *закреплены*, либо *свободны*. В первом случае не получается ничего нового по сравнению с натянутым шнуром. Во втором случае на концах стержня должны быть *пучности смещения*, а все остальное остается по-старому. Наконец, возможен случай, когда один конец стержня закреплен, а второй свободен. В этом случае основному колебанию соответствует длина волны, равная четверти длины стержня. Длины волн прочих собственных колебаний определяются формулой

$$l = \frac{\lambda}{4} + \frac{\lambda}{2}\, n, \tag{140.3}$$

где n — целое число. На рис. 333 *а*, *б* приведены первые два собственных колебания для стержня с закрепленными и свободными концами, а на рис. 333 *в* — для стержня, один конец которого закреплен, а другой свободен. Различные типы собственных колебаний в натянутом шнуре можно наблюдать, если конец M длинного резинового жгута закрепить в стене, а другой конец N взять в руку и привести в колебание (рис. 334). При надлежащей частоте колебаний руки и натяжении жгута удается возбудить основное колебание и несколько

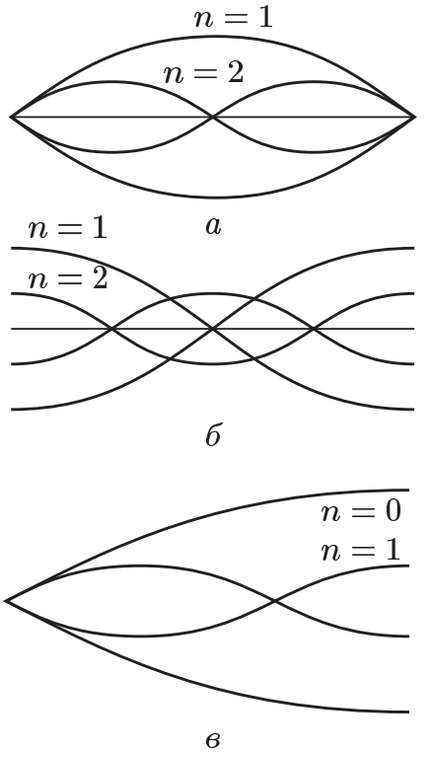


Рис. 333

его первых гармоник. При этом рука не находится в узле M', а несколько смещена относительно него. Из рис. 334 видно, что отношение OA/NP тем больше, чем меньше это смещение. Это значит, что при неизменной амплитуде колебаний руки колебания в жгуте возбудятся тем сильнее, чем ближе точка приложения силы руки к узлу стоячей волны. Бесконечно малой силой можно возбудить колебания конечной амплитуды, если силу приложить бесконечно близко к узлу. Но в этом случае на длине шнура укладывается целое число полуволн, а частота приложенной силы совпадает с одной из собственных частот шнура. Поэтому указанное сильное возбуждение колебаний есть не что иное, как явление резонанса.

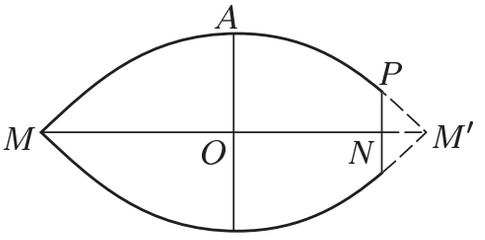


Рис. 334

3. Все сказанное о стоячих волнах в шнуре и стержнях относится и к *электромагнитным волнам.* В этом случае, однако, волна характеризуется не одним вектором, а двумя взаимно перпендикулярными векторами **E** и **H**. Пусть волна распространяется в положительном направлении оси X и представляется уравнениями

$$E_y = E_0 \cos(\omega t - kx),$$

$$H_z = H_0 \cos(\omega t - kx).$$

Волну, распространяющуюся в обратном направлении, можно получить отсюда, если изменить знаки у k и одного из векторов **E** или **H**, например магнитного. Это дает

$$E_y = E_0 \cos(\omega t + kx), \quad H_z = -H_0 \cos(\omega t + kx).$$

В результате суперпозиции с предыдущей волной получится

$$E_y = 2E_0 \cos kx \cos \omega t, \quad H_z = 2H_0 \sin kx \sin \omega t. \tag{140.4}$$

Это и есть *стоячая электромагнитная волна.* Она состоит из двух стоячих волн: *электрической* и *магнитной.* Мы видим, что колебания электрического поля сдвинуты по фазе относительно колебаний

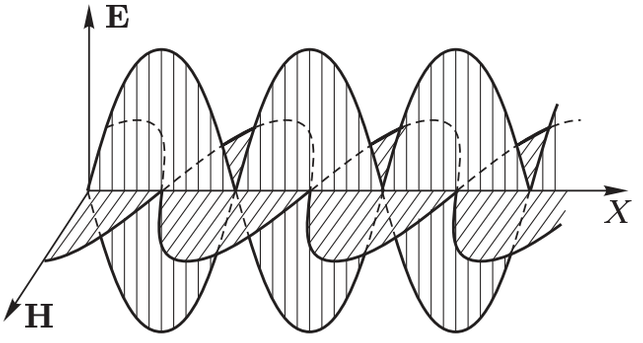


Рис. 335

магнитного поля на $\pi/2$. Кроме того, пучности электрического поля совпадают с узлами магнитного поля, а узлы — с пучностями (рис. 335). Вектор Пойнтинга обращается в нуль в узлах (а следовательно, и в пучностях) как электрического, так и магнитного поля. Поэтому электромагнитная энергия не переходит ни через один из этих узлов. Ее движение ограничено колебаниями между узлом (пучностью) электрического поля и пучностью (узлом) магнитного поля.

§ 141. Поле излучения диполя Герца

1. Простейшей системой, излучающей электромагнитные волны, является точечный диполь, дипольный момент которого быстро меняется (колеблется) во времени. Такой диполь называется *диполем Герца*, по имени ученого, впервые рассчитавшего его электромагнитное поле. Задача об излучении диполя Герца в теории излучающих систем имеет особое значение. Дело в том, что всякую реальную излучающую систему — антенну, по которой течет переменный ток, — можно мысленно разложить на элементы тока, каждый из которых излучает, как диполь Герца. Путем суперпозиции электромагнитных полей таких элементов можно получить электромагнитное поле всей излучающей системы. Предпошлем решению задачи о поле излучения диполя Герца несколько вспомогательных математических формул.

2. Рассмотрим функцию $f = f(\mathbf{r}, t - r/v)$, где r — расстояние от начала координат, $\mathbf{r}$ — радиус-вектор, проведенный из того же начала, v — некоторая постоянная скорость. Введем обозначение $t' = t - r/v$. Частные производные функции f по x при постоянных t и t' связаны соотношением

$$\left(\frac{\partial f}{\partial x}\right)_t = \left(\frac{\partial f}{\partial x}\right)_{t'} + \left(\frac{\partial f}{\partial t'}\right)_r \left(\frac{\partial t'}{\partial x}\right)_t.$$

Отсюда с учетом соотношения $(\partial r/\partial x) = x/r$ получаем

$$\left(\frac{\partial f}{\partial x}\right)_t = \left(\frac{\partial f}{\partial x}\right)_{t'} - \frac{x}{rv}\,\frac{\partial f}{\partial t},$$

так как по смыслу производных

$$\left(\frac{\partial f}{\partial t'}\right)_r = \frac{\partial f}{\partial t}.$$

В дальнейшем для сокращения значок t будем опускать, а вместо значка t' будем ставить штрих над знаком производной по координате, если дифференцирование производится при постоянном t'. При таком соглашении

$$\frac{\partial f}{\partial x} = \frac{\partial' f}{\partial x} - \frac{x}{rv}\,\frac{\partial f}{\partial t}.$$

Таким образом, операторы $\partial/\partial x$ и $\partial'/\partial x$ связаны соотношением

$$\frac{\partial}{\partial x} = \frac{\partial'}{\partial x} - \frac{x}{vr}\,\frac{\partial}{\partial t}, \tag{141.1}$$

а операторы ∇ и ∇' — соотношением

$$\nabla = \nabla' - \frac{r}{vr}\frac{\partial}{\partial t}. \tag{141.2}$$

Умножая оператор ∇ сначала скалярно, а затем векторно на векторную функцию $\mathbf{A}\left(\mathbf{r}, t - \dfrac{r}{v}\right)$, получим

$$\operatorname{div}\mathbf{A} = \operatorname{div}'\mathbf{A} - \frac{(\mathbf{r}\dot{\mathbf{A}})}{vr}, \tag{141.3}$$

$$\operatorname{rot}\mathbf{A} = \operatorname{rot}'\mathbf{A} - \frac{[\mathbf{r}\dot{\mathbf{A}}]}{vr}. \tag{141.4}$$

Используя соотношения

$$\operatorname{div}\mathbf{r} = 3, \quad \operatorname{div}(\varphi\mathbf{A}) = \varphi\operatorname{div}\mathbf{A} + (\mathbf{A}\operatorname{grad}\varphi),$$
$$\operatorname{rot}\mathbf{r} = 0, \quad \operatorname{rot}(\varphi\mathbf{A}) = \varphi\operatorname{rot}\mathbf{A} + [\operatorname{grad}\varphi\cdot\mathbf{A}]$$

и предполагая, что вектор $\mathbf{g}$ зависит только от аргумента t', из формул (141.3) и (141.4) найдем

$$\operatorname{div}\frac{\mathbf{g}}{r^n} = -n\frac{(\mathbf{gr})}{r^{n+2}} - \frac{(\dot{\mathbf{g}}\mathbf{r})}{vr^{n+1}}, \quad \operatorname{div}\frac{(\mathbf{gr})\mathbf{r}}{r^{n+2}} = -(n+2)\frac{(\mathbf{gr})}{r^{n+2}} - \frac{(\dot{\mathbf{g}}\mathbf{r})}{vr^{n+1}}, \tag{141.5}$$

$$\operatorname{rot}\frac{\mathbf{g}}{r^n} = n\frac{[\mathbf{gr}]}{r^{n+2}} + \frac{[\dot{\mathbf{g}}\mathbf{r}]}{vr^{n+1}}, \quad \operatorname{rot}\frac{(\mathbf{gr})\mathbf{r}}{r^{n+2}} = \frac{[\mathbf{gr}]}{r^{n+2}}.$$

3. После этих математических отступлений обратимся к задаче о поле излучения диполя Герца. Будем предполагать, что диполь неподвижен и помещен в однородную среду с диэлектрической и магнитной проницаемостями ε и μ. Начало координат поместим в точке нахождения диполя. Если бы дипольный момент $\mathbf{p}$ был постоянным, то вектор индукции $\mathbf{D}$ определялся бы формулой электростатики

$$\mathbf{D} = \frac{3(\mathbf{pr})}{r^5}\mathbf{r} - \frac{\mathbf{p}}{r^3}.$$

На малых расстояниях от диполя эта формула верна и в тех случаях, когда дипольный момент $\mathbf{p}$ меняется во времени. Но на больших расстояниях r формула не может быть верной, так как на прохождение таких расстояний электромагнитному возмущению, распространяющемуся со скоростью v, требуется конечное время r/v, в течение которого дипольный момент $\mathbf{p}$ может заметно измениться. Конечная скорость распространения электромагнитных возмущений, очевидно, будет учтена, если в предыдущей формуле значение дипольного момента $\mathbf{p}$ брать не в момент наблюдения t, а в более ранний момент $t' = t - r/v$, когда произошло излучение сигнала, дошедшего от диполя до точки наблюдения. Однако выражение, полученное таким путем, не будет удовлетворять уравнениям Максвелла. В самом деле, если вектор $\mathbf{D}$ меняется во времени, то это означает наличие тока смещения, пронизывающего контур окружности MN (рис. 336) и всякий другой контур, лежащий в плоскости этой окружности. Ток смещения возбудит магнитное поле с силовыми линиями, обвивающимися вокруг вектора $\mathbf{p}$. Возникнет

переменный магнитный поток через всякий контур, расположенный в меридиональной плоскости OPQ, а следовательно, и электрическое поле, лежащее в той же плоскости. Это поле в свою очередь даст ток смещения через всякую окружность, обвивающуюся вокруг оси вектора $\mathbf{p}$. В результате этого возникнет добавочное магнитное поле. Последнее в свою очередь возбудит добавочное электрическое поле и т. д. Из соображений симметрии ясно, что вектор $\mathbf{D}$ полного поля будет лежать в меридиональных плоскостях, а силовые линии вектора $\mathbf{H}$ будут иметь форму коаксиальных окружностей, центры которых лежат на оси вектора $\mathbf{p}$.

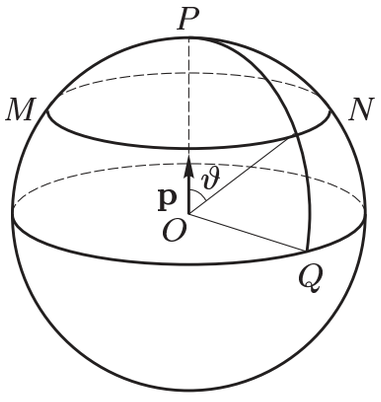


Рис. 336

4. Начнем с вычисления вектора $\mathbf{D}$. На малых расстояниях от диполя выражение для $\mathbf{D}$ должно переходить в соответствующую формулу электростатики и, следовательно, убывать обратно пропорционально кубу расстояния r. К электростатическому выражению надо добавить слагаемые, зависящие от производных вектора $\mathbf{p}$ по времени и убывающие более медленно. Разложив все слагаемые вдоль $\mathbf{r}$ и $\mathbf{p}$, будем искать $\mathbf{D}$ в виде

$$\mathbf{D}=\left[\frac{3(\mathbf{pr})}{r^5}\,\mathbf{r}-\frac{\mathbf{p}}{r^3}\right]+\left[a\frac{(\dot{\mathbf{p}}\mathbf{r})}{vr^4}\,\mathbf{r}+b\frac{\dot{\mathbf{p}}}{vr^2}\right]+\left[f\frac{(\ddot{\mathbf{p}}\mathbf{r})}{v^2r^3}\,\mathbf{r}+h\frac{\ddot{\mathbf{p}}}{v^2r}\right], \quad (141.6)$$

где a, b, f, h — постоянные коэффициенты. Они безразмерны благодаря введению в знаменатели множителей v и v^2. Члены нулевой и положительных степеней по r входить не могут, так как они не удовлетворяют условию на бесконечности, где электромагнитное поле должно обращаться в нуль. Точно так же не могут входить члены с более высокими отрицательными степенями по r, так как в противном случае не получилось бы предельное статическое поле при $r\to 0$. Предполагается, что значения $\mathbf{p}$ и его производных по времени берутся не в момент времени t, а в более ранний момент $t'=t-r/v$. Этим учтено, что электромагнитное поле распространяется со скоростью v. Конечно, возможность представления $\mathbf{D}$ в форме, приведенной выше, еще не доказана этими рассуждениями. Доказательство получится лишь после того, как будет показано, что выражения для всех векторов электромагнитного поля удовлетворяют системе уравнений Максвелла. Это делается ниже.

Вычислив по формулам (141.5) дивергенцию $\mathbf{D}$, получим

$$\operatorname{div}\mathbf{D}=-2(1+b)\frac{(\dot{\mathbf{p}}\mathbf{r})}{vr^4}-(a+b-f+h)\frac{(\ddot{\mathbf{p}}\mathbf{r})}{v^2r^3}-(f+h)\frac{(\dddot{\mathbf{p}}\mathbf{r})}{v^3r^2}=0.$$

Так как векторы $\dot{\mathbf{p}}$, $\ddot{\mathbf{p}}$, $\dddot{\mathbf{p}}$ независимы, то коэффициенты этого выражения должны обращаться в нуль. Это приводит к соотношениям

$$b=-1,\quad h=-f,\quad a-2f=1. \quad (141.7)$$

Соотношение $h = -f$ показывает, что последний член в выражении (141.6) есть вектор, перпендикулярный к $\mathbf{r}$, в чем легко убедиться, умножая этот вектор скалярно на $\mathbf{r}$. Отсюда следует, что поток этого вектора через любой участок сферы с центром в начале координат равен нулю.

Для определения коэффициентов a и f установим сначала вид выражения для магнитного поля $\mathbf{H}$. Вычислим поток $\Phi_{\text{эл}}$ вектора $\mathbf{D}$ через участок сферы, ограниченный окружностью MN (см. рис. 336). Для этого выражение (141.6) умножаем на единичный вектор $\mathbf{r}/r$ и интегрируем по поверхности указанного участка. Учитывая при этом соотношения (141.7), получим

$$\Phi_{\text{эл}} = \left(\frac{2p}{r} + \frac{a-1}{v}\,\dot{p}\right)\pi \sin^2\vartheta.$$

После этого поле $\mathbf{H}$ найдется из уравнения

$$\oint H\,dl = \frac{1}{c}\,\dot{\Phi}_{\text{эл}},$$

если в качестве контура интегрирования взять ту же окружность MN. Это дает

$$H = \left(\frac{\dot{p}}{cr^2} + \frac{a-1}{2cvr}\,\ddot{p}\right)\sin\vartheta,$$

если в векторной форме

$$\mathbf{H} = \frac{[\dot{\mathbf{p}}\mathbf{r}]}{cr^3} + \frac{a-1}{2cvr^2}\,[\ddot{\mathbf{p}}\mathbf{r}]. \tag{141.8}$$

Первое слагаемое здесь соответствует закону Био и Савара. Уравнение $\operatorname{div}\mathbf{H} = 0$ удовлетворяется автоматически, поскольку магнитное поле аксиально симметрично, а силовые линии его являются окружностями с центрами, расположенными на оси диполя (см. § 53). Однако коэффициент a все еще не определен. Для его определения и одновременно завершения всего доказательства надо воспользоваться последним уравнением Максвелла. Запишем его в виде

$$\operatorname{rot}\mathbf{D} = -\frac{\varepsilon\mu}{c}\,\dot{\mathbf{H}}. \tag{141.9}$$

Вычислив $\operatorname{rot}\mathbf{D}$ с использованием формул (141.5), найдем

$$\operatorname{rot}\mathbf{D} = (a-3)\frac{[\dot{\mathbf{p}}\mathbf{r}]}{vr^4} - \frac{[\ddot{\mathbf{p}}\mathbf{r}]}{v^2r^3} - f\frac{[\dddot{\mathbf{p}}\mathbf{r}]}{v^3r^2}.$$

С другой стороны, формула (141.8) с учетом соотношения $c^2 = \varepsilon\mu v^2$ дает

$$-\frac{\varepsilon\mu}{c}\dot{\mathbf{H}} = -\frac{[\ddot{\mathbf{p}}\mathbf{r}]}{v^2r^3} - \frac{a-1}{2v^3r^2}\,[\dddot{\mathbf{p}}\mathbf{r}].$$

Путем сравнения с предыдущим выражением находим $a = 3$, $f = 1$. Эти значения удовлетворяют и последнему уравнению системы (141.7).

Таким образом, все условия задачи удовлетворены, и поле диполя Герца можно представить в окончательном виде:

$$\mathbf{D} = \left[\frac{3(\mathbf{pr})}{r^5}\,\mathbf{r} - \frac{\mathbf{p}}{r^3}\right]_{t-r/v} + \left[\frac{3(\dot{\mathbf{p}}\mathbf{r})}{vr^4}\,\mathbf{r} - \frac{\dot{\mathbf{p}}}{vr^2}\right]_{t-r/v} + \left[\frac{(\ddot{\mathbf{p}}\mathbf{r})}{v^2r^3}\,\mathbf{r} - \frac{\ddot{\mathbf{p}}}{v^2r}\right]_{t-r/v},$$

$$\mathbf{H} = \frac{1}{cr^3}\,[\dot{\mathbf{p}}\mathbf{r}]_{t-r/v} + \frac{1}{cvr^2}\,[\ddot{\mathbf{p}}\mathbf{r}]_{t-r/v}. \tag{141.10}$$

Индекс $t - r/v$, поставленный внизу справа от каждого слагаемого, напоминает о том, что значения дипольного момента $\mathbf{p}$ и его производных по времени следует брать не в момент времени t, а в более ранний момент времени $t - r/v$. Такое «запаздывание» поля колеблющегося диполя означает, что электромагнитное поле не стоит на месте, а распространяется в виде возмущения со скоростью v.

При выводе формул (141.10) мы для краткости выражались так, как если бы векторы $\dot{\mathbf{p}}$ и $\ddot{\mathbf{p}}$ имели то же направление, что и сам вектор $\mathbf{p}$. В действительности это обстоятельство нигде не использовалось. Поэтому окончательные формулы (141.10) справедливы независимо от того, совпадают ли направления всех трех векторов $\mathbf{p}$, $\dot{\mathbf{p}}$, $\ddot{\mathbf{p}}$ или эти направления различны.

5. Поле, представляемое формулами (141.10), можно разложить на три части в зависимости от того, как они убывают с расстоянием от диполя r. При малых r, когда запаздывание несущественно, основную роль играет первое слагаемое. Оно меняется обратно пропорционально кубу расстояния r. Его мгновенное значение совпадает с полем статического диполя момента $\mathbf{p}$. Затем идут два члена: один электрический, другой магнитный, зависящие от $\dot{\mathbf{p}}$ и убывающие обратно пропорционально квадрату расстояния r. Магнитный член на малых расстояниях от диполя соответствует закону Био и Савара и представляет собой магнитное поле элемента тока $\dot{\mathbf{p}}$. Электрический член не допускает какой-либо простой интерпретации. Наконец, последние два члена, из которых один также электрический, а другой магнитный, зависят от второй производной дипольного момента. Эти члены убывают с расстоянием всего медленнее, а именно обратно пропорционально первой степени расстояния r. На больших расстояниях r эти члены становятся настолько большими по сравнению со всеми остальными членами, что последними можно полностью пренебречь. Тогда говорят о *волновой зоне* излучающего диполя. Таким образом, если пренебречь всеми более быстро убывающими слагаемыми, то *для поля в волновой зоне можно написать*

$$\mathbf{D} = \left[\frac{(\ddot{\mathbf{p}}\mathbf{r})}{v^2r^3}\,\mathbf{r} - \frac{\ddot{\mathbf{p}}}{v^2r}\right]_{t-r/v} = \frac{1}{v^2r^3}\,[[\ddot{\mathbf{p}}\mathbf{r}]\mathbf{r}]_{t-r/v},$$

$$\mathbf{H} = \frac{1}{cvr^2}\,[\ddot{\mathbf{p}}\mathbf{r}]_{t-r/v}. \tag{141.11}$$

Отсюда следует

$$\mathbf{D} = \frac{c}{v}[\mathbf{HN}], \quad \mathbf{H} = \frac{v}{c}[\mathbf{ND}], \tag{141.12}$$

где $\mathbf{N} = \mathbf{r}/r$ — единичный вектор, направленный вдоль радиуса-вектора $\mathbf{r}$. Мы видим, что векторы $\mathbf{D}$ и $\mathbf{H}$ взаимно перпендикулярны. Оба они перпендикулярны к радиусу-вектору $\mathbf{r}$ (рис. 337). Вектор $\mathbf{D}$ лежит в плоскости меридиана, а вектор $\mathbf{H}$ направлен вдоль параллели, если в качестве полярной оси взять направление вектора $\ddot{\mathbf{p}}$ (в момент времени $t - v/r$). Полученное поле в волновой зоне называется *сферической электромагнитной волной* [1]. Таким образом, соотношения между электрическим и магнитным векторами в волновой зоне такие же, как и для плоской электромагнитной волны (см. § 139). Это вполне естественно, так как на больших расстояниях от излучающего источника малый участок сферического волнового фронта должен вести себя практически как плоский.

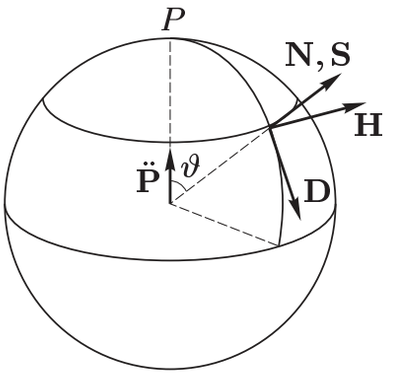


Рис. 337

6. Вектор плотности потока электромагнитной энергии равен

$$\mathbf{S} = \frac{c}{4\pi}[\mathbf{EH}] = \frac{c}{4\pi\varepsilon}[\mathbf{DH}].$$

Подставляя сюда значение $\mathbf{D}$ из (141.12), получим

$$\mathbf{S} = \frac{c^2}{4\pi\varepsilon v}\mathbf{H}^2\mathbf{N},$$

или с учетом выражения (141.11) для $\mathbf{H}$

$$\mathbf{S} = \frac{\sin^2\vartheta}{4\pi\varepsilon v^3 r^2}\ddot{\mathbf{p}}^2_{t-r/v}\mathbf{N}, \tag{141.13}$$

где ϑ — полярный угол, т. е. угол между векторами $\ddot{\mathbf{p}}$ и $\mathbf{r}$.

Выражение (141.13) показывает, что колеблющийся диполь излучает электромагнитную энергию в окружающее пространство. Поток энергии направлен *вдоль радиуса*, его плотность *обратно пропорциональна квадрату расстояния* r. Излучение *не изотропно*, его максимум приходится на угол $\vartheta = 90°$. В направлении вектора $\ddot{\mathbf{p}}$ излучение отсутствует. «Диаграмма направленности» излучения диполя Герца представлена на рис. 338. Чтобы найти интегральную мощность излучения, т. е. полную энергию — $d\mathscr{E}/dt$, излучаемую диполем в единицу

[1] Это название не совсем точно. Сферическая симметрия относится не к электромагнитному полю диполя, а только к *форме волновых фронтов*, т. е. поверхностей, до которых одновременно доходит возмущение: волновые фронты являются сферами, центры которых находятся в точке нахождения диполя в момент излучения.

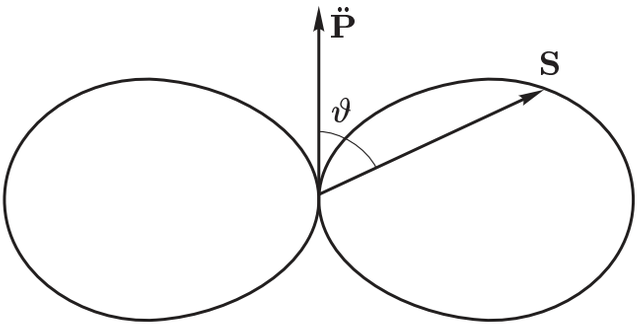


Рис. 338

времени, умножим модуль выражения (141.13) на элемент поверхности сферы $2\pi r^2 \sin\vartheta\, d\vartheta$ и проинтегрируем по всем направлениям. Таким путем найдем

$$-\frac{d\mathscr{E}}{dt} = \frac{2}{3\varepsilon v^3}\,\ddot{\mathbf{p}}^2_{t-r/v}. \tag{141.14}$$

7. Во всех предыдущих вычислениях относительно характера изменения дипольного момента $\mathbf{p}$ во времени не вводилось никаких предположений. Допустим теперь, что вектор $\mathbf{p}$ совершает синусоидальные колебания $\mathbf{p} = \mathbf{p}_0 e^{i\omega t}$. Тогда $\dot{\mathbf{p}} = i\omega\mathbf{p}$, $\ddot{\mathbf{p}} = -\omega^2\mathbf{p}$. Замена времени t на $t - r/v$ преобразует экспоненциальный множитель в $e^{i\omega(t-r/v)} = e^{i(\omega t - kr)}$, где $k = \omega/v$ — волновое число. В результате из формул (141.10) получаем

$$\mathbf{D} = \left[\left(\frac{3(\mathbf{p}_0\mathbf{r})}{r^5}\,\mathbf{r} - \frac{\mathbf{p}_0}{r^3}\right) + ik\left(\frac{3(\mathbf{p}_0\mathbf{r})}{r^4}\,\mathbf{r} - \frac{\mathbf{p}_0}{r^2}\right) - k^2\left(\frac{(\mathbf{p}_0\mathbf{r})}{r^3}\,\mathbf{r} - \frac{\mathbf{p}_0}{r}\right)\right] e^{i(\omega t - kr)},$$

и аналогично для $\mathbf{H}$. Это выражение интересно в том отношении, что оно позволяет точно сформулировать, когда можно пользоваться электростатическим приближением, а когда достаточно ограничиться волновым полем, убывающим обратно пропорционально r. Действительно, отношение каждого члена полученной суммы к предыдущему порядка kr. Если $kr \ll 1$, то справедливо *электростатическое приближение*. Это — *условие квазистационарности*, которое также можно представить в виде $r \ll \lambda$, где $\lambda = 2\pi/k$ — длина волны. Если же $kr \gg 1$, т. е. $r \gg \lambda$, то можно ограничиться только *волновым полем*. Условие $r \gg \lambda$ и определяет *волновую зону* колеблющегося диполя Герца.

Формула (141.14) для синусоидально колеблющегося диполя принимает вид

$$-\frac{d\mathscr{E}}{dt} = \frac{2\omega^4}{3\varepsilon v^3}\,p_0^2\cos^2(\omega t - kr). \tag{141.15}$$

Если ее усреднить по периоду колебаний T, то для средней мощности излучения получится

$$-\frac{d\mathscr{E}}{dt} = \frac{\omega^4}{3\varepsilon v^3}\,\mathbf{p}_0^2. \tag{141.16}$$

Мощность излучения пропорциональна *четвертой степени частоты* ω. По этой причине для увеличения мощности излучающих систем пользуются *короткими волнами.* Так, частота переменного тока городской сети $\nu = 50$ Гц, а длина волны $\lambda = c/\nu = 6 \cdot 10^7$ км. Такие токи практически не излучают. Но если $\lambda = 1$ км, то при той же амплитуде мощность излучения увеличится в $(6 \cdot 10^7)^4 \approx 10^{31}$ раз!

8. Полученные результаты можно использовать для нахождения электромагнитного поля одиночного движущегося точечного заряда e. С этой целью допустим, что диполь состоит из двух точечных зарядов: e и $-e$, из которых второй *бесконечно тяжелый.* Второй заряд можно считать неподвижным. Он создает только электростатическое поле по закону Кулона. Если последнее вычесть из полного поля диполя, то останется только поле движущегося заряда e. Представим дипольный момент в виде $\mathbf{p} = e(\mathbf{R} - \mathbf{R}')$, где $\mathbf{R}$ — радиус-вектор заряда e, а $\mathbf{R}'$ — заряда $-e$. Так как $\mathbf{R}' = \text{const}$, то отсюда получаем $\dot{\mathbf{p}} = e\dot{\mathbf{R}} = e\mathbf{v}$, $\ddot{\mathbf{p}} = e\dot{\mathbf{v}}$, где $\mathbf{v}$ — скорость, а $\dot{\mathbf{v}}$ — ускорение заряда e. Для излучаемой мощности находим

$$-\frac{d\mathscr{E}}{dt} = \frac{2e^2}{3\varepsilon v^3}\,\dot{\mathbf{v}}^2. \tag{141.17}$$

Таким образом, *электрический заряд излучает электромагнитную энергию, если он движется ускоренно.*

Понятно, что формула (141.17) справедлива только в *нерелятивистском приближении,* поскольку при ее выводе использовалась нерелятивистская кинематика.

ЗАДАЧИ

1. Найти выражение для поля излучения в однородной среде точечного магнитного диполя с магнитным моментом $\mathfrak{M}$.

Р е ш е н и е. Из уравнений Максвелла

$$\operatorname{rot}\mathbf{H} = \frac{1}{c}\,\frac{\partial \mathbf{D}}{\partial t}, \quad \operatorname{rot}\mathbf{E} = -\frac{1}{c}\,\frac{\partial \mathbf{B}}{\partial t}$$

и предельного условия при $r \to 0$ следует, что искомое поле получится из формулы (141.10) путем замены

$$\mathbf{p} \to \mathfrak{M}, \quad \mathbf{D} \to \mathbf{B}, \quad \mathbf{H} \to -\mathbf{E}.$$

2. По бесконечно длинному соленоиду течет переменный ток. Магнитное поле внутри соленоида определяется формулой $H = (4\pi/c)i$, где i — ток, приходящийся на единицу длины соленоида. Если на соленоид надеть проволочный виток, то магнитный поток через него будет меняться и в витке возникнет индукционный ток. Но магнитное поле в месте нахождения витка равно нулю, и не видно причины для возникновения индукционного тока. Разъяснить этот парадокс.

Ответ. Магнитное поле переменного тока вне соленоида не равно нулю (см. следующую задачу).

3. Рассчитать напряженности магнитного и электрического полей во внешнем пространстве бесконечно длинного и бесконечно тонкого прямолинейного соленоида, по которому течет переменный ток (см. предыдущую задачу). Ограничиться квазистационарным приближением, предполагая, что расстояние до оси соленоида R мало по сравнению с длиной волны λ и велико по сравнению с поперечными размерами соленоида.

Решение. Воспользуемся формулами для электромагнитного поля точечного магнитного диполя. Статические члены нам не понадобятся, а членами, содержащими $\ddot{\mathfrak{M}}$, которые существенны в волновой зоне, мы пренебрежем. Кроме того, в требуемом приближении можно пренебречь запаздыванием. С этими оговорками

$$\mathbf{B} = \frac{3(\ddot{\mathfrak{M}}\mathbf{r})}{vr^4}\,\mathbf{r} - \frac{\ddot{\mathfrak{M}}}{vr^2}, \quad \mathbf{E} = -\frac{1}{cr^3}\,[\ddot{\mathfrak{M}}\mathbf{r}]. \tag{141.18}$$

Теперь разобьем весь бесконечный соленоид на бесконечно короткие цилиндрики с высотой dx. Каждый из них представляет собой точечный магнитный диполь с дипольным моментом $d\mathfrak{M} = \dfrac{Si}{c}\,dx$, где S — площадь поперечного сечения соленоида. Вычислив по предыдущим формулам магнитное и электрическое поле каждого такого кольцевого тока и проинтегрировав по длине бесконечного соленоида, получим

$$B = \frac{\pi S}{2vcR}\,\frac{di}{dt}, \quad E = -\frac{2S}{Rc^2}\,\frac{di}{dt}$$

(предполагается, что $\mu = 1$). Магнитное поле направлено параллельно оси соленоида. Электрические силовые линии имеют форму коаксиальных окружностей, центры которых расположены на оси соленоида. При интегрировании по длине бесконечного соленоида статические члены, входящие в выражение для электромагнитного поля точечного магнитного диполя, дают нуль. Поэтому они и опущены в (141.18).

§ 142. Демонстрационные опыты с электромагнитными волнами

1. Электромагнитные волны впервые были получены и исследованы на опыте Генрихом Герцем в 1887–1888 гг. *Вибратор Герца*, с помощью которого он возбуждал электромагнитные волны, состоял из двух одинаковых металлических стержней V, V, разделенных искровым промежутком (рис. 339 слева). Обе половины вибратора соединялись с индуктором I. Когда напряжение на искровом промежутке достигало пробойного значения, искровой промежуток пробивался электрической искрой, замыкавшей обе половины вибратора. В вибраторе возникали затухающие собственные электрические колебания высокой частоты с узлами тока на его концах. Наиболее интенсивным было основное собственное колебание с пучностью тока посередине вибратора и с длиной волны, равной приблизительно удвоенному расстоянию между его концами. Собственные колебания с высшими частотами были

слабы и практически не играли роли. Для предотвращения ухода быстропеременных токов из вибратора в катушку индуктора последний присоединялся к вибратору через дроссели D. Такой вибратор, как уже указывалось в начале предыдущего параграфа, можно рассматривать как систему точечных диполей Герца, излучения которых складываются друг с другом и дают поле излучения всего вибратора.

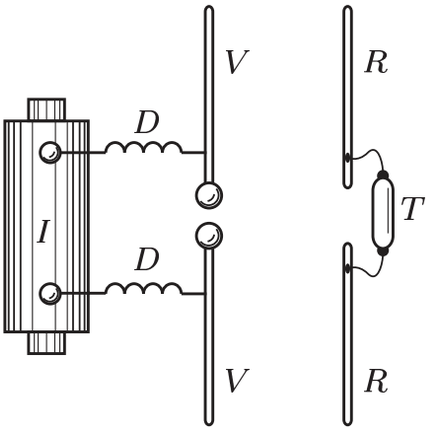


Рис. 339

Для обнаружения электромагнитных волн Герц пользовался *резонаторами* различной формы. Наиболее простым и удобным является *прямой открытый резонатор* (рис. 339 справа). По форме и размерам он должен быть тождествен с излучающим вибратором, чтобы совпадали их собственные частоты. Когда электромагнитная волна достигает резонатора, в нем возбуждаются вынужденные высокочастотные электрические токи. Они наиболее интенсивны при совпадении собственных частот вибратора и резонатора (резонанс). О появлении таких токов Герц судил по проскакиванию слабой электрической искорки в малом зазоре в середине резонатора или по свечению миниатюрной газоразрядной трубки T, включенной между обеими половинами резонатора.

2. Недостаток аппаратуры Герца состоит в том, что свободные колебания в его вибраторе *быстро затухают* и обладают малой мощностью. От этого недостатка свободны *ламповые генераторы*, позволяющие получать колебания *практически любой мощности* и притом *правильной синусоидальной формы.* Они могут быть использованы вместо индуктора I для возбуждения незатухающих колебаний в вибраторе. С этой целью можно, например, между половинами вибратора VV ввести один или несколько витков медной проволоки, индуктивно связанных с катушкой самоиндукции L генератора (рис. 340).

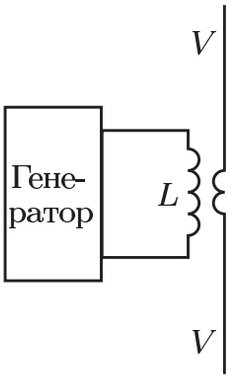


Рис. 340

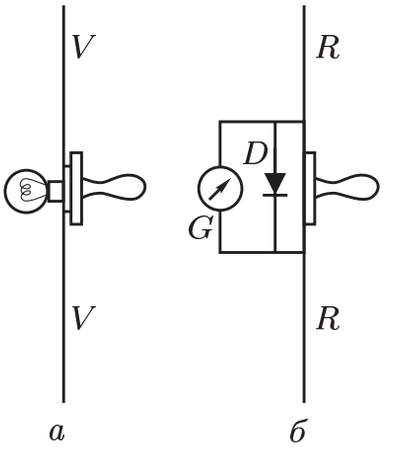


Рис. 341

Для обнаружения излучения вибратора в промежутке между концами приемного резонатора можно включить маленькую электрическую лампочку (рис. 341 *а*). Этот способ особенно удобен в демонстрационных опытах, когда расстояния между вибраторами и резонатором невелики и поэтому колебания, возникающие в резонаторе, достаточно сильны. Для обнаружения более слабых колебаний в промежуток между половинами резонатора можно ввести *кристаллический детектор* D и к его концам присоединить гальванометр постоянного тока $Г$ (рис. 341 *б*). Сопротивление детектора зависит от направления тока. Когда ток через детектор течет в пропускном направлении, то сопротивление детектора мало. Мало и напряжение на детекторе. Если же ток течет в запорном направлении, то напряжение на детекторе сильно возрастает. Поэтому в случае переменного тока на детекторе появляется постоянная составляющая напряжения, а через гальванометр течет ток *постоянного направления.* Этот способ применяется в большинстве опытов, описываемых ниже. Вместо детектора можно взять *когерер*, т. е. стеклянную трубочку, наполненную оксидированными металлическими опилками. Действие его разъясняется на примере следующей демонстрации.

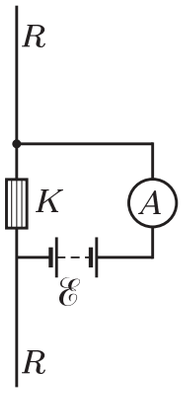


Рис. 342

3. Искровой вибратор, возбуждаемый индуктором (см. рис. 339 слева), помещается в соседней комнате, резонатор с когерером — в аудитории. В цепь когерера K включена батарея с амперметром (рис. 342). В обычном состоянии сопротивление опилок когерера очень велико, и амперметр не обнаруживает тока. При включении индуктора электромагнитные волны, излучаемые вибратором, попадая на когерер, вызывают искорки, проскакивающие между опилками. Происходит *спекание опилок*, и сопротивление их резко уменьшается (в сотни раз). В результате через амперметр начинает идти ток от батареи, не прекращающийся и при выключении индуктора. Если, однако, слегка постучать по когереру, то спекание опилок исчезает, сопротивление когерера сильно возрастает, и ток через амперметр прекращается. Амперметр снова обнаружит ток, если включить индуктор и тем самым заставить вибратор излучать электромагнитные волны.

4. В следующей демонстрации колебания в вибраторе возбуждаются ламповым генератором с длиной волны $\lambda = 3$ м и частотой $\nu = 10^8$ Гц по схеме рис. 340. Прием осуществляется резонатором с лампочкой (см. рис. 341 *а*). Длина вибратора и резонатора $\lambda/2 = 1{,}5$ м. Резонатор и вибратор устанавливаются на некотором расстоянии параллельно друг к другу и перпендикулярно к линии AB, соединяющей их центры, как указано на схематическом рис. 343. В этом случае лампочка горит ярко. Если поворачивать резонатор вокруг оси AB, то накал лампочки уменьшается, совсем пропадая при повороте на 90°. То же происходит, если поворачивать резонатор вокруг оси,

перпендикулярной к плоскости рисунка, — накал пропадает, когда резонатор устанавливается параллельно линии AB. Дело в том, что электромагнитные волны, излучаемые вибратором, *линейно поляризованы*: электрический вектор волны в точке B параллелен оси вибратора. Токи в резонаторе возбуждает только составляющая электрического вектора, направленная *вдоль оси резонатора*. Она, а следовательно, и возбуждаемый ею ток пропорциональны косинусу угла α между осями вибратора и резонатора. Следовательно, джоулево тепло, выделяющееся в лампочке, должно быть пропорционально квадрату того же косинуса. Если приемный резонатор перемещать параллельно самому себе вдоль линии AB, то при удалении от вибратора накал лампочки уменьшается не монотонно, а проходит через максимумы и минимумы. Дело в том, что волны, излучаемые вибратором, отражаются от стен аудитории и в ней устанавливается некоторое подобие стоячей волны с узлами и пучностями. Лампочка горит ярче в пучностях электрического поля и слабее в узлах.

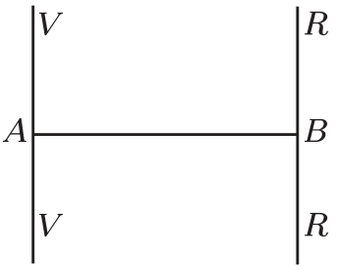


Рис. 343

5. Используя отражение волн от металлических поверхностей, Герц получал *направленные волны*, подобные параллельным пучкам света. Изгибанием металлических листов изготовлялись зеркала, имеющие форму параболических цилиндров. Их размеры в несколько раз превосходили длину электромагнитной волны (чтобы получить правильное отражение, не очень сильно искаженное дифракцией). Вдоль фокальной линии одного зеркала помещался небольшой вибратор, вдоль фокальной линии другого — приемный резонатор. Электромагнитные волны, исходящие от вибратора, после отражения от зеркала распространялись в виде направленного пучка параллельно главной оптической оси этого зеркала. Если зеркала были обращены отражающими поверхностями друг к другу, а их главные оптические оси совпадали, то этот направленный пучок отражался от второго зеркала, собирался в его фокусе и вызывал сильные электрические колебания резонатора (рис. 344, вибратор V и резонатор R установлены перпендикулярно к плоскости рисунка). Если же второе зеркало было смещено в сторону, то пучок не попадал на него, и колебания в резонаторе не возникали. Этот опыт доказывает *прямолинейное распространение электромагнитных волн*. Если же поворачивать одно из зеркал (вместе с вибратором или резонатором) вокруг общей главной оптической оси VR, то интенсивность приема будет убывать и обратится в нуль, когда вибратор и резонатор установятся взаимно перпендикулярно. Толкование этого опыта такое же, что и опыта, описанного в п. 4.

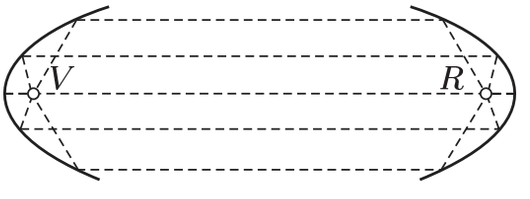


Рис. 344

6. Помещая между зеркалами на пути распространяющегося пучка (см. рис. 344) лист стекла, эбонита и других диэлектриков, а также металлические листы, можно убедиться, что все диэлектрики хорошо пропускают электромагнитные волны, а металлы не пропускают. Такое поведение металлов объясняется двумя обстоятельствами, тесно связанными между собой. Во-первых, *металлы сильно отражают электромагнитные волны.* Во-вторых, электрическое поле электромагнитной волны, частично проникшей в металл, возбуждает в нем электрические токи, энергия которых переходит в джоулево тепло, так что на протяжении толщи металла электромагнитная волна *полностью поглощается.*

7. Вместо сплошного металлического листа можно взять сетку из параллельных медных прямолинейных проводов (рис. 345) Если такую сетку поставить на пути пучка электромагнитных волн и притом так, чтобы проволоки были параллельны вибратору и резонатору, то электромагнитные волны через нее практически не проходят. В этом случае вдоль проволок возбуждаются электрические токи и происходит поглощение электромагнитной энергии с выделением джоулева тепла. Если же проволоки сетки перпендикулярны к вибратору и резонатору, то они будут перпендикулярны и к электрическому вектору **E**. В этом случае ток в проволоках не возникает и волна проходит через сетку практически так же, как через диэлектрическую пластинку. Описанный опыт дает новое доказательство линейной поляризации волн, излучаемых вибратором (см. п. 4) Заметим также, что проволочная сетка действует *как сплошная пластинка*, если расстояние между проволоками мало по сравнению с длиной волны. Проволоки делают только такую «пластинку» *анизотропной.* Этот вопрос будет подробно разобран в т. IV нашего курса — в оптике.

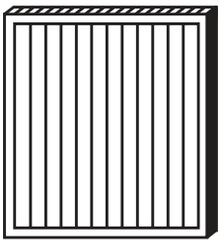


Рис. 345

Поворотом зеркал вокруг их общей оптической оси установим вибратор и резонатор взаимно перпендикулярно. Приема не будет. Если же между вибратором и резонатором ввести проволочную сетку под некоторым углом к ним, то прием возобновится. Для понимания опыта разложим электрический вектор перед сеткой на две составляющие: параллельно и перпендикулярно к проволокам. Вторая составляющая будет пропущена сеткой. А так как ее электрическое поле имеет составляющую вдоль оси резонатора, то в нем возбудятся колебания.

8. Направляя электромагнитную волну от параболического зеркала вибратора на плоское зеркало AB (металлический лист, рис. 346), можно наблюдать *отражение волны.* Оказалось, что (как и для световых лучей) угол падения волны равен углу отражения, а нормали к падающей и отраженной волнам лежат в одной плоскости с нормалью к поверхности зеркала. Кроме того, легко убедиться, что металлы значительно лучше отражают электромагнитные волны, чем диэлектрики.

9. Можно наблюдать и *преломление электромагнитных волн*, изготовив большую призму из диэлектрика (парафин, сера и т. п., рис. 347).

Таким путем можно убедиться в справедливости *закона преломления Снеллиуса*: отношение синуса угла падения к синусу угла преломления не зависит от угла падения. Более того, опыт подтвердил соотношение

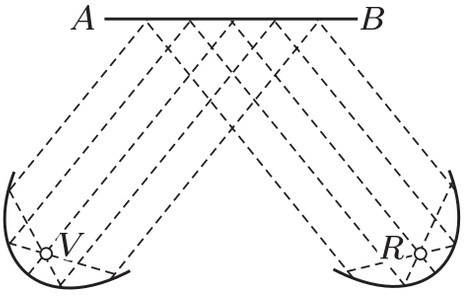


Рис. 346

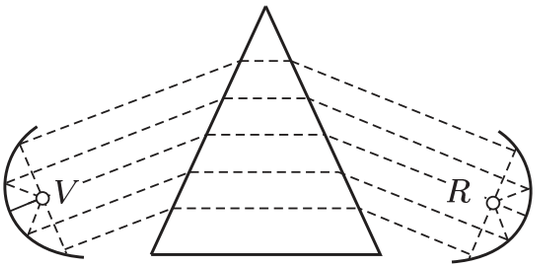


Рис. 347

$n = \sqrt{\varepsilon}$ между показателем преломления среды n и диэлектрической проницаемостью ε. Это соотношение было теоретически установлено Максвеллом. Оно будет подробно рассмотрено в т. IV.

10. Можно получить и *стоячие электромагнитные волны.* Для этого волну от параболического зеркала вибратора надо нормально направить на плоское металлическое зеркало. От наложения отраженной волны на падающую образуется система стоячих волн с узлом электрического поля на поверхности зеркала (см. § 145). Положение пучностей и узлов электрического поля в такой волне исследуется с помощью резонатора обычным путем. Расстояние между соседними пучностями (или узлами) равно половине длины волны. Измерив это расстояние, можно найти длину волны λ. В опытах Герца длина волны была 60 см, а высота параболических зеркал 2 м.

11. В анизотропных кристаллах должно наблюдаться *двойное преломление электромагнитных волн*, аналогичное двойному преломлению световых лучей. Этот вывод был сделан Максвеллом. Явление двойного преломления будет подробно изучено нами в т. IV. Здесь же заметим, что вывод Максвелла был экспериментально проверен П. Н. Лебедевым (1866–1912) в 1895 г. Для опытов необходимо было получить очень короткие электромагнитные волны, так как размеры кристаллов обычно не превышают нескольких сантиметров. С помощью миниатюрного искрового вибратора Лебедев получил волны с длиной около 6 мм — в сто раз короче волн Герца. С помощью таких волн он воспроизвел в кристаллах серы все явления двойного преломления, которые обычно наблюдают со световыми лучами. В частности, из кристаллов серы с эбонитовой прокладкой он изготовил прибор, вполне аналогичный *призме Николя*, применяемой в оптике для получения *поляризованного света.*

Опыты Герца сыграли решающую роль для доказательства и быстрого признания электродинамики Максвелла и электромагнитной

теории света. Через семь лет после этих опытов (1895) электромагнитные волны нашли применение в *беспроволочной телеграфии и телефонии* (*радио*). Пионером в этой области был наш соотечественник А. С. Попов.

§ 143. Волны вдоль проводов

1. Рассмотрим два одинаковых параллельных провода, в которых с помощью генератора могут возбуждаться переменные токи высокой частоты. Такие два провода называются *системой Лехера.* Связь проводов с генератором может быть либо *емкостной*, либо *индуктивной* (рис. 348). Примем, что *по отношению к поперечным размерам системы выполнено условие квазистационарности.* Это значит, что расстояние между проводами должно быть весьма мало по сравнению с длиной волны. В то же время будем предполагать провода длинными — на их длине должно укладываться по меньшей мере несколько волн. Поэтому электрические токи в проводах *не квазистационарны*, сила тока $I(x)$, а также линейная плотность электрического заряда $q(x)$ существенно меняются вдоль проводов (ось X направлена параллельно проводам). В силу предполагаемой симметрии ток $I(x)$, текущий вдоль одного из проводов, равен и противоположно направлен току, текущему напротив него вдоль другого провода. Аналогичное утверждение относится к величинам зарядов на проводах и их знакам. Электрическое напряжение между проводами, измеренное вдоль перпендикуляра к ним, будем обозначать через $V(x)$. Указание «пути», вдоль которого измеряется напряжение, т. е. интеграл $\int E\,dl$, необходимо потому, что в переменном электромагнитном поле значение этого интеграла *зависит от формы пути.*

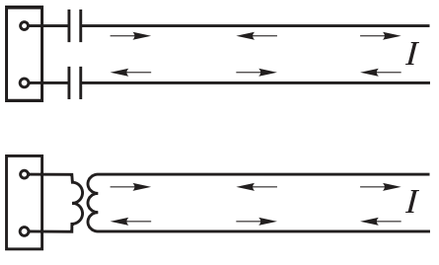


Рис. 348

При изложении теории распространения волн вдоль проводов мы не будем пользоваться уравнением Максвелла с током смещения, а изберем более простой путь, которым пользовался еще Кирхгоф задолго до появления максвелловской теории. Наше рассмотрение применимо также к задаче о распространении волн вдоль кабеля, состоящего из двух коаксиальных цилиндрических проводов: наружного полого и помещенного внутри него сплошного, пространство между которыми заполнено однородным диэлектриком. Наружный проводник играет роль одного провода системы Лехера, внутренний — другого. Однако оно совсем неприменимо к линии, состоящей только из одного провода.

Рис. 349

2. Возьмем на одном из проводов системы Лехера бесконечно малый участок dx (рис. 349). Через конец A за время dt внутрь рассмат- риваемого участка входит

электрический заряд $I(x)dt$, а через конец D выходит заряд $I(x + dx)dt$. Избыток входящего электричества над выходящим составляет

$$[I(x) - I(x + dx)]dt = -\frac{\partial I}{\partial x}\, dx dt.$$

Ту же величину можно представить в виде $\dot{q}dxdt$. Таким образом,

$$\dot{q} = -\frac{\partial I}{\partial x}. \tag{143.1}$$

Это уравнение выражает закон сохранения электрического заряда.

Применим теперь к контуру $ADCB$ уравнение $\oint E\, dl = -\frac{1}{c}\dot{\Phi}\, dx$, где $\Phi(x)\, dx$ — магнитный поток, пронизывающий этот контур. Имеем

$$\int\limits_{DC} E\, dl = V(x + dx),$$

$$\int\limits_{BA} E\, dl = -V(x),$$

$$\int\limits_{DC+BA} E\, dl = V(x + dx) - V(x) = \frac{\partial V}{\partial x}\, dx,$$

$$\int\limits_{AD+CB} E\, dl = RI\, dx,$$

где $R\, dx$ — суммарное сопротивление элементов проводов AD и CB. Следовательно,

$$\frac{\partial V}{\partial x} + RI = -\frac{1}{c}\dot{\Phi}. \tag{143.2}$$

Величины q, Φ и R суть заряд, магнитный поток и сопротивление единицы длины рассматриваемой *двухпроводной линии.* В дальнейшем всюду будем предполагать, что сопротивление R равно нулю. Используем теперь условие квазистационарности по отношению к поперечным размерам линии. При выполнении этого условия можно ввести емкость C и индуктивность L *единицы длины линии.* Эти величины определяются соотношениями

$$q = CV, \quad \Phi = LI. \tag{143.3}$$

Исключив из уравнений (143.1) и (143.2) величины q и Φ, при $R = 0$ получим

$$\frac{\partial I}{\partial x} = -C\frac{\partial V}{\partial t}, \quad \frac{\partial V}{\partial x} = -\frac{L}{c^2}\frac{\partial I}{\partial t}. \tag{143.4}$$

Эти уравнения *формально тождественны* с уравнениями (139.3). Поэтому все следствия, полученные из уравнений (139.3), остаются справедливыми и для системы Лехера. Надо только сделать замену $H \to I$, $E \to V$, $\varepsilon \to cC$, $\mu \to L/c$. Таким путем приходим к выводу, что напряжение и ток распространяются вдоль проводов в виде волны со

скоростью

$$v = \frac{c}{\sqrt{LC}}. \tag{143.5}$$

В бегущей волне напряжение и ток связаны соотношением

$$V = \pm W I, \tag{143.6}$$

где

$$W = \frac{1}{c}\sqrt{\frac{L}{C}}. \tag{143.7}$$

Величина W называется *волновым сопротивлением линии.* Знак плюс относится к волне, распространяющейся в *положительном*, а минус — в *отрицательном* направлении оси X. Аналогия формулы (143.6) с законом Ома чисто внешняя, так как V есть напряжение *между проводами*, т. е. вдоль прямой, перпендикулярной к току, тогда как в законе Ома речь идет о напряжении *вдоль провода*, по которому течет ток.

Для тонких цилиндрических проводов радиуса a

$$L = 2\mu \ln(d/a), \quad C = \frac{\varepsilon}{2\ln(d/a)}, \tag{143.8}$$

где d — расстояние между проводами. Поэтому формула (143.5) преобразуется в

$$v = c/\sqrt{\varepsilon\mu}, \tag{143.9}$$

т. е. v совпадает со скоростью распространения волн в *свободном пространстве.*

3. Возможность распространения волн, к которой мы пришли на основе уравнений (143.1) и (143.2), может показаться неожиданной. Мы не вводили тока смещения, благодаря которому и возникает распространение электромагнитного поля в пространстве с конечной скоростью. Уравнения, которыми мы пользовались, не отличаются от уравнений домаксвелловской электродинамики с ее представлением о мгновенности распространения взаимодействий между зарядами и токами. Ответ на этот вопрос заключается в следующем. Из уравнения Максвелла с током смещения (82.1) или (82.1а), как известно, вытекает уравнение непрерывности, выражающее закон сохранения электрического заряда (см. § 82). Для решения некоторых частных задач это последнее уравнение может заменить уравнение Максвелла (82.1). Так обстоит дело в случае системы Лехера из двух одинаковых параллельных проводов. Но уравнение непрерывности *не специфично* для теории Максвелла, оно справедливо также и в теории действия на расстоянии. Вот почему теория действия на расстоянии привела в рассматриваемом вопросе к правильному решению — волне, распространяющейся с конечной скоростью. Однако необходимо заметить, что в этой теории речь идет не о распространении электромагнитного поля, а о распространении *электрических зарядов*, точнее, *состояния*

электризации вдоль проводов. Возможность этого вполне совместима с представлением о действии на расстоянии, как видно из приведенного анализа. Электромагнитные же волны, распространяющиеся с конечной скоростью *в вакууме* или *в диэлектрической среде*, специфичны *только для максвелловской теории.* Разумеется, замена уравнения Максвелла с током смещения уравнением непрерывности допустима далеко не всегда. Даже и задача о системе Лехера, которую мы разобрали, для своего строгого и полного решения требует использования всей системы уравнений Максвелла.

4. Выше о форме колебаний и волн в системе Лехера не вводилось никаких предположений. Допустим теперь, что колебания и волны *синусоидальны.* Если волна *бегущая*, то ток I и напряжение V колеблются в *одинаковых фазах.* Это непосредственно следует из соотношения (143.6), поскольку волновое сопротивление W — величина *вещественная.* В бегущей волне электрический и магнитный векторы перпендикулярны к проводам, а вектор Пойнтинга параллелен им. Поэтому возникает *поток энергии параллельно проводам*, как это и должно быть в бегущей волне.

В неограниченной линии, возбуждаемой с одного конца, всегда возникала бы бегущая волна. Пусть теперь линия *ограничена.* Тогда на ее концах в любой момент времени должны соблюдаться определенные граничные условия. Если концы обоих проводов *свободны*, то на этих концах должен обращаться в нуль *электрический ток* I. Если же линия *закорочена* (т. е. ее концы соединены проводом с пренебрежимо малым сопротивлением R_1), то на конце линии должно обращаться в нуль напряжение V. Действительно, согласно закону Ома ток в соединительном проводе равен $I = V/R_1$. Если бы напряжение V на конце линии оставалось конечным, то при $R_1 \to 0$ получился бы бесконечно большой ток I, что физически невозможно. Бегущая волна не удовлетворяет ни одному из этих граничных условий. Природа здесь, как и во всех аналогичных случаях, находит простой выход из создавшегося положения: достигнув конца линии, волна *отражается* и бежит в обратном направлении. От наложения падающей и отраженной волн в линии возникает *стоячая волна.* Нет необходимости для такой волны повторять рассуждения § 140, проведенные для стоячих волн в свободном пространстве. Достаточно привести результаты. Колебания тока и напряжения в стоячей волне сдвинуты по фазе на $\pi/2$. В случае незамкнутой линии на ее конце возникает узел тока и пучность напряжения. Если же линия закорочена, то появляется пучность тока и узел напряжения. Узлы тока являются пучностями напряжения, а пучности тока — узлами напряжения. Посередине между двумя пучностями тока находится пучность напряжения, посередине между двумя узлами тока — узел напряжения и т. д. Как в тех, так и в других узлах обращается в нуль один из векторов $\mathbf{E}$ или $\mathbf{H}$, а с ним и вектор Пойнтинга $\mathbf{S} = (c/4\pi)[\mathbf{EH}]$. Электромагнитная энергия совершает колебательное движение между пучностью (узлом) тока и соседним узлом (пучностью) напряжения, но она *не может переходить* с одной стороны узла (пучности) на

другую. Никакого направленного переноса энергии вдоль всей линии не происходит, как и должно быть в стоячей волне. Наиболее сильные вынужденные колебания тока и напряжения в линии возникают при тех же условиях, как и во всякой стоячей волне. Если, например, на обоих концах линии — пучность тока (напряжения), то это происходит тогда, когда в линии укладывается целое число полуволн.

5. Для демонстрации стоячих волн в системе Лехера берут два длинных и толстых голых провода и натягивают их от одного конца аудитории до другого. Провода располагаются в горизонтальной плоскости на расстоянии около 10 см друг от друга. Линия питается с одного конца ламповым генератором с длиной волны 3 м. Колебательный контур генератора имеет конденсатор переменной емкости, что позволяет настраивать линию в резонанс. Второй конец линии может быть либо свободным, либо закороченным. В качестве индикаторов для исследования стоячих волн применяются *неоновые лампочки*, включаемые между обоими проводами линии, а также небольшие *проволочные витки* с включенными в них лампочками от карманных фонариков. Витки располагаются горизонтально между проводами, как указано на рис. 350. Такие индикаторы лишь незначительно искажают поле

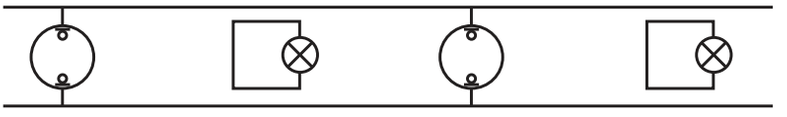


Рис. 350

стоячей волны. Неоновая лампочка загорается в тех местах, где есть электрическое поле, создающее на ней напряжение, превосходящее потенциал зажигания. Она светится наиболее ярко в пучностях и гаснет в узлах напряжения. Электрические лампочки, напротив, горят наиболее ярко в пучностях и гаснут в узлах тока. Действительно, когда виток помещен в пучность тока, он пронизывается максимальным магнитным потоком. Так как этот магнитный поток переменный, то в витке возбуждается индукционный ток, и лампочка горит наиболее ярко. Если же виток находится в узле тока, то магнитный поток равен нулю, и индукционный ток не возникает. С помощью этих индикаторов легко исследовать распределение пучностей и узлов тока и напряжения.

6. Стоячие волны можно демонстрировать также в катушках длиной 2–3 м, состоящих из однослойных витков тонкой проволоки (общая длина проволоки несколько километров). Один конец катушки присоединяют к ламповому генератору (длина волны $\lambda \approx 300$ м), другой замыкают. В проволоке возникают стоячие волны тока, а в окружающем пространстве — переменные электромагнитные поля. Пучности электрического поля можно наблюдать, если параллельно катушке расположить длинную газосветную стеклянную трубку. Газ светится неравномерно, расстояния между максимумами свечения составляют обычно 20–30 см. Можно взять стеклянную трубку в форме кольца,

наполненную неоном, и надеть ее на катушку. При перемещении вдоль катушки трубка периодически зажигается и гаснет. Можно также пользоваться трубками различной формы и наблюдать их свечение при перемещении вокруг катушки. Произвести количественные расчеты в этих случаях затруднительно.

ЗАДАЧА

Провода лехеровой системы соединены между собой мостиком из катушки индуктивности, омического сопротивления и конденсатора (рис. 351). Комплексное сопротивление (импеданс) мостика равно Z. Провода и однородные среды, в которые они погружены, по разные стороны мостика могут быть разными, так что волновое сопротивление линии перед мостиком равно W, а за мостиком W'. Если на мостик падает волна, то она частично отражается, частично проходит дальше. Определить у мостика отраженную (I_r) и прошедшую (I_d) волны тока, а также ток через мостик I, если ток падающей волны равен I_e.

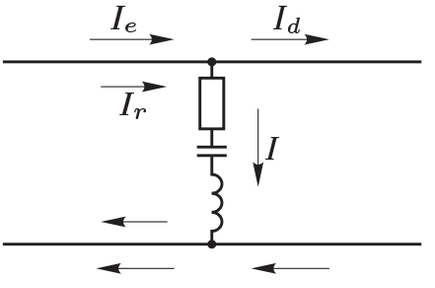


Рис. 351

Р е ш е н и е. Напряжение в падающей, отраженной и прошедшей волнах у мостика:

$$V_e = WI_e, \quad V_r = -WI_r, \quad V_d = W'I_d$$

(минус во второй формуле потому, что отраженная волна идет назад). По первому правилу Кирхгофа

$$I_e + I_r = I + I_d.$$

Напряжение между концами мостика можно представить следующими тремя выражениями: ZI, V_e+V_r, V_d. Приравнивая их и выражая напряжения через токи, получим два уравнения:

$$ZI = W(I_e - I_r), \quad ZI = W'I_d.$$

Решая эти уравнения совместно с предыдущим, найдем

$$\frac{I_r}{I_e} = \frac{WW' + (W - W')Z}{\Delta}, \quad \frac{I_d}{I_e} = \frac{2WZ}{\Delta}, \quad \frac{I}{I_e} = \frac{2WW'}{\Delta},$$

где

$$\Delta = WW' + (W + W')Z.$$

Если мостик поставлен на конец линии, то прошедшей волны не будет ($W' = \infty$). В этом случае

$$I_r = \frac{W - Z}{W + Z}\, I_e, \quad I = \frac{2W}{W + Z}\, I_e.$$

Отсюда легко получить амплитуды и фазы токов I_e и I для закороченной линии ($Z = 0$) и для линии со свободными концами ($Z = \infty$). В обоих случаях отражение полное, но с разными фазами. Если сопротивление мостика чисто реактивное ($Z = iX$), то $|I_e| = |I_r|$: отражение также полное, наличие реактивного сопротивления сказывается только на положении узлов и пучностей. При $Z = W$ получаем $I_r = 0$, т. е. волна совсем не отражается. Ток, подойдя к мостику по одному из проводов, проходит через мостик, попадает в другой провод и возвращается к началу линии.

§ 144. Свойства быстропеременных токов. Скин-эффект

1. В металлах из-за их высокой проводимости основную роль играет *ток проводимости*, по сравнению с которым током смещения можно полностью пренебречь. Благодаря этому проникновение магнитного поля в металл в математическом отношении вполне аналогично *процессу диффузии.*

Напомним сначала простейшие уравнения, которыми описывается процесс диффузии частиц вещества. Пусть частицы помещены в длинную прямолинейную трубу, направленную параллельно оси X (рис. 352). Поперечное сечение трубы для простоты будем считать равным единице. Концентрация частиц n является функцией координаты x и времени t. Если $j(x)$ — плотность диффузионного тока, то приращение в единицу времени числа частиц в слое, заштрихованном на рисунке, представится разностью $j(x) - j(x+dx) = -(\partial j/\partial x)\,dx$. То же приращение равно $(\partial n/\partial t)\,dx$. Таким образом,

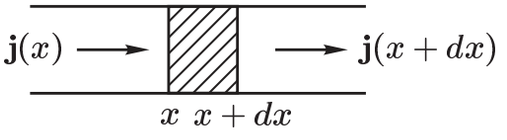


Рис. 352

$$\frac{\partial j}{\partial x} = -\dot{n}, \quad j = -D\frac{\partial n}{\partial x}, \tag{144.1}$$

где D — коэффициент диффузии. Выделим теперь в трубе столб вещества длиной l. Пусть в начальный момент времени на левом конце этого столба концентрация n отлична от нуля, а на правом обращается в нуль. Допустим также, что концентрация n равна нулю во всех точках трубы, расположенных правее выделенного столба. Плотность диффузионного тока будет $j \sim Dn/l$. Ту же величину можно представить выражением $j = nv$, где v — скорость, с которой вещество при диффузии распространяется вдоль трубы. Отсюда находим $v \sim D/l$. Время τ, за которое вещество диффундирует на расстояние l, будет $\tau \approx l/v \approx l^2/D$, и, следовательно,

$$l \sim \sqrt{D\tau}. \tag{144.2}$$

2. Перейдем теперь к *задаче о проникновении электромагнитного поля в металл.* В пренебрежении током смещения этот процесс описывается уравнениями

$$\operatorname{rot}\mathbf{H} = \frac{4\pi}{c}\,\mathbf{j}, \quad \operatorname{rot}\mathbf{E} = -\frac{1}{c}\,\dot{\mathbf{B}}.$$

Пользуясь соотношениями

$$\mathbf{j} = \lambda\mathbf{E} \quad \text{и} \quad \mathbf{B} = \mu\mathbf{H},$$

исключим из этих уравнений векторы $\mathbf{E}$ и $\mathbf{H}$. Получим

$$\mathbf{j} = \frac{c}{4\pi\mu}\,\operatorname{rot}\mathbf{B}, \quad \operatorname{rot}\mathbf{j} = -\frac{\lambda}{c}\,\dot{\mathbf{B}}.$$

Допустим, что все величины зависят только от *одной координаты* x, и для этого случая запишем полученные уравнения в координатной форме. Поступая так же, как при выводе формул (139.2), убедимся, что уравнения распадутся на две однотипные группы. В одну группу войдут уравнения, содержащие j_y и B_z, в другую — j_z и B_y. Поскольку эти две группы уравнений *независимы*, достаточно исследовать одну из них. Поэтому с самого начала можно предположить, что ток $\mathbf{j}$ течет параллельно оси Y, а магнитное поле $\mathbf{B}$ направлено вдоль оси Z. В таком случае

$$\frac{\partial j}{\partial x} = -\frac{\lambda}{c}\,\dot{B}, \quad j = -\frac{c^2}{4\pi\mu\lambda}\,\frac{\partial}{\partial x}\left(\frac{\lambda}{c}\,B\right). \tag{144.3}$$

Эти уравнения переходят в (144.1), если произвести замену $n \to (\lambda/c)B$ и положить

$$D = \frac{c^2}{4\pi\mu\lambda}. \tag{144.4}$$

Таким образом, распространение электромагнитного поля в металле описывается *уравнениями диффузии*, причем *роль коэффициента диффузии играет величина* (144.4).

Допустим теперь, что по плоской поверхности металла начинает течь переменный ток, периодически меняющийся во времени с периодом T. Через половину периода $T/2$ ток изменит направление на противоположное. За это время, как видно из формулы (144.2), магнитное поле, а с ним и электрический ток успеют проникнуть внутрь металла на глубину

$$l \sim \sqrt{D\frac{T}{2}} \sim c\sqrt{\frac{T}{8\pi\mu\lambda}} \sim \frac{c}{\sqrt{8\pi\mu\lambda\nu}}, \tag{144.5}$$

где $\nu = 1/T$ — частота переменного тока.

Таким образом, плотность тока в металле *убывает с глубиной.* Ток течет практически только в поверхностном слое металла, толщина которого порядка l. Это явление называется *скин-эффектом* (от английского слова «скин», что означает «кожа»), а величина l — *глубиной проникновения тока* (или магнитного поля) в металл. Глубина проникновения убывает с частотой, как $1/\sqrt{\nu}$. Возьмем в качестве примера медь. Для нее $\lambda = 5{,}4\cdot 10^{17}$ с$^{-1}$ (величину μ принимаем равной единице). Тогда для «городского» тока $\nu = 50$ Гц, и по формуле (144.5) находим $l \approx 0{,}7$ см~ 1 см. При $\nu = 5000$ Гц $l \sim 0{,}1$ см, при $\nu = 5\cdot 10^5$ Гц $l \sim 0{,}01$ см и т. д. Таким образом, токи высокой частоты текут только в тонком поверхностном слое металла. В целях упрощения вычислений мы рассмотрели явление скин-эффекта для случая бесконечного металла, ограниченного плоскостью. Но скин-эффект, конечно, имеет место и в тех случаях, когда токи текут по проводам. Только в этом случае выражение для глубины проникновения будет отличаться от (144.5) множителем, зависящим от радиуса провода. Впрочем, этот множитель порядка единицы, а потому для понимания явления он никакой существенной роли не играет.

3. Дадим теперь качественное объяснение скин-эффекта. Рассмотрим цилиндрический провод, по которому течет переменный ток I (рис. 353). Возьмем в меридиональном сечении провода прямоугольный контур $MM'N'N$, одна из сторон которого MM' проходит по оси провода. Пусть в рассматриваемый момент ток I течет вверх. Магнитные силовые линии будут обвиваться вокруг оси провода, как указано на рис. 353. Допустим теперь, что ток I, а с ним и магнитный поток Φ через контур $NMM'N'$ возрастают. Тогда возникает электрическое поле индукции $\mathbf{E}_{\text{инд}}$, направленное на NN' вверх, а на MM' — вниз. Это поле будет усиливать ток на периферии NN' и ослаблять на оси MM'. Если же ток I убывает, то поле $\mathbf{E}_{\text{инд}}$ изменит направление — теперь ток будет усиливаться на оси MM' и ослабляться на периферии NN'. Таким образом, каково бы ни было изменение тока I, индукционный ток способствует этому изменению на периферии и препятствует на оси провода. В результате этого амплитуда установившихся колебаний тока на NN' становится больше, чем на оси MM'. Чем дальше от оси провода, тем больше амплитуда колебаний тока, так как магнитный поток Φ через контур $MM'N'N$ увеличивается с возрастанием расстояния MN. Такое перераспределение тока по сечению провода и есть скин-эффект.

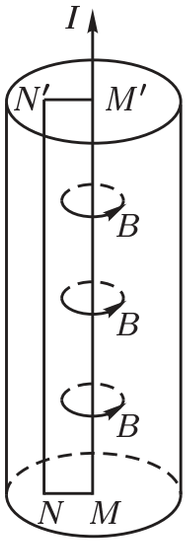


Рис. 353

Концентрация тока на поверхности провода приводит к тому, что сплошной провод начинает вести себя как полый с толщиной стенок, равной приблизительно глубине проникновения l. От этого сопротивление провода увеличивается, а индуктивность уменьшается. Поэтому для быстропеременных токов провода выгодно делать полыми, в виде труб, а не в виде сплошных проволок, так как внутренние приосевые части таких проволок в этих случаях бесполезны. Для более конкретного представления о влиянии скин-эффекта на сопротивление проводов приведем следующие данные. При частоте переменного тока $\nu = 50$ Гц сопротивление медного провода диаметром 2 см по сравнению с сопротивлением постоянному току увеличивается приблизительно на 3 %, а для проволоки диаметром 2 мм — всего на 0,0003 %. Для быстропеременных токов с частотой $\nu = 10^6$ Гц и провода с диаметром 2 мм сопротивление увеличивается почти в 7 раз. Если проволока не прямая, а навита, например, на катушку, то распределение тока по сечению проволоки становится несимметричным: плотность тока больше на стороне проволоки, обращенной внутрь катушки.

4. При быстрых электрических колебаниях магнитное поле меняется также очень быстро. Поэтому индукционные действия быстропеременных токов могут быть очень сильными. Их можно еще больше усилить, если воспользоваться явлением резонанса. На этом основано действие высокочастотного *резонансного трансформатора Тесла* (1856–1943). Это воздушный трансформатор, первичная обмотка которого L_1 состоит из небольшого числа витков толстой проволоки

и входит в состав колебательного контура, содержащего конденсатор (лейденскую банку) C и искровой промежуток F. Колебательный контур через дроссели D соединен с индуктором I (рис. 354). Вторичной обмоткой L_2, которая помещается внутри первичной, служит спираль, состоящая из многих витков тонкой проволоки, намотанной в один ряд на изолирующий цилиндр. Обе обмотки, а также отдельные витки вторичной обмотки должны быть хорошо изолированы друг от друга. Вторичная обмотка поэтому часто парафинируется и погружается в масло. Когда пробивается искровой промежуток F, в первичной обмотке возникают высокочастотные электрические колебания, а во вторичной — индукционные токи той же частоты. Катушка L_2 обладает собственными частотами, и поэтому колебания высокочастотных индукционных токов особенно сильны при совпадении одной из этих частот (обычно основной) с собственной частотой колебательного контура. В колебаниях с основной частотой посередине катушки возникает пучность тока, электрическое поле максимально на краях. Амплитуда колебаний напряжения между концами вторичной обмотки получается значительно больше амплитуды колебаний напряжения на конденсаторе первичной обмотки. В этом можно убедиться с помощью выражения для электрической энергии конденсатора $W_e = CV^2/2$, если принять во внимание, что электромагнитная энергия из первичной обмотки переходит во вторичную, и пренебречь при этом потерями энергии. Действительно, емкость вторичной обмотки много меньше емкости конденсатора первичной цепи. Поэтому напряжение на ней будет больше. Аналогично убеждаемся, что амплитуда тока во вторичной обмотке получится значительно меньше, чем в первичной [1].

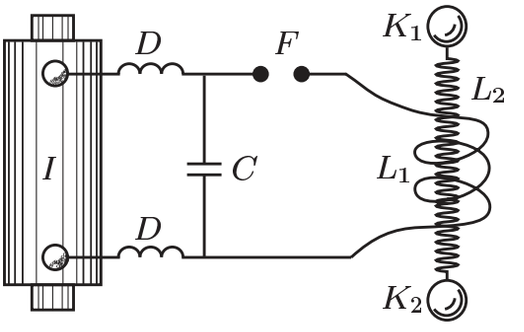


Рис. 354

Сильное электрическое поле вблизи полюсов K_1 и K_2 вторичной обмотки очень эффектно проявляется в темноте в виде энергичных разрядов в воздух с этих полюсов. При приближении к полюсам K_1 и K_2 заземленных проводников образуются длинные электрические искры. Тесла получал искры длиной до двух метров, что соответствует напряжению более миллиона вольт. Искры можно извлекать и из

[1]) Аргументация подобного рода, хотя она и часто применяется, отличается неясностью. Вторичная обмотка характеризуется определенными емкостью и индуктивностью. Эти понятия применимы для постоянных и квазистационарных полей, тогда как в трансформаторе Тесла мы имеем дело с высокочастотными полями, для которых условие квазистационарности не выполняется. Неясным остается и смысл напряжения, поскольку в переменных полях интеграл $\int E\,dl$ зависит от пути интегрирования. Строгое (и понятное) рассмотрение должно основываться на системе уравнений Максвелла. Однако такового, по-видимому, не существует.

других частей вторичной обмотки, но они слабее. Вблизи середины обмотки искры не возникают (узел напряжения). Газоразрядные трубки светятся даже на значительном расстоянии от трансформатора Тесла.

Быстропеременные токи от трансформатора Тесла, несмотря на их высокое напряжение, не опасны для человека, а наоборот, ими даже пользуются теперь в лечебных целях. Человек может без боли выдерживать переменный ток с частотой 50 Гц до 0,1 А, тогда как при частоте 10^5 Гц можно довести силу тока до 0,8 А без заметного сокращения мускулов. (Эти пределы для разных людей разные.) Можно произвести следующий опыт. Кусок голой медной проволоки, служащий ручкой, обвивается два-три раза вокруг нарезного цоколя обыкновенной лампочки накаливания. Экспериментатор становится на изолирующую скамейку, удерживая лампочку за другой конец проволоки. Второй контакт лампочки он подносит к верхнему полюсу вторичной обмотки трансформатора Тесла, нижний полюс которой заземлен. Хотя цепь лампочки и не замкнута (экспериментатор стоит на изолирующей скамейке), лампочка ярко горит. Это происходит потому, что за половину периода лампочка и тело экспериментатора заряжаются электричеством определенного знака, а в следующую половину периода — таким же электричеством, но противоположного знака. Ввиду того, что колебания высокочастотные (порядка 10^6 Гц и выше), такая перезарядка происходит очень часто, и возникает переменный ток, достаточный для накала нити лампы. Через тело экспериментатора проходят при этом токи в несколько ампер. Такие токи, если бы они были постоянными, вызвали бы сильные физиологические действия и были бы очень опасны для жизни. Однако в описанном опыте токи высокочастотные. Экспериментатор не ощущает их, так как они протекают только в тонком поверхностном слое его кожи и не заходят вглубь, где расположена нервная система и жизненно важные органы человека. Можно даже подносить к полюсу вторичной обмотки палец руки. Искры, проскакивающие между полюсом и пальцем, никакого заметного болевого ощущения не вызывают.

ЗАДАЧА

В § 47 было указано, что для измерения электростатических полей на моделях в электролитической ванне применяется переменное напряжение. Выяснить пределы применимости такого метода.

Р е ш е н и е. Прежде всего должно соблюдаться условие квазистационарности. При наличии проводимости жидкости оно означает, что размеры ванны L должны быть малы не только по сравнению с длиной электромагнитной волны в электролите Λ, но и по сравнению с глубиной l проникновения электромагнитного поля в электролит из-за скин-эффекта. (Длина волны обозначается через Λ, λ применяется для обозначения удельной проводимости электролита.) Кроме того, должно быть выполнено условие применимости метода сечений. При обосновании этого метода используется то обстоятельство, что в электролите на его поверхности должна обращаться в нуль нормальная составляющая электрического поля, что в случае переменных полей может соблюдаться только приближенно. Пусть электролит граничит с воздухом.

Тогда на границе раздела должны соблюдаться условия

$$-D_n + D_{0n} = 4\pi\sigma, \quad j_n = \partial\sigma/\partial t,$$

где $\mathbf{D}$ и $\mathbf{D}_0$ — векторы индукции в электролите и в воздухе, σ — поверхностная плотность зарядов, j — плотность тока в электролите (нормаль проведена от электролита к воздуху). Введем электрические поля $\mathbf{E}$ и $\mathbf{E}_0$ и предположим, что они $\sim e^{i\omega t}$. Тогда

$$E_{0n} - \varepsilon E_n = 4\pi\sigma, \quad \lambda E_n = i\omega\sigma,$$

откуда

$$E_n = \frac{E_0 n}{\varepsilon - i\cdot 4\pi\lambda/\omega}, \quad D_n = \frac{E_{0n}}{1 - i\cdot 4\pi\lambda/\omega\varepsilon}.$$

Скачок нормальной составляющей вектора $\mathbf{D}$ обусловлен поверхностными зарядами на границе электролита. Эти заряды ослабляют поле электродов внутри и усиливают вне электролита. При выполнении условия $|D_n| \ll \ll E_{0n}$, поверхностные заряды почти полностью компенсируют нормальную составляющую поля электродов в электролите на его границе. Для полной компенсации надо было бы очень мало изменить поверхностную плотность σ. Тогда мы получили бы точно $E_n = 0$, и для такого искаженного поля условие применимости метода сечений было бы выполнено. Но, очевидно, малое изменение плотности σ мало изменит и поле внутри электролита. Поэтому можно считать, что при соблюдении неравенства

$$\varepsilon \ll 4\pi\lambda/\omega$$

нормальная составляющая E_n на поверхности электролита практически равна нулю. Это неравенство и есть условие применимости метода сечений. Используя формулу (144.5), а также выражение $\Lambda = 2\pi c/(\omega\sqrt{\varepsilon})$, неравенство можно записать в виде $l \ll \Lambda$. Окончательно условие применимости метода электролитической ванны можно записать в виде

$$L \ll l \ll \Lambda. \tag{144.6}$$

Подставляя выражения для l и Λ, получаем

$$\nu \ll \frac{c}{L\sqrt{\varepsilon}}, \tag{144.7}$$

$$\frac{\nu\varepsilon}{8\pi} \ll \lambda \ll \frac{c^2}{8\pi\nu L^2}. \tag{144.8}$$

При $L = 1$ м, $\varepsilon = 81$ из (144.7) получаем $\nu \ll 3\cdot 10^7$ Гц независимо от электропроводности электролита. Для предельной частоты $3\cdot 10^7$ Гц условие (144.8) дает

$$9{,}7\cdot 10^7 \ll \lambda \ll 12\cdot 10^7 \text{ с}^{-1}.$$

Отсюда видно, что метод электролитической ванны (при $L \sim 1$ м) уже неприменим на частотах порядка 10^7 Гц. Для обычно применяемых частот $\nu \sim 10^3$ Гц имеем $3\cdot 10^3 \ll \lambda \ll 3\cdot 10^{12}$ с$^{-1}$, или $10^{-9} \ll \lambda \ll 1$ Ом$^{-1}\cdot$см$^{-1}$.

§ 145. Давление и импульс электромагнитных волн

1. Максвеллом теоретически было показано, что электромагнитные волны, отражаясь или поглощаясь в телах, на которые они падают, оказывают на них *давление.* Это давление есть результат воздействия магнитного поля волны на электрические токи, возбуждаемые электрическим полем той же волны, а иногда также воздействия элек-

трического поля на заряды, индуцируемые в веществе тем же полем. Рассмотрим, например, бегущую плоскую электромагнитную волну в однородной среде. Если среда поглощающая, т. е. обладает проводимостью, то электрическое поле волны возбуждает в ней электрический ток с плотностью $\mathbf{j} = \lambda\mathbf{E}$. Вследствие этого на единицу объема среды действует сила $\mathbf{f} = (1/c)[\mathbf{jB}] = (\lambda/c)[\mathbf{EB}]$, направленная в сторону распространения волны. Эта сила и вызывает давление электромагнитной волны. При отсутствии поглощения ($\lambda = 0$) $\mathbf{f} = 0$, т. е. распространение электромагнитной волны в этом случае не связано ни с каким давлением на среду.

2. Для вычисления давления электромагнитных волн и уяснения его происхождения рассмотрим сначала частный случай. Пусть плоская электромагнитная волна, распространяющаяся в вакууме, падает на плоскую границу идеально проводящего металла (рис. 355). При отражении волна изменяет направление. При этом должно измениться на противоположное направление одного из векторов $\mathbf{E}$ или $\mathbf{H}$. Легко видеть, что это произойдет с вектором $\mathbf{E}$. Действительно, так как вторая среда идеально проводящая ($\lambda = \infty$), в ней электрическое поле должно обращаться в нуль. Иначе, в силу закона Ома $\mathbf{j} = \lambda\mathbf{E}$, в среде возникли бы электрические токи с бесконечной плотностью, что физически невозможно. Так как тангенциальные составляющие элек-

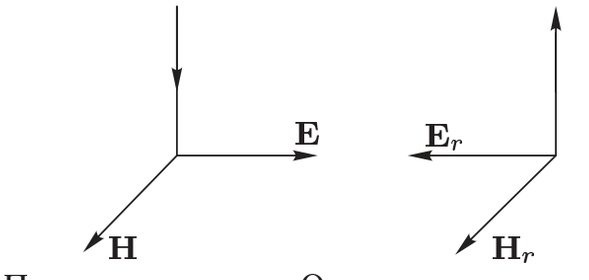


Рис. 355

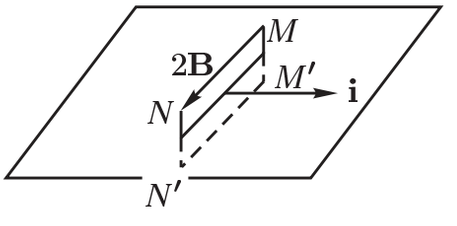


Рис. 356

трического поля непрерывны, то на границе раздела электрический вектор должен обращаться в нуль и в первой среде. Но поле в первой среде складывается из поля $\mathbf{E}$ падающей и поля $\mathbf{E}_r$ отраженной волн. Поэтому на границе должно быть $\mathbf{E} + \mathbf{E}_r = 0$, т. е. $\mathbf{E}_r = -\mathbf{E}$, что и требовалось доказать. Напротив, на той же границе магнитный вектор отраженной волны будет $\mathbf{H}_r = \mathbf{H}$, а результирующее магнитное поле в первой среде $\mathbf{H} + \mathbf{H}_r = 2\mathbf{H}$. Таким образом, магнитный вектор при переходе через границу раздела претерпевает скачок, равный $2\mathbf{H}$. Это означает, что по поверхности металла в направлении электрического вектора $\mathbf{E}$ течет поверхностный ток с линейной плотностью $\mathbf{i}$ (рис. 356). Величина этой плотности найдется по теореме о циркуляции, если применить последнюю к контуру $MNN'M'$. Это дает $2H = 4\pi i/c$, откуда $i = cH/2\pi$.

При вычислении силы, действующей на элементарную площадку dS поверхности тела с током $i\,dS$, надо соблюдать осторожность. Дело

в том, что эта сила определяется магнитным полем $\mathbf{H}_{\text{внеш}}$, *внешним по отношению к самому току* $\mathbf{i}\,dS$. Внешнее поле на поверхности площадки dS, очевидно, одинаково по обе ее стороны; оно непрерывно. Собственное же магнитное поле тока $\mathbf{i}\,dS$ претерпевает разрыв. Если со стороны вакуума его обозначить через $\mathbf{H}_{\text{соб}}$, то со стороны металла, ввиду симметрии, оно будет $-\mathbf{H}_{\text{соб}}$. Применение теоремы о циркуляции к контуру $MNN'M'$ дает $H_{\text{соб}} = 2\pi i/c = H$. Вычитая это значение из полного поля $2H$ вне металла, получим $\mathbf{H}_{\text{внеш}} = \mathbf{H}$. (Тот же результат можно получить из условия, что внутри металла внешнее поле должно уничтожать собственное поле.) Сила, действующая со стороны внешнего поля на ток $\mathbf{i}\,dS$, направлена внутрь металла, т. е. это есть *сила давления.* Давление на единицу поверхности металла будет

$$\mathscr{P} = \frac{1}{c}\,\overline{iH}_{\text{внеш}} = \frac{1}{2}\,\overline{H^2}, \tag{145.1}$$

где черта означает усреднение по времени. Ввиду равенства $E = H$ можно также написать

$$\mathscr{P} = \frac{1}{2\pi}\,\overline{EH} = \frac{1}{4\pi}\,(\overline{E^2} + \overline{H^2}) = 2\overline{w}, \tag{145.2}$$

где $\overline{w}$ — средняя плотность электромагнитной энергии падающей волны. Таким образом, при нормальном падении электромагнитной волны на идеально отражающую поверхность металла металл испытывает давление, равное *удвоенной средней плотности энергии падающей волны.*

3. Совершенно так же может быть разобран случай *наклонного падения волны.* Вычисления здесь будут несколько сложнее. Появится разрыв не только тангенциальных составляющих магнитного поля, но и нормальных составляющих электрического поля. Последний разрыв означает, что на поверхности металла возникнут *электрические заряды,* и надо принимать во внимание силы, действующие на эти заряды со стороны электрического поля. Таким образом, давление электромагнитной волны в этом случае имеет *двойное происхождение.* Оно складывается из *«магнитной силы»*, с которой магнитное поле действует на поверхностные токи металла, и из *«электрической силы»*, с которой на поверхностные заряды того же металла действует электрическое поле. Мы не будем производить эти вычисления, а воспользуемся более общим подходом к вопросу о давлении электромагнитных волн.

4. Будем предполагать опять, что волна падает нормально на поверхность идеального металла. Допустим, что поле волны заполняет цилиндр высотой c с площадью основания, равной единице. Ось цилиндра совпадает с направлением распространения волны. Такая волна будет падать на металл в течение секунды. Поскольку она оказывает давление $\mathscr{P}$ на поверхность металла, последний за это время приобретет импульс $\mathbf{I}_{\text{вещ}} = \mathscr{P} = \overline{EH}/2\pi$, или в векторной форме $I_{\text{вещ}} = \overline{[\mathbf{EH}]}/2\pi$. В замкнутой системе, состоящей из металла и электромагнитного поля, получилось бы *нарушение закона сохранения импульса,* если бы импульсом обладало *только вещество.* Импульс указанной системы

может сохраняться только при условии, что *электромагнитная волна также обладает импульсом*: металл приобретает импульс за счет импульса, передаваемого ему электромагнитной волной. Для вычисления импульса падающей электромагнитной волны $\mathbf{I}_{эл}$ замечаем, что при отражении ее импульс *не меняется по модулю, но меняет направление на противоположное*, т. е. изменение электромагнитного импульса в этом процессе равно $\Delta\mathbf{I}_{эл} = -\mathbf{I}_{эл} - (+\mathbf{I}_{эл}) = -2I_{эл}$, тогда как для вещества $\Delta\mathbf{I}_{вещ} = \mathbf{I}_{вещ}$. Закон сохранения импульса требует $\Delta\mathbf{I}_{эл} + \Delta\mathbf{I}_{вещ} = 0$, откуда $\mathbf{I}_{эл} = \mathbf{I}_{вещ}/2 = \overline{[\mathbf{EH}]}/4\pi$. Разделив это выражение на длину c цилиндра, получим средний электромагнитный импульс единицы объема, т. е. среднюю плотность электромагнитного импульса

$$\overline{\mathbf{g}}_{эл} = \frac{1}{4\pi c}\,\overline{[\mathbf{EH}]} = \frac{1}{c^2}\,\overline{\mathbf{S}}, \tag{145.3}$$

где $\mathbf{S}$ — вектор Пойнтинга. При выводе предполагалось, что волна падает нормально на поверхность металла. Однако это обстоятельство не может отразиться на окончательном результате (145.3), так как плотность импульса $\mathbf{g}_{эл}$ есть характеристика *только самой электромагнитной волны* и не может зависеть от тел, с которыми она взаимодействует. Полученные результаты согласуются с тем, что было сказано об электромагнитном количестве движения в § 84.

5. Покажем на примере, как следует пользоваться формулой (145.3) для вычисления сил, с которыми излучение действует на тело. Пусть электромагнитная волна, распространяющаяся в направлении единичной нормали $\mathbf{N}$, частично отражается в направлении нормали $\mathbf{N}'$, а частично проходит во вторую среду и там поглощается (рис. 357).

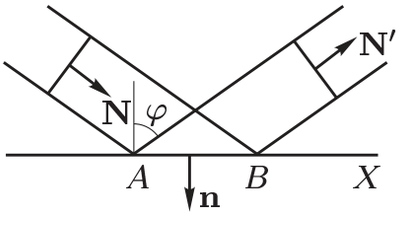


Рис. 357

Если площадь AB, на которую падает волна, равна единице, а угол падения φ, то поперечные сечения падающего и отраженного пучков будут равны $\cos\varphi$ каждый. Возьмем длины пучков равными c. Тогда импульс, передаваемый излучением телу в одну секунду, будет $\mathbf{I} = \cos\varphi(\overline{w}\mathbf{N} - \overline{w}'\mathbf{N}') = \overline{w}\cos\varphi(\mathbf{N} - \rho\mathbf{N}')$, где $\overline{w}$ и $\overline{w}'$ — средние плотности энергии падающей и отраженной волн, а ρ — *коэффициент отражения*. Излучение действует на единичную площадку AB на границе тела с силой $\mathbf{f} = \mathbf{I}$. Проецируя ее на нормаль и к поверхности тела, находим давление излучения

$$\mathscr{P} = \overline{w}\cos^2\varphi(1+\rho), \tag{145.4}$$

а проецируя на ось X — среднюю касательную силу, действующую на площадку AB:

$$\tau = \overline{w}\sin\varphi\cos\varphi(1-\rho). \tag{145.5}$$

При $\rho = 1$ и при нормальном падении получаем прежний результат $\mathscr{P} = 2\overline{w}$. Если же среда полностью поглощает падающее излучение ($\rho = 0$), то $\mathscr{P} = \overline{w}$, т. е. в этом случае давление вдвое меньше.

6. Чтобы составить представление о давлении излучения, рассчитаем его для солнечного излучения вблизи земной поверхности. Как показали измерения, средняя плотность потока энергии в этом случае $\overline{S} = 2$ кал/(см$^2 \cdot$ мин)$= 1{,}4 \cdot 10^3$ Вт/м^2. Для давления излучения на полностью поглощающую поверхность, перпендикулярную к излучению, находим $\mathscr{P} = \overline{S}/c = 4{,}7 \cdot 10^{-6}$ Па, а на полностью отражающую $9{,}4 \cdot 10^{-6}$ Па. Несмотря на ничтожные значения этих величин, экспериментальное доказательство существования давления электромагнитных волн было впервые получено на волнах света в классических опытах П. Н. Лебедева. Лебедев в 1900 г. доказал существование светового давления на твердые тела, а в 1910 г. — и на газы. Результаты этих опытов оказались в согласии с электромагнитной теорией света. Впрочем, давление излучения не всегда столь мало. Если с помощью линзы сфокусировать на поверхности монеты пучок света от лазера, то световое давление пробивает монету, оставляя в ней маленькую дырочку (диаметром в несколько десятых миллиметра). Давление излучения громадно внутри горячих звезд и играет существенную роль при их взрывах. Когда температура в звезде достигает 10^8 кэВ (такие температуры достигаются также при взрывах атомных и водородных бомб), давление излучения становится того же порядка, что и давление плазмы, из которой состоит звезда.

§ 146. Принципы радиосвязи

1. Всякий провод, по которому течет переменный ток, излучает электромагнитные волны. Однако если ток замкнут и выполнено условие квазистационарности, то *излучения практически не будет.* Действительно, разобьем провод на элементы тока $I\,d\mathbf{l}$. Каждый из этих элементов излучает, как точечный диполь, производная дипольного момента которого по времени определяется выражением $d\dot{\mathbf{p}} = I\,d\mathbf{l}$. Если выполнено условие квазистационарности, то все эти дипольные моменты колеблются почти в одинаковых фазах, а потому при рассмотрении поля в волновой зоне (а только такое поле и представляет интерес в вопросах радиосвязи) все их можно считать как бы сосредоточенными в одной точке. Следовательно, весь виток с током будет излучать, как точечный диполь, для которого $\dot{\mathbf{p}} = \oint d\dot{\mathbf{p}} = \oint I\,d\mathbf{l}$. В случае квазистационарных токов величина I одна и та же вдоль всего витка, а потому $\dot{\mathbf{p}} = I\oint d\mathbf{l} = 0$, так как для всякого замкнутого контура $\oint d\mathbf{l} = 0$. При более точном рассмотрении замкнутый виток с током можно разбить на две половины и каждую из них заменить точечным диполем. Тогда получится система двух точечных диполей *с равными, но противоположно направленными моментами*, сдвинутыми один относительно другого на расстояние, малое по сравнению с длиной волны. Теоретически такая система излучает, но практически излучения нет, так как в волновой зоне поле излучения меняется с расстоянием, как $\dfrac{d}{dr}\left(\dfrac{1}{r}\right) \sim \dfrac{1}{r^2}$, т. е. убывает слишком быстро. Эти рассуждения

применимы и к колебательному контуру, а также ко всякой «закрытой» системе, т. е. к системе с замкнутыми или почти замкнутыми токами. Для получения интенсивного излучения надо перейти к *«открытым»* системам, в которых токи не замкнуты. Теоретически наилучшей системой является прямолинейный провод, в котором возбуждаются (незамкнутые) переменные токи. Излучатель такого типа был использован уже в классических опытах Герца. В радиотехнике излучающей системой служит *антенна* — незамкнутый провод или система проводов, подвешенных высоко над землей, по которым текут переменные токи.

Низкочастотные токи, применяемые в электротехнике, не годятся для целей радиосвязи по двум причинам. Во-первых, потому, что они излучают *очень слабо.* (Мощность излучения пропорциональна четвертой степени частоты, см. § 141.) Во-вторых, для возбуждения сильных колебаний в антенне используется явление резонанса, а следовательно, размеры антенны должны быть *очень большими.* Например, если излучающей антенной служит прямолинейный провод, то для возбуждения основного колебания длина провода должна быть $l = \lambda/2$. При частоте $\nu = 1000$ Гц это дает $l = 150$ км. Поэтому в радиотехнике излучающие антенны питаются *токами высокой частоты.* В радиовещании пользуются частотами, лежащими приблизительно в пределах от 10^5 до 10^8 Гц, чему соответствуют длины волн от 3 км до 3 м. Для решения специальных задач, где существенна острая направленность излучения, применяют дециметровые и сантиметровые волны (частоты до 10^{10} Гц и выше).

Принцип радиосвязи очень прост. В антенне передающей радиостанции, настроенной в резонанс с генератором (рис. 358 *а*), возбуждаются сильные высокочастотные токи. Электромагнитные волны, излу-

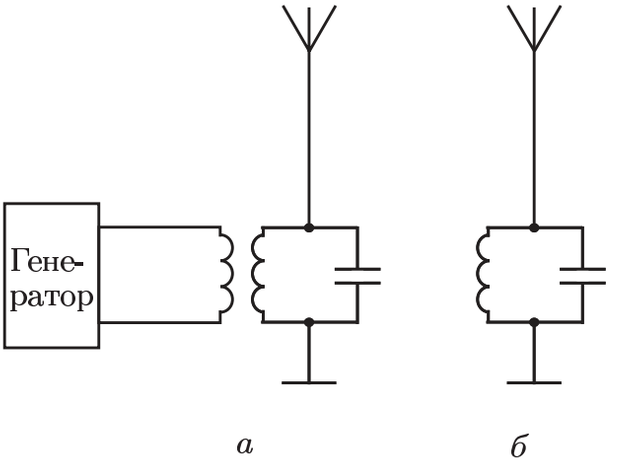


Рис. 358

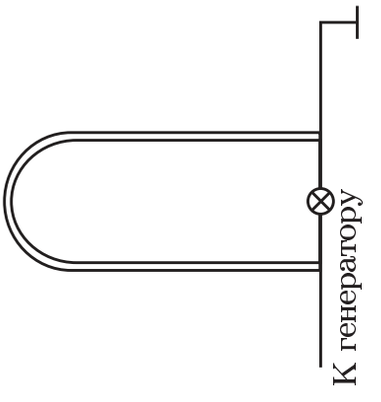


Рис. 359

чаемые этой антенной, достигая приемной антенны (рис. 358 *б*), также настроенной в резонанс с генератором, в свою очередь возбуждают в ней токи той же частоты, которые могут быть усилены и использованы. Для настройки антенн конденсатор можно включать не только параллельно катушке индуктивности (рис. 358), но и последовательно

с ней (рис. 359). При параллельном включении увеличивается общая емкость системы, а следовательно, ее собственная частота уменьшается. При последовательном соединении, наоборот, емкость уменьшается, а частота увеличивается.

2. Однако чисто синусоидальные волны высокой частоты, излучаемые антенной, не только *не несут никакой информации*, но и *не воспринимаются ухом человека*, если их преобразовать в звуковые волны той же частоты. Для передачи информации, например человеческой речи или музыки, требуются *низкочастотные сигналы*, с частотами примерно от 100 до нескольких тысяч герц. Для передачи телеграфных сигналов требуемые частоты еще ниже. Непосредственная передача низкочастотных сигналов радиоволнами тех же частот, как указывалось выше, невозможна из-за невозможности их генерации. Эта трудность устраняется в радиотехнике тем, что *передача осуществляется синусоидальными волнами высокой частоты, измененными низкочастотными сигналами.* Такое изменение называется *модуляцией*, а сами волны — *модулированными* (см. § 128). Модуляция состоит в изменении во времени *амплитуды*, *частоты* или *фазы колебания.* Во всем последующем изложении имеется в виду наиболее простая, *амплитудная*, *модуляция.*

Если из студии радиостанции не производится передача, то колебания тока в передающей антенне синусоидальны:

$$I = I_0 \sin \omega t \tag{146.1}$$

(рис. 360 *а*). Во время передачи (речь, музыка перед микрофоном) эти колебания преобразуются в

$$I = I_0[1 + f(t)] \sin \omega t, \tag{146.2}$$

где функция $f(t)$ зависит от формы передаваемого (модулирующего) сигнала [$|f(t)| < 1$]. Она называется *модулирующей функцией.* Простейшая модулирующая функция соответствует передаче чисто музыкального тона, например звучания камертона. В этом случае $f(t) = \alpha \sin \Omega t$ (рис. 360 *б*). Постоянная Ω называется *частотой*, а постоянная α — *глубиной модуляции.* (Начальную фазу мы взяли равной нулю, так как она не играет принципиальной роли.) Подставляя $f(t)$ в формулу (146.2), представим модулированное колебание в виде

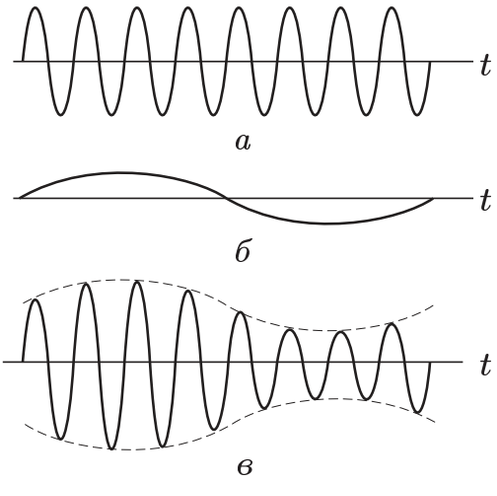


Рис. 360

$$I = I_0[1 + \alpha \sin \Omega t] \sin \omega t \tag{146.3}$$

(рис. 360 *в*). Преобразовывая это выражение, получим

$$I = I_0 \sin \omega t + \frac{\alpha I_0}{2} \cos (\omega - \Omega)t - \frac{\alpha I_0}{2} \cos (\omega + \Omega)t. \tag{146.4}$$

Если модулирующая функция $f(t)$ не синусоидальна, но периодична, то ее можно разложить в ряд Фурье и к каждому члену этого разложения применить рассуждение, приведенное выше. Тогда вместо двух частот $\omega - \Omega$ и $\omega + \Omega$ слева и справа от ω появится несколько боковых частот. Они образуют *боковые полосы*, состоящие из дискретного набора частот. Если же модулирующая функция не периодична, то она разлагается в интеграл Фурье. В этом случае боковые полосы слева и справа от ω содержат не только бесконечное, но и *непрерывное* множество частот. Примером могут служить модулирующая функция и модулированные колебания в случае телеграфных знаков, передаваемых по азбуке Морзе (рис. 361). Существенно, что все боковые частоты расположены *очень близко* от «несущей» частоты ω. Все колебания с такими частотами являются *высокочастотными*, а потому они пригодны для излучения радиоволн. Приемник должен принимать *не только колебания несущей частоты, но и колебания примыкающих к ней боковых полос.* Иначе вся информация, содержащаяся в передаваемом сигнале, будет утеряна. Поэтому настройка приемника *не должна быть особенно острой.*

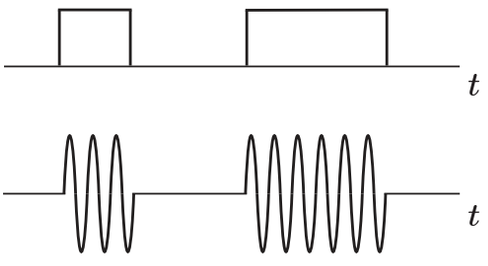


Рис. 361

3. Для осуществления модуляции применяются устройства, образующие *произведения*, а не сумму подводимых к ним напряжений. Такие устройства принадлежат к классу нелинейных систем. На рис. 362 приведена простейшая схема генератора радиотелефонного передатчика с модуляцией на сетку. Колебательный контур включен в цепь анода. Модулирующее напряжение возникает в цепи микрофона M и повышается трансформатором. Сопротивление сеточного конденсатора C_g для высокочастотных токов генератора мало, а сопротивление вторичной обмотки трансформатора благодаря ее большой индуктивности, наоборот, велико. Поэтому токи высокой частоты не ответвляются в модулирующую микрофонную цепь, и генератор работает практически так же, как и в отсутствие этой цепи. Иначе ведут себя микрофонные токи, идущие от вторичной обмотки трансформатора. Это токи низкой частоты. Для них конденсатор C_g оказывает очень большое сопротивление, а потому напряжение трансформатора оказывается целиком приложенным между сеткой и катодом лампы.

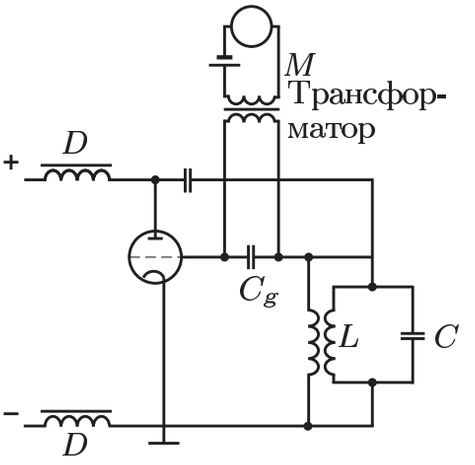


Рис. 362

Это напряжение и модулирует выходное напряжение генератора. Действительно, анодный ток I есть функция анодного V_1 и сеточного V_2 напряжений: $I = I(V_1, V_2)$. Разложим эту функцию в ряд Тейлора по обоим аргументам, оборвав разложение на членах второй степени. Получим

$$I = a_0 + a_1 V_1 + a_2 V_2 + a_{11} V_1^2 + a_{22} V_2^2 + a_{12} V_1 V_2.$$

Последний член пропорционален *произведению* подводимых напряжений. Он и вызывает модуляцию анодного напряжения, а следовательно, и напряжения на конденсаторе или катушке индуктивности колебательного контура. Допустим, например, что $V_1 = A_1 \sin \omega t, V_2 = A_2 \sin \Omega t$. Тогда

$$I = a_1 A_1 \left(1 + \frac{a_{12}}{a_1} A_2 \sin \Omega t\right) \sin \omega t + \\ + (a_0 + a_2 A_2 \sin \Omega t + a_{11} A_1^2 \sin^2 \omega t + a_{22} A_2^2 \sin^2 \Omega t).$$

Первая скобка и есть требуемое модулированное колебание. Его можно выделить с помощью резонансного контура, настроенного (но не слишком остро) на частоту ω. Из приведенного анализа следует, что для модуляции колебаний лампа обязательно должна работать в таком режиме, чтобы ее вольт-амперная характеристика была *нелинейна.* Модулированные колебания передаются в антенный контур, связанный индуктивно с катушкой L колебательного контура.

4. Обратимся теперь к приему модулированных колебаний. Под действием модулированной волны передающей антенны в приемной антенне возникают такие же модулированные колебания, только более слабые. Однако даже после предварительного усиления они еще не годятся для непосредственного воспроизведения посланного сигнала. Дело в том, что модулированные колебания, если их разложить на синусоидальные составляющие, будут содержать *только высокие частоты*, с которыми мембрана телефона (благодаря своей инерционности) не может колебаться с заметными амплитудами. Если даже изготовить малоинерционную мембрану (что в принципе возможно), то все равно возбуждаемые ею звуковые колебания (с частотами 10^5–10^8 Гц) будут находиться *вне пределов слышимости человеческого уха.* Поэтому принятые модулированные колебания подвергаются в приемнике *демодуляции*, или *детектированию*, превращаясь в колебания низких частот, соответствующих частотам посланного сигнала.

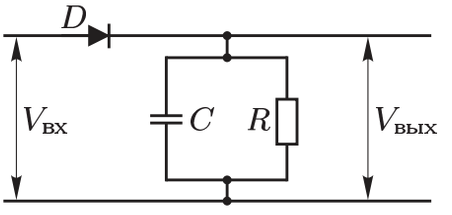


Рис. 363

Демодуляция состоит в том, что высокочастотные колебания выпрямляются (с помощью *кристаллического детектора* или *электронной лампы*), а затем *сглаживаются* цепью, обладающей подходящим временем релаксации (постоянной времени). На рис. 363 изображена одна из возможных схем

детектирования. Выпрямление производится с помощью кристаллического детектора D, сглаживание — с помощью цепи из параллельно соединенных конденсатора C и сопротивления R. Допустим, например, что модулирующий сигнал радиопередатчика состоит из одинаковых равноотстоящих прямоугольных импульсов длительностью t_1 каждый, следующих друг за другом через равные промежутки времени t_2 (рис. 364 *а*). На вход приемника подается модулированный сигнал (рис. 364 *б*). Детектор D «срезает» верхнюю часть этого сигнала. Если бы не было конденсатора C, то на сопротивлении R появилось бы переменное напряжение V_R, представляемое кривой на рис. 364 *в*. В течение времени t_1 оно состояло бы из большого числа равноотстоящих одинаковых микроимпульсов длительностью $T/2$ каждый, где T — период высокочастотных колебаний. В последующий промежуток времени t_2 оно было бы равно нулю. При наличии конденсатора C большая часть тока от первого микроимпульса пойдет на конденсатор, заряжая его до напряжения, малого по сравнению с высотой микроимпульса. Такое же напряжение появится на сопротивлении R. В промежутках между микроимпульсами заряд конденсатора и напряжение на нем убывают экспоненциально по закону $e^{-t/\tau}$, где $\tau = RC$ — время релаксации (постоянная времени) RC-цепи. Если τ велико по сравнению с промежутком времени Δt между последовательными микроимпульсами, то за время Δt заряд и напряжение на конденсаторе при разряде изменятся мало. Второй микроимпульс сигнала немного повысит это напряжение. За ним последует второй разряд и т. д. Этот процесс будет продолжаться до тех пор, пока повышение напряжения от микроимпульса при заряде конденсатора не сделается равным последующему понижению того же напряжения при разряде. С этого момента на сопротивлении R установится практически постоянное напряжение. Когда через систему пройдет последний микроимпульс, напряжение на выходе начнет экспоненциально убывать к нулю. Напряжение появится вновь, когда детектор пропустит второй импульс, и т. д. Если время установления напряжения мало по сравнению с t_1 и t_2, то напряжение на выходе будет иметь вид, представленный на рис. 364 *г*. При надлежаще выбранных параметрах системы форма выходного импульса практически не будет отличаться от формы низкочастотного сигнала, посланного передающей радиостанцией. Так же обстоит дело и в случае передачи низкочастотных сигналов непрямоугольной формы.

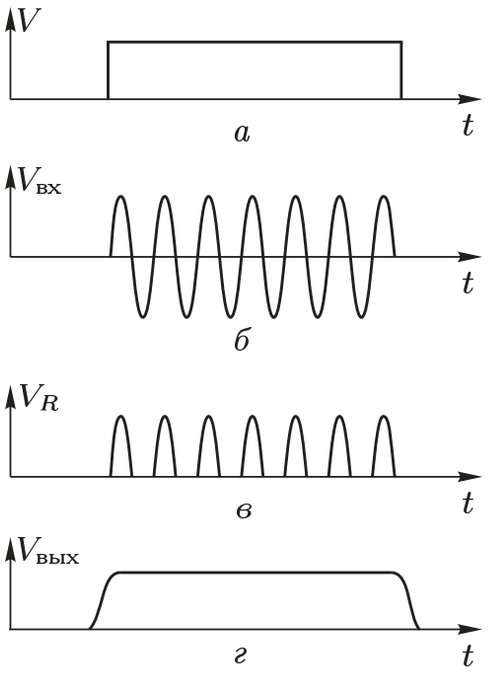


Рис. 364

Для выпрямления колебаний при детектировании по схеме рис. 363 кристаллический детектор можно заменить двухэлектродной электронной лампой — *диодом*. Однако вместо диода обычно применяют *трехэлектродную лампу* (*триод*), которая *не только выпрямляет, но и усиливает колебания*. Соответствующая схема *с сеточной демодуляцией* показана на рис. 365. Ясно, что для получения выпрямляющего действия необходимо, чтобы рабочая точка электронной лампы лежала *в нелинейной части характеристики*, например у ее нижнего загиба.

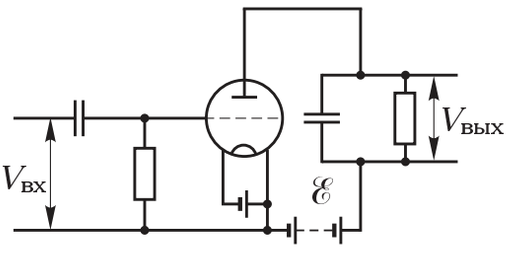


Рис. 365

5. Теперь можно понять действие радиоприемника. Высокочастотные колебания, возбужденные в приемной антенне, попадают сначала в *усилитель высокой частоты*. Входной колебательный контур этого усилителя должен быть *настраиваемым*, чтобы можно было выделять волну определенной радиостанции. Настройка, однако, *не должна быть слишком острой*, чтобы практически одинаково усиливались колебания с несущей частотой и колебания с частотами боковых полос (см. п. 2). Усиленные колебания высокой частоты попадают в *демодулятор* и там преобразуются в *колебания низких частот*, содержащихся в передаваемом сигнале. Эти низкочастотные колебания *опять усиливаются* и попадают в *громкоговоритель* или *другой индикатор*. Приемник, работающий по такому принципу, называется *приемником прямого усиления*. Блок-схема такого приемника приведена на рис. 366. Следует отметить, однако, что в радиовещании обычно применяются так называемые *супергетеродинные приемники*, работающие по несколько измененной схеме (см. п. 7).

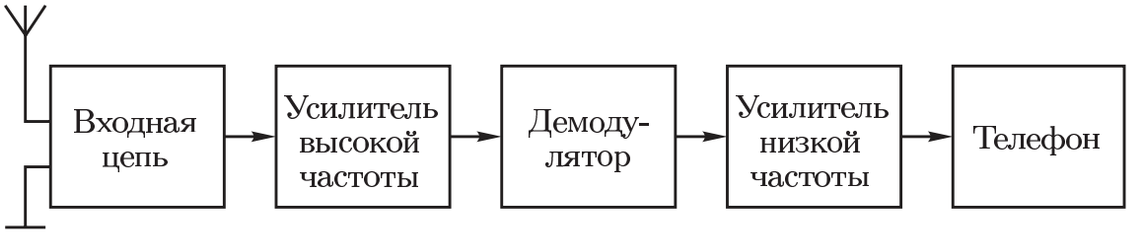


Рис. 366

6. Скажем сначала о *гетеродинном приеме*, применяемом в *радиотелеграфии*. В настоящее время телеграфные сигналы азбуки Морзе принимают почти исключительно с помощью телефона, «на слух». Для этого импульсы высокочастотных колебаний не подвергают демодуляции, а *превращают их в колебания звуковой частоты*. С этой целью на одну из ламп приемника подают два переменных высокочастотных напряжения: напряжение сигнала $V = a\sin\omega t$ и напряжение $V_1 = a_1\sin\omega_1 t$ от *гетеродина*, т. е. местного маломощного генератора, смонтированного в приемнике. Если лампа работает в нелинейном

режиме, то происходит умножение этих напряжений и образование колебаний с комбинированными частотами $\omega - \omega_1$ и $\omega + \omega_1$ (см. п. 3). Величину ω_1 подбирают так, чтобы частота $\omega - \omega_1$ лежала в области звуковых частот, воспринимаемых ухом. Если радиостанция работает, но не передает сигналов, то в телефоне слышен *непрерывный свист* (музыкальный тон с частотой $\omega - \omega_1$). Если же станция начинает передавать телеграфные сигналы, то непрерывный свист сменяется последовательностью *длинных к коротких* звуков. Натренированное ухо телеграфиста воспринимает длинный звук как тире, а короткий как точку. Среди других преимуществ гетеродинный прием обладает еще одним преимуществом, разъясняемым на следующем примере. Допустим, что на приемник действуют одновременно волны от нескольких радиостанций с практически совпадающими амплитудами и настолько близкими частотами, что их невозможно выделить путем резонанса. Гетеродинный прием в некоторых случаях делает возможным и это. Пусть, например, таких волн три с частотами 99 000, 100 000 и 101 000 Гц, а частота колебаний гетеродина 101 000 Гц. Тогда последнюю волну мы совсем не услышим, так как после гетеродинирования ей соответствует колебание с нулевой частотой. Первым же двум волнам соответствуют частоты в 2000 и 1000 Гц, отличающиеся на целую октаву. Такая разница в высоте звука легко различима, и при некотором навыке можно принимать радиотелеграмму одной волны, не обращая внимания на другую.

7. Недостаток описанного выше приемника прямого усиления состоит в том, что при очень высоких частотах он не позволяет получать большие усиления сигналов. Этому препятствуют неизбежные *паразитные емкости* (емкости электронных ламп, емкости соединительных проводов и пр.), которые при очень высоких частотах оказывают току ничтожные сопротивления и потому *шунтируют лампы*. Имеются и другие причины, связанные с процессами внутри самих электронных ламп. Применять же демодуляцию непосредственно после поступления принимаемого сигнала в антенну, отказавшись вовсе от усиления высокой частоты, нельзя, так как в этом случае при усилении низкой частоты усиливался бы не только сигнал, но и различные *помехи*, которые особенно велики при низких частотах. От указанного недостатка в значительной степени свободен *супергетеродинный приемник*, наиболее распространенный в настоящее время. Основная идея этого приемника состоит *в преобразовании несущей частоты ω в частоту более низкую*. Блок-схема супергетеродинного приемника изображена на рис. 367. Высокочастотное модулированное колебание $V = a[1 + f(t)] \sin \omega t$ подводится к одному, а колебание $V_1 = a_1 \sin \omega_1 t$ от местного гетеродина — к другому электроду специальной лампы C. Если лампа C работает в нелинейном режиме, то происходит *умножение напряжений* V и V_1 с образованием колебаний типа

$$[1 + f(t)] \cos (\omega - \omega_1)t, \tag{146.5}$$

$$[1 + f(t)] \cos (\omega + \omega_1)t, \tag{146.6}$$

которые модулированы той же функцией $f(t)$, что и исходное колебание в передающей антенне. Однако несущая частота ω в них заменена на $\omega - \omega_1$ и $\omega + \omega_1$. Этот процесс преобразования частоты в радио-

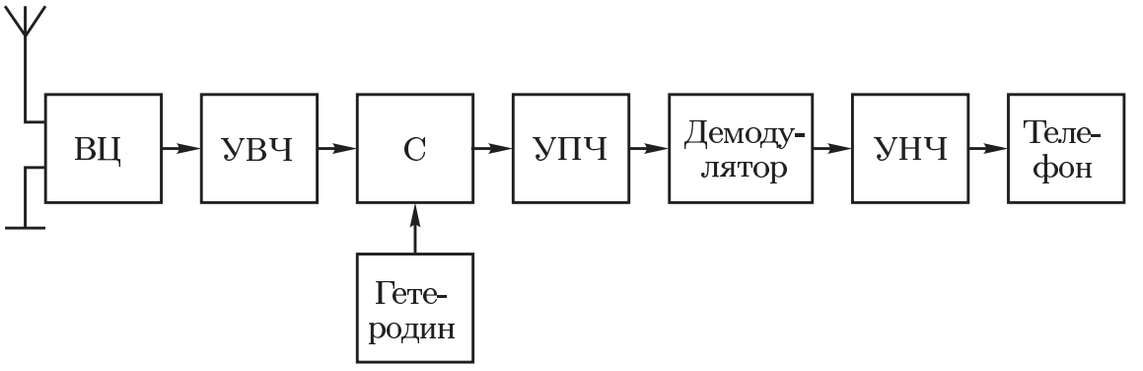


Рис. 367

технике называется *смешением частот*, а лампа, осуществляющая его, — *смесительной лампой* или *смесителем*. В супергетеродинном приемнике представляет интерес колебание (146.5) с частотой $\omega - \omega_1$, называемой *промежуточной частотой*. В отличие от простого гетеродинного приема, промежуточную частоту выбирают *достаточно высокой*, чтобы колебания с этой частотой *не воспринимались ухом*. Колебание (146.5) выделяют резонансным способом и поступают с ним так же, как с исходным колебанием $[1+f(t)]\sin\omega t$ в приемнике прямого усиления, т. е. усиливают в усилителе промежуточной частоты (УПЧ), демодулируют, усиливают в усилителе низкой частоты (УНЧ) и, наконец, направляют в телефон или другой индикатор. Заметим еще, что при настройке на разные станции с изменением емкости конденсатора входного контура приемника изменяется и емкость конденсатора колебательного контура гетеродина таким образом, что промежуточная частота $\omega - \omega_1$ *остается неизменной*. С этой целью подвижные пластины обоих конденсаторов насаживают на общую ось.

§ 147. Распространение электромагнитных волн в волноводах

1. В современной радиотехнике и ее приложениях большое значение имеют *волноводы*, вдоль которых могут распространяться электромагнитные волны. Волновод представляет собой трубу, обычно с металлическими стенками и постоянного поперечного сечения. Труба может быть и криволинейной, но мы ограничимся рассмотрением случая *прямолинейных волноводов*. Внутри волновода обычно находится воздух, который практически можно рассматривать как вакуум. Однако в целях общности мы будем предполагать, что волновод заполнен изотропной однородной средой с диэлектрической проницаемостью ε и магнитной проницаемостью μ. Также в целях общности будем считать, что эти

величины зависят от частоты электромагнитного поля ω (дисперсия). Частный случай вакуума получится из общих формул при $\varepsilon = \mu = 1$.

В нашем рассмотрении волновод будет считаться бесконечно длинным. Это допустимо, когда длина волновода велика по сравнению с длиной волны в той среде, которой он заполнен, а также по сравнению с поперечным сечением самого волновода.

При рассмотрении теории волновода нам придется ссылаться на некоторые элементарные понятия и результаты, изложенные в следующих томах курса. Эти ссылки будут указаны в соответствующих местах.

2. Направим координатную ось z вдоль волновода. Если она проходит внутри волновода, то ее условно можно назвать *осью самого волновода.* Для волновода правильной геометрической формы (например, цилиндрической) за ось волновода естественно принять его *геометрическую* ось, т. е. ось симметрии волновода. Координатные оси x и y выберем перпендикулярно к оси z.

Если вдоль волновода может распространяться бегущая электромагнитная волна частоты ω, то ввиду полной эквивалентности всех поперечных сечений волновода фаза волны не должна зависеть от координат x и y, а должна быть лишь функцией координаты z. Это значит, что электромагнитное поле в волноводе должно иметь вид

$$E = E_0(x, y)e^{i(\omega t - k_z z)}, \quad H = H_0(x, y)e^{i(\omega t - k_z z)}, \tag{147.1}$$

где для незатухающей волны k_z — вещественная постоянная. Амплитуды колебаний E_0 и H_0, разумеется, могут меняться от точки к точке в каждом поперечном сечении волновода. Это и отмечено в формулах (147.1).

Фазовая скорость $v_\text{ф}$ электромагнитного поля одна и та же вдоль всех прямых, параллельных оси волновода. Она определяется формулой

$$v_\text{ф} = \omega / k_z \tag{147.2}$$

(см. т. IV, § 8).

Операция дифференцирования по времени выражения типа (147.1) сводится к умножению уравнений на $i\omega$. Поэтому уравнения Максвелла запишутся в виде

$$\operatorname{rot} \mathbf{H} = i\frac{\omega\varepsilon}{c}\mathbf{E}, \quad \operatorname{rot} \mathbf{E} = -i\frac{\omega\mu}{c}\mathbf{H}.$$

Ввиду независимости ε и μ от координат уравнения

$$\operatorname{div} \mathbf{E} = \operatorname{div} \mathbf{H} = 0$$

следуют из этих двух уравнений, а потому они автоматически принимаются во внимание. Пользуясь затем формулами

$$\operatorname{rot} \mathbf{H} = \operatorname{rot} \mathbf{H}_0 \cdot e^{i(\omega t - k_z z)} - i[\mathbf{k}_z \mathbf{H}], \quad \operatorname{rot} \mathbf{E} = \ldots,$$

запишем уравнения Максвелла в координатной форме:

$$\frac{\partial H_z}{\partial y} + ik_z H_y = i\frac{\omega\varepsilon}{c} E_x, \quad \frac{\partial E_z}{\partial y} + ik_z E_y = -i\frac{\omega\mu}{c} H_x,$$

$$ik_z H_x + \frac{\partial H_z}{\partial x} = -i\frac{\omega\varepsilon}{c} E_y, \quad ik_z E_x + \frac{\partial E_z}{\partial x} = i\frac{\omega\mu}{c} H_y, \tag{147.3}$$

$$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = i\frac{\omega\varepsilon}{c} E_z, \quad \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = -i\frac{\omega\mu}{c} H_z.$$

Для определения электромагнитного поля внутри волновода к этим уравнениям надо прибавить *граничные условия*, которым должны удовлетворять векторы $\mathbf{E}$ и $\mathbf{H}$ на внутренних стенках волновода. Мы будем рассматривать простейший и наиболее важный случай, когда эти стенки *идеально проводящие*. Тогда сформулированная задача, как показывает более подробное математическое исследование, при определенных частотах допускает два типа решений.

3. Первый тип. В этом случае $H_z = 0$, $E_z \neq 0$. Такие волны называются волнами *электрического типа*. Для них уравнения (147.3) переходят в

$$k_z H_y = \frac{\omega\varepsilon}{c} E_x, \quad \frac{\partial E_z}{\partial y} + ik_z E_y = -i\frac{\omega\mu}{c} H_x,$$

$$k_z H_x = -\frac{\omega\varepsilon}{c} E_y, \quad ik_z E_x + \frac{\partial E_z}{\partial x} = i\frac{\omega\mu}{c} H_y, \tag{147.4}$$

$$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = i\frac{\omega\varepsilon}{c} E_z, \quad \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = 0.$$

Из этих уравнений все компоненты поля можно выразить через E_z. Например, из первого и пятого уравнений (147.4) получим

$$E_x = \frac{ck_z}{\omega\varepsilon} H_y = \frac{ck_z}{\omega\varepsilon}\left(\frac{ck_z}{\omega\mu} E_x - \frac{c}{i\omega\mu}\frac{\partial E_z}{\partial x}\right),$$

откуда и находится выражение для E_x через $\partial E_z/\partial x$. Аналогично получается и для остальных компонент электромагнитного поля. При этом удобно ввести обозначения

$$\alpha^2 = k^2 - k_z^2, \tag{147.5}$$

$$k^2 = \frac{\omega^2}{c^2}\varepsilon\mu. \tag{147.6}$$

Тогда

$$E_x = -\frac{ik_z}{\alpha^2}\frac{\partial E_z}{\partial x}, \quad H_x = i\frac{\omega\varepsilon}{c\alpha^2}\frac{\partial E_z}{\partial y},$$

$$E_y = -\frac{ik_z}{\alpha^2}\frac{\partial E_z}{\partial y}, \quad H_y = -i\frac{\omega\varepsilon}{c\alpha^2}\frac{\partial E_z}{\partial x}, \tag{147.7}$$

$$H_z = 0.$$

Таким образом, задача сводится к нахождению единственной функции $E_z(x, y)$. Для определения этой функции используется уравнение

$$\frac{\partial^2 E_z}{\partial x^2} + \frac{\partial^2 E_z}{\partial y^2} = -\alpha^2 E_z. \tag{147.8}$$

Оно является следствием уравнения, которому должно удовлетворять поле E и каждая его компонента. В частности,

$$\frac{\partial^2 E_z}{\partial x^2} + \frac{\partial^2 E_z}{\partial y^2} + \frac{\partial^2 E_z}{\partial z^2} + k^2 E_z = 0. \tag{147.9}$$

Кроме того, ввиду (147.1) справедливо уравнение

$$\frac{\partial^2 E_z}{\partial z^2} + k_z^2 E_z = 0, \quad \frac{\partial^2 E_z}{\partial x^2} + \frac{\partial^2 E_z}{\partial y^2} + (k^2 - k_z^2) E_z = 0,$$

откуда, если использовать обозначение (147.5), и получается уравнение (147.8).

Электрическое поле внутри идеально проводящей стенки равно нулю. Поэтому из равенства тангенциальных компонент электрического поля на границе раздела следует, что на внутренней поверхности стенки волновода $E_z = 0$. Получилась краевая задача для дифференциального уравнения (147.8). Из математики известно, что подобная краевая задача может иметь конечные, однозначные и непрерывные решения не для всех, а только для избранных значений постоянной α^2.

Избранные значения (α^2) *называются собственными значениями оператора* $\partial^2/\partial x^2 + \partial^2/\partial y^2$, а соответствующие им решения — *собственными функциями* рассматриваемой краевой задачи (см. т. V, § 22). Вид собственных функций зависит от формы и размеров поперечного сечения волновода. Для волновода прямоугольной формы собственные функции находятся элементарно (см. задачу 1 к этому параграфу). Для цилиндрического волновода собственные функции выражаются через функции Бесселя. Однако мы не будем заниматься нахождением собственных функций, а ограничимся рассмотрением только общих свойств волновода, не зависящих от его поперечного сечения.

4. Собственные значения двумерного оператора отрицательны, а поэтому величина α^2 положительна. Докажем эту теорему для более общего случая трехмерного оператора

$$\Delta = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2,$$

частным случаем которого является рассматриваемый двумерный оператор. Запишем уравнение (147.9) в виде

$$\Delta E_z + k^2 E_z = 0$$

и умножим обе части его на E_z^* — величину, комплексно сопряженную по отношению к E_z. Тогда, принимая во внимание тождество

$$\operatorname{div}(E_z^* \operatorname{grad} E_z) = E_z^* \Delta E_z^* + \operatorname{grad} E_z^* \operatorname{grad} E_z,$$

получим

$$\operatorname{div}(E_z^* \operatorname{grad} E_z) - \operatorname{grad} E_z^* \operatorname{grad} E_z + k^2 E_z E_z^*.$$

Мы предполагаем, что область, в которой определена функция E_z, окружена замкнутой металлической оболочкой, на стенках которой $E_z = 0$. Поэтому при интегрировании по объему, ограниченному этой оболочкой, интеграл от дивергенции, в силу теоремы Гаусса, обратится в нуль. В результате получится

$$k^2 = \int \operatorname{grad} E_z^* \operatorname{grad} E_z \, dV \Big/ \int E_z^* E_z \, dV,$$

откуда и видно, что величина k^2 существенно положительна.

Доказательство без существенных изменений распространяется и на случай двумерного уравнения (147.8). Только вместо трехмерной замкнутой области надо рассматривать двумерную область, ограниченную замкнутой кривой, получающейся в результате поперечного сечения волновода. Заметим между прочим, что теорема верна и в том случае, когда граничное условие $E_z = 0$ заменено на условие $\partial E_z/\partial n = 0$, где дифференцирование производится по нормали к замкнутой поверхности или замкнутой линии, ограничивающей рассматриваемую область изменения E_z.

5. Изложенное можно получить и из других соображений, которые лучше уясняют физический принцип работы волноводов. Рассмотрим сначала неограниченную однородную среду с диэлектрической и магнитной проницаемостью соответственно $\varepsilon(\omega)$ и $\mu(\omega)$. В такой среде могут распространяться плоские незатухающие волны определенной частоты ω:

$$E,\ H \sim e^{i(\omega t - \mathbf{kr})},$$

где волновой вектор $\mathbf{k}$ может иметь любое направление, но его длина определяется соотношением (147.8). Если фиксировать ось z, то волновой вектор каждой плоской волны можно представить в виде $\mathbf{k} = \mathbf{k}_z + \mathbf{k}_\perp$, где $\mathbf{k}_z$ и $\mathbf{k}_\perp$ — компоненты этого вектора вдоль оси z и перпендикулярно к ней соответственно. Таким образом, всякая незатухающая монохроматическая волна частоты ω может быть представлена в виде

$$\sim e^{ik_\perp z_\perp} e^{i(\omega t - k_z z)}.$$

При отсутствии затухания суперпозицией волн такого типа с постоянными коэффициентами можно получить любое решение уравнений Максвелла в среде.

Допустим теперь, что среда заключена внутри волновода. Роль волновода сводится к тому, что из всех волн указанного выше типа он выделяет волны, удовлетворяющие граничному условию $E_z = 0$. Это выделение сводится к тому, что в каждом поперечном сечении волновода могут получиться только такие системы стоячих волн, каждая из которых удовлетворяет указанному условию. Все волны каждой из таких систем характеризуются одним и тем же значением волнового

числа $k_\perp$, которое в силу уравнения (147.8) тождественно с параметром α. В этом причина того, почему параметр α может принимать только дискретный ряд значений. Волновое число k_z также может принимать только допустимые дискретные значения, которые определяются соотношением $k^2 = k_z^2 + \alpha^2$, тождественным с соотношением (147.5). Из изложенного видно, что при отсутствии затухания оба числа k_z^2 и α^2 положительны, причем $k_z^2 < k^2$.

6. Каждая волна в волноводе характеризуется определенным значением параметра α. При этом волны электрического типа не поперечны относительно электрического вектора **E**, так как продольная компонента E_z обращается в нуль только на стенках волновода, а в остальных местах $E_z \neq 0$. Кроме того, из соотношения $k_z^2 < k^2$ следует $1/k_z > 1/k$, или после умножения на ω получим $\omega/k_z > \omega/k$, или $v_\text{ф} > v$, где $v = c/\sqrt{\varepsilon\mu}$ — фазовая скорость электромагнитной волны в *неограниченной среде.* Таким образом, при одной и той же частоте фазовая скорость волны в волноводе $v_\text{ф} = c/k_z$, а с ней и длина волны $\lambda_\text{вол} = 2\pi v/\omega$ больше соответствующих величин в неограниченной среде. В частности, если внутри волновода вакуум, то всегда $v_\text{ф} > c$, $\lambda_\text{вол} > \lambda_\text{вак}$.

Не все фазовые скорости допустимы, а только такие, которые соответствуют допустимым значениям параметра α. Таким образом, волновод является *фильтром*, пропускающим из непрерывного спектра частот только избранные или собственные частоты. Наконец, при фиксированном значении параметра α невозможно, чтобы $k_z > k$, т. е. $\lambda_\text{среды} = \lambda_\text{вол} = 2\pi/\alpha$. Это значит, что при заданном α волновод не пропускает волны, длина которых $\lambda_\text{среды}$ меньше некоторого верхнего предела $2\pi/\alpha$.

7. Второй тип. В этом случае $E_z = 0$, $H_z \neq 0$, т. е. волны опять не поперечны, но уже относительно вектора **H**. Такие волны называются волнами магнитного типа. Их свойства вполне аналогичны волнам электрического типа, а потому нет необходимости их подробно рассматривать. Укажем только на существенные моменты при исследовании волн магнитного типа. Прежде всего все компоненты поля выражаются через H_z по формулам

$$\begin{gathered} E_x = -i\frac{\omega\mu}{c\alpha^2}\frac{\partial H_z}{\partial y}, \quad H_x = -i\frac{k_z}{\alpha^2}\frac{\partial H_z}{\partial x}, \\ E_y = i\frac{\omega\mu}{c\alpha^2}\frac{\partial H_z}{\partial x}, \quad H_y = -i\frac{k_z}{\alpha^2}\frac{\partial H_z}{\partial y}, \\ E_z = 0, \end{gathered} \tag{147.10}$$

причем H_z удовлетворяет уравнению

$$\frac{\partial^2 H_z}{\partial x^2} + \frac{\partial^2 H_z}{\partial y^2} = -\alpha^2 H_z, \tag{147.11}$$

отличающемуся от (147.8) только обозначениями. Существенное отличие состоит в изменении *граничных условий*, которому должны удовлетворять решения уравнения (147.11). При идеальной проводимости стенок волновода граничные условия в рассматриваемом случае сводятся к обращению в нуль тангенциальных компонент вектора $\mathbf{E}$, в частности к обращению в нуль E_x и E_y. Иными словами, на стенках волновода должны обращаться в нуль производные $\partial H_z/\partial x$ и $\partial H_z/\partial y$, как это видно из (147.10). Но через эти производные можно выразить производную $\partial H_z/\partial n$ по нормали к поверхности волновода:

$$\frac{\partial H_z}{\partial n} = \frac{\partial H_z}{\partial x}\,\frac{\partial x}{\partial n} + \frac{\partial H_z}{\partial y}\,\frac{\partial y}{\partial n}.$$

Поэтому в более компактной форме граничные условия сводятся к обращению в нуль производной по нормали $\partial H_z/\partial n$ на стенках волновода. Но это обстоятельство не меняет общих выводов, которые были сделаны ранее относительно волн электрического типа. Только допустимые числовые значения параметра α в уравнении (147.11) получаются иными, чем в случае уравнения (147.8). Поэтому нет и необходимости подробно исследовать случай электромагнитных волн магнитного типа.

8. Выведем теперь соотношение между фазовой $v_\text{ф} = \omega/k_z$ и групповой $v_r = d\omega/dk_z$ скоростями распространения волн в волноводе, одинаково справедливое для волн электрического и магнитного типа. С этой целью возьмем одно из собственных значений α. Если изменять частоту ω, то одновременно у рассматриваемой волны будут меняться k и k_z. Поэтому, дифференцируя соотношение (147.5) при постоянном α и принимая во внимание (147.6), получим

$$\varepsilon\mu\frac{\omega d\omega}{c^2} + \frac{\omega^2}{2c^2}\,\frac{d(\varepsilon\mu)}{d\omega}\,d\omega = k_z\,dk_z,$$

откуда

$$v_\text{ф}v_r\left[1 + \frac{\omega}{2\varepsilon\mu}\,\frac{d(\varepsilon\mu)}{d\omega}\right] = \frac{c^2}{\varepsilon\mu} = v^2, \tag{147.12}$$

где $v = c/\sqrt{\varepsilon\mu}$ — фазовая скорость волны в безграничной среде. Если среда, заполняющая волновод, не обладает дисперсией, то

$$v_\text{ф}v_r = v^2. \tag{147.13}$$

В частности, если средой в волноводе является вакуум, то $v_\text{ф}v_r = c^2$. Фазовая скорость не может быть скоростью сигнала, и на нее требования теории относительности не распространяются. Поэтому всегда $v_\text{ф} > c$. Но тогда из последнего соотношения следует, что всегда $v_r < c$, как этого и требует теория относительности, поскольку группа волн, распространяющаяся с групповой скоростью, может служить сигналом.

9. Групповая скорость при определенных ограничениях является также *средней скоростью энергии*, переносимой волной. Докажем эту

теорему в предположении, что среда непоглощающая, но диспергирующая. Зависимость ε и μ от температуры принимать во внимание не будем. Тогда внутренняя энергия среды будет совпадать с ее свободной энергией. Для средней по времени плотности электромагнитной энергии электродинамика дает

$$\overline{W} = \frac{1}{8\pi}\left[\frac{d(\varepsilon\omega)}{d\omega}\,EE^* + \frac{d(\mu\omega)}{d\omega}\,HH^*\right] + \text{ компл. сопр.} \tag{147.14}$$

(см. т. IV, § 88). Что касается вектора плотности потока электромагнитной энергии, то он во всех случаях определяется вектором Пойнтинга S. Его среднее значение для монохроматического поля дается выражением

$$\mathbf{S} = \frac{c}{16\pi}\,[\mathbf{E}\mathbf{H}^*] + \text{ компл. сопр.} \tag{147.15}$$

Средняя скорость распространения электромагнитной энергии определяется формулой

$$v_{\text{эл}} = \frac{\overline{S}}{\overline{W}}. \tag{147.16}$$

Для лучшего уяснения вопроса совпадение скоростей $v_{\text{эл}}$ и v_r докажем сначала не для волновода, а для более простого случая *неограниченной однородной среды.* Если в такой среде распространяется плоская волна в направлении единичного вектора $\mathbf{N}$, то поля $\mathbf{E}$ и $\mathbf{H}$ связаны соотношениями

$$\sqrt{\varepsilon}\,\mathbf{E} = -\sqrt{\mu}\,[\mathbf{N}\mathbf{H}]; \quad \sqrt{\mu}\,\mathbf{H} = [\mathbf{N}\mathbf{E}].$$

Пользуясь этими выражениями, из (147.14), получим

$$\overline{W} = \frac{1}{16\pi}\left\{\frac{d(\omega\varepsilon)}{d\omega} + \frac{\varepsilon}{\mu}\,\frac{d(\omega\mu)}{d\omega}\right\} EE^*.$$

Для преобразования выражения в фигурных скобках запишем выражение (147.6) в виде $k^2 = (1/c^2)(\omega\varepsilon)(\varepsilon\mu)$ и продифференцируем его по ω. Тогда

$$2k\frac{dk}{d\omega} = \frac{\omega\mu}{c^2}\,\frac{d(\omega\varepsilon)}{d\omega} + \frac{\omega\varepsilon}{c^2}\,\frac{d(\omega\mu)}{d\omega},$$

откуда

$$\frac{d(\omega\varepsilon)}{d\omega} + \frac{\varepsilon}{\mu}\,\frac{d(\omega\mu)}{d\omega} = 2k\frac{c^2}{\omega\mu}\,\frac{dk}{d\omega} = 2\frac{\omega}{c}\sqrt{\varepsilon\mu}\,\frac{c}{\omega\mu}\,\frac{dk}{d\omega} = 2c\sqrt{\frac{\varepsilon}{\mu}}\,\frac{dk}{d\omega}.$$

Таким образом,

$$\overline{W} = \frac{c}{8\pi}\sqrt{\frac{\varepsilon}{\mu}}\,\frac{dk}{d\omega}.$$

Далее, выражение (147.15) переходит в

$$\overline{\mathbf{S}} = \frac{c}{8\pi}\sqrt{\frac{\varepsilon}{\mu}}\,(\mathbf{E}\mathbf{E}^*)\mathbf{N}.$$

В результате получается

$$v_{\text{эл}} = \frac{\overline{S}}{\overline{W}} = \frac{d\omega}{dk} = v_r.$$

10. Распространим теперь полученный результат на случай электромагнитной волны в *волноводе*. Не нарушая общности, ограничимся волнами электрического типа. В случае волновода в формуле (147.16) величины $\overline{S}$ и $\overline{W}$ следует заменить на средние значения полного потока энергии через поперечное сечение волновода и энергии, приходящейся на единицу его длины. Средний поток энергии в единицу времени через поперечное сечение волновода определяется выражением

$$\int \overline{S}_z \, df = \frac{c}{16\pi} \int (E_x H_y^* + E_y H_x^*) \, df + \text{компл. сопр.},$$

причем интегрирование ведется по площади поперечного сечения волновода. Подставляя вместо E и H их выражения из (147.7), получим

$$\int \overline{S}_z \, df = \frac{\omega \varepsilon k_z}{8\pi\alpha^4} \int \left(\frac{\partial E_z}{\partial x} \frac{\partial E_z^*}{\partial x} + \frac{\partial E_z}{\partial y} \frac{\partial E_z^*}{\partial y} \right) df. \tag{147.17}$$

Средняя энергия, приходящаяся на единицу длины волновода,

$$\int \overline{W} \, df =$$

$$= \frac{1}{16\pi} \int \frac{1}{\alpha^4} \left[k_z^2 \frac{d\omega\varepsilon}{d\omega} + \frac{\omega^2\varepsilon^2}{c^2} \frac{d(\omega\mu)}{d\omega} \right] \left(\frac{\partial E_z}{\partial x} \frac{\partial E_z^*}{\partial x} + \frac{\partial E_z}{\partial y} \frac{\partial E_z^*}{\partial y} \right) df +$$

$$+ \frac{1}{16\pi} \int \frac{d(\omega\varepsilon)}{d\omega} E_z E_z^* \, df.$$

Очевидно,

$$\int \left(\frac{\partial E_z}{\partial x} \frac{\partial E_z^*}{\partial x} + \frac{\partial E_z}{\partial y} \frac{\partial E_z^*}{\partial y} \right) df =$$

$$= \int \left[\frac{\partial}{\partial x} \left(E_z \frac{\partial E_z^*}{\partial x} \right) + \frac{\partial}{\partial y} \left(E_z \frac{\partial E_z^*}{\partial y} \right) \right] df - \int E_z \left(\frac{\partial^2 E_z^*}{\partial x^2} + \frac{\partial^2 E_z^*}{\partial y^2} \right) df.$$

К первому интегралу в правой части применим математическую теорему Гаусса. Тогда для этого интеграла получим

$$\oint E_z \left(\frac{\partial E_z}{\partial x} n_x + \frac{\partial E_z}{\partial y} n_y \right) df = \oint E_z \frac{\partial E_z^*}{\partial n} \, dl,$$

где интегрирование ведется по замкнутому контуру — линии, получающейся при нормальном сечении волновода, а dl означает элемент длины этого контура. Так как на контуре $E_z = 0$, то рассматриваемый интеграл обращается в нуль.

В случае волны магнитного типа вместо E_z надо брать H_z, но соответствующий интеграл также обращается в нуль, так как на контуре

$\partial H_z^*/\partial n = 0$. Только в этом пункте доказательство для волн магнитного типа отличается от доказательства для волн электрического типа. Второй же интеграл в правой части преобразуется при помощи (147.8), в результате чего получим

$$\int \left(\frac{\partial E_z}{\partial x} \frac{\partial E_z^*}{\partial x} + \frac{\partial E_z^*}{\partial y} \frac{\partial E_z}{\partial y} \right) df = -\alpha^2 \int E_z E_z^* \, df.$$

Умножим это выражение на $\dfrac{1}{\alpha^2} \dfrac{d(\omega\varepsilon)}{d\omega}$ и прибавим его к правой части (147.8). Тогда, принимая во внимание (147.6), найдем

$$\int \overline{W} \, df = \frac{\omega^2 \varepsilon}{16\pi\alpha^4 c^2} \left[\mu \frac{d(\omega\varepsilon)}{d\omega} + \varepsilon \frac{d(\varepsilon\mu)}{d\omega} \right] \int \left(\frac{\partial E_z}{\partial x} \frac{\partial E_z^*}{\partial x} + \frac{\partial E_z}{\partial y} \frac{\partial E_z^*}{\partial y} \right) df. \tag{147.18}$$

Поделив (147.17) на (147.18), получим выражение для средней скорости распространения энергии вдоль волновода

$$v_{\text{эл}} = \frac{c}{v_{\text{ф}}} \frac{1}{\varepsilon\mu \left(1 + \dfrac{\omega}{2\varepsilon\mu} \dfrac{d(\varepsilon\mu)}{d\omega} \right)}. \tag{147.19}$$

Сравнивая эту формулу с (147.12), убеждаемся, что средняя скорость распространения волны в волноводе совпадает с групповой скоростью.

ЗАДАЧИ

1. Найти собственные значения и функции уравнений (147.8) и (147.11) в волноводе прямоугольного поперечного сечения при соответствующих граничных условиях.

Р е ш е н и е. Ищем решение уравнения (147.8) с разделяющимися переменными, т. е. решение в виде

$$E_z = \varphi(x)\psi(y).$$

После подстановки этого выражения в уравнение (147.8) получим

$$\frac{1}{\varphi(x)} \frac{d^2\varphi}{dx^2} + \frac{1}{\psi(y)} \frac{d^2\psi}{dy^2} = -\alpha^2.$$

Первое слагаемое здесь зависит только от x, а второе — только от y. Сумма этих слагаемых есть постоянная. Это возможно тогда и только тогда, когда эти слагаемые сами постоянны. Обозначив их через α_x^2 и α_y^2, получим

$$\frac{1}{\varphi(x)} \frac{d^2\varphi}{dx^2} = -\alpha_x^2,$$

$$\frac{1}{\psi(y)} \frac{d^2\psi}{dy^2} = -\alpha_y^2,$$

причем $\alpha_x^2 + \alpha_y^2 = \alpha^2$. Решения этих уравнений, удовлетворяющие граничным условиям $\varphi(0) = \varphi(a) = \psi(0) = \psi(b) = 0$, где a и b — поперечные размеры волновода вдоль осей x и y соответственно, имеют вид

$$\varphi(x) = \sin \frac{n_x \pi x}{a}, \quad \psi(y) = \sin \frac{n_y \pi y}{b}, \quad n_x, n_y = 1, 2, 3, \ldots$$

Таким образом, собственными функциями будут

$$E_z = \sin\frac{n_x \pi x}{a}\,\sin\frac{n_y \pi y}{b}, \tag{147.20}$$

а собственными значениями

$$\alpha^2 = \alpha_x^2 + \alpha_y^2 = \pi^2\left(\frac{n_x^2}{a^2} + \frac{n_y^2}{b^2}\right). \tag{147.21}$$

В случае волн магнитного вида собственные функции

$$H_z = \cos\frac{n_x \pi x}{a}\,\cos\frac{n_y \pi y}{b},$$

а собственные значения такие же, как и в первом случае.

2. Пусть волновод заполнен однородной средой с $\mu(\omega) = 1$. При каком законе дисперсии $\varepsilon = \varepsilon(\omega)$ связь между фазовой скоростью и скоростью движения энергии в волноводе принимает вид

$$v_{\text{эл}} v_{\text{ф}} = c^2. \tag{147.22}$$

Р е ш е н и е. Для этого необходимо и достаточно, чтобы выражение (147.19) переходило в (147.22), т. е.

$$\varepsilon\left(1 + \frac{\omega}{2\varepsilon}\,\frac{d\varepsilon}{d\omega}\right) = 1$$

или

$$(\varepsilon - 1)\,d\omega + \frac{\omega}{2}\,d\varepsilon = 0.$$

Интегрируя это соотношение, найдем

$$\varepsilon - 1 = \frac{\text{const}}{\omega^2}. \tag{147.23}$$

Таким законом дисперсии обладает ионизированный газ — плазма, причем в этом случае const отрицательна.

ПРИЛОЖЕНИЯ

Важнейшие формулы электродинамики в системе СИ

Для удобства читателя номер формулы совпадает с номером соответствующей формулы основного текста книги в гауссовой системе СГС.

$$\mathbf{F} = q\mathbf{E}. \qquad (2.2)$$

$$F = \frac{1}{4\pi\varepsilon_0}\,\frac{q_1 q_2}{r_{12}^2}. \qquad (3.2)$$

$$\mathbf{E} = \frac{1}{4\pi\varepsilon_0}\,\frac{q}{r^3}\,\mathbf{r}. \qquad (3.4)$$

$$E = \frac{\sigma}{2\varepsilon_0}. \qquad (3.7)$$

$$\mathbf{E} = \frac{1}{4\pi\varepsilon_0}\left[\frac{3(\mathbf{pr})}{r^5}\,\mathbf{r} - \frac{\mathbf{p}}{r^3}\right]. \qquad (4.3)$$

$$\mathbf{M} = [\mathbf{pE}]. \qquad (4.4)$$

$$\mathbf{F} = (\mathbf{p}\nabla)\mathbf{E}. \qquad (4.7)$$

$$\Phi \equiv \oint (\mathbf{E}\,d\mathbf{S}) = \frac{1}{\varepsilon_0}\,q. \qquad (5.5)$$

$$E = \begin{cases} \rho x/\varepsilon_0 & \text{внутри пластинки,} \\ \rho a/\varepsilon_0 & \text{вне пластинки.} \end{cases} \qquad (6.2)$$

$$\mathbf{E} = \frac{1}{4\pi\varepsilon_0}\,\frac{q\mathbf{r}}{a^3} = \frac{1}{3\varepsilon_0}\,\rho\mathbf{r}. \qquad (6.5)$$

$$E = \frac{\varkappa}{2\pi\varepsilon_0}\,\frac{1}{r}. \qquad (6.6)$$

$$\mathbf{E} = \frac{1}{2\varepsilon_0}\,\rho\mathbf{r}. \qquad (6.7)$$

$$E_{2n} - E_{1n} = \frac{\sigma}{\varepsilon_0}. \qquad (6.9)$$

$$f = \frac{\varepsilon_0}{2}\,(E_2^2 - E_1^2)\mathbf{n}. \qquad (6.15)$$

$$E = \frac{\sigma}{\varepsilon_0}. \qquad (6.16)$$

$$\operatorname{div}\mathbf{E} = \frac{1}{\varepsilon_0}\,\boldsymbol{\rho}. \qquad (7.3)$$

$$\sigma_{\text{пол}} = (\mathbf{Pn}) = P_n. \qquad (12.2)$$

$$q_{\text{пол}} = -\oint P_n\,dS = -\oint (\mathbf{P}\,d\mathbf{S}). \qquad (12.3)$$

$$\mathbf{D} = \varepsilon_0\mathbf{E} + \mathbf{P}. \qquad (13.3)$$

$$\oint D_n\,dS = q. \qquad (13.4)$$

$$\rho_{\text{пол}} = -\operatorname{div}\mathbf{P}. \qquad (13.6)$$

$$\mathbf{D} = \frac{1}{4\pi}\,\frac{q}{r^3}\,\mathbf{r}. \qquad (13.10)$$

$$\mathbf{D} = \sigma\mathbf{n}. \qquad (14.3)$$

$$\mathbf{P} = \varepsilon_0\alpha\mathbf{E}. \qquad (15.1)$$

$$\mathbf{D} = \varepsilon_0\varepsilon\mathbf{E}. \qquad (15.2)$$

$$\varepsilon = 1 + \alpha. \qquad (15.3)$$

$$\mathbf{E}^{(i)} = -\frac{1}{3\varepsilon_0}\,\mathbf{P}. \qquad (16.1)$$

$$\mathbf{E}' = \mathbf{E} + \frac{1}{3\varepsilon_0}\,\mathbf{P}. \qquad (16.7)$$

$$\mathbf{E} = -\operatorname{grad}\varphi = -\nabla\varphi. \qquad (18.5)$$

$$\varphi = \frac{1}{4\pi\varepsilon_0\varepsilon}\,\frac{q}{r}. \qquad (19.1)$$

$$\operatorname{div}\operatorname{grad}\varphi = -\frac{1}{\varepsilon_0\varepsilon}\,\rho. \qquad (22.2)$$

$$\mathbf{p} = 4\pi\varepsilon_0 a^3 \mathbf{E}. \qquad (23.5)$$

$$C = 4\pi\varepsilon_0\varepsilon a. \qquad (26.2)$$

$$C = 4\pi\varepsilon_0\varepsilon \frac{R_1 R_2}{R_2 - R_1}. \qquad (26.5)$$

$$C = \frac{\varepsilon_0\varepsilon S}{d}. \qquad (26.6)$$

$$C = \frac{2\pi\varepsilon_0\varepsilon l}{\ln(b/a)}. \qquad (26.7)$$

$$C = \frac{2\pi\varepsilon_0\varepsilon l}{\ln(4h^2/ab)}. \qquad (26.8)$$

$$W = \frac{q^2}{2C} = \frac{1}{2}\, q\varphi = \frac{1}{2}\, C\varphi^2. \qquad (28.3)$$

$$w = \frac{1}{2}\, \mathbf{ED} = \frac{\varepsilon_0\varepsilon E^2}{2} = \frac{D^2}{2\varepsilon_0\varepsilon}. \qquad (29.7)$$

$$w = \int \mathbf{E}\, d\mathbf{D}. \qquad (29.8)$$

$$U = \frac{1}{4\pi\varepsilon_0}\, \frac{q_1 q_2}{r_{12}}. \qquad (30.1)$$

$$dU = T\, dS + \mathbf{E}\, d\mathbf{D}. \qquad (31.6)$$

$$d\Psi = -S\, dT + \mathbf{E}\, d\mathbf{D}. \qquad (31.7)$$

$$d\Phi = -S\, dT - \mathbf{D}\, d\mathbf{E}. \qquad (31.8)$$

$$dI = T\, dS - \mathbf{D}\, d\mathbf{E}. \qquad (31.9)$$

$$\Psi = \frac{\varepsilon_0\varepsilon E^2}{2} + \Psi_0 = \frac{D^2}{2\varepsilon_0\varepsilon} + \Psi_0. \qquad (31.12)$$

$$U = \left(1 + \frac{T}{\varepsilon}\, \frac{\partial\varepsilon}{\partial T}\right) \frac{\varepsilon_0\varepsilon E^2}{2} + \\ + U_0(T, \tau). \qquad (31.14)$$

$$f = -w = -\frac{D^2}{2\varepsilon_0\varepsilon} = -\frac{\varepsilon_0\varepsilon E^2}{2}. \qquad (32.8)$$

$$f = (\mathscr{P} - \mathscr{P}_0) - \left(\varepsilon + \tau\frac{\partial\varepsilon}{\partial\tau}\right) \frac{\varepsilon_0 E^2}{2}. \qquad (32.9)$$

$$f = \frac{\varepsilon_0}{2}\, (\varepsilon_1 - \varepsilon_2) E^2. \qquad (33.4)$$

$$f = \frac{D^2}{2\varepsilon_0\varepsilon} \left(\frac{1}{\varepsilon_2} - \frac{1}{\varepsilon_1}\right). \qquad (33.5)$$

$$\mathscr{T} = -\mathscr{P} + \frac{\varepsilon_0 E^2}{2} \left(\varepsilon + \tau\frac{\partial\varepsilon}{\partial\tau}\right). \qquad (33.8)$$

$$\Pi = \mathscr{P} + \frac{\varepsilon_0 E^2}{2} \left(\varepsilon - \tau\frac{\partial\varepsilon}{\partial\tau}\right). \qquad (33.9)$$

$$\mathbf{f} = -\operatorname{grad}\mathscr{P} + \rho\mathbf{E} - \frac{\varepsilon_0 E^2}{2}\, \operatorname{grad}\varepsilon + \\ + \frac{\varepsilon_0}{2}\, \operatorname{grad}\left(\tau\frac{\partial\varepsilon}{\partial\tau}\, E^2\right). \qquad (34.5)$$

$$\mathbf{p} = \varepsilon_0\beta\mathbf{E}. \qquad (35.1)$$

$$\beta = 4\pi a^3. \qquad (35.2)$$

$$\frac{\varepsilon - 1}{\varepsilon + 2} = \frac{1}{3}\, n\beta. \qquad (35.15)$$

$$\mathbf{P} = \frac{np_0^2}{3kT}\, \mathbf{E}. \qquad (36.5)$$

$$\alpha = \frac{np_0^2}{3kT\varepsilon_0}. \qquad (36.6)$$

$$\varepsilon = 1 + \frac{np_0^2}{3kT\varepsilon_0}. \qquad (36.7)$$

$$\mathbf{j} = ne\mathbf{u}. \qquad (40.1)$$

$$\mathbf{j} = \lambda\mathbf{E}. \qquad (41.1)$$

$$\lambda = \frac{ne^2}{m}\, \overline{\tau}. \qquad (42.12)$$

$$Q = (\mathbf{jE}) = \lambda\mathbf{E}^2 = \frac{1}{\lambda}\, \mathbf{j}^2. \qquad (42.22)$$

$$\varphi_1 - \varphi_2 + \mathscr{E} = IR. \qquad (44.5)$$

$$R = \frac{\varepsilon\varepsilon_0}{C\lambda}. \qquad (46.1)$$

$$\tau = RC = \frac{\varepsilon_0 \varepsilon}{\lambda}. \quad (48.8)$$

$$\mathbf{E} = q(\mathbf{E} + [\mathbf{vB}]). \quad (49.3)$$

$$d\mathbf{F} = [\mathbf{jB}]\, dV. \quad (49.4)$$

$$d\mathbf{F} = I[d\mathbf{l}\, \mathbf{B}]. \quad (49.6)$$

$$\mathbf{B} = \frac{\mu_0 q}{r^3}\, [\mathbf{vr}]. \quad (50.2)$$

$$\mathbf{F}_{12} = \frac{\mu_0 q_1 q_2}{r_{12}^3}\, [\mathbf{v}_2[\mathbf{v}_1 \mathbf{r}_{12}]]. \quad (50.4)$$

$$\mathbf{B} = \frac{\mu_0}{4\pi} \int \frac{[\mathbf{jr}]}{r^3}\, dV. \quad (50.10)$$

$$\mathbf{B} = \oint \frac{\mu_0}{4\pi}\, \frac{I[d\mathbf{l}\, \mathbf{r}]}{r^3}. \quad (50.11)$$

$$B = \frac{\mu_0 I}{2\pi R}. \quad (51.1)$$

$$F = \frac{\mu_0}{2\pi R}\, I_1 I_2 l. \quad (51.2)$$

$$B = \frac{\mu_0 a^2 I}{2r^3}. \quad (51.5)$$

$$B = \frac{\mu_0}{4\pi}\, i\Omega. \quad (51.7)$$

$$B = \mu_0 i. \quad (51.8)$$

$$\mathbf{M} = [\mathfrak{M}\mathbf{B}]. \quad (52.1)$$

$$\mathfrak{M} = I\mathbf{S}. \quad (52.2)$$

$$\oint \mathbf{B}\, d\mathbf{s} = \mu_0 I. \quad (55.4)$$

$$B = \frac{NI}{2\pi R}. \quad (55.10)$$

$$\operatorname{rot} \mathbf{B} = \mu_0 \mathbf{j}. \quad (56.1)$$

$$\mathfrak{m} = I\mathbf{n}. \quad (57.2)$$

$$\mathbf{F} = (\mathfrak{M}\nabla)\mathbf{B}. \quad (57.3)$$

$$\overline{\boldsymbol{\omega}} = -\frac{e}{m}\, \mathbf{B}. \quad (57.4)$$

$$r = \frac{v}{|\omega|} = \frac{mv}{|e|B}. \quad (57.5)$$

$$i_m = I. \quad (58.6)$$

$$\mathbf{j}_m = \operatorname{rot} \mathbf{I}. \quad (59.4)$$

$$\mathbf{H} = \frac{\mathbf{B}}{\mu_0} - \mathbf{I}. \quad (59.5)$$

$$\oint \mathbf{H}\, d\mathbf{l} = I. \quad (59.6)$$

$$\operatorname{rot} \mathbf{H} = \mathbf{j}. \quad (59.7)$$

$$[\mathbf{nH}_2] - [\mathbf{nH}_1] = \mathbf{i}. \quad (60.3)$$

$$\mathbf{I} = \varkappa \mathbf{H}. \quad (61.1)$$

$$\mathbf{B} = \mu_0 \mu \mathbf{H}. \quad (61.2)$$

$$\mu = 1 + \varkappa. \quad (61.3)$$

$$A_{12} = I(\Phi_2 - \Phi_1). \quad (62.2)$$

$$\mathscr{E}^{\text{инд}} = -\frac{d\Phi}{dt}. \quad (64.1)$$

$$\mathbf{E}' = \mathbf{E} + [\mathbf{VB}]. \quad (66.6)$$

$$\mathbf{B}' = \mathbf{B} - \varepsilon_0 \mu_0 [\mathbf{VE}]. \quad (66.7)$$

$$\Phi = LI. \quad (68.1)$$

$$L = \frac{\mu_0 \mu N^2 S}{l}. \quad (68.2)$$

$$\tau = \frac{L}{R}. \quad (68.10)$$

$$W_m = \frac{1}{2} \sum I_i \Phi_i. \quad (69.5)$$

$$w_m = \frac{1}{2}\, \mu_0 \mu H^2 = \frac{1}{2}\, HB = \frac{B^2}{2\mu_0 \mu}. \quad (70.3)$$

$$F = \frac{SB^2}{2\mu_0} \left(\frac{1}{\mu_2} - \frac{1}{\mu_1} \right) = \frac{\mu_0 S}{2} (\mu_2 H_2^2 - \mu_1 H_1^2). \quad (72.5)$$

$$F = \frac{\mu_0 S H^2}{2} (\mu_1 - \mu_2) = \frac{S}{2\mu_0} \left(\frac{B_1^2}{\mu_1} - \frac{B_2^2}{\mu_2} \right). \quad (72.7)$$

$$T = \mathscr{P} = \frac{\mu_0 \mu H^2}{2} = \frac{HB}{2} = \frac{B^2}{2\mu_0 \mu}. \tag{72.8}$$

$$\delta Q = dU - \mathbf{H}\, d\mathbf{B}. \tag{73.1}$$

$$dU = T\, dS + \mathbf{H}\, d\mathbf{B}. \tag{73.2}$$

$$d\Psi = -S\, dT + \mathbf{H}\, d\mathbf{B}. \tag{73.3}$$

$$d\Phi = -S\, dT - \mathbf{B}\, d\mathbf{H}. \tag{73.4}$$

$$dI = T\, dS - \mathbf{B}\, d\mathbf{H}. \tag{73.5}$$

$$\Psi = \frac{\mu\mu_0}{2} H^2 + \Psi_0 = \frac{B^2}{2\mu\mu_0} + \Psi_0. \tag{73.7}$$

$$U = \left(\mu + T\frac{\partial \mu}{\partial T}\right)\frac{\mu_0 H^2}{2} + U_0(T, \tau). \tag{73.8}$$

$$dT = -\frac{TB}{\mu_0 \mu^2 c_B}\,\frac{d\varkappa}{dT}\, dB. \tag{73.10}$$

$$dT = -\frac{\mu_0 T H}{c_H}\,\frac{d\varkappa}{dT}\, dH. \tag{73.11}$$

$$\Gamma = \frac{\mathfrak{M}}{L} = -\frac{e}{2m}. \tag{75.1}$$

$$\mathfrak{M} = -\frac{e\hbar}{2m}\, n. \tag{75.2}$$

$$\mathbf{I} = -\frac{\mu_0 N Z e^2}{6m}\,\overline{R^2}\,\mathbf{H}. \tag{76.3}$$

$$\varkappa = -\frac{\mu_0 Z N e^2}{6m}\,\overline{R^2}. \tag{76.4}$$

$$\mu = 1 - \frac{\mu_0 N Z e^2}{6m}\,\overline{R^2}. \tag{76.5}$$

$$\varkappa = \frac{\mu_0 n \mathfrak{M}^2}{3kT}. \tag{77.5}$$

$$\frac{d\mathbf{j}}{dt} = -\frac{1}{\mu_0 \Lambda^2}\,\mathbf{E}. \tag{80.2}$$

$$\Lambda = \left(\frac{m}{\mu_0 n_s e^2}\right)^{1/2}. \tag{80.3}$$

$$\psi_s + \frac{1}{2\mu_0}\,\mathbf{B}_k^2 = \psi_n. \tag{80.6}$$

$$\mathbf{B} = \mathbf{B}_0 + \frac{3\mu_0}{4\pi}\,\frac{(\mathfrak{M}\mathbf{r})}{r^5}\,\mathbf{r} - \frac{\mu_0}{4\pi}\,\frac{\mathfrak{M}}{r^3}. \tag{80.8}$$

$$\mathfrak{M} = -\frac{2\pi a^3}{\mu_0}\,\mathbf{B}. \tag{80.9}$$

$$\mathbf{j}_{\text{см}} = \dot{\mathbf{D}}. \tag{81.6}$$

$$\mathbf{j}_{\text{см}} = \varepsilon_0 \dot{\mathbf{E}} + \dot{\mathbf{P}}. \tag{81.9}$$

$$\oint\limits_L \mathbf{H}\, d\mathbf{l} = \int\limits_S \left(\mathbf{j} + \frac{\partial \mathbf{D}}{\partial t}\right) d\mathbf{S}. \tag{82.1}$$

$$\oint\limits_L \mathbf{E}\, d\mathbf{l} = -\int \frac{\partial \mathbf{B}}{\partial t}\, d\mathbf{S}. \tag{82.2}$$

$$\oint (\mathbf{D}\, d\mathbf{S}) = \int \rho\, dV. \tag{82.3}$$

$$\oint (\mathbf{B}\, d\mathbf{S}) = 0. \tag{82.4}$$

$$\operatorname{rot} \mathbf{H} = \mathbf{j} + \frac{\partial \mathbf{D}}{\partial t}. \tag{82.1a}$$

$$\operatorname{rot} \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}. \tag{82.2a}$$

$$\operatorname{div} \mathbf{D} = \rho. \tag{82.3a}$$

$$\operatorname{div} \mathbf{B} = 0. \tag{82.4a}$$

$$D_{2n} - D_{1n} = \sigma. \tag{82.5}$$

$$[\mathbf{n}\mathbf{H}_2] - [\mathbf{n}\mathbf{H}_1] = \mathbf{i}. \tag{82.8}$$

$$\mathbf{D} = \varepsilon_0 \varepsilon \mathbf{E}. \tag{82.10}$$

$$\mathbf{B} = \mu_0 \mu \mathbf{H}. \tag{82.11}$$

$$\mathbf{j} = \lambda \mathbf{E}. \tag{82.12}$$

$$v = \frac{1}{\sqrt{\varepsilon_0 \varepsilon \mu_0 \mu}}. \tag{83.2}$$

$$\mathbf{E} = -[\mathbf{v}\mathbf{B}], \quad \mathbf{H} = [\mathbf{v}\mathbf{D}]. \tag{83.3}$$

$$\varepsilon_0 \varepsilon E^2 = \mu_0 \mu H^2. \tag{83.4}$$

$$\mathbf{S} = [\mathbf{E}\mathbf{H}]. \tag{84.3}$$

$$w = \int (\mathbf{E}\, d\mathbf{D} + \mathbf{H}\, d\mathbf{B}). \tag{84.7}$$

$$w = \frac{1}{2}\,(\mathbf{ED} + \mathbf{HB}). \qquad (84.9)$$

$$m\dot{\mathbf{v}} = e(\mathbf{E} + [\mathbf{vB}]). \qquad (86.2)$$

$$\mathbf{v}_{\text{д}} = \frac{[\mathbf{EB}]}{B^2}. \qquad (86.3)$$

$$\mathbf{v}_{\text{д}} = \frac{1}{B^2 e}\,[\mathbf{FB}]. \qquad (86.4)$$

$$\mathbf{v}_{\text{д}} = \frac{mv_\perp^2}{2eB^2}\,\frac{\partial B}{\partial N}\,\mathbf{b}. \qquad (87.2)$$

$$\mathbf{v}_{\text{д}} = \frac{mv_\perp^2}{2eBR}\,\mathbf{b}. \qquad (87.3)$$

$$\mathbf{v}_{\text{д}} = \frac{mv_\parallel^2}{eBR}\,\mathbf{b}. \qquad (87.4)$$

$$\mathbf{V} = v_\parallel \mathbf{h} + \frac{1}{B^2}\,[\mathbf{EB}] + \\ + \frac{m}{eBR}\left(v_\parallel^2 + \frac{1}{2}\,v_\perp^2\right)\mathbf{b}. \qquad (87.5)$$

$$\mathfrak{M} = -\frac{mv_\perp^2}{2B}\,\mathbf{h}. \qquad (88.2)$$

$$\Phi = \frac{m^2 v_\perp^2}{e^2 B}. \qquad (88.3)$$

$$\mathbf{g} = \frac{[\mathbf{EH}]}{c}. \qquad (91.1)$$

$$m_{\text{эл}} = \frac{e^2}{6\pi\varepsilon_0 a c^2}. \qquad (91.4)$$

$$R = \frac{1}{ne}. \qquad (98.4)$$

$$R = \frac{3\pi}{8ne}. \qquad (98.5)$$

$$C = \frac{4\varepsilon_0}{9l^2}\sqrt{\frac{2e}{m}}. \qquad (101.17)$$

$$D = \sqrt{\frac{2\varepsilon_0 kT}{ne^2}}. \qquad (121.1)$$

$$\lambda = \frac{(4\pi\varepsilon_0)^2 (3kT)^{3/2}}{\pi Z e^2 L\sqrt{m}}. \qquad (121.3)$$

$$D = \sqrt{\frac{\varepsilon_0 kT}{2ne^2}}. \qquad (121.8)$$

$$\omega_0 = \sqrt{\frac{ne^2}{m\varepsilon_0}}. \qquad (123.9)$$

$$v = \frac{B}{\sqrt{\mu_0\rho}}. \qquad (138.6)$$

$$\mathbf{D} = \frac{1}{4\pi}\left[\frac{3(\mathbf{pr})}{r^5}\,\mathbf{r} - \frac{\mathbf{p}}{r^3}\right]_{t-r/v} + \\ + \frac{1}{4\pi}\left[\frac{3(\dot{\mathbf{p}}\mathbf{r})}{vr^4}\,\mathbf{r} - \frac{\dot{\mathbf{p}}}{vr^2}\right]_{t-r/v} + \\ + \frac{1}{4\pi}\left[\frac{(\ddot{\mathbf{p}}\mathbf{r})}{v^2 r^3}\,\mathbf{r} - \frac{\ddot{\mathbf{p}}}{v^2 r}\right]_{t-r/v},$$

$$\mathbf{H} = \frac{1}{4\pi r^3}\,[\dot{\mathbf{p}}\mathbf{r}]_{t-r/v} + \\ + \frac{1}{4\pi v r^2}\,[\ddot{\mathbf{p}}\mathbf{r}]_{t-r/v}. \qquad (141.10)$$

$$\mathbf{D} = \frac{1}{4\pi}\left[\frac{(\ddot{\mathbf{p}}\mathbf{r})}{v^2 r^3}\,\mathbf{r} - \frac{\ddot{\mathbf{p}}}{v^2 r}\right]_{t-r/v} = \\ = \frac{1}{4\pi v^2 r^3}\,[[\ddot{\mathbf{p}}\mathbf{r}]\mathbf{r}]_{t-r/v},$$

$$\mathbf{H} = \frac{1}{4\pi v r^2}\,[\ddot{\mathbf{p}}\mathbf{r}]_{t-r/v}. \qquad (141.11)$$

$$\mathbf{S} = \frac{\sin^2\vartheta}{16\pi^2\varepsilon_0\varepsilon v^3 r^2}\,\ddot{\mathbf{p}}^2_{t-r/v}\mathbf{N}. \qquad (141.13)$$

$$-\frac{d\mathscr{E}}{dt} = \frac{1}{6\pi\varepsilon_0\varepsilon v^3}\,\ddot{\mathbf{p}}^2_{t-r/v}. \qquad (141.14)$$

$$-\frac{d\mathscr{E}}{dt} = \frac{\omega^4}{12\pi\varepsilon_0\varepsilon v^3}\,\mathbf{p}_0^2. \qquad (141.16)$$

$$-\frac{d\mathscr{E}}{dt} = \frac{e^2}{6\pi\varepsilon_0\varepsilon v^3}\,\dot{v}^2. \qquad (141.17)$$

$$D = \frac{1}{\mu_0\mu\lambda}. \qquad (144.4)$$

$$l \sim \frac{1}{\sqrt{2\mu_0\mu\lambda\nu}}. \qquad (144.5)$$

Фундаментальные физические константы по ГСССД 1–87

Константа	Обозначение	Числовое значение	Размерность и единица физической величины
Скорость света в вакууме	c	299 792 458	м/с
Магнитная постоянная	μ_0	$4\pi \cdot 10^{-7} =$ $= 12{,}566\ 370\ 614\ldots$	10^{-7} Н/А^2
Электрическая постоянная	ε_0	$(\mu_0 c^2)^{-1} =$ $= 8{,}854\ 187\ 817\ldots$	10^{-12} Ф/м
Гравитационная постоянная	G	6,672 59(85)	10^{-11} м^3/(кг·с^2)
Постоянная Планка	h	6,626 075 5(40)	10^{-34} Дж·с
в электрон-вольтах $h/\{e\}$		4,135 669 2(12)	10^{-15} эВ·с
$h/2\pi$	$\hbar$	1,054 572 66(63)	10^{-34} Дж·с
в электрон-вольтах $\hbar\{e\}$		6,582 1220(20)	10^{-16} эВ·с
Планковская масса $(\hbar c/G)^{1/2}$	m_P	2,176 71(14)	10^{-8} кг
Элементарный заряд	e	1,602 177 33(49)	10^{-19} Кл
Магнетон Бора $e\hbar/2m_e$	μ_B	9,274 0154(31)	10^{-24} Дж/Тл
Боровский радиус $\alpha/(4\pi R_\infty)$	a_0	0,529 177 249(24)	10^{-10} м
Масса покоя электрона	m_e	9,109 389 7(54)	10^{-31} кг
Отношение заряда электрона к его массе	$-\dfrac{e}{m_e}$	−1,758 819 62(53)	10^{11} Кл/кг
Классический радиус электрона $\alpha^2 a_0$	r_e	2,817 940 92(38)	10^{-15} м
Масса покоя протона	m_p	1,672 6231(10)	10^{-27} кг
Отношение массы протона к массе электрона	$\dfrac{m_p}{m_e}$	1836,152 701(37)	
Масса покоя нейтрона	m_n	1,674 9286(10)	10^{-27} кг
Масса покоя дейтрона	m_d	3,343 5860(20)	10^{-27} кг
Постоянная Авогадро	N_A	6,022 1367(36)	10^{23} моль$^{-1}$
Атомная единица массы 1 а.е.м. $= \dfrac{1}{12}\, m(^{12}\mathrm{C}) \equiv m_{\text{а.е.м.}}$	а.е.м.	1,660 5420(10)	10^{-27} кг
Постоянная Фарадея	F	96 485,309(29)	Кл/моль
Универсальная газовая постоянная	R	8,314 510(70)	Дж/(моль·К)
Постоянная Больцмана R/N_A	k	1,380 658(12)	10^{-23} Дж/К

Именной указатель

Абрагам 350
Абрикосов 326
Авенариус 456, 466
Авогадро 183, 300, 308, 397, 418, 645
Альфвен 584
Ампер 9, 211, 221, 234–236, 253, 279, 302
Аррениус 393, 401
Астон 495

Бардин 327
Баркгаузен 316
Барнет 310
Био 212, 215, 216, 223, 229, 272, 273, 331, 595, 596
Боголюбов 327
Богуславский 441
Бозе 327
Бойс 470
Больцман 147, 148, 150, 183, 186, 191, 303, 307, 416, 432, 438, 510, 645
Бор 13, 140, 242, 296–298, 303, 304, 308, 425, 474, 645
Браун 260, 460

Валашек 162
Ван-де-Грааф 55
Ван-Лёвен 296
Вант-Гофф 393
Вебер Вильгельм 10, 217, 344
Вейнер 162
Вейсс 165, 169, 296, 310–315
Видеман 186, 423, 508
Вильсон 134, 382
Вирсма 291
Вольта 406, 450, 453, 455
Вул 162

Гамильтон 24
Гаусс 10, 27, 28, 30–33, 35, 38–44, 48, 51, 60–63, 84, 88, 89, 92, 104, 111, 117, 135, 176, 201, 217, 223, 228, 234, 255, 256, 331, 340, 633, 637
Гейзенберг 140, 314
Гель 496
Гельмгольц 133, 138, 216, 258, 398
Герц Генрих 10, 11, 345, 474, 592–600
Герц Густав 473–476
Гинзбург 169, 169
Гитторф 399, 401
Гольдман 162
Гольдстейн 498
Грин 10
Гротгус 391, 393
Гук 154, 158, 515
Гухер 475

Даниэль 405–407
Дебай 291, 302
Де Гааз 291, 309, 553
Дешман 438
Джиок 291
Джорджи 362
Джоуль 178, 184, 192, 196, 392, 460
Дирак 185, 186, 307, 327, 337, 416, 418, 419, 436
Дорфман 314
Друде 186, 418, 423
Дэви 397, 503
Дэвис 475

Жуковский 308

Зеебек 454, 455, 460
Зелени 484
Зоммерфельд 186, 418, 423

Иоффе 383, 384, 462
Ирншоу 44–45, 51

Кавендиш 17, 34
Камерлинг-Оннес 317
Карлейль 397
Кеттеринг 413
Кённингам 385
Кирхгоф 10, 197–199, 545, 606, 611
Клаузиус 146, 150, 392, 463–466
Кобеко 162
Кольрауш Рудольф 217, 344
Кольрауш Фридрих 399, 401
Кориолис 189, 268, 299
Крукс 483, 497
Кулон 9, 17–19, 21, 22, 31–35, 39, 40, 92, 100, 189, 213, 224, 234, 314, 331, 333, 356, 470, 599
Кундт 159
Купер 327
Курчатов 162
Кюри Жак 151, 155, 162
Кюри Пьер 151, 155, 161, 162, 165, 169–172, 290, 291, 296, 304, 305, 307, 312–315, 317

Ландау 170, 316, 327, 426
Ланжевен 158, 159, 296, 302–304, 307, 311–313
Лаплас 9, 85, 86, 94, 176, 200, 204
Лармор 299, 301, 302
Лаутон 34
Лебедев 605, 621

Лекланше 407
Ленгмюр 440, 441, 504
Ленц 178, 184, 192, 196, 260–263, 273, 283, 301, 341, 392, 460
Леру 461, 464
Лехер 606–608
Ле Шателье 157, 260, 460
Липпман 155
Лондон Ганс 321, 322
Лондон Фриц 321, 322
Лоренц 13, 46, 70, 145, 146, 168, 186, 211, 257, 263, 265, 268, 355, 377, 380, 412, 415
Лоусон 510

Мак-Клюнг 481
Маклаков 560
Максвелл 7, 11–14, 17, 34, 40, 60, 70, 129, 130, 133, 138, 175, 176, 217, 224, 263–265, 272, 331, 332, 334–337, 344, 355, 387, 438, 514, 545, 585, 593–595, 605, 617
Мандельштам 412, 552, 572
Мейсснер 153, 320–321
Менделеев 476
Мерц 165
Милликен 382–386
Миткевич 503
Морзе 624, 627
Мосотти 146, 150

Неель 316
Нейман Карл 10
Нейман Франц 10
Нернст 305–307, 402
Николь 605
Никольсон 397
Ньютон 9, 13, 31, 45, 345, 354, 582

Огава 162
Оксенфельд 320–321
Ом 175–180, 183, 184, 189, 192–198, 201, 269, 273, 318, 327, 392, 403, 411, 436, 444, 478, 540, 543
Оствальд 394
Остроградский 10, 41, 117, 135, 234

Папалекси 412, 552, 572
Паули 307, 327, 419, 425
Пашен 492, 493
Пельтье 454, 460–463, 466
Петров 503
Планк 297, 390, 475, 645
Плимптон 34
Поггендорф 199
Пойнтинг 348, 349, 351, 358, 592, 609, 620, 636
Поккельс 162
Поль 20, 455
Попов 12, 345, 606
Пристли 34
Пуазейль 239
Пуассон 9, 85, 440, 511

Рауль 393
Резерфорд 13, 480, 481, 484
Рике 411
Ричардсон 438
Роговский 267, 268
Розинг 311
Роуланд 216

Савар 212, 215, 216, 223, 229, 272, 273, 331, 595, 596
Симон 291
Скотт 413
Снеллиус 605
Соломон 162
Стокс 233, 281, 384–386
Столетов 292
Стоней 398
Стюарт 178, 412, 413

Таунсенд 382, 480, 486–492, 500
Тесла 614, 615
Толмен 178, 412, 413
Томсон Вильям (Кельвин) 464, 519
Томсон Дж. Дж. 13, 378, 381
Томсон Элиу 261, 549
Тэт 456, 466

Уитстон 198, 199
Умов 348

Фарадей 10–12, 53–56, 86, 88, 97, 100, 106, 130, 256, 264, 397, 418, 645
Фейнман 83
Ферми 183, 185, 186, 307, 327, 416, 416–426, 434–437, 451, 512
Фохт 159
Франклин 52
Франк 473–475
Франц 10, 186, 423, 508
Френкель 314, 427
Фуко 262, 263, 295
Фурье 309, 529, 536–540, 624

Хаблютцель 165
Хевисайд 355
Холл 414–418, 434

Шриффер 327

Эдисон 435
Эйлер 527
Эйнштейн 191, 309, 327, 386, 389, 553
Эйхенвальд 216
Эрстед 212, 251

Юнг 156

Предметный указатель

Автоколебания 560
Адиабатический инвариант 376
Аккумулятор, емкость 410
— свинцовый 409
— щелочной 410
Акцепторы 433
Ампер (единица тока) 195, 217
Ампера закон 211
— теорема 234
Ампер-витки 574
Амплитуда 519, 522, 587
Анодный ток 443
Антенна 622
Антикатод 498
Антисегнетоэлектричество 172
—, точка Кюри 172
Антиферромагнетизм 316
Антиферромагнетики 291, 316
Апериодический процесс 523

Баллистическая постоянная 525
Баллистический гальванометр 524
Баркгаузена эффект 316
Барнета опыт 310
Биения 550
Био–Савара закон 215, 595, 596
Боковые полосы 624
Бора и Ван-Лёвен теорема 296
— магнетон 297

Ван-де-Граафа генератор 55
Вебер (единица магнитного потока) 272
Ведущий центр 368
Вейсса постоянная 311
Векторная диаграмма 525
Векторные линии 21
Векторный метод 527
Взаимодействие переменных токов 549
Видемана–Франца закон 186, 423, 423, 508
Вильсона камера 134
Внутреннее (дифференциальное) сопротивление лампы 444
Волновая зона 596
Волновое уравнение 581
— число 587
Волновой вектор 588
Волны альфвеновские 584
— синусоидальные (монохроматические) 587
— стоячие 588, 589, 610
— сферические 597
Вольт (единица напряжения) 195
Вольтова дуга 503
Вольты закон последовательных контактов 450
— ряд 450
Восприимчивость магнитная 249, 358
Время затухания 522, 530
— инерционное 179
— релаксации 206
— свободного пробега 180
— установления 530
Вторичные электроны 446
Вырождение 581

Газ вырожденный 420
— — полностью
— — сильно 421
Гальванический элемент 405–409
Гамильтона оператор «набла» 24
Гамма 217
Гармоники 590
Гармонический осциллятор 518
Гаусс (единица напряженности магнитного поля) 217
Гаусса–Остроградского формула 41
Гаусса способ 255
— теорема 27, 31, 60
— —, дифференциальная форма 38, 39
— —, для магнитных полей 244
Геля закон 496
Генри (единица индуктивности) 272
Герца вибратор 600
— диполь 592–600
— —, излучение 599
— резонатор 601
Гетеродин 627
Гинзбурга теория 169
Гиромагнитное отношение 297
Гиромагнитные явления 308
Гистерезис, двойные петли 172
— диэлектрический 168
— магнитный 249, 293
— —, теплота 295
—, петля 168, 293
Глубина модуляции 538, 623
Градиент 72
Граничные условия 61, 247, 248, 338, 590
Гроттуса гипотеза 391

Двойной электрический слой 403, 435

Дебаевская длина (радиус) 505, 511
Дебаевский потенциал 512
Действие на расстоянии 9
Дейтерий 509
Декремент 523
Демодуляция (детектирование) 625
— сеточная 627
Детектор 602, 625
Джоуля–Ленца закон 184, 196, 392, 392
Диамагнетизм 298
Диамагнетик 250, 291
Дивергенция 39, 43
Динатронный эффект 446
Диод вакуумный 435
Диполь электрический 22
Диссоциации степень (коэффициент) 394
Дифференцирующие и интегрирующие ячейки 517
Диффузия амбиполярная 506
Диэлектрик (изолятор) 426, 430
Диэлектрическая проницаемость 357
— — абсолютная 357
— — относительная 357
Длина волны 588
Добротность 523, 532
Домены диэлектрические 167
— магнитные 315
Доноры 433
Дорфмана опыт 314
Дрейф 368, 178
— магнитный 371–374
— центробежный 374
— электрический 366, 369–371
Дрейфовая теория 369
Дроссель 541
Друде формула 186
Дырки 431

Емкость 97
— двух параллельных проволок 101
— единицы длины линии 607
— конденсатора плоского 99
— — слоистого 102
— — цилиндрического 100, 103
— — шарового 98
— эллипсоида 104

Задача многих тел 429
— одного тела 429
Заряд магнитный 224, 337
— удельный 378
Заряды индукционные 50
— поляризационные 57
— свободные 57
— связанные 57
Защита магнитная 250
— электростатическая 54
Зеебека явление 454, 460, 466
Зонд капельный 81
— пламенный 81
— радиоактивный 81
— электрический 80
Зонная теория 429

Излучение бетатронное 510
— магнитотормозное 510
— рекомбинационное 476
— тормозное 510
— ускоренно движущегося заряда 599
Импеданс 540
Импульс электромагнитный 350
Индуктивности коэффициенты 275
Индуктивность взаимная 275
— единицы длины линии 607
— провода 271
— соленоида 271
—, теорема взаимности 276
Индукции вектор 60
Индукция магнитная 247
— электромагнитная 256
— —, максвелловская трактовка 263
— —, механическая аналогия 268
— —, основной закон 258
— —, — —, дифференциальная форма 265
— —, фарадеева трактовка 256
— —, электродвижущая сила 257
Ионизация 427, 470
—, метод электронного удара 473
— объемная 470
— поверхностная 470
—, потенциал 473
— ударная 487
—, энергия 427
Ионосфера 506
Ирншоу теорема 44
Искра конденсированная 500
Искровые каналы 499

Катод оксидный 439
— сложный 439, 448
Катодное падение 494, 496
— — аномальное 497
— — нормальное 496
— пятно 504
Катодные лучи 497
Квазинейтральность 480
Квазистационарность 206, 520
Квантовые состояния 419
Кеттеринга и Скотта опыт 413
Кённингама поправка 385

Кирхгофа правила 197, 198
— — для переменных токов 545
Классический радиус электрона 389
Колебания 513
— вынужденные 530
— гармонические 518, 527
— затухающие 522
— модулированные 537, 550
— нормальные 580
— основные 590
— параметрические 568
— плазменные 521
—, процессы установления 530
— релаксационные 568
— с двумя степенями свободы 577
—, самовозбуждение 562
— свободные 516, 530
— связанные 577
— собственные 590
Колебательный контур 513
Колесо Франклина 52
Кольрауша закон 401
Комплексные обозначения 527
Комптоновская длина волны 390
Контактная разность потенциалов внешняя 452
— — — внутренняя 452
— — —, компенсационный метод измерения 417
Концентрация эквивалентная 399
Кориолиса сила 268
Корона 501
— отрицательная 501
— положительная 501
Коэффициенты емкости 105
— —, симметрия 110
— индукции 105
— —, симметрия 278
— потенциальные 104
— —, симметрия 110
Кулон (единица заряда) 18, 217
Кулона закон 17, 18, 217, 331, 356
Кулоновский логарифм 507
Куперовские пары 327
Кюри–Вейсса закон 165, 313
— — температура 165
— точка 165
— — антиферромагнитная 317
— — диэлектрическая 161
— — магнитная 296

Лампа трехэлектродная 442
Ламповый выпрямитель 442
— генератор 561
— усилитель 444
— —, коэффициент усиления 445
Ланжевена теория магнетизма 302–304
— функция 304
Лармора частота 299, 365
Ларморовский радиус 237, 365
Лебедева опыты 621
Ленгмюра–Богуславского закон трех вторых 441
Ленца правило 260, 341
Лехера система 606
Ле Шателье–Брауна принцип 260, 460
Линии тока 20
Лондонов теория сверхпроводимости 321

Магнетик 243
Магнитная гидродинамика 506
— постоянная 356
— проницаемость 357
— — абсолютная 357
— — относительная 360
— термоизоляция 509
Магнитное поле, напряженность 210, 247
— — прямого провода 228
— — соленоида 229
— тороидальной катушки 231
Магнитный листок 235
— момент тока 221
Магнитокалорический эффект 290
Магнитомеханические явления 308
Макроскопическое поле (макрополе) 47, 243
Маклакова молоток 561
Максвелл (единица магнитного потока) 272
Максвелла уравнения 336
— — в рационализированной форме (СИ) 355
Максвелловские натяжения и давления 129, 287
Мандельштама и Папалекси опыт 412, 552
— — параметрическая машина 572
Масса и энергия в теории относительности 350, 386
— электромагнитная 388
Масс-селектор 238
Масс-спектрография 382
Масс-спектрометрия 382
Материальные уравнения 339
Мейсснера–Оксенфельда эффект 320
Металлы 430
Метод масляных капель Милликена 382

— натекания 477
— парабол 381
— постоянного отклонения 477
Мировой эфир 12
Модулирующая функция 623
Модуль пьезоэлектрический 154
Моды 580
Молекулы полярные 57
Молекулярное поле Вейсса 311
Молния 499
Моль 397
Момент диполя 22
Монополи 337
Мосотти–Клаузиуса формула 146
Мощность переменного тока 548

Намагничивание спонтанное 292, 312
Намагничивания вектор 244
Напряжение магнитное 267
— разложения электролита 408
— электрическое 194
Нееля закон 316
Нернста метод 402
Нормальные координаты 580
Носители тока 173
— — неосновные 434
— — основные 434

Обертоны 590
Область полярная 161
Обратной связи катушка 561
Огни святого Эльма 502
Ом (единица сопротивления) 195
Ома закон 175, 194, 194, 392, 403, 411, 436, 444, 478, 540, 543
— — для переменных токов 540
Основной период 536
Оствальда закон разведения 394
Открытая система 622
Отражение волн 604, 609
— заряженных частиц 378

Пара электрон–позитрон 16, 390
Парамагнетизм, объяснение 301
— свободных электронов 307, 425
Парамагнетик 249
Паули принцип 419
Пашена закон 493
Пельтье коэффициент 460
— крест 460
— теплота 460
— явление 460
Пентод 446
Переводной коэффициент 358
Перезарядка 498
Перенапряжение при электролизе 409
Переходные процессы 547
Пироэлектрический эффект 159, 160
Пироэлектричество 159, 160
Плазма 505
—, время релаксации 508
— двухтемпературная 508
—, коллективные колебания 506
— неизотермическая 509
— полностью ионизованная 506
—, проводимость 506
— слабо ионизованная 506
— умеренно ионизованная 506
Планетарная модель атома 45
Платина платинированная 404
Плотность тока 173
— — линейная 218
— электричества поверхностная 21
— — объемная 38
Поггендорфа метод 199
Подвижность ионов 398
— —, метод Зелени 484
— —, — переменного поля 484
— — по Кольраушу 399
— частицы 182
Пойнтинга вектор 349
Положительные лучи (каналовые) 498
Полупроводники 426, 430
—, выпрямляющее действие 466
Полупроводники смешанные 434
— p-типа 434
— n-типа 434
Поля безвихревые 233
— бездивергентные (соленоидальные) 224
— вихревые 233
— потенциальные (консервативные) 69
Поляризации вектор 58
Поляризация вакуума 390
— диэлектриков 56
Поляризуемость 63, 141, 164, 358, 358
Порошковых фигур метод 316
Потенциал бесконечно длинного цилиндра 77
— векторный 281, 282
— зажигания 567
— плоскопараллельной пластинки 77
— погасания 567
— прямолинейной нити 77
— равномерно заряженной плоскости 76
— — — сферы 77
— точечного диполя 77
— — заряда 77
— электрический 71

— электродный 404
— — нормальный 405
Поток вектора 27
— импульса 350
— магнитный 223
— —, теорема сохранения 283
— энергии 347
Преломление волн 604
— — двойное 605
Принципы радиосвязи 621–629
Пробное тело 14
Пробный виток 223
— заряд 14
Пробой вакуума 390
— газа 492
Проводимость 175
— дырочная 432
— примесная 432
— собственная 432
— электронная 432
Произведения подводимых напряжений 624
Проницаемость диэлектрическая 63
— магнитная 249
— сетки 444
Пучности 589
Пьезоэлектрический эффект 151–155
Пьезоэлектричество 151–159

Работа выхода 434
Радиоприемник прямого усиления 627
— супергетеродинный 627
Размагничивающий фактор 252
Разряд дуговой 500, 502
—, зажигание 492
— искровой 499
— —, стримерная теория 500
— коронный 501
— несамостоятельный 478, 491
— самостоятельный 479, 492
— сложный 495
— таунсендовский 489
— тлеющий 494
— — аномальный 497
Раствор нормальный 404
Резонанс 532
— параметрический 570
Резонансная кривая 532
— — амплитудная 535
— — фазовая 535
— —, ширина (полуширина) 532, 532
Рекомбинация 427
—, коэффициент 472
—, —, измерение 480
—, —, —, методы Мак-Клюнга 481
–, —, —, — переключения 481
—, —, —, — Резерфорда 480
Рентгеновские лучи 497
— трубки 497
Рике опыт 411
Ричардсона–Дешмана формула 438
Роговского пояс 267
Ротор вектора 233, 234
Ртутная лампа 504

Самоиндукции коэффициент 271
Сантиметр (единица емкости) 98
— (единица индуктивности) 272
Сверхпроводимость 317
Сверхпроводник 317
— второго рода 320, 326
—, глубина проникновения 321
—, двухжидкостная модель 318
Сверхпроводник, домены 325
—, критическая температура 317
—, критический ток 324
—, критическое поле 322
—, первого рода 320, 326
Свечение анодное 495
— положительное 495
— рекомбинационное 495
— тлеющее 495
Свободная энергия и силы 123–129
Связанные маятники 578
Сегнетова соль 162
Сегнетоэлектрик 162
Сегнетоэлектричество 161–172
Секунда (единица удельной проводимости) 175
— обратная (единица удельного сопротивления) 175
Сетка защитная (противодинатронная) 446
— управляющая 442
— экранная 446
Сеточный ток 443
Сила коэрцитивная 294
— Лоренца 211, 355, 412
— тока 193
— электрического изображения 434
— электродвижущая 193
Силовые линии квазизамкнутые 71
— — магнитные 21
— —, преломление 65
— — электрические 19
Силы амперовы 212
— обменные 314
— пондеромоторные 123, 137
— сторонние 189
— электродинамические 9
— электромагнитные 9

— электрострикционные 133
Система единиц абсолютная 354
— — гауссова 18, 210
— — международная (СИ) 354
— — Хевисайда–Лоренца 377
— — электромагнитная (СГСМ) 18
— — электростатическая (СГСЭ) 18
Скин-эффект 613
Скорость дрейфовая (упорядоченная) 173
— света 360
Смесительная лампа (смеситель) 629
Смешение частот 629
Смещения возможные (виртуальные) 124
— действительные 124
Соединение конденсаторов 101
Соленоид 219
Соотношение неопределенностей 389
Сопротивление активное 544
— волновое 608
— емкостное 542
— заземления 202
— индуктивное 541
— комплексное 540
— омическое 540
— полное 544
— реактивное 544
— удельное 175
Спин электрона 242, 298
Стокса теорема 233
Страты 495
Стример 500, 501
Суперпозиции принцип 529

Таунсенда коэффициенты α, β и γ 487, 489
— опыт 487
— теория 488
Телесный угол 224–227
Темное пространство астоново 495
— — гитторфово 495
- — — катодное 495
— — круксово 495
— — фарадеево 495
Температура вырождения электронного газа 421
Теорема о циркуляции магнитного поля 228, 232, 246
Теплоемкость электронного газа 421
Термодинамика диэлектриков 119–123
— магнетиков 289–291
Термоэлектрический ряд 455
Термоэлектричество 454
—, коэффициент термоэлектродвижущей силы 456
—, — Томсона 465
—, теория Клаузиуса 464
—, — — Томсона 466
—, теплота Томсона 464
—, точка инверсии 456
Тетрод 446
Титанат бария 163
Ток насыщения 436, 479
— полный 332
— поляризации 335
— смещения 332
— термоэлектрический 454
Токи индукционные 256
— молекулярные 243
— намагничивания 243
— поверхностные 218
— проводимости 243
Томсона Вильяма формула 519
— — явление 454, 464
Томсона Элиу опыт 261, 549
Точечный заряд 17
Транзистор 565
Трансформатор 572
—, передача электрической энергии на расстояние 576
— Тесла 614
—, ток холостого хода 574
Триод 442
— полупроводниковый (транзистор) 565
Тритий 509

Удельный заряд 378
Узлы 589
Уитстона мостик 198
Умова вектор 348
Умова–Пойнтинга теорема 348
Управляемые термоядерные реакции 509
Управляющее напряжение 444
Уравнение колебательного контура 513, 517
— непрерывности (неразрывности) 175

Фаза колебаний 519
— — начальная 519
— неполярная 161
— полярная 161
Фазовое пространство 420
Фарад 98
Фарадея законы электролиза 397
— постоянная 397, 645
— теорема 53
— цилиндр 54
Ферми граница 420
Ферромагнетизм 291–296

Ферромагнетики 291
Флюксметр 267
Фотоэлектрический эффект 449
Фотоэлектронный умножитель 448
Френкеля–Гейзенберга объяснение ферромагнетизма 314
Фуко токи 262
Функция периодическая 536
Фурье интеграл 540
Фурье коэффициенты 536
— теорема 536

Характеристика анодная 443
—, крутизна 444
— сеточная 443
Химический эквивалент 397
Холла коэффициент 417, 418
— постоянная 416
— эффект 414
— явление 414

Циклотронная частота 237, 365
Циклотронный радиус 365
Циркуляция 70

Частота биений 550
— колебаний 519
— модуляции 538, 550
— несущая 538, 550
— основная 536
— плазменная 521
— циклическая (круговая) 519
Числа переноса 401

Эйнштейна и Де Гааза опыт 309, 553
Эквипотенциальные поверхности 73
Экстраток замыкания 273
— размыкания 274
Электрет 64
Электрическая дуга Петрова 503
— — —, кратер 503
— постоянная 356
Электрические изображения 87
— — в плоскости 87
— — — сфере 88
Электрический заряд 14
— —, закон сохранения 174
— —, термоэлектронный 435
— ток 173
Электрическое поле бесконечного цилиндра 35
— — бесконечной пластинки 31
— — — плоскости 32
— — Земли 82
— —, напряженность 14
— — параллельных плоскостей 37
— — равномерно поляризованного шара 66, 67
— — сферической полости 68
— — шара 32
— — эллипсоида 92–97
Электрод нормальный водородный 404
Электродинамическая постоянная 213
Электролиз 391
—, вторичные продукты 397
—, —, реакции 391
—, первичные продукты 397
—, —, реакции 391
Электролитическая ванна 203–206, 616
— —, метод сечений 204
Электролиты 391, 394, 426
Электромагнитное поле 10
Электромагнитные волны (возмущения) 342–343, 586, 621
Электрометр квадратный 476
— струнный 476
Электронная теория поляризации 140
— — — неполярных диэлектриков 140–146
— — — полярных диэлектриков 146–151
Электронные лавины 487
Электроны валентные 427
— свободные 427
Электропроводность эквивалентная 399
Электрострикция 133
Электрофильтр 502
Элемент гальванический 191
Элементарные законы 9
Эмиссионная способность 436
Эмиссия автоэлектронная (холодная) 449
— вторичная 447
— термоэлектронная 381, 435
Эмиттеры 566, 448
Энергетические зоны кристалла 428, 428
— уровни 419, 427, 433
Энергия взаимная 117
— и силы 284
— магнитная 276
— поляризации диэлектриков 122
— магнетиков 290
— свободная 110, 120, 278
— электрическая 110
— электромагнитная 347
Эрстеда опыт 212
Эффективное напряжение 548
Эффективный ток 548

Учебное издание

СИВУХИН Дмитрий Васильевич

ОБЩИЙ КУРС ФИЗИКИ

Том 3

ЭЛЕКТРИЧЕСТВО

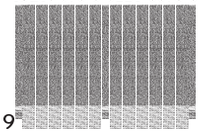
Редактор *О.В. Салецкая*
Оригинал-макет: *В.В. Худяков*
Оформление переплета: *А.А. Логунов*

Подписано в печать 20.12.05. Формат 60×90/16. Бумага офсетная.
Печать офсетная. Усл. печ. л. 41. Уч.-изд. л. 45,1. Тираж 3000 экз.
Заказ №

Издательская фирма «Физико-математическая литература»
МАИК «Наука/Интерпериодика»
117997, Москва, ул. Профсоюзная, 90
E-mail: fizmat@maik.ru, fmlsale@maik.ru;
http://www.fml.ru

Отпечатано в ПФ «Полиграфист»
160001, г. Вологда, ул. Челюскинцев, 3
Тел.: (8172) 72-07-92, 72-61-75, 72-60-63; факс: (8172) 76-00-49, 72-71-11
E-mail: forma@pfpoligrafist.com

ISBN 978-5-9221-0673-3

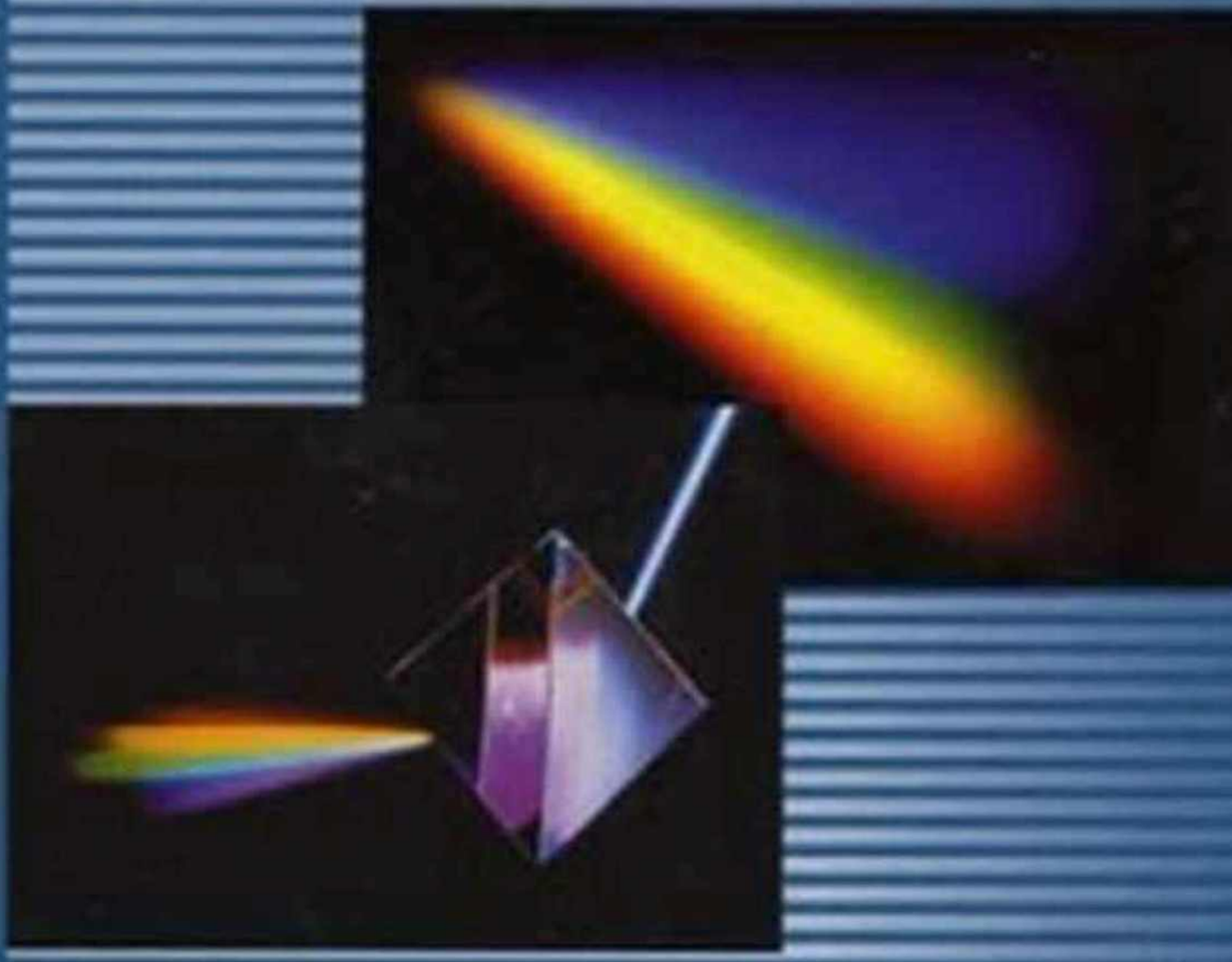


ОБЩИЙ КУРС ФИЗИКИ

Д.В. Сивухин

ОПТИКА

ТОМ IV



УДК 535
ББК 22.34
С34

Сивухин Д. В. **Общий курс физики.** Учеб. пособие: Для вузов. В 5 т. Т. IV. **Оптика**. — 3-е изд., стереот. — М.: ФИЗМАТЛИТ, 2005. — 792 с. — ISBN 5-9221-0228-1.

Четвертый том курса физики, широко известного у нас в стране и за рубежом. Книга написана на основе лекций, которые в течение ряда лет читались автором студентам Московского физико-технического института. Основное внимание уделено выяснению физического смысла и содержания основных законов и понятий оптики, установлению границ применимости этих законов, развитию у студентов навыков физического мышления и умения ставить и решать конкретные задачи.

Второе издание четвертого тома вышло в 1985 г.

Для студентов физических и математических факультетов университетов, физико-технических и инженерно-физических институтов, а также вузов, где физика является основной дисциплиной.

ISBN 5-9221-0228-1 (Т. IV)
ISBN 5-9221-0229-X

© ФИЗМАТЛИТ, 1985, 2002, 2005

ОГЛАВЛЕНИЕ

Предисловие 7

ГЛАВА I

ВВЕДЕНИЕ

§ 1. Предмет оптики 9
§ 2. Геометрическая оптика 12
§ 3. Эволюция представлений о природе света 20
§ 4. Искривление световых лучей в неоднородных сферах 33
§ 5. Плоские электромагнитные волны 37
§ 6. Предельный переход от волновой оптики к геометрической . . . 44
§ 7. Принцип Ферма 49
§ 8. Групповая скорость 57

ГЛАВА II

ГЕОМЕТРИЧЕСКАЯ ТЕОРИЯ ОПТИЧЕСКИХ ИЗОБРАЖЕНИЙ

§ 9. Понятие оптического изображения 67
§ 10. Преломление на сферической поверхности. Сферические зеркала и тонкие линзы 73
§ 11. Общие свойства центрированных оптических систем 78
§ 12. Сложение центрированных систем. Толстые линзы 89
§ 13. Ограничение лучей при помощи диафрагм 96
§ 14. Астигматические пучки лучей. Каустика 101
§ 15. Геометрические аберрации центрированных систем 106
§ 16. Хроматическая аберрация 113
§ 17. Условие отсутствия дисторсии 120
§ 18. Условие синусов Аббе 122
§ 19. Теорема косинусов. Стигматические изображения широкими пучками лучей 129
§ 20. Об абсолютно оптических инструментах 135
§ 21. Глаз и зрение 139
§ 22. Фотометрические понятия и единицы 152
§ 23. Яркость и освещенность оптического изображения. Нормальное увеличение 163
§ 24. Оптические инструменты 172
§ 25. Электрические и магнитные линзы 190

ГЛАВА III
ИНТЕРФЕРЕНЦИЯ СВЕТА

§ 26. Общие сведения об интерференции . . . 199
§ 27. Классические интерференционные опыты . . . 210
§ 28. Влияние размеров источника света. Пространственная когерентность . . . 217
§ 29. Спектральное разложение . . . 225
§ 30. Влияние немонохроматичности света . . . 230
§ 31. Корреляция и когерентность света . . . 234
§ 32. Теорема Ван-Циттера–Цернике . . . 240
§ 33. Интерференция в пленках и пластинках . . . 242
§ 34. Интерферометр Жамена . . . 249
§ 35. Интерферометр Майкельсона . . . 256
§ 36. Многолучевая интерференция . . . 258
§ 37. Стоячие световые волны . . . 267
§ 38. Излучение Вавилова–Черенкова . . . 271

ГЛАВА IV
ДИФРАКЦИЯ СВЕТА

§ 39. Принцип Гюйгенса–Френеля. Зоны Френеля . . . 277
§ 40. Дифракция на оси от круглого отверстия и экрана. Зонная пластинка . . . 284
§ 41. Метод Френеля решения дифракционных задач. Дифракция Фраунгофера и Френеля . . . 291
§ 42. Зоны Шустера и спираль Корню . . . 298
§ 43. Принцип Гюйгенса в формулировке Кирхгофа . . . 304
§ 44. Дифракция Фраунгофера на щели . . . 309
§ 45. Дифракция Фраунгофера на отверстиях . . . 315
§ 46. Дифракционная решетка . . . 320
§ 47. Дифракционная решетка как спектральный прибор . . . 329
§ 48. Эшелон Майкельсона и интерференционные спектральные приборы . . . 334
§ 49. Разрешающая способность призмы . . . 339
§ 50. Действие спектрального аппарата на световые импульсы . . . 344
§ 51. Вогнутая отражательная решетка . . . 350
§ 52. Дифракция на решетке как краевая задача . . . 353
§ 53. Примеры на применение метода Рэлея . . . 359
§ 54. Голография . . . 363
§ 55. Световое поле вблизи фокуса . . . 374
§ 56. Разрешающая способность телескопа и микроскопа . . . 378
§ 57. Теория и демонстрационные опыты Аббе . . . 388
§ 58. Телескоп без объектива. Получение изображений с помощью малых отверстий . . . 395
§ 59. Фазовый контраст . . . 399
§ 60. Измерение угловых диаметров звезд . . . 402
§ 61. Дифакция на двумерных и трехмерных решетках. Дифракция рентгеновских лучей . . . 407

ГЛАВА V
ОТРАЖЕНИЕ И ПРЕЛОМЛЕНИЕ СВЕТА

§ 62. Поляризованный и естественный свет 420
§ 63. Число независимых граничных условий в электромагнитной теории света 423
§ 64. Геометрические законы отражения и преломления волн 426
§ 65. Формулы Френеля 430
§ 66. Полное отражение 436
§ 67. Прохождение света через плоскопараллельную пластинку. Просветление оптики 441
§ 68. Распространение света в среде с точки зрения молекулярной оптики 450
§ 69. Вывод формул Френеля в молекулярной оптике 455
§ 70. Отступления от формул Френеля 459

ГЛАВА VI
ОПТИКА МЕТАЛЛОВ

§ 71. Уравнения Максвелла и волны в металлах 466
§ 72. Геометрические законы отражения и преломления света на границе металла 469
§ 73. Формулы Френеля. Измерение оптических констант металлов . 472
§ 74. Аномальный скин-эффект и эффективная диэлектрическая проницаемость 478

ГЛАВА VII
КРИСТАЛЛООПТИКА

§ 75. Плоские волны в кристаллах 481
§ 76. Оптически одноосные кристаллы 483
§ 77. Поляризационные устройства 491
§ 78. Анализ поляризованного света 499
§ 79. Интерференция поляризованных лучей 508
§ 80. Нормальные скорости и поляризация волн в двуосных кристаллах 520
§ 81. Лучи, волновые нормали и связь между ними 528
§ 82. Коническая рефракция 537
§ 83. Замечания об отражении и преломлении света на границе кристаллов 543

ГЛАВА VIII
МОЛЕКУЛЯРНАЯ ОПТИКА

§ 84. Классическая теория дисперсии света 548
§ 85. Понятие о дисперсионной формуле квантовой механики 559
§ 86. Методы экспериментального исследования аномальной дисперсии 564
§ 87. Дисперсия плазмы 569
§ 88. Средняя плотность электромагнитной энергии в диспергирующих средах 573
§ 89. Поглощение света и уширение спектральных линий 576
§ 90. Двойное преломление света в электрическом и магнитном полях 583
§ 91. Линейный электрооптический эффект Поккельса 594

§ 92. Эффект Зеемана 596
§ 93. Понятие об эффекте Штарка 602
§ 94. Вращение плоскости поляризации 604
§ 95. Магнитное вращение плоскости поляризации 611
§ 96. Временная и пространственная дисперсия. Теория естественной оптической активности 617
§ 97. О тепловых флуктуациях 626
§ 98. Рассеяние света 632
§ 99. Явление Мандельштама–Бриллюэна 643
§ 100. Комбинационное рассеяние света 649

ГЛАВА IX

ТЕОРИЯ ОТНОСИТЕЛЬНОСТИ

§ 101. Принцип относительности в ньютоновской и релятивистской механике 655
§ 102. Опыт Майкельсона 659
§ 103. Независимость скорости света от движения источника 665
§ 104. Понятие одновременности 667
§ 105. Преобразование координат и времени в теории относительности 671
§ 106. Лоренцово сокращение длины и замедление времени 680
§ 107. Эффект Допплера и аберрация света 688
§ 108. Эффект Допплера в акустике и теории эфира 695
§ 109. Замедление хода часов в гравитационном поле 699
§ 110. Сложение скоростей в теории относительности 702
§ 111. Релятивистская механика 706

ГЛАВА X

ТЕПЛОВОЕ ИЗЛУЧЕНИЕ

§ 112. Равновесное излучение в полости 713
§ 113. Закон Кирхгофа 717
§ 114. Формула Кирхгофа–Клаузиуса 721
§ 115. Закон Стефана–Больцмана 723
§ 116. Теорема и закон смещения Вина 726
§ 117. Формула Рэлея–Джинса 731
§ 118. Формула Планка 737
§ 119. Спонтанное и индуцированное излучение 743

ГЛАВА XI

ЛАЗЕРЫ И НЕЛИНЕЙНАЯ ОПТИКА

§ 120. Принципы работы лазера 748
§ 121. Рубиновый лазер 754
§ 122. Гелий-неоновый лазер 762
§ 123. Нелинейная поляризация среды 765
§ 124. Первое приближение. Оптическое детектирование. Генерация вторых гармоник, суммарной и разностной частот 768
§ 125. Второе приближение. Самофокусировка 773
§ 126. Параметрическая генерация света 776

Именной указатель 780
Предметный указатель 783

ПРЕДИСЛОВИЕ

Этот — четвертый — том общего курса физики посвящен физической оптике и является естественным продолжением предыдущего тома, в котором излагается учение об электрических и магнитных явлениях. Физическая оптика рассматривается в нем преимущественно с волновой (конечно, электромагнитной) точки зрения. Вопросы квантовой оптики затрагиваются лишь частично. Дается представление о фотонах и процессе излучения как о квантовом переходе атомных систем из одного энергетического состояния в другое. Это необходимо для введения понятия индуцированного излучения и объяснения принципов работы лазеров. Однако систематическое изложение основ квантовой оптики, в той мере, в какой это возможно сделать в рамках общей физики, а также относящихся сюда квантовых явлений (фотоэффект, эффект Комптона, спектральные закономерности, люминесценция, эффект Зеемана, эффект Штарка и пр.), предполагается дать в пятом томе, где будет излагаться атомная физика в широком смысле этого слова.

Вопросы геометрической оптики собраны в первых двух главах курса, чтобы в дальнейшем можно было ссылаться на них при изложении интерференции, дифракции и других разделов физической оптики. Геометрическая оптика излагается не как математическая, а как физическая дисциплина — как приближенный предельный случай волновой оптики. Тем самым четко определяются границы ее применимости. С целью простоты в основу обоснования геометрической оптики положено скалярное волновое уравнение. Хотя в общем случае неоднородной среды оно и неверно, но даже в этом случае при рассмотрении предельного перехода к геометрической оптике оно приводит к правильным результатам. Конечно, на основе скалярного уравнения ничего нельзя сказать относительно вращения плоскости поляризации луча в неоднородной среде. Для этого надо было бы положить в основу векторные уравнения Максвелла. Но это, ничего не меняя в идейном отношении, потребовало бы довольно громоздких вычислений. Существенно, что скалярное волновое уравнение правильно передает основные закономерности распространения волн не только в однородных, но и в неоднородных средах. Геометрическая же оптика получается из него в предельном случае коротких волн, длины которых пренебрежи-

мо малы по сравнению с характерными размерами, определяющими распространение света в среде.

Впрочем, обоснование геометрической оптики при всей его важности может быть опущено при первом чтении книги.

После изложения основ геометрической оптики во второй главе излагается геометрическая теория оптических изображений – главным образом с принципиальной стороны. Большая часть материала этой главы при первом чтении может быть также опущена. Достаточно ограничиться минимумом: получением оптических изображений в параксиальном приближении и основными понятиями фотометрии. К остальным вопросам можно обращаться по мере надобности в процессе изучения физической оптики.

Основное содержание курса составляет физическая оптика, изложение которой начинается с главы III. В нее входят также специальная теория относительности, краткое изложение принципов работы оптических квантовых генераторов (лазеров) и элементов нелинейной оптики. Как и в предыдущих томах курса, главное внимание здесь обращено на выяснение физического смысла и содержания оптических явлений, на связь между ними и с общими принципами физики. Автор стремился к идейной простоте изложения, но старался избегать вульгаризации.

Вопросы истории, экспериментальное обоснование законов оптики, применения ее в технике и других науках затронуты лишь постольку, поскольку это необходимо для уяснения основных явлений и принципов оптики. Из-за недостатка места даже описанию демонстраций в этом томе уделено меньше внимания, чем в предыдущих томах курса. Впрочем, демонстрации, осуществленные в Московском физико-техническом институте моими лекционными ассистентами Е.Н. Морозовым, М.И. Маклаковым, В.П. Молчановым, В.А. Кузнецовой, которым я глубоко благодарен, сыграли немалую роль при написании и этого тома курса физики.

Рукопись этого тома была тщательно просмотрена рецензентами — профессором С.С. Герштейном, профессором И.С. Горбанем и сотрудниками руководимой им кафедры экспериментальной физики Киевского государственного университета им. Т.Г. Шевченко, а также доктором физико-математических наук Г.П. Пекой. Их справедливые критические замечания учтены при окончательном редактировании книги. Всем этим лицам автор приносит глубокую благодарность.

Д.В. Сивухин

ГЛАВА I

ВВЕДЕНИЕ

§ 1. Предмет оптики

Оптика, точнее — *физическая оптика*, есть раздел физики, изучающий свойства и физическую природу света, а также его взаимодействие с веществом. Под светом понимают не только *видимый свет*, но и примыкающие к нему широкие области спектра электромагнитного излучения — *инфракрасную* и *ультрафиолетовую*. Различные участки спектра электромагнитного излучения отличаются друг от друга *длиной волны* λ и *частотой* ν — величинами, характеризующими не только *волновые*, но и *квантовые свойства* электромагнитного излучения. Электромагнитный спектр принято делить на *радиоволны*, *инфракрасное*, *видимое*, *ультрафиолетовое*, *рентгеновское* и *гамма-излучения*. Эти участки спектра различаются не по своей физической природе, а *по способу генерации и приема излучения*. Поэтому между ними нет резких переходов, сами участки перекрываются, а границы между ними условны.

Радиоволнами называются электромагнитные излучения, длины волн которых превосходят примерно 0,1 мм. Их принято делить на: 1) *сверхдлинные волны* с длиной волны $\lambda > 10$ км (частота $\nu < 30$ кГц); 2) *длинные волны* ($\lambda = 10 - 1$ км, $\nu = 30 - 300$ кГц); 3) *средние волны* ($\lambda = 1$ км $- 100$ м, $\nu = 300$ кГц $- 3$ МГц); 4) *короткие волны* ($\lambda = 100 - 10$ м, $\nu = 3 - 30$ МГц); 5) *ультракороткие волны* ($\lambda < 10$ м, $\nu > 30$ МГц). Последние в свою очередь принято подразделять на *метровые*, *дециметровые*, *миллиметровые* и *субмиллиметровые* или *микрометровые*. Волны с длиной $\lambda < 1$ м ($\nu > 300$ МГц) принято также называть *микроволнами* или *волнами сверхвысоких частот* (СВЧ). Из-за больших значений λ распространение радиоволн можно рассматривать *феноменологически* без учета атомистического строения среды. Исключение составляют только самые

короткие радиоволны, примыкающие к инфракрасному участку спектра. Практически не сказываются и квантовые свойства радиоизлучения.

Видимое, инфракрасное и ультрафиолетовое излучения составляют так называемую *оптическую область спектра* в широком смысле этого слова. Выделение такой области обусловлено не только близостью соответствующих участков спектра, но и сходством методов и приборов, применяющихся для ее исследования и разработанных исторически главным образом при изучении видимого света (линзы и зеркала для фокусировки излучения, призмы, дифракционные решетки, интерференционные приборы для исследования спектрального состава излучения и пр.). *Оптический спектр* занимает диапазон от условной границы инфракрасного излучения ($\lambda = 2$ мм, $\nu = 1{,}5 \cdot 10^{11}$ Гц) до условной коротковолновой границы ультрафиолета ($\lambda = 10^{-6}$см $= 10$ нм, $\nu = 3 \cdot 10^{16}$ Гц), что составляет примерно 18 октав [1]). Видимое излучение занимает приблизительно одну октаву ($\lambda = 400 - 760$ нм), ультрафиолет — 5 октав ($\lambda = 10 - 400$ нм), инфракрасное излучение — 11 октав ($\lambda = 760$ нм – 2 мм). В оптической области спектра частоты ν уже перестают быть малыми по сравнению с собственными частотами атомов и молекул, а длины волн большими по сравнению с молекулярными размерами и межмолекулярными расстояниями. Благодаря этому в этой области становятся существенными явления, обусловленные атомистическим строением вещества. По той же причине, наряду с волновыми, проявляются и квантовые свойства света. Энергия светового кванта определяется выражением

$$\mathcal{E} = h\nu, \tag{1.1}$$

где $h = 6{,}63 \cdot 10^{-27}$ эрг·с — *постоянная Планка*. Полезно заметить, что для длины волны $\lambda = 1000$ нм энергия соответствующего кванта составляет $\mathcal{E} = 1{,}23$ эВ, или приблизительно один электрон-вольт. На концах видимого спектра ($\lambda_{\text{кр}} = 760$ нм, $\lambda_{\text{фл}} = 400$ нм) для энергии кванта формула (1.1) дает $\mathcal{E}_{\text{кр}} \approx \approx 1{,}6$ эВ, $\mathcal{E}_{\text{фл}} \approx 3$ эВ.

В области рентгеновского и гамма-излучения на первый план выступают *квантовые свойства излучения*. Рентгеновское излучение возникает при торможении быстрых заряженных частиц (электронов, протонов и пр.), а также в результате процессов, происходящих внутри электронных оболочек атомов. Гамма-излучение появляется в результате процессов, происходящих

[1]) Октавой называется интервал частот между произвольной частотой ω и ее гармоникой 2ω.

внутри атомных ядер, а также в результате превращения элементарных частиц. Оно появляется и при торможении быстрых заряженных частиц. Характерные энергии гамма-квантов — порядка одного или нескольких МэВ. Границы областей рентгеновского и гамма-излучения, в особенности коротковолновые, могут быть определены лишь весьма условно. Для общей ориентировки можно принять, что энергия рентгеновских квантов лежит в пределах 20 эВ – 1 МэВ ($\lambda = 50 - 10^{-3}$ нм), а энергия гамма-квантов — больше 0,1 МэВ ($\lambda < 10^{-2}$ нм).

Следует заметить, что волновые и квантовые закономерности являются общими для всего спектра электромагнитного излучения. Только, в зависимости от длины волны, на первый план выступают разные явления, разные методы исследования и разные практические применения. Поэтому на оптику нельзя смотреть как на замкнутую дисциплину, изучающую только оптическую область спектра, отделенную от других областей резкими границами. Закономерности и результаты, найденные в этих и других областях, могут оказаться применимыми в оптической области спектра и обратно.

Практическое значение оптики и ее влияние на другие отрасли знания исключительно велики. Изобретение телескопа и спектроскопа открыло перед человеком удивительнейший и богатейший мир явлений, происходящих в необъятной Вселенной. Изобретение микроскопа произвело революцию в биологии. Фотография помогла и продолжает помогать чуть ли не всем отраслям науки. Одним из важнейших элементов научной аппаратуры является линза. Без нее не было бы микроскопа, телескопа, спектроскопа, фотоаппарата, кино, телевидения и т.п. Не было бы очков, и многие люди, которым перевалило за пятьдесят лет, были бы лишены возможности читать и выполнять многие работы, связанные со зрением.

Область явлений, изучаемая физической оптикой, весьма обширна. Оптические явления теснейшим образом связаны с явлениями, изучаемыми в других разделах физики, а оптические методы исследования относятся к наиболее тонким и точным. Поэтому неудивительно, что оптике на протяжении длительного времени принадлежала ведущая роль в очень многих фундаментальных исследованиях и развитии основных физических воззрений. Достаточно сказать, что обе основные физические теории текущего столетия — теория относительности и теория квантов — зародились и в значительной степени развились на почве оптических исследований. Изобретение лазеров открыло новые широчайшие возможности не только в оптике, но и ее приложениях в различных отраслях науки и техники.

§ 2. Геометрическая оптика

1. Простейшие оптические явления, например возникновение теней и получение изображений в оптических приборах, могут быть поняты в рамках так называемой *геометрической оптики.* В основу формального построения последней можно положить четыре закона, установленных опытным путем: 1) *закон прямолинейного распространения света*; 2) *закон независимости световых пучков*; 3) *закон отражения* и 4) *закон преломления света.* Для понимания более сложных явлений нужна уже *физическая оптика*, рассматривающая эти явления в связи с *физической природой света.* Физическая оптика позволяет, в частности, не только вывести все законы геометрической оптики, но и установить *границы их применимости.* Без знания этих границ формальное применение законов геометрической оптики может в конкретных случаях привести к результатам, противоречащим наблюдаемым явлениям. Поэтому нельзя ограничиться чисто формальным построением геометрической оптики, а необходимо смотреть на нее как на *раздел физической оптики.* Обоснование геометрической оптики будет дано постепенно на протяжении нашего курса. Сейчас же ограничимся формулировкой перечисленных четырех опытных законов.

2. Согласно закону прямолинейного распространения, *свет в прозрачной однородной среде распространяется по прямым линиям.* Опытным доказательством этого закона могут служить резкие тени, отбрасываемые непрозрачными телами, освещаемыми точечными источниками света, т.е. источниками, размеры которых весьма малы по сравнению с размерами освещаемого тела и расстоянием до него. Непрозрачный предмет AB (рис. 1),

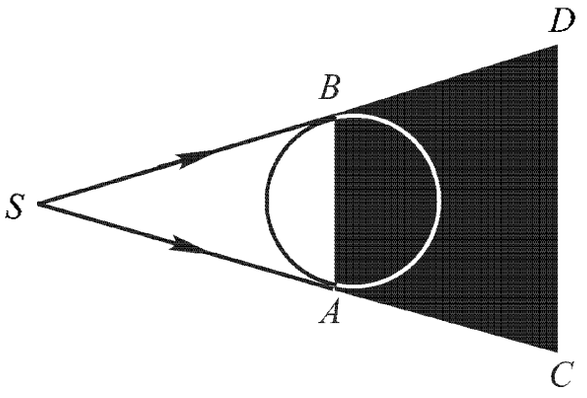


Рис. 1

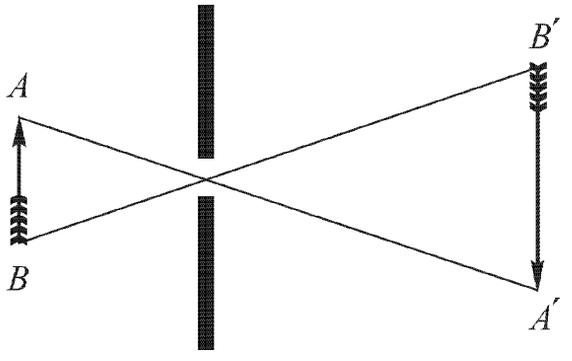


Рис. 2

поставленный на пути светового пучка от точечного источника S, не пропускает свет в пространство за этим предметом, ограниченное боковой поверхностью конуса SCD, касающейся краев предмета AB. Однако на распределение света вне этого

пространства присутствие предмета AB не оказывает никакого влияния. Это и значит, что распространение света происходит вдоль прямых линий [1]). Другое доказательство дает общеизвестный способ получения изображений светящихся предметов в камере с малым отверстием (камере-обскуре, рис. 2).

Наблюдаются отступления от закона прямолинейного распространения света. Рассмотрим, например, тень от резкого края непрозрачного предмета. Если источник света точечный, то, согласно этому закону, следовало бы ожидать, что на экране получится совершенно резкий переход от света к тени. На самом деле возникает переходная область, в которой освещенность меняется непрерывно и не монотонно: в ней наблюдаются «дифракционные полосы». Аналогичное положение имеет место в камере с малым отверстием. Если отверстие недостаточно мало, то изображение предмета размыто, так как светящаяся точка изображается в виде светлого кружка. Можно было бы ожидать, что при уменьшении отверстия размеры кружка будут уменьшаться, а отчетливость изображения увеличиваться, хотя само изображение и получится менее ярким. На самом деле это оправдывается лишь до известного предела, а при дальнейшем уменьшении отверстия резкость изображения снова начинает ухудшаться. Когда диаметр отверстия $\sim 10^{-3}$ мм, а источник света точечный, получается практически равномерная освещенность экрана.

Приведенные примеры показывают, что свет, подобно звуку, огибает препятствия, стоящие на пути его распространения. Это явление называется *дифракцией*.

Если среда мутная, например туман, то из-за дифракции прямолинейное распространение света сопровождается его *рассеянием в стороны*.

3. Закон независимости световых пучков состоит в том, что *распространение всякого светового пучка в среде совершенно не зависит от того, есть в ней другие пучки света или нет.* Световой пучок, прошедший через какую-либо область пространства, выходит из нее одним и тем же, независимо от того, заполнена она другим светом или не заполнена. Так, изображение на сетчатке глаза не изменится, если свет, образующий это изображение, будет на своем пути проходить через боковые пучки света, не попадающие в глаз. Границы применимости закона

[1]) Протяженный источник ведет себя как совокупность точечных источников, каждому из которых соответствует свое распределение светового поля на экране. В результате наложения таких картин наряду с тенями возникают полутени, т.е. области более или менее плавного перехода от освещенных частей экрана к менее освещенным и совсем темным.

независимости световых пучков лучше рассматривать в связи с теорией света (см. § 3, 5 и гл. XI).

Закон независимости световых пучков необходимо дополнить утверждением, определяющим совместное действие световых пучков при их наложении друг на друга. Оно состоит в том, что *освещенность экрана, создаваемая несколькими световыми пучками, равна сумме освещенностей, создаваемых каждым пучком в отдельности.* Нарушения справедливости этого утверждения имеют место в явлениях *интерференции света.*

4. На основе законов прямолинейного распространения и независимости световых пучков сложилось представление о *световых лучах. В математическом смысле луч есть линия, вдоль которой распространяется свет.* Это — математическая абстракция. *О существовании луча в таком смысле можно говорить лишь постольку, поскольку он входит в состав светового пучка,* содержащего бесконечное множество лучей. Реальное существование имеют не математические лучи и бесконечно тонкие пучки света, а *пучки конечного поперечного сечения,* вырезаемые, например, диафрагмами. Поэтому *под лучом в физическом смысле этого слова мы будем понимать конечный, но достаточно узкий световой пучок, который еще может существовать изолированно от других пучков.* Луч, выделенный какой-либо диафрагмой, не может быть бесконечно длинным, так как из-за дифракции распространение света сопровождается увеличением поперечных размеров луча, т.е. его боковым уширением и размытием. Чем длиннее луч, тем больше это дифракционное уширение. О луче можно говорить только тогда, когда уширение мало по сравнению с поперечными размерами самого луча. Для этого длина луча не должна превышать определенного предела, величина которого тем больше, чем шире луч (см. § 6, формулу (6.17)).

5. Когда луч достигает плоской границы раздела двух прозрачных сред, он частично проходит во вторую среду (преломляется), частично возвращается обратно (отражается). Закон отражения света был известен еще грекам. Он утверждает, что *падающий и отраженный лучи лежат в одной плоскости с нормалью к границе раздела в точке падения* (эта плоскость называется *плоскостью падения*), *причем угол падения* φ *равен углу отражения* φ' (рис. 3).

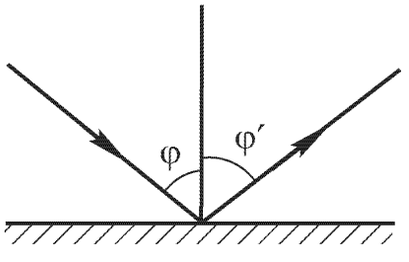


Рис. 3

6. Закон преломления был установлен экспериментально в 1621 г. голландским ученым Снеллиусом (1580–1626) и опубликован только после его смерти. Позднее Декарт (1596–1650) в

1637 г. опубликовал тот же закон, не ссылаясь на Снеллиуса. Знал ли Декарт работы Снеллиуса — этот вопрос остался открытым, хотя он и был предметом многочисленных дискуссий. Декарт получил закон преломления Снеллиуса, пользуясь аналогией между преломлением света и прохождением упругого шара (мяча) через границу раздела воздуха с водой. Его рассуждения были неубедительны и крайне туманны, но окончательный результат, к которому он пришел, оказался верным.

Согласно закону преломления Снеллиуса, *преломленный луч лежит в плоскости падения, причем отношение синуса угла падения* φ (рис. 4) *к синусу угла преломления* ψ *для рассматриваемых сред зависит только от длины световой волны, но не зависит от угла падения, т.е.*

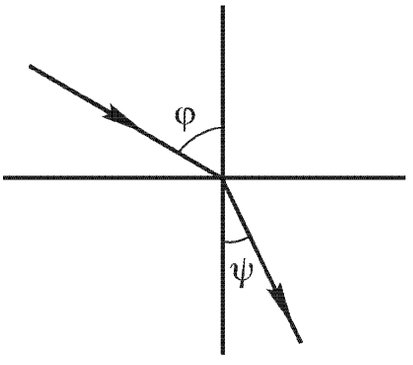


Рис. 4

$$\frac{\sin\varphi}{\sin\psi} = n_{21}. \qquad (2.1)$$

Постоянная величина n_{21} называется *относительным показателем* или *коэффициентом преломления второй среды относительно первой. Показатель преломления среды относительно вакуума называют абсолютным показателем (коэффициентом) преломления этой среды.* Его будем обозначать через n, снабжая эту букву, если требуется, соответствующими индексами. Например, n_1 — показатель преломления первой, а n_2 — второй сред. Ради краткости величину n обычно называют просто показателем (коэффициентом) преломления среды, т.е. опускают прилагательное «абсолютный».

Относительный показатель преломления n_{21} выражается через абсолютные показатели n_1 и n_2 соотношением

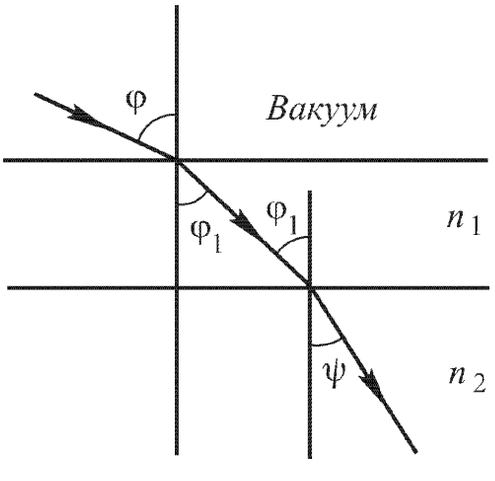


Рис. 5

$$n_{21} = n_2/n_1. \qquad (2.2)$$

Это соотношение можно получить путем предельного перехода. Пусть световой луч падает из вакуума на плоскопараллельную пластинку с показателем преломления n_1, а затем попадает в среду с показателем преломления n_2 (рис. 5). Для преломления на границах пластинки можно написать

$$\frac{\sin\varphi}{\sin\varphi_1} = n_1, \quad \frac{\sin\varphi_1}{\sin\psi} = n_{21}.$$

Перемножая написанные равенства, получим

$$\frac{\sin\varphi}{\sin\psi} = n_{21}n_1. \tag{2.3}$$

Это соотношение справедливо, какова бы ни была толщина пластинки. Оно остается верным и в предельном случае, когда толщина пластинки стремится к нулю. Но тогда свет будет преломляться так, как если бы никакой пластинки вообще не было. Поэтому должно быть $\sin\varphi/\sin\psi = n_2$. Сравнение этого результата с предыдущим и приводит к соотношению (2.2).

Слабая сторона приведенного рассуждения состоит в следующем. Показатель преломления есть *макроскопическая характеристика среды*. Когда толщина пластинки, разделяющей среды 1 и 2, становится порядка атомных размеров, ее уже нельзя рассматривать как непрерывную среду, так что понятие показателя преломления теряет смысл. Однако окончательный результат (2.2) остается верным. Он подтверждается опытом и в дальнейшем при рассмотрении теории отражения и преломления света будет выведен с различных точек зрения (см. § 3, 64).

С учетом соотношения (2.2) закон преломления можно записать в симметричной форме:

$$n_1 \sin\varphi = n_2 \sin\psi. \tag{2.4}$$

Из формулы (2.2) следует также:

$$n_{21} = 1/n_{12}. \tag{2.5}$$

7. Если $n_{21} < 1$, то может оказаться, что величина $\sin\psi$, формально вычисленная по формуле (2.1), начнет превосходить единицу, т.е. $\varphi/n_{21} > 1$. Соответствующего угла преломления не существует. Поэтому преломленный луч не возникает, а свет отражается полностью. Это явление называется *полным отражением*. Угол падения, при котором оно возникает, определяется условием $\varphi \geqslant \varphi_0$, причем

$$\sin\varphi_0 = n_{21}. \tag{2.6}$$

Величина φ_0 называется *предельным углом полного отражения*. Полное отражение будет исследовано в § 66.

8. Если поверхность тела, на которую падают световые лучи, не плоская, а кривая, то ее можно мысленно разбить на малые площадки, считая каждую из них плоской. Тогда ход лучей можно найти по законам отражения и преломления, изложенным выше, а затем выполнить предельный переход к гладкой поверхности, устремляя к нулю размеры каждой площадки. Однако этот прием применим только тогда, когда кривизна поверхности не превышает некоторого предела, так как в противном случае начинают проявляться отступления от законов правильного отражения и преломления и наступает дифракция.

Шероховатые поверхности дают не правильное, а *рассеянное*, или *диффузное*, отражение и преломление света. Только благодаря этому поверхность тела становится видимой. Абсолютно зеркальная поверхность невидима, видны только отраженные от нее лучи, попадающие от источников света, расположенных вне зеркала, т.е. видны только сами эти источники света.

ЗАДАЧИ

1. Два плоских зеркала *1* и *2* наклонены друг к другу, образуя двугранный угол α (рис. 6). Падающий луч, лежащий в плоскости, перпендикулярной к ребру двугранного угла, отражается сначала от одного, а затем от другого зеркала. Показать, что в результате этих двух отражений луч отклоняется на угол δ, величина которого не зависит от направления падающего света. Вычислить угол δ.

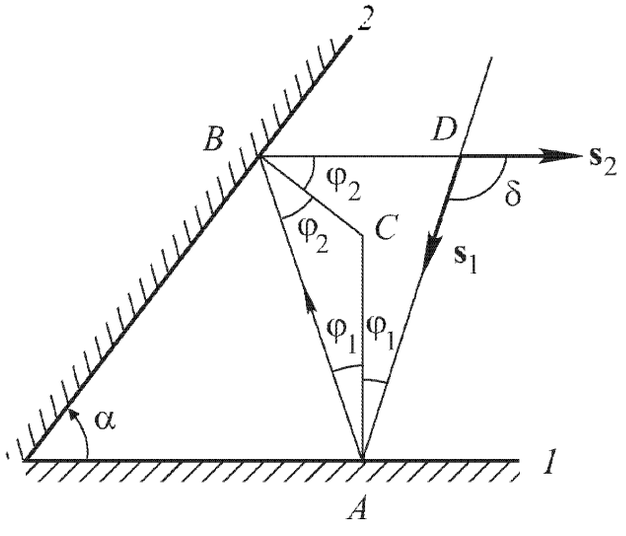


Рис. 6

Р е ш е н и е. Как видно из треугольника ABD, $\delta = 2(\varphi_1 + \varphi_2)$, а из треугольника ABC $\varphi_1 + \varphi_2 = \alpha$. Поэтому $\delta = 2\alpha$. Результат справедлив при любом α и любом направлении падающего света, если только угол α отсчитывать от зеркала *1*, вращая его по кратчайшему пути к зеркалу *2*, а угол δ — от направления падающего луча s_1 к направлению выходящего луча s_2, производя вращение в том же направлении.

2. Показать, что луч света, последовательно отражающийся от трех взаимно перпендикулярных зеркал, меняет свое направление на противоположное.

Р е ш е н и е. Пусть $\mathbf{s}_0$, $\mathbf{s}_1$, $\mathbf{s}_2$, $\mathbf{s}_3$ — единичные векторы падающего и отраженных лучей, а $\mathbf{N}_1$, $\mathbf{N}_2$, $\mathbf{N}_3$ — единичные нормали к отражающим плоскостям зеркал. Тогда

$$\mathbf{s}_1 - \mathbf{s}_0 = -2(\mathbf{N}_1\mathbf{s}_0)\mathbf{N}_1,$$
$$\mathbf{s}_2 - \mathbf{s}_1 = -2(\mathbf{N}_2\mathbf{s}_1)\mathbf{N}_2,$$
$$\mathbf{s}_3 - \mathbf{s}_2 = -2(\mathbf{N}_3\mathbf{s}_2)\mathbf{N}_3.$$

Из первого уравнения скалярным умножением на $\mathbf{N}_2$ и $\mathbf{N}_3$ получаем

$$(\mathbf{N}_2\mathbf{s}_1) = (\mathbf{N}_2\mathbf{s}_0), \quad (\mathbf{N}_3\mathbf{s}_1) = (\mathbf{N}_3\mathbf{s}_0).$$

Аналогично, скалярное умножение второго уравнения на $\mathbf{N}_3$ дает

$$(\mathbf{N}_3\mathbf{s}_2) = (\mathbf{N}_3\mathbf{s}_1) = (\mathbf{N}_3\mathbf{s}_0).$$

Почленно складывая все три уравнения и учитывая найденные соотношения, получим

$$\mathbf{s}_3 - \mathbf{s}_0 = -2(\mathbf{N}_1\mathbf{s}_0)\mathbf{N}_1 - 2(\mathbf{N}_2\mathbf{s}_0)\mathbf{N}_2 - 2(\mathbf{N}_3\mathbf{s}_0)\mathbf{N}_3 = -2\mathbf{s}_0,$$

откуда $\mathbf{s}_3 = -\mathbf{s}_0$.

Полученный результат лежит в основе устройства уголкового отражателя, применяющегося для изменения направления распространения света на противоположное. Уголковый отражатель можно получить, отсекая от стеклянного куба с посеребренными гранями трехгранный угол плоскостью, перпендикулярной к пространственной диагонали куба. Всякий луч, вступивший внутрь так полученной пирамиды через ее основание, испытав отражения от трех ее боковых граней, выйдет через то же основание, изменив свое направление на противоположное. Действительно, преломление на основании пирамиды, испытываемое падающим лучом, не играет роли, поскольку оно полностью компенсируется преломлением на том же основании при выходе луча из пирамиды.

3. Луч света проходит через ряд однородных сред, разграниченных плоскостями, параллельными между собой. Показать, что направление луча в последней среде (если луч проникает в нее) зависит только от угла падения и от показателей преломления первой и последней сред. В частности, если показатели преломления первой и последней сред одинаковы, то луч войдет в последнюю среду параллельно тому направлению, которое он имел в первой среде.

4. В преломляющей призме (рис. 7) $\angle C = \angle D$, $\angle A = \alpha$, $\angle B = 2\alpha$. Световой луч вступает в призму через грань BD, находясь в ее главном сечении,

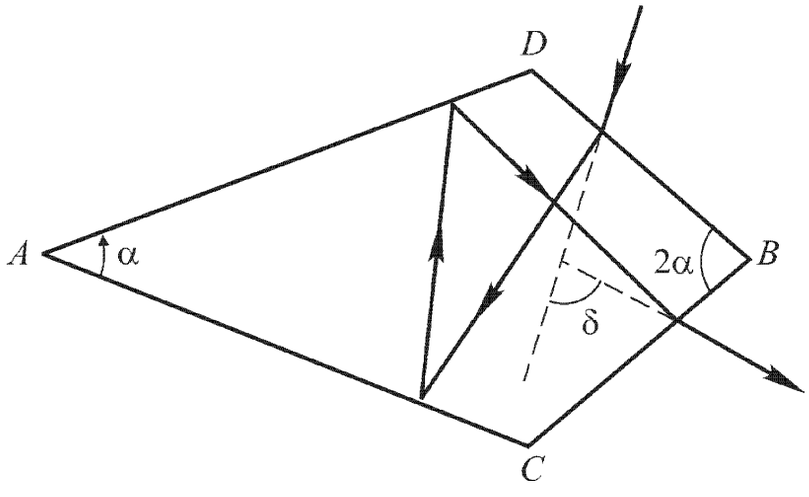


Рис. 7

т.е. в плоскости, перпендикулярной к преломляющим ребрам призмы, а затем последовательно отражается от граней AC и AD, выходя наружу через грань BC. Показать, что угол δ отклонения вышедшего луча от исходного направления не зависит от угла падения. Вычислить угол δ.

О т в е т. $\delta = 2\alpha$.

5. Исследовать преломление светового луча в главном сечении трехгранной призмы. Определить угол отклонения светового луча δ от исходного направления и его наименьшее значение $\delta_{\text{мин}}$.

Р е ш е н и е. Как видно из треугольника CDE (рис. 8),

$$\delta = (\varphi_1 - \psi_1) + (\varphi_2 - \psi_2).$$

Далее, из четырехугольника $ACFD$, у которого углы C и D — прямые, имеем: $A + F = \pi$, а из треугольника CDF $(\psi_1 + \psi_2) + F = \pi$, так что $\psi_1 + \psi_2 = A = \text{const}$, где A — преломляющий угол призмы. Следовательно,

$$\delta = \varphi_1 + \varphi_2 - A. \tag{2.7}$$

Для нахождения минимума угла отклонения δ удобно за независимую переменную принять угол преломления ψ_1, так как при этом будет достигнута симметрия и упрощение выкладок. Для первой производной получаем

$$\frac{d\delta}{d\psi_1} = \frac{d\varphi_1}{d\psi_1} + \frac{d\varphi_2}{d\psi_1} = \frac{d\varphi_1}{d\psi_1} - \frac{d\varphi_2}{d\psi_2},$$

или, на основании закона преломления,

$$\frac{d\delta}{d\psi_1} = n\left(\frac{\cos\psi_1}{\cos\varphi_1} - \frac{\cos\psi_2}{\cos\varphi_2}\right). \tag{2.8}$$

Аналогично, для второй производной

$$\frac{d^2\delta}{d\psi_1^2} = n\left(\frac{d}{d\psi_1}\frac{\cos\psi_1}{\cos\varphi_1} - \frac{d}{d\psi_1}\frac{\cos\psi_2}{\cos\varphi_2}\right) = n\left(\frac{d}{d\psi_1}\frac{\cos\psi_1}{\cos\varphi_1} + \frac{d}{d\psi_2}\frac{\cos\psi_2}{\cos\varphi_2}\right) =$$

$$= n\sum_{i=1}^{2}\frac{\sin\psi_i}{\cos^3\psi_i}(n^2\cos^2\psi_i - \cos^2\varphi_i).$$

Если $n > 1$, то $d^2\delta/d\psi_1^2 > 0$, а потому кривая $\delta = \delta(\psi_1)$ во всех точках обращена выпуклостью вниз. Отсюда следует, что угол δ достигает минимума при $\varphi_1 = \varphi_2$, т.е. при симметричном ходе луча через призму. Других минимумов (и вообще экстремумов) быть не может.

При симметричном ходе луча $\varphi = {}^1\!/_2(A + \delta_{\text{мин}})$, $\psi = {}^1\!/_2 A$, и следовательно,

$$n = \frac{\sin[(A + \delta_{\text{мин}})/2]}{\sin(A/2)}. \tag{2.9}$$

На этой формуле основан удобный метод измерения показателя преломления.

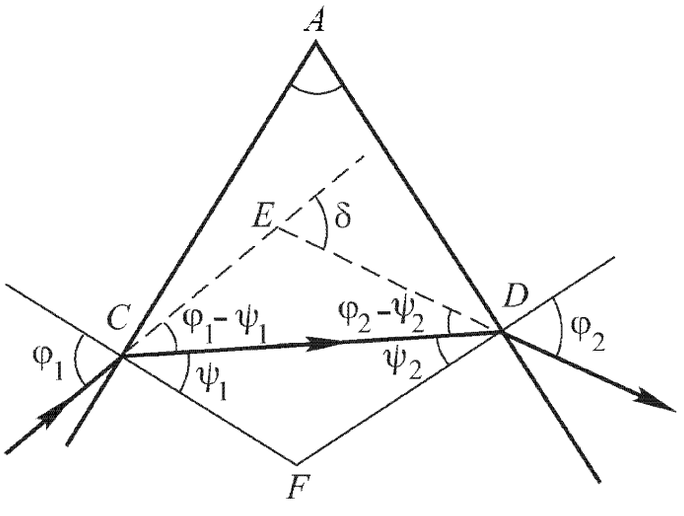


Рис. 8

6. Световой луч падает на боковую грань призмы под малым углом. Преломляющий угол призмы A мал (рис. 8). Вычислить угол отклонения луча δ, ограничиваясь членами первой степени по A. Вычислить также угол наименьшего отклонения $\delta_{\text{мин}}$ с точностью до членов порядка A^3 включительно.

Р е ш е н и е. $\delta = (\varphi_1 - \psi_1) + (\varphi_2 - \psi_2) \approx (n\psi_1 - \psi_1) + (n\psi_2 - \psi_2) = (n-1)(\psi_1 + \psi_2)$, или, на основании соотношения $\psi_1 + \psi_2 = A$,

$$\delta = (n-1)A. \tag{2.10}$$

В рассматриваемом приближении угол δ не зависит от угла падения. В высших приближениях это уже не так. В третьем приближении из формулы (2.9), разлагая $\delta_{\text{мин}}$ в ряд, находим

$$\delta_{\text{мин}} = (n-1)A\left[1 + \frac{n(n+1)}{24}A^2\right]. \tag{2.11}$$

7. Удобный метод измерения показателя преломления твердых тел состоит в следующем. Плоскопараллельная пластинка из исследуемого вещества рассматривается в микроскоп. Сначала микроскоп устанавливается

для наблюдения верхней поверхности пластинки. Затем смещают тубус микроскопа вниз, пока не будет отчетливо видна нижняя поверхность пластинки. (Для удобства наблюдения на поверхностях пластинки можно сделать метки.) Показать, что если смещение тубуса равно h, а толщина пластинки l, то показатель преломления пластинки можно найти по формуле

$$n = l/h.$$

Метод годится и для измерения показателей преломления жидкостей.

§ 3. Эволюция представлений о природе света

1. Пифагор (около 580–500 до н.э.) считал, что предметы становятся видимыми благодаря мельчайшим частицам, испускаемым ими и попадающим в глаз наблюдателя. Декарт полагал, что свет — это сжатие, распространяющееся в идеальной упругой среде (эфире), заполняющей мировое пространство и промежутки между частицами тел. Однако последовательно провести эту точку зрения Декарт не смог, при выводе законов отражения и преломления он пользовался представлением о свете как о потоке частиц. Гук (1635–1703) также считал, что свет представляет собой импульсы сжатия, распространяющиеся мгновенно или с очень большими скоростями. (Скорость света была определена только в 1676 г. Олафом Рёмером (1644–1710) из наблюдений затмений спутников Юпитера.) Несколькими годами ранее Гука чешский монах Марци (1595–1667) и итальянский монах Гримальди (1618–1663) также пришли к мысли, что свет представляет собою быстро распространяющиеся волны. Подобные отрывочные высказывания о природе света были усовершенствованы и развиты в более систематические теории Исааком Ньютоном (1643–1727), с одной стороны, и Христианом Гюйгенсом (1629–1695), с другой.

Основные оптические работы Ньютона докладывались в Лондонском Королевском обществе (1671–1675 гг.) и были напечатаны в трудах этого общества. Содержание этих сообщений, наряду с другими оптическими исследованиями Ньютона, было заимствовано из его «Лекций по оптике» [1]), которые он читал в Кембридже в 1669–1671 гг. Спустя много лет (в 1704 г.) вышла знаменитая книга Ньютона «Оптика» [2]). В оптике Ньютон сосредоточил основное внимание на экспериментальных исследованиях и стремился при изложении этого, как и других разделов

[1]) *Исаак Ньютон.* Лекции по оптике. Перевод, комментарии и редакция академика С.И. Вавилова. — М.: Издательство АН СССР, 1946.

[2]) *И. Ньютон.* Оптика или трактат об отражениях, преломлениях, изгибаниях и цветах света. Перевод с примечаниями С.И. Вавилова. — М.: Гостехиздат, 1954.

физики, отделить достоверные факты от сомнительных гипотез. Экспериментальные исследования Ньютона по оптике вплоть до XIX века не имели себе равных. Его «Оптика» служила основным источником, из которого черпали материал все учебники того времени («Лекции по оптике» были опубликованы после смерти Ньютона и по ряду причин остались почти неизвестными). Что касается природы света, то к этому вопросу Ньютон подходил значительно более осторожно, чем все остальные ученые, включая Гюйгенса. Он предпочитал с самого начала строить чисто феноменологическую теорию, освобожденную от произвольных гипотез. Все, что относится к физической природе света, он выделил в особый раздел, содержащий 31 «вопрос», которым заканчивается его «Оптика». Здесь после тщательного рассмотрения различных исключающих друг друга возможностей Ньютон пришел к заключению, что факты говорят больше в пользу корпускулярной теории, чем волновой. Поэтому он, правда не без колебаний и сомнений, отдал предпочтение корпускулярной теории.

Гюйгенс в своих высказываниях о природе света не был столь сдержанным и осторожным, как Ньютон. Он не сомневался в волновой природе света. Свои воззрения он изложил на заседании Парижской Академии наук (1678 г.), а затем более полно в сочинении «Трактат о свете» [1]), опубликованном в 1690 г. В этом сочинении впервые в истории оптики была сделана попытка изложить с единой — волновой — точки зрения распространение, отражение, преломление и двойное лучепреломление света.

2. Перейдем к более подробной характеристике корпускулярной и волновой теорий света.

Согласно корпускулярной теории, свет состоит из мельчайших частиц, или корпускул, испускаемых светящимися телами. С этой точки зрения прямолинейное распространение света сводится к закону инерции. Для истолкования закона независимости световых пучков надо было ввести предположение, что средние расстояния между корпускулами в световых пучках настолько велики, что корпускулы практически не взаимодействуют между собой; случаи сближения, в которых проявляется такое взаимодействие, крайне редки и при существующей точности эксперимента ускользают от наблюдения.

Отражение и преломление света корпускулярная теория объясняла силами притяжения и отталкивания, действующими на световые корпускулы в очень тонком приграничном слое вблизи границы раздела сред, на которую падает свет. Внутри приграничного слоя путь световой корпускулы искривляется, по выхо-

[1]) *Х. Гюйгенс.* Трактат о свете. — М.; Л.: ОНТИ, 1935.

де из него корпускула движется снова прямолинейно и равномерно, но уже в другом направлении. Если корпускула отражается, то она возвращается в первую среду с прежним значением скорости. Если же корпускула проходит во вторую среду, то величина ее скорости изменяется (рисунки 9 и 10). Ввиду тонкости

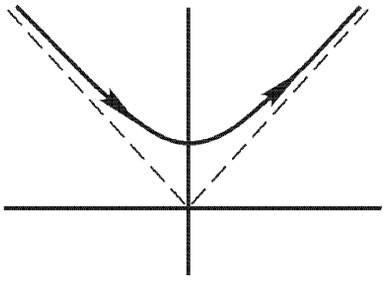


Рис. 9

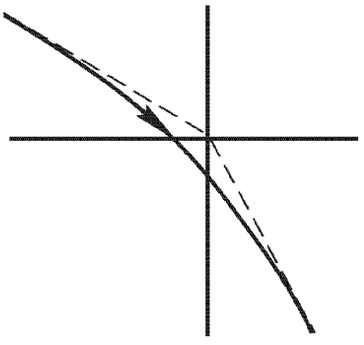


Рис. 10

приграничного слоя явление воспринимается так, как если бы на границе сред происходил резкий излом траектории корпускулы (как показано на рисунках 9 и 10 штриховыми линиями).

Для вывода количественных законов отражения и преломления света надо было ввести предположение, что силы, действующие на световую корпускулу в приграничном слое, *нормальны к границе раздела сред.* Такие силы меняют только нормальные скорости световой корпускулы, оставляя касательные без изменения. Рассмотрим, например, преломление света. Обозначим через v_1 скорость световой корпускулы в первой среде, а через v_2 — во второй. В силу равенства касательных составляющих этих скоростей $v_1 \sin\varphi = v_2 \sin\psi$ (рис. 11), или

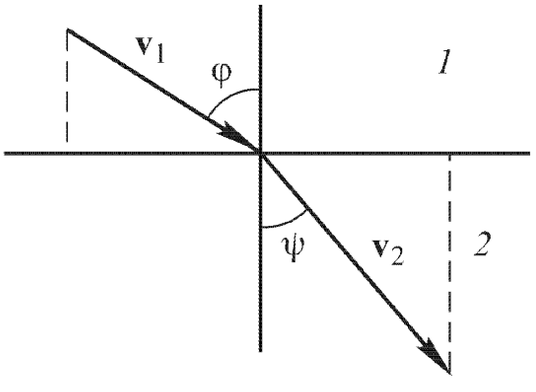


Рис. 11

$$\frac{\sin\varphi}{\sin\psi} = \frac{v_2}{v_1}. \qquad (3.1)$$

Эта формула была получена Ньютоном. *Из нее следует, что в сильнее преломляющих средах скорость света должна быть больше, чем в менее преломляющих.* Однако чтобы из формулы (3.1) вывести закон преломления Снеллиуса, необходимо добавочное предположение, что отношение скоростей света v_2 и v_1 для световых корпускул одного и того же типа постоянно, т.е. определяется только свойствами сред *1* и *2*, в которых корпускулы движутся, но не зависит от того, каким путем они туда попали. Если использовать экспериментальный факт, что в вакууме скорость света одна и та же, то для выполнения этого условия

необходимо потребовать, чтобы все силы, действующие на световые корпускулы в приграничных слоях, были *потенциальны*. (Заметим, что цвета света тогда можно объяснить, например, различными размерами корпускул.) Однако то обстоятельство, что скорость световой корпускулы зависит только от среды, в которую она испускается, но не зависит от способа испускания, представляется малопонятным с точки зрения классической механики. В этом отношении корпускулярная теория уступает волновой теории, поскольку в последней скорость света, естественно, есть характеристика *только самой среды*, в которой он распространяется.

При справедливости введенных предположений показатель преломления среды *2* относительно среды *1* определяется выражением

$$n_{21} = v_2/v_1. \tag{3.2}$$

В частности, для абсолютного показателя преломления получаем

$$n = v/c, \tag{3.3}$$

где c — скорость света в вакууме. Следовательно, $n_1 = v_1/c$, $n_2 = v_2/c$, а потому

$$n_{21} = n_2/n_1, \tag{3.4}$$

что совпадает с формулой (2.2). Однако это обстоятельство нельзя рассматривать как аргумент в пользу корпускулярной теории, так как такое же соотношение следует и из волновой теории.

3. Согласно *волновой теории*, свет представляет собой *волны*, распространяющиеся в гипотетической всепроникающей среде, — *мировом* или *световом эфире* — заполняющей все мировое пространство и промежутки между мельчайшими частицами тел. Если колебания частиц эфира малы, то уравнения, описывающие распространение волн, будут *линейны и однородны*. В этом случае справедлив *принцип суперпозиции волн*, являющийся в волновой теории математическим выражением закона независимости световых пучков.

Значительно сложнее обстояло дело с объяснением прямолинейного распространения света. Гюйгенс применил к этой проблеме принцип, названный позднее его именем. Если камень бросить в воду, то от места падения, как из центра, побегут по поверхности воды круговые волны, или возмущения. Этот процесс продолжается и после того, как камень упадет на дно, т.е. исчезнет причина, породившая первоначальное возмущение. Отсюда следует, что непосредственной причиной распространяющегося волнового процесса является не камень, а то первоначальное возмущение движения воды, которое он вызвал. Подобные рассуждения навели Гюйгенса на мысль, что *каждая точка волнового*

возмущения является источником сферических волн, распространяющихся от нее во все стороны. Они получили название *вторичных* или *элементарных волн Гюйгенса*. Результирующее волновое возмущение можно рассматривать как наложение вторичных волн. В этом и состоит знаменитый *принцип Гюйгенса*. Он справедлив не только для световых, но и для *любых* волновых процессов и в приведенной формулировке полностью сохранил свое значение до настоящего времени. Его значение далеко выходит за рамки сравнительно узкой задачи о прямолинейном распространении света. Действительно, принцип Гюйгенса указывает на принципиальную возможность определения волнового возмущения во всем пространстве, если известны все вторичные волны, порожденные в более ранний момент времени.

Но Гюйгенс не рассматривал задачу в столь общей постановке. Он дополнил формулировку своего принципа, приведенную выше, замечанием, что отдельные вторичные волны слишком слабы и что заметное световое действие они производят только на их *огибающей*. После этого дополнения принцип Гюйгенса потерял свою общность и превратился лишь в приближенный *геометрический рецепт* для построения волновых фронтов, т.е. поверхностей, до которых дошло световое возмущение. Он стал непригоден для количественного расчета светового поля, в частности распределения интенсивности последнего в пространстве.

4. Обратимся теперь к объяснению прямолинейного распространения света, данному Гюйгенсом. Пусть точечный источник света S (рис. 12), расположенный перед непрозрачным экраном с отверстием AB, излучает сферическую волну. Если в момент времени t волна дойдет до краев отверстия AB, то оно вырежет из волнового фронта сферический участок ACB. Построим из каждой точки этого участка, как из центра, вторичные волны радиусом $c\,dt$. Огибающая всех этих вторичных волн оборвется на краях сферического участка A_1B_1, за пределы которого проникнут только отдельные вторичные волны. Действие таких волн, согласно предположению Гюйгенса, пренебрежимо мало, а потому волновой фронт возмущения в момент $t+dt$ ограничится только сферическим участком A_1B_1. Взяв этот фронт за исходный, можно таким же построением найти волновой фронт A_2B_2 в более поздний момент времени, и т.д. Построение показывает, что волновое возмущение будет резко обрываться на поверхности телесного угла с вершиной S, вырезаемого

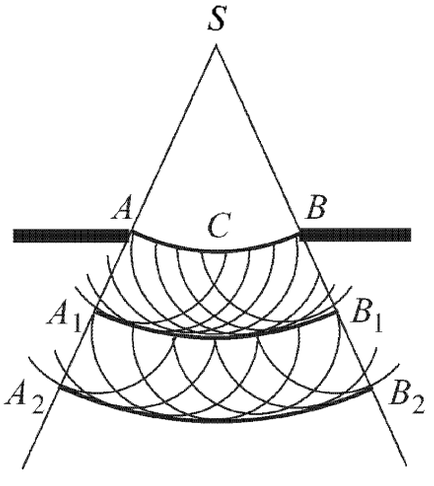


Рис. 12

диафрагмой AB. А это и означает прямолинейное распространение света.

Рассуждение Гюйгенса лишено доказательной силы. В нем доказывается то, что по существу содержится в исходных предположениях. А само построение Гюйгенса есть только *один из возможных рецептов построения волнового фронта*, согласующийся с представлением о распространении света вдоль лучей. Непонятно, почему объяснение Гюйгенса применимо к световым, но не применимо к звуковым волнам, которые, как известно, огибают препятствия, стоящие на их пути.

Естественно, что такое «объяснение» не могло удовлетворить Ньютона. Кроме того, Ньютон, как и Гюйгенс, не видел, как в рамках волновой теории можно объяснить поляризацию света, открытую Гюйгенсом в двойном лучепреломлении. Это действительно невозможно для продольных волн, какими, согласно Гюйгенсу, является свет. Только такие волны и были известны в физике того времени. Представления о поперечных волнах еще не существовало. Эти трудности казались Ньютону настолько существенными, что его симпатии оказались на стороне корпускулярной теории.

Проблема прямолинейного распространения света есть *частный случай проблемы дифракции* и может быть решена до конца только в рамках последней. Дифракция света была открыта Гримальди и независимо от него несколько позднее Гуком. Ньютон много занимался экспериментальными исследованиями дифракции света. Но Гюйгенс в «Трактате о свете» почему-то полностью обошел молчанием это явление. Кроме того, ему осталась неизвестной периодичность световых процессов (в отличие от Ньютона, который первый подметил ее). Гюйгенс писал, что свет, подобно звуку, распространяется *сферическими поверхностями*, и именно такие поверхности называл *волнами*. Он специально подчеркивал, что удары, возбуждающие световые возмущения в центрах волн, совершаются совершенно беспорядочно, а потому не следует думать, что сами волны следуют друг за другом на равных расстояниях. В этом отношении высказывания Гюйгенса примыкают к более ранним представлениям Декарта и Гука. Понятие длины волны нигде не встречается в теории Гюйгенса, а без этого невозможно установить, при каких условиях (приближенно) справедлив закон прямолинейного распространения света.

5. Перейдем к вопросу об отражении и преломлении света в волновой теории Гюйгенса. Пусть плоская волна падает на плоскую границу раздела двух сред. В некоторый момент в точке A волновой фронт AB достигнет границы раздела сред (рис. 13). В этот момент из A начнут распространяться вторичные волны Гюйгенса: одна в первую, другая во вторую среду. В точках E

и D аналогичные волны возникнут несколько позднее. Согласно принципу Гюйгенса, от наложения таких вторичных волн в первой среде образуется *отраженная*, а во второй — *преломленная* волны. Огибающая вторичных волн во второй среде FGD есть плоскость, определяющая волновой фронт преломленной волны. Аналогично строится волновой фронт и отраженной волны (на рис. 13 он не указан). Таково объяснение отражения и преломления света в волновой теории Гюйгенса. Оно не раскрывает детальный механизм возникновения отраженной и преломленной волн. Для этого надо было бы явно использовать физическую природу световых волн, о которой в эфирной теории Гюйгенса, в сущности, ничего не говорится.

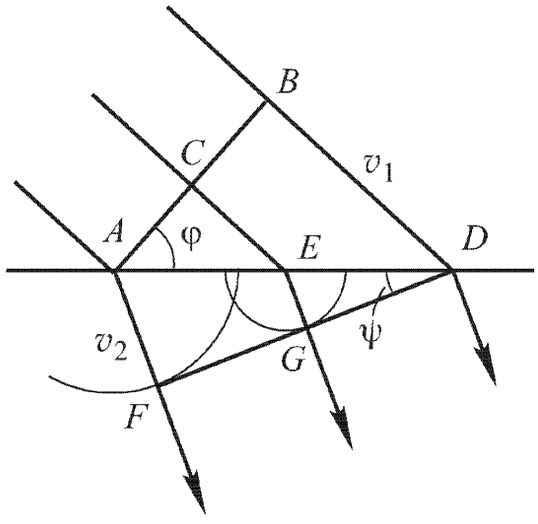


Рис. 13

Геометрические законы отражения и преломления, однако, *совершенно не зависят* от физической природы волн и от конкретного механизма отражения и преломления. Они одинаковы в любой волновой теории. Действительно, падающая волна возбуждает возмущение, бегущее вдоль границы раздела со скоростью $AD = BD/\sin\varphi = v_1/\sin\varphi$ (если воспользоваться надлежащими единицами), где v_1 — скорость света в первой среде. Но отраженная и преломленная волны порождаются падающей волной и поэтому бегут вместе с ней вдоль границы раздела с той же скоростью. Следовательно, можно написать также, что $AD = v_1/\sin\varphi' = v_2/\sin\psi$, где v_2 — скорость света во второй среде, φ' — угол наклона фронта отраженной волны к границе раздела сред (не показанный на рис. 13). В результате получается

$$\frac{\sin\varphi}{v_1} = \frac{\sin\varphi'}{v_1} = \frac{\sin\psi}{v_2}. \tag{3.5}$$

Этими соотношениями определяются *направления фронтов* отраженной и преломленной волн. А так как в плоской волне световые лучи перпендикулярны к волновым фронтам, то те же соотношения определяют также *направления отраженных и преломленных лучей*. Легко видеть, что φ есть угол падения, φ' — угол отражения, ψ — угол преломления. Из (3.5) следует, что $\varphi = \varphi'$ (закон отражения) и

$$\frac{\sin\varphi}{\sin\psi} = \frac{v_1}{v_2} \qquad \text{(закон преломления)}. \tag{3.6}$$

В противоположность корпускулярной теории (см. (3.1)), *волновая теория приводит к заключению, что скорость света в более преломляющей среде меньше, чем в менее преломляющей.* И это заключение справедливо независимо от того, какова физическая природа световых волн.

Для относительного показателя преломления волновая теория дает

$$n_{21} = v_1/v_2, \tag{3.7}$$

а для абсолютного

$$n = c/v, \tag{3.8}$$

т.е. выражения, обратные соответствующим выражениям в корпускулярной теории (3.2) и (3.3). Из (3.7) и (3.8) снова получается формула (3.4), которая, следовательно, справедлива как в корпускулярной, так и в волновой теориях.

6. В вопросе об отражении и преломлении света между корпускулярной и волновой теориями имеется и другое различие. Опыт показывает, что световой пучок, падающий на границу раздела сред, разделяется на два пучка: отраженный и преломленный (исключением является случай полного отражения). В волновой теории объяснение этого факта не встречает затруднений. Но как истолковать его с точки зрения корпускулярной? Корпускула всегда ведет себя как целое. Она не может разделиться на две части, из которых одна отражается, а другая проходит во вторую среду. В противном случае падающий, отраженный и прошедший свет был бы окрашен различно. Для преодоления этой трудности Ньютон ввел идею о *приступах легкого отражения и легкого прохождения*, в которых периодически может находиться световая корпускула. Если корпускула подходит к границе сред в приступе легкого отражения, то она отталкивается и отражается. Если же это случится в приступе легкого прохождения, то произойдет притяжение и прохождение корпускулы во вторую среду.

Идея «приступов», как заметил Я.И. Френкель (1894–1952), напоминает современные представления о *световых квантах — фотонах.* Согласно этим представлениям, отражение и прохождение фотонов через границу раздела сред управляется *статистическими законами*: существует определенная вероятность, что фотон отразится, и определенная вероятность, что он пройдет во вторую среду. Сам Ньютон, конечно, не раскрыл физический механизм приступов легкого отражения и легкого прохождения. Однако он ставил вопрос, не являются ли эти приступы результатом обратного воздействия каких-то быстрых волн, возбуждаемых в среде световыми корпускулами. Идея приступов навязывалась также *периодичностью световых процессов*, подмеченной Ньютоном при исследовании интерференционного

явления «ньютоновых колец». Эту периодичность Ньютон также пытался объяснить, дополнив корпускулярную теорию волновыми представлениями.

7. Физики XVIII и начала XIX веков, забыв о колебаниях и сомнениях Ньютона, приняли корпускулярную теорию света. Сторонники волновой теории насчитывались единицами. Правда, среди них были Эйлер (1707–1783), Ломоносов (1711–1765) и Франклин (1706–1790), выдвигавшие возражения против корпускулярной и приводившие аргументы в пользу волновой теории света. Эйлер, например, утверждал, что если бы корпускулярная теория была справедлива, то из-за излучения света масса Солнца заметно уменьшалась бы, а это сказалось бы на движении планет. Франклин указывал, что по корпускулярной теории свет должен был бы оказывать давление на освещаемые тела, тогда как все попытки обнаружить это давление оканчивались безрезультатно. Слабость этих возражений состояла в том, что они не содержали количественных оценок ожидаемых эффектов. Впрочем, согласно современным представлениям о взаимосвязи между массой и энергией, потеря массы из-за излучения, а также величина светового давления совершенно не зависят от того, является ли свет потоком частиц или волновым процессом. Эти и аналогичные возражения, как показало последующее развитие физики, не могли решить спор между корпускулярной и волновой теориями света. Нужны были новые экспериментальные факты.

8. Начиная с XIX века, положение стало складываться в пользу волновой теории благодаря работам Юнга (1773–1829) и в особенности Френеля (1788–1827), систематически исследовавших явления интерференции и дифракции света. На основе волновых представлений была создана стройная теория этих явлений, выводы и предсказания которой полностью согласовывались с экспериментом. Объяснение прямолинейного распространения света содержалось в этой теории как частный случай. Были открыты и исследованы новые оптические явления: поляризация света при отражении (Малюс, 1808) и преломлении (Малюс и Био, 1811), угол полной поляризации (Брюстер, 1815), интерференция поляризованных лучей (Френель и Араго, 1816), количественные законы и теория отражения и преломления света (Френель, 1821), двойное преломление сжатым стеклом (Брюстер, 1815), двуосные кристаллы (Брюстер, 1815), законы и теория распространения света в двуосных кристаллах (Френель, 1821), вращение плоскости поляризации в кварце (Араго, 1811) и жидкостях (Био, 1815; оба явления исследовались далее Био, Брюстером и др.). Юнг (1807) измерил на опыте длину световой волны. Оказалось, что волны красного света длиннее, чем синего и фиолетового. Тем самым в волновой теории было дано экс-

периментально обоснованное объяснение цветов света, которое связывало это явление с длиной световой волны. (Такое объяснение предлагалось еще Эйлером, но он не мог указать, длина каких волн больше — красных или синих.) Юнг (1817) высказал также мысль о *поперечности световых волн*. К такому же заключению независимо от него пришел Френель (1821) и обосновал это заключение путем исследования поляризации света и интерференции поляризованных лучей. Все эти факты и в особенности явления интерференции и дифракции света находили непринужденное объяснение в рамках волновой теории света. Корпускулярная теория не могла противопоставить ничего эквивалентного и к началу 30-х годов XIX века была оставлена.

Смертельный удар корпускулярной теории в ее ньютоновской форме был нанесен в 1850 г. К этому времени Физо (1819–1896) и Фуко (1819–1868) впервые измерили скорость света лабораторными методами. Как мы указывали (см. пп. 2 и 5), по корпускулярной теории скорость света в воде больше, а по волновой теории меньше, чем в вакууме. В 1850 г. Фуко и независимо от него Физо и Бреге сравнили обе скорости. Опыт оказался в согласии с волновой и в противоречии с корпускулярной теориями света. Физики XIX века восприняли это как решающий опыт, окончательно доказавший неправильность корпускулярной теории света.

9. Волновая теория, конечно, не могла считаться полной, пока не была установлена *природа световых колебаний*, или колебаний мирового эфира, как говорили физики девятнадцатого (и отчасти первой четверти двадцатого) века. Они не сомневались, что эфир подчиняется обычным законам механики Ньютона и к нему применимы такие понятия, как плотность, упругость, пространственное перемещение, скорость, ускорение и пр. Они пытались вывести строение и свойства эфира из наблюдаемых явлений и экспериментально установленных законов оптики. Поперечность световых волн заставила приписать мировому эфиру свойства твердой среды. Это породило ряд трудностей, в частности в вопросе об отражении и преломлении света (подробнее см. § 63). Нет необходимости останавливаться на этих трудностях и попытках их преодоления в теории эфира. Все это уже давно потеряло актуальность и сохранило лишь исторический интерес.

В 60-х годах XIX века Максвеллом были установлены общие законы электромагнитного поля, которые привели его к заключению, что *свет — это электромагнитные волны* (см. т. III, гл. IV). Подтверждением такой точки зрения в то время были открытие Фарадеем в 1846 г. вращения плоскости поляризации света в магнитном поле и совпадение скорости света в вакууме с электродинамической постоянной, установленное на

опыте в 1856 г. Вебером и Кольраушем (см. т. III, § 51 и 83). После известных опытов Герца (1887–1888 гг., см. т. III, § 142) электромагнитная природа света быстро получила признание. Уже в первом десятилетии XX века она из гипотезы превратилась в твердо установленный факт. Световые колебания были отождествлены с колебаниями электромагнитного поля. Оптика превратилась в раздел учения об электрических и магнитных явлениях.

Электромагнитная теория света устранила трудности, с которыми столкнулась теория упругого твердого эфира. Однако физики XIX века считали, что она дала не настоящее, а только *символическое* решение вопроса о природе света. Они смотрели на нее как на формальную схему, уравнения которой правильно передают количественные соотношения между различными величинами и явлениями, но символы, входящие в эти уравнения, еще не получили отчетливого физического истолкования. Считалось, что уравнения Максвелла должны составлять математический остов будущей более полной физической теории электромагнитных и световых явлений, но сами по себе они такой физической теории еще не составляют; последняя станет возможной лишь после того, как будут найдены механические свойства эфира.

Однако эти надежды не оправдались. Световой (или, лучше, электромагнитный) эфир упрямо отказывался обнаружить свои «механические свойства». В частности, потерпели неудачу настойчивые попытки обнаружить движение Земли относительно эфира, начатые Майкельсоном (1852–1931) в 1881 г. При изучении оптических и электродинамических явлений в движущихся средах обнаружились и другие расхождения теории эфира с опытом. Это привело Эйнштейна (1879–1955) в 1905 г. к теории относительности. Гипотеза механического эфира была оставлена [1]). Пошатнулась вера в «механическую картину мира», к построению которой так настойчиво стремились физики

[1]) Необходимо заметить, однако, что в вакууме нет обычного вещества, как оно понимается в химии. Но вакуум не есть пустота в буквальном смысле этого слова. Его заполняют физические поля (гравитационное, электромагнитное, ядерное и пр.). Они, наряду с обычным химическим веществом, являются различными формами материи. В вакууме могут происходить различные физические процессы. Примером может служить *поляризация вакуума*, т.е. рождение пар электрон–позитрон в сильных электрических полях. Можно было бы не возражать по существу против употребления термина «эфир» в смысле носителя этих физических свойств «пустого» пространства. Возражение относится к представлению об эфире как о жидкой, твердой, упругой или какой-либо другой среде, наделенной механическими свойствами. Однако в современной физике предпочитают не пользоваться термином «эфир» в указанном смысле, а употребляют термин «вакуум».

XIX века. И все последующее развитие науки привело физиков к убеждению в невозможности сведения всех явлений природы к механике. Утверждать противоположное — значит предъявлять к природе необоснованные требования. Поэтому, когда в современной волновой теории говорят, что свет — это колебания электромагнитного поля, то на это уже не смотрят как на формальное утверждение, а считают, что сами эти колебания не сводятся к чему-то «более простому и наглядному».

10. Важным этапом в развитии максвелловской электродинамики было введение в нее *атомистических представлений*, что было систематически проведено в *электронной теории* Г.А. Лорентца (1853–1928). В теории Максвелла вещество характеризовалось *феноменологическими постоянными* — диэлектрической и магнитной проницаемостями и удельной электрической проводимостью. В электронной теории эти макроскопические постоянные получили истолкование с атомистической точки зрения. В результате этого не только получили принципиальное, хотя и недостаточное, объяснение многие известные электродинамические и оптические явления (например, *дисперсия* и *абсорбция света*), но были открыты и объяснены и другие явления (*явления Керра*, *Зеемана*, *Фарадея*, *Коттона–Мутона*, *молекулярное рассеяние света* и т.д.).

Однако классическая физика и, в частности, электронная теория оказались недостаточными для истолкования явлений *атомного масштаба*. Потребовалось введение *квантовых представлений*. Необходимость и плодотворность последних обнаружилась ранее всего при изучении проблемы распределения энергии в спектре черного излучения, т.е. температурного излучения абсолютно черного тела. Применение к этой проблеме принципов классической физики приводило к глубоким противоречиям с опытом. Планк (1858–1947) в конце 1900 г. получил согласующуюся с опытом формулу для распределения энергии в спектре черного излучения. При этом он ввел чуждое классической физике представление, что излучение и поглощение света осуществляется не непрерывно, а *конечными порциями*, или *квантами энергии*, причем величина кванта определяется выражением (1.1). Для решения проблемы черного излучения Планку достаточно было принять, что этот квантовый характер излучения и поглощения света относится к *статистическим процессам*. Через пять лет Эйнштейн показал, что его необходимо распространить и на *элементарные процессы*. Согласно Эйнштейну, не только излучение и поглощение, но и *распространение света* в пространстве происходят конечными порциями — квантами света, обладающими определенной энергией и определенным импульсом. Так возродилось представление о частицах света, названных позднее *фотонами*.

Гипотеза фотонов позволила прежде всего объяснить загадочные закономерности в явлениях *фотоэффекта*, совершенно непонятные с точки зрения классической волновой теории света. Существование импульса у фотонов было доказано открытием в 1923 г. *эффекта Комптона* — изменения длины волны при рассеянии рентгеновского излучения. Гипотеза световых квантов позволила понять *химические действия света* и их закономерности. Квантовый характер излучения и поглощения света был использован Бором (1885–1962) для объяснения *спектральных закономерностей.*

Но как согласовать корпускулярные представления о свете с результатами опытов Фуко и Физо (см. п. 8)? Эти опыты вне всякого сомнения опровергают корпускулярную теорию света в ее ньютоновской форме. Приходится поэтому признать, что *к световым корпускулам классические представления о движении неприменимы.* Интерференция и дифракция света доказывают, что в этих явлениях свет ведет себя как *волны.* Фотоэффект, комптоновское рассеяние рентгеновских лучей и пр. с неменьшей убедительностью доказывают, что здесь свет действует как *частицы.* Вообще, *явления распространения света правильно описываются в рамках волновых теорий, а для описания взаимодействия света и вещества необходимы корпускулярные представления.* Этот «*дуализм волн и частиц*» надо рассматривать как экспериментальный факт, и поэтому полная теория света должна быть не корпускулярной и не волновой, а *корпускулярно-волновой.*

Такое парадоксальное положение еще более расширилось после того, как Дэвиссоном (1881–1958) и Джермером (1896–1971) в 1927 г. была открыта *дифракция электронов.* Оказалось, что *волновые свойства присущи и частицам обычного вещества*, — идея, развивавшаяся французским физиком де Бройлем (р. 1892) за несколько лет до открытия дифракции электронов. Развитие квантовой механики позволило частично объяснить возникшее парадоксальное положение ценой отказа от основного положения классической физики — *принципа причинности в форме детерминизма.* А исследования в области *физики высоких энергий* (иначе называемой *физикой элементарных частиц*) показали, что если энергия частиц превосходит их энергию покоя, то частицы могут рождаться, исчезать или превращаться друг в друга. В этом отношении они ведут себя подобно фотонам, которые могут излучаться или поглощаться. В квантовой электродинамике фотоны рассматриваются как *кванты электромагнитного поля.* Поэтому в физике высоких энергий целесообразно говорить об *электронно-позитронном*, *мезонном*, *нуклонном* и прочих полях, квантами которых являются электроны, позитроны, мезоны, протоны, нейтроны и т.д.

Таким образом, вопрос о природе света стал частью более общей проблемы строения вещества.

§ 4. Искривление световых лучей в неоднородных средах

1. В неоднородных средах представление о распространении света вдоль лучей сохраняется, но сами лучи становятся *криволинейными*. Действительно, рассмотрим среду, состоящую из плоскопараллельных слоев с постоянными показателями преломления (рис. 14), меняющимися скачкообразно от слоя к слою. Световой луч, преломляясь на границах слоев, примет форму ломаной линии. Будем неограниченно увеличивать число слоев, устремляя к нулю их толщины и скачки показателя преломления. Тогда в пределе показатель преломления среды станет меняться в пространстве непрерывно, а луч перейдет в кривую с непрерывно изменяющейся касательной. Допустим теперь, что показатель преломления меняется в пространстве как угодно, но не слишком резко (см. следующий параграф). Проведем в среде поверхности равного показателя преломления. В пределах каждого достаточно малого объема эти поверхности можно считать плоскими, а среду плоскослоистой, к которой применимо рассуждение, приведенное выше.

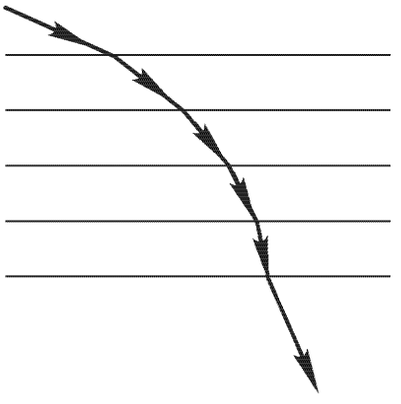


Рис. 14

Из изложенного следует, что геометрическую форму луча можно однозначно определить из закона Снеллиуса путем предельного перехода. Но закон Спеллиуса получается не только в волновой, но и в корпускулярной теории Ньютона. Поэтому при определении формы светового луча можно рассуждать так, как если бы свет состоял из ньютоновых корпускул, а показатель преломления n определялся формулой (3.3). Поскольку преломление определяется относительным показателем преломления, абсолютный показатель можно заменить величиной, ему пропорциональной. Ради краткости можно просто положить $n = v$. Скорость корпускулы v однозначно определяется уравнением сохранения энергии, а потому v можно рассматривать как известную функцию координат. Таким образом, *траектория корпускулы в потенциальном поле сил геометрически совпадает с лучом света в среде*, показатель преломления n которой численно равен v. Для этого, конечно, необходимо, чтобы исходные направления этих двух кривых были одинаковы. Эта формаль-

ная аналогия между движением частицы и распространением светового луча позволяет перенести результаты, полученные в световой оптике, в *электронную микроскопию*, где роль световых лучей выполняют электроны, движущиеся в потенциальных электрических полях.

Воспользуемся отмеченной аналогией для вычисления радиуса кривизны R светового луча. Нормальное ускорение корпускулы определяется формулой

$$\frac{mv^2}{R} = F_N = -\frac{\partial U}{\partial N},$$

где F_N — составляющая действующей силы $\mathbf{F}$ вдоль единичного вектора главной нормали $\mathbf{N}$, а U — потенциальная энергия корпускулы. Дифференцируя вдоль нормали $\mathbf{N}$ уравнение энергии $\frac{1}{2}mv^2 + U = \text{const}$, получаем $F_N = mv\,\partial v/\partial N$, а потому

$$\frac{1}{R} = \frac{1}{v}\frac{\partial v}{\partial N}.$$

Заменяя v на n, находим выражение для кривизны луча:

$$\frac{1}{R} = \frac{1}{n}\frac{\partial n}{\partial N} = \frac{p}{\partial N}(\ln n). \tag{4.1}$$

При этом ускорение корпускулы, а с ним и сила $\mathbf{F}$ не имеют составляющей вдоль бинормали $\mathbf{b}$ к траектории, т.е. $F_b = -\partial U/\partial b = mv\,\partial v/\partial b = 0$. Отсюда $\partial v/\partial b = 0$, а потому также $\partial n/\partial b = 0$. Значит, вектор $\operatorname{grad} n$ лежит в соприкасающейся плоскости светового луча. Поэтому из всех направлений, перпендикулярных к лучу, направление главной нормали $\mathbf{N}$ характеризуется самым быстрым изменением показателя преломления среды. Это значит, что *в неоднородной среде луч изгибается в сторону наиболее быстрого изменения показателя преломления*. Если среда однородна ($n = \text{const}$), то кривизна $1/R$ обращается в нуль, т.е. световые лучи прямолинейны.

2. Плотность земной атмосферы, а с ней и показатель преломления убывают с высотой. Этим объясняется ряд явлений, связанных с искривлением световых лучей. К ним относится, например, *астрономическая рефракция*, т.е. кажущееся поднятие небесного светила из-за искривления световых лучей в земной атмосфере. Если светило стоит высоко над горизонтом, то при расчете этого явления мы не сделаем заметной ошибки, считая поверхность Земли плоской. Но так поступать нельзя, когда светило находится вблизи горизонта. В этом случае надо учитывать сферичность земной поверхности.

Пренебрежем малыми боковыми градиентами показателя преломления воздуха n и будем считать, что n зависит только от высоты над земной поверхностью или, что то же самое, от расстояния r от центра земного шара O (рис. 15). Световой луч AM от небесного светила будет лежать в вертикальной плоскости, проходящей через это светило и глаз наблюдателя. На основании (4.1)

$$\frac{1}{R} = \frac{d\alpha}{ds} = \frac{\partial}{\partial N}(\ln n) = \frac{d(\ln n)}{dr}\frac{\partial r}{\partial N} = -\frac{d(\ln n)}{dr}\sin\beta,$$

где α — так называемое *зенитное расстояние*, т.е. переменный угол между вертикалью места MZ и касательной к лучу, а s — длина луча, отсчитываемая от глаза наблюдателя M. Как видно из рис. 15,

$$\frac{d\alpha}{ds} = \frac{d\alpha}{dr}\frac{dr}{ds} = \frac{d\alpha}{dr}\cos\beta,$$

а потому

$$\frac{1}{\operatorname{tg}\beta}\frac{d\alpha}{dr} = -\frac{d}{dr}(\ln n). \qquad (4.2)$$

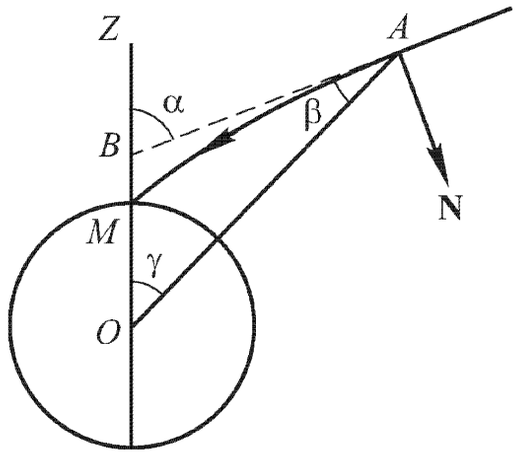


Рис. 15

Далее, $r\,d\gamma = ds\sin\beta$, откуда

$$\frac{1}{\operatorname{tg}\beta}\frac{d\gamma}{dr} = \frac{1}{r}\frac{ds}{dr}\cos\beta = \frac{1}{r}.$$

Вычитая это равенство из (4.2) и принимая во внимание, что $\alpha - \gamma \approx \beta$, получим

$$\frac{d\beta}{\operatorname{tg}\beta} = -d\ln n - \frac{dr}{r} = -d\,\ln(nr).$$

Интегрирование этого уравнения дает

$$nr\sin\beta = n_0 r_0 \sin\alpha_0, \qquad (4.3)$$

где нулем обозначены величины n, r, α в точке M (в которой $\alpha \equiv \beta$). Вычислив отсюда $\operatorname{tg}\beta$ и подставив его значение в (4.2), найдем

$$\alpha_\infty - \alpha_0 = -n_0 r_0 \sin\alpha_0 \int_{r_0}^{\infty} \frac{d\ln n}{dr}\frac{dr}{\sqrt{n^2r^2 - n_0^2 r_0^2\sin^2\alpha_0}}. \qquad (4.4)$$

Здесь α_0 — видимое зенитное расстояние светила в точке M, а α_∞ — угол между асимптотой луча и вертикалью в той же точке. Разность $\alpha_\infty - \alpha_0$ называется *рефракцией*. Для ее вычисления и служит формула (4.4). Зная зависимость плотности воздуха ρ от высоты над земной поверхностью, находят n по формуле $(n-1)/\rho = \text{const}$, после чего производят численное интегрирование в формуле (4.4).

Если светило находится не слишком близко от горизонта, то кривизной земной поверхности можно пренебречь. Тогда рефракция не зависит от характера изменения показателя преломления с высотой и может быть вычислена по формуле

$$\sin\alpha_\infty = n_0\sin\alpha_0 \qquad (4.5)$$

(см. задачу 3 к § 2). Ошибка, даваемая этой формулой, быстро убывает с уменьшением α. Уже при $\alpha_0 = 80^\circ$ она не превосходит 3 %. При $\alpha_0 < 75^\circ$ формула (4.5) дает достаточную точность.

Если светило находится на горизонте ($\alpha = 90^\circ$), то средняя рефракция при 10 °C и 760 мм рт. ст. составляет $35'24''$. Она быстро уменьшается по мере поднятия светила над горизонтом. Уже при $\alpha_0 = 89^\circ$ средняя рефракция уменьшается до $24'37''$. Этим объясняется сплюснутая форма Солнца при восходе и заходе.

С рефракцией связано некоторое удлинение дня. Для средних широт оно составляет в среднем 3–4 минуты.

Рефракция зависит от длины волны. Поэтому при заходе Солнца сначала должны исчезать красные и желтые лучи, а оставшийся сегмент солнечного диска должен окрашиваться на одну-две секунды в зеленый или даже синий цвет. При восходе Солнца, наоборот, должна сначала появляться кратковременная зеленая вспышка. Это явление «зеленого луча» наблюдается на море, да и то крайне редко, так как необходимы исключительно спокойная атмосфера и ясная погода.

3. Иногда вблизи земной поверхности из-за сильного нагревания или охлаждения возникают большие градиенты показателя преломления воздуха. Тогда шарообразность Земли можно не учитывать. Если градиент n направлен вертикально, то можно воспользоваться формулой (4.3), полагая в ней $\beta = \alpha$, $r = r_0$. В результате получится

$$n \sin \alpha = n_0 \sin \alpha_0 = \text{const}. \tag{4.6}$$

Допустим ради определенности, что световой луч распространяется вверх под малым углом к горизонту, а показатель преломления $n = n(z)$ убывает с высотой z. Может случиться, что на некоторой высоте угол α обратится в $90°$. Тогда касательная к лучу станет горизонтальной. Высота $z = h$, на которой это может произойти, определится из соотношения

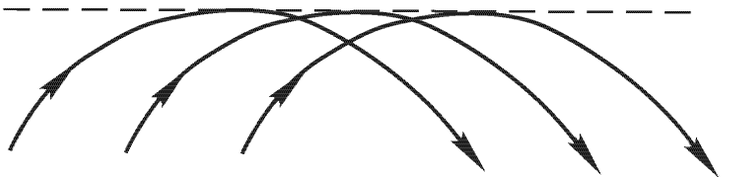


Рис. 16

$$n(h) = n_0 \sin \alpha_0.$$

Так как показатель преломления возрастает вниз, то, достигнув высоты $z = h$, луч в дальнейшем должен загнуться книзу. На указанной высоте происходит нечто аналогичное полному отражению (рис. 16). Такое явление может возникнуть при сильно аномальном распределении плотности воздуха по высоте и является причиной различного рода *миражей*, наблюдаемых в атмосфере.

Рис. 17

Обычно наблюдается *верхний* или *нижний мираж*. При верхнем мираже, помимо самих предметов, видны их изображения, расположенные сверху; при нижнем мираже изображение получается ниже самого предмета. Нижний мираж наблюдается в пустынях и в степях в теплое время года, когда прилегающий к земной поверхности слой воздуха сильно нагрет, а его плотность и показатель преломления быстро возрастают с высотой. Из каждой точки предмета в глаз наблюдателя всегда попадают прямые лучи, т.е. лучи, не испытавшие полного отражения в воздухе; им соответствует обычное — прямое — изображение предмета. Но при больших градиентах показателя преломления могут также попасть лучи, испытавшие полное отражение. Они дают обратное изображение предмета, как в зеркале (рис. 17). При этом лучи, выходящие из различных точек предмета, претерпевают полное

отражение на несколько разных высотах; точка M, в которой отражается луч, выходящий из A, лежит несколько выше точки N, в которой отражается луч, выходящий из B. Наблюдатель видит два изображения: прямое и обратное. Создается иллюзия водной поверхности, в которой, как в зеркале, видно изображение неба.

Аналогично объясняется и верхний мираж. Он наблюдается зимой в холодных странах, когда вблизи земной поверхности образуется холодный слой воздуха, в котором показатель преломления быстро убывает с высотой. В горах, хотя и очень редко, наблюдается *боковой мираж*, связанный с изменением показателя преломления воздуха в боковом направлении. Вообще, в зависимости от характера распределения показателя преломления воздуха, мираж проявляется в весьма разнообразных и часто причудливых формах.

ЗАДАЧИ

1. Показать, что в пренебрежении кривизной земной поверхности справедлива формула

$$\alpha_\infty - \alpha_0 = (n_0 - 1)\operatorname{tg}\alpha_0. \tag{4.7}$$

2. Определить, при каких градиентах температуры T воздуха возможен нижний мираж.

Р е ш е н и е. Показатель преломления, а с ним и плотность воздуха ρ должны увеличиваться с высотой h (т.е. $d\rho/dh > 0$). Воспользуемся уравнением состояния $\rho = \mu P/(RT)$, где P — давление, μ — относительная молекулярная масса воздуха, а R — универсальная газовая постоянная. Из него находим

$$\frac{1}{\rho}\frac{d\rho}{dh} = \frac{1}{P}\frac{dP}{dh} - \frac{1}{T}\frac{dT}{dh}.$$

При механическом равновесии воздуха $dP/dh = -\rho g$, где g — ускорение свободного падения. В результате получаем

$$\frac{dT}{dh} < \frac{\mu g}{R}. \tag{4.8}$$

Используя соотношение $c_P - c_v = R/\mu$, этому условию можно придать вид

$$\frac{dT}{dh} < -\frac{g}{c_P - c_v} \approx -0{,}025 \text{ К/м}, \tag{4.9}$$

где c_P и c_v — удельные теплоемкости воздуха при постоянном давлении и постоянном объеме. Такое распределение температур *конвективно неустойчиво*, так как для конвективной устойчивости необходимо

$$\frac{dT}{dh} \geqslant -\frac{g}{c_P} \approx -0{,}01 \text{ К/м}$$

(см. т. II, § 121).

§ 5. Плоские электромагнитные волны

1. С электромагнитными волнами мы встречались уже в томе III нашего курса (гл. X). Выведем еще раз основные результаты, полученные в § 139 указанного тома, предполагая здесь,

что волны — *монохроматические*. Будем представлять их в комплексной форме. В связи с этим полезно еще раз обратиться к § 126 того же тома, где введены комплексные обозначения и изложены правила оперирования с ними. По причинам, которые выяснятся в ходе изложения, монохроматические волны и их комплексная форма имеют важнейшее значение в учении о волнах вообще, а в особенности в оптике.

Допустим, что волны распространяются в однородном незаряженном диэлектрике. Применим к ним фундаментальные уравнения Максвелла

$$\operatorname{rot}\mathbf{H}=\frac{1}{c}\frac{\partial\mathbf{D}}{\partial t},\quad \operatorname{rot}\mathbf{E}=-\frac{1}{c}\frac{\partial\mathbf{B}}{\partial t} \tag{5.1}$$

и материальные уравнения

$$\mathbf{D}=\varepsilon\mathbf{E},\quad \mathbf{B}=\mu\mathbf{H}. \tag{5.2}$$

Пусть волна — плоская и монохроматическая. Запишем ее в комплексном виде

$$\mathbf{E}=\mathbf{E}_0e^{i(\omega t-\mathbf{kr})},\quad \mathbf{H}=\mathbf{H}_0e^{i(\omega t-\mathbf{kr})}, \tag{5.3}$$

где ω — круговая частота, $\mathbf{k}$ — волновой вектор, а амплитуды $\mathbf{E}_0$ и $\mathbf{H}_0$ постоянны. Дифференцируя по времени, получаем $\partial\mathbf{E}/\partial t=i\omega\mathbf{E}$, т.е. операция дифференцирования в этом случае сводится к умножению на $i\omega$. Аналогично, дифференцирование по координатам x, y, z сводится к умножению на $-ik_x$, $-ik_y$, $-ik_z$. Заметив это и обозначая координатные орты через $\mathbf{e}_x$, $\mathbf{e}_y$, $\mathbf{e}_z$, получаем

$$\operatorname{rot}\mathbf{H}=\begin{vmatrix}\mathbf{e}_x & \mathbf{e}_y & \mathbf{e}_z\\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z}\\ H_x & H_y & H_z\end{vmatrix}=-i\begin{vmatrix}\mathbf{e}_x & \mathbf{e}_y & \mathbf{e}_z\\ k_x & k_y & k_z\\ H_x & H_y & H_z\end{vmatrix}=-i[\mathbf{kH}]$$

и аналогично для $\operatorname{rot}\mathbf{E}$. В результате уравнения (5.1) перейдут в

$$[\mathbf{kH}]=-\frac{\omega}{c}\mathbf{D},\quad [\mathbf{kE}]=\frac{\omega}{c}\mathbf{B}. \tag{5.4}$$

Введем единичный вектор $\mathbf{N}$ нормали к фронту волны и скорость распространения последнего в направлении этой нормали — так называемую *нормальную скорость* v. Тогда

$$\mathbf{k}=\frac{\omega}{v}\mathbf{N}, \tag{5.5}$$

и предыдущие соотношения перейдут в

$$\mathbf{D}=-\frac{c}{v}[\mathbf{NH}],\quad \mathbf{B}=\frac{c}{v}[\mathbf{NE}]. \tag{5.6}$$

Отсюда видно, что векторы $\mathbf{E}$, $\mathbf{H}$, $\mathbf{v}$ в плоской электромагнитной волне взаимно перпендикулярны. Их взаимное расположе-

ние представлено на рис. 18. Перпендикулярность векторов $\mathbf{E}$ и $\mathbf{H}$ к вектору $\mathbf{v}$, или, что то же, к направлению распространения волны, означает, что *электромагнитные волны поперечны*. Таким образом, проблема поперечности световых волн, с которой не могли справиться теории механического эфира, совсем не возникает в электромагнитной теории света.

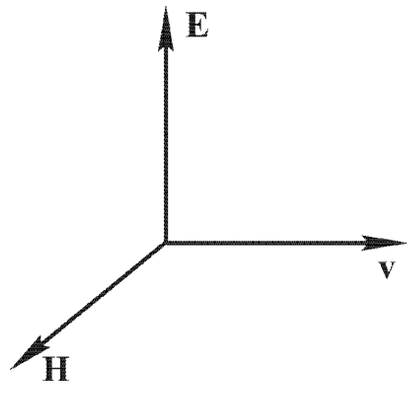


Рис. 18

Следует отметить, что между взаимно перпендикулярными составляющими, на которые можно разложить комплексные амплитуды $\mathbf{E}_0$ и $\mathbf{H}_0$, может существовать какая-то разность фаз φ. Если она равна нулю или π, то вектор $\mathbf{E}$ во всех точках пространства и во все моменты времени будет находиться в одной и той же плоскости, проходящей через направление нормали $\mathbf{N}$, а вектор $\mathbf{H}$ — в плоскости, к ней перпендикулярной. Тогда волну называют *линейно-* или *плоскополяризованной*. Плоскость $(\mathbf{E}, \mathbf{N})$, в которой лежат векторы $\mathbf{E}$ и $\mathbf{N}$, называется *плоскостью колебаний* или *плоскостью поляризации волны* [1]). Если φ не равна нулю или π, то возникает так называемая *эллиптическая поляризация*. Исследованием ее мы займемся в главах V, VI и VII.

Из уравнений (5.1) можно определить и скорость электромагнитной волны v. С этой целью запишем эти уравнения в скалярной форме:

$$D = \frac{c}{v} H, \quad B = \frac{c}{v} E, \tag{5.7}$$

или

$$\varepsilon E = \frac{c}{v} H, \quad \mu H = \frac{c}{v} E. \tag{5.8}$$

Отсюда после почленного перемножения и сокращения на EH получаем для v и показателя преломления $n = c/v$ следующие выражения:

$$v = c/\sqrt{\varepsilon\mu}, \tag{5.9}$$

$$n = \sqrt{\varepsilon\mu}. \tag{5.10}$$

Последнее соотношение называется *законом Максвелла*. Для немагнитных сред ($\mu = 1$) оно переходит в

$$n = \sqrt{\varepsilon}. \tag{5.11}$$

[1]) Раньше плоскостью поляризации называли плоскость $(\mathbf{H}, \mathbf{N})$, перпендикулярную к $(\mathbf{E}, \mathbf{N})$. Но этот термин в таком понимании выходит из употребления.

В вакууме $v = c$, т.е. v совпадает с *электродинамической постоянной* c. Тем самым раскрывается глубокий физический смысл открытия В. Вебера (1804–1891) и Р. Кольрауша (1809–1858), впервые измеривших эту постоянную в 1856 г. (см. т. III, § 51, 83).

2. Обратимся теперь к экспериментальной проверке соотношения (5.11). В табл. 1 сопоставлены экспериментально измеренные значения n и $\sqrt{\varepsilon}$ для ряда веществ (показатели преломления относятся к желтой линии натрия). Для газов, приведенных в этой таблице, закон Максвелла (5.11) хорошо согласуется с опытом. Для жидких углеводородов согласие хуже. Для воды и спиртов, а также для большинства других твердых и жидких тел наблюдаются резкие нарушения соотношения (5.11). Однако в этом нет ничего неожиданного. Дело в том, что значения ε, приведенные в табл. 1, относятся к *статическим электрическим полям*, а значения n — к *электромагнитным полям световых волн*, частоты которых порядка $5 \cdot 10^{14}$ Гц. Диэлектрическая проницаемость ε обусловлена *поляризацией диэлектрика*, т.е. смещением заряженных частиц внутри атомов и молекул под действием внешнего электрического поля. Для правильного сопоставления надо брать значения ε, измеренные в электрических полях *тех же частот*. Действительно, атомы и молекулы обладают собственными частотами, так что амплитуды (и фазы) вынужденных колебаний электронов и ядер, из которых они состоят, зависят от частоты внешнего электрического поля. Особенно сильную зависимость следует ожидать в тех случаях, когда частота внешнего поля близка к одной из собственных частот атомов или молекул (резонанс!). В результате возникает зависимость показателя преломления вещества от частоты световой волны — так называемая *дисперсия света*.

Таблица 1

Вещество	n	$\sqrt{\varepsilon}$	Вещество	n	$\sqrt{\varepsilon}$
Воздух	1,000292	1,000302	Толуол	1,499	1,549
Азот	1,000299	1,000307	Четыреххлористый	1,461	1,523
Кислород	1,000270	1,000273	углерод		
Водород	1,000139	1,000139	Бензол	1,501	1,511
Углекислота	1,000499	1,000485	Сероуглерод	1,629	1,626
Гелий	1,000035	1,000037	Парафин	1,422	1,405
Окись углерода	1,000335	1,000346	Вода	1,33	9,00
Аммиак	1,000385	1,000385	Метиловый спирт	1,34	5,70
Закись азота	1,000507	1,000547	Этиловый спирт	1,36	5,10

Аналогично, намагниченность вещества (парамагнетизм и ферромагнетизм) возникает в результате *поворотов магнитных моментов атомов и молекул* под действием внешнего магнитного поля. Однако в полях столь высоких частот, которые лежат в оптической области спектра, атомы и молекулы не успевают поворачиваться за времена порядка периода световых колебаний. Диамагнитный же эффект, принципиально имеющий место во всех веществах, пренебрежимо мал. Поэтому в полях указанных частот намагничивание вещества практически не происходит. Вот почему в оптических явлениях, за редкими исключениями, магнитные свойства вещества не проявляются, и можно пользоваться формулой (5.11) вместо более общей формулы (5.10).

Таким образом, расхождения закона Максвелла (5.10) или (5.11) с опытом происходят за счет *нарушения материальных уравнений* (5.2). Такие уравнения справедливы не всегда, а только для *монохроматических полей*, причем *ε и μ являются функциями частоты электромагнитного поля*, различными для различных веществ. Чтобы отметить это обстоятельство, величины $\varepsilon(\omega)$ и $\mu(\omega)$ часто называют *динамическими диэлектрической и магнитной проницаемостями*, в отличие от *статических проницаемостей*, в которые они переходят при $\omega = 0$. Лишь в области сравнительно длинных электромагнитных волн (превышающих примерно 1 см) функции $\varepsilon(\omega)$ и $\mu(\omega)$ становятся постоянными для всех веществ. Поэтому в оптике электромагнитное поле приходится разлагать на монохроматические составляющие, что всегда возможно, согласно математической теореме Фурье (см. т. III, § 128). Предполагая, что выполняется принцип суперпозиции, эти монохроматические составляющие можно рассматривать *независимо друг от друга*. Таким путем можно исследовать распространение электромагнитных волн любого спектрального состава. Функции можно разлагать не только по синусам и косинусам, но и по бесконечному множеству других, «полных систем» функций. Однако выполнение материальных уравнений (5.2) для монохроматических полей, а также многие другие причины делают в оптике разложение полей на монохроматические составляющие *физически выделенным* среди множества других математически возможных разложений. Изложенные соображения, как и соображения, излагаемые в следующем пункте, имеют, конечно, общее значение, а не только для плоских электромагнитных волн.

3. Соблюдение материальных уравнений (5.2) (для монохроматических полей) предполагает, конечно, что величина ε не зависит от $\mathbf{E}$. Благодаря этому уравнения поля *линейны и однородны*. Однако это справедливо только в *слабых полях*, т.е. таких полях, напряженность которых весьма мала по сравнению с на-

пряженностями внутриатомных и внутримолекулярных полей (10^7–10^8 В/см). В *сильных полях*, где это условие не соблюдается, уравнения поля в средах становятся *нелинейными*, что ведет к *нарушению принципа суперпозиции*, а следовательно, и закона независимости световых пучков. Оптику слабых электромагнитных полей (в указанном выше смысле) называют *линейной оптикой*, а оптику сильных полей — *нелинейной оптикой*.

До недавнего времени источники света, которыми располагала оптика, позволяли получать световые пучки лишь «слабой» интенсивности с максимальными электрическими полями 0,1–10 В/см (см. задачу к этому параграфу). Нелинейные эффекты (за редкими исключениями) в этих случаях совсем незаметны. Случай сильных электрических полей в оптике считался чисто умозрительным и долгое время не исследовался. Начиная с 1960 г. — года изобретения *оптических квантовых генераторов* (*лазеров*) — положение изменилось. Эти источники света позволяют получать световые волны, в которых электрические поля достигают 10^5–10^7 В/см. Такие поля уже не могут считаться пренебрежимо малыми по сравнению с внутриатомными и внутримолекулярными полями. В них наблюдаются качественно новые «*нелинейные явления*» и притом не как малые поправки к «линейным эффектам», а как *явления крупного масштаба*, получившие уже важные практические применения. О нелинейной оптике мы будем говорить в гл. XI. Во всех остальных главах излагается линейная оптика.

4. Для некоторых целей требуется более подробное исследование системы уравнений (5.4), которое мы и произведем. Запишем эту систему с учетом материальных уравнений (5.2):

$$[\mathbf{kH}] = -\frac{\omega\varepsilon}{c}\,\mathbf{E}, \quad [\mathbf{kE}] = \frac{\omega\mu}{c}\,\mathbf{H}. \tag{5.12}$$

Подставляя $\mathbf{H}$ из второго уравнения в первое, получим

$$\frac{c}{\omega\mu}\,[\mathbf{k}[\mathbf{kE}]] = -\frac{\omega\varepsilon}{c}\,\mathbf{E}.$$

Так как векторы $\mathbf{k}$ и $\mathbf{E}$ взаимно перпендикулярны, то после раскрытия двойного векторного произведения и сокращения на $\mathbf{E}$ отсюда найдем

$$\mathbf{k}^2 = \frac{\omega^2}{c^2}\,\varepsilon\mu, \tag{5.13}$$

или, пренебрегая намагничиванием вещества,

$$\mathbf{k}^2 = \frac{\omega^2}{c^2}\,\varepsilon. \tag{5.14}$$

Таково условие совместности уравнений (5.12). Так как по самому определению волнового вектора $k = \omega/v$, то из (5.13) следует: $v = c/\sqrt{\varepsilon\mu}$, т.е. прежний результат (5.9).

Однако условиям (5.13) или (5.14) можно удовлетворить и в тех случаях, когда волновой вектор $\mathbf{k}$ *комплексный*, т.е. имеет вид

$$\mathbf{k} = \mathbf{k}' - i\mathbf{k}'', \tag{5.15}$$

где векторы $\mathbf{k}'$ и $\mathbf{k}''$ — вещественные. В этом случае волну называют *неоднородной*, в отличие от *однородной волны*, у которой волновой вектор $\mathbf{k}$ вещественный. Электрический вектор неоднородной волны можно представить в виде

$$\mathbf{E} = \mathbf{E}_0 e^{-\mathbf{k}''\mathbf{r}} e^{i(\omega t - \mathbf{k}'\mathbf{r})}. \tag{5.16}$$

Это выражение можно рассматривать как волну, амплитуда которой равна $\mathbf{E}_0 e^{-\mathbf{k}''\mathbf{r}}$. Она экспоненциально убывает в направлении вектора $\mathbf{k}''$. Для такой волны можно говорить о *поверхностях равных амплитуд* и *поверхностях равных фаз.* Поверхности равных амплитуд суть плоскости, перпендикулярные к вектору $\mathbf{k}''$. Поверхности равных фаз также плоскости, но перпендикулярные к вектору $\mathbf{k}'$. *В непоглощающей изотропной среде плоскости равных амплитуд и равных фаз взаимно перпендикулярны.* Для доказательства подставим выражение (5.15) в формулу (5.14). Отделяя вещественную часть от мнимой, найдем

$$\mathbf{k}'^2 - \mathbf{k}''^2 = \frac{\omega^2}{c^2}\varepsilon, \quad (\mathbf{k}'\mathbf{k}'') = 0. \tag{5.17}$$

Из второго соотношения видно, что векторы $\mathbf{k}'$ и $\mathbf{k}''$, а следовательно и плоскости равных амплитуд и фаз, взаимно перпендикулярны. Для поглощающих сред это утверждение, вообще говоря, не справедливо.

Чтобы получить более отчетливое представление о неоднородной волне, запишем ее в вещественной форме. Направим ось X вдоль вектора $\mathbf{k}'$, а ось Z — вдоль вектора $\mathbf{k}''$. Кроме того, положим $E_{0x} = A_x \exp(i\delta_x)$ и аналогично для E_{0y} и E_{0z}. Тогда неоднородную волну (5.16) в вещественной форме можно представить так:

$$\begin{aligned} E_x &= A_x e^{-k''z}\cos(\omega t - k'x + \delta_x),\\ E_y &= A_y e^{-k''z}\cos(\omega t - k'x + \delta_y),\\ E_z &= A_z e^{-k''z}\cos(\omega t - k'x + \delta_z). \end{aligned}$$

Плоскости равных фаз распространяются в направлении оси X со скоростью $v' = \omega/k'$. Она меньше скорости однородных волн $c/\sqrt{\varepsilon}$, поскольку $k' > \omega\sqrt{\varepsilon}/c$, как это следует из (5.17). Амплитуда убывает в направлении оси Z. Когда $z \to -\infty$, амплитуда неоднородной волны возрастает неограниченно. Поэтому плоские неоднородные волны в неограниченной среде существовать не могут. Но они могут при определенных условиях

возникать вблизи границы среды. Так, например, такие волны возникают в оптически менее плотной среде при полном отражении света (см. § 66). Поле неоднородной волны заметно лишь в пограничном слое, толщина которого порядка длины волны. По этой причине неоднородные волны называют также *поверхностными волнами.*

ЗАДАЧА

Оценить напряженность поля солнечного излучения вблизи земной поверхности, если величина солнечной постоянной составляет около 2 кал·см$^{-2}$ мин$^{-1}$ = 1,39·10^6 эрг·см$^{-2}$с$^{-1}$. *Солнечной постоянной* называется количество энергии, попадающей от Солнца (при его среднем удалении от Земли) за единицу времени на единицу площади земной поверхности, перпендикулярной к излучению (при отсутствии абсорбции в атмосфере).

Р е ш е н и е. Волну, излучаемую Солнцем, у земной поверхности можно считать плоской. В такой волне $E = H$, поскольку для вакуума $\varepsilon = \mu = 1$. Плотность потока энергии $\frac{c}{4\pi}EH = \frac{c}{4\pi}E^2$. Приравнивая среднее значение этой величины значению солнечной постоянной, получим

$$\frac{c}{4\pi}\overline{E^2} = 1{,}39 \cdot 10^6, \text{ откуда } \overline{E^2} = 5{,}85 \cdot 10^{-4}, \ \sqrt{\overline{E^2}} = 0{,}024 \text{ СГСЭ} = 7{,}2 \text{ В/см}.$$

§ 6. Предельный переход от волновой оптики к геометрической

1. Геометрическая оптика является приближенным предельным случаем, в который переходит волновая оптика, когда длина световой волны стремится к нулю. Чтобы показать это, надо было бы исходить из уравнений Максвелла в неоднородных средах. Однако такой путь приводит к громоздким вычислениям. Мы поступим иначе. Среду, в которой распространяется свет, будем считать прозрачной и однородной. Предполагая сначала, что она изотропна, исключим из уравнений (5.1) и (5.2) вектор $\mathbf{H}$. С этой целью первое уравнение (5.1) дифференцируем по t, а от обеих частей второго возьмем операцию rot, воспользовавшись при этом векторной формулой

$$\operatorname{rot}\operatorname{rot}\mathbf{E} = \operatorname{grad}\operatorname{div}\mathbf{E} - \Delta\mathbf{E} \ ^{1}), \tag{6.1}$$

где Δ — оператор Лапласа в прямоугольной системе координат, т.е.

$$\Delta \equiv \nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \tag{6.2}$$

[1]) Эту формулу легко получить, записав левую часть в виде $[\nabla[\nabla\mathbf{E}]]$ и раскрыв по обычному правилу двойное векторное произведение. При этом надо только помнить, что векторы ∇ и $\mathbf{E}$ нельзя переставлять. Таким путем получаем: $\nabla(\nabla\mathbf{E}) - \nabla^2\mathbf{E}$, т.е. правую часть (6.1).

Из полученных таким образом соотношений легко исключить **H**. В результате получится

$$\Delta\mathbf{E} - \frac{1}{v^2}\frac{\partial^2\mathbf{E}}{\partial t^2} = 0, \tag{6.3}$$

где v определяется прежним выражением (5.9). Уравнение (6.3) называется *волновым*. Такому же уравнению удовлетворяет и вектор **H**.

Для неоднородных сред уравнение (6.3) усложняется. Однако если интересоваться *только интенсивностью волн*, отвлекаясь от их *поляризации*, то оказывается, что в предельном случае геометрической оптики уравнение (6.3) приводит к правильным результатам [1]). Поэтому даже в случае неоднородных сред предельный переход к геометрической оптике можно выполнить на основе волнового уравнения

$$\Delta E - \frac{1}{v^2}\frac{\partial^2 E}{\partial t^2} = 0, \tag{6.4}$$

в котором E означает длину вектора **E**, а скорость v считается известной функцией координат. Хотя такой путь и не вполне удовлетворителен, но на нем проще уяснить метод, применяемый при обосновании геометрической оптики. Результаты, к которым мы придем, применимы не только к световым, но и *ко всем другим волнам*, например акустическим или волнам де Бройля в квантовой механике.

2. Предполагая волну монохроматической, запишем ее в виде

$$E = a(\mathbf{r})e^{i(\omega t - k_0\Phi)}, \tag{6.5}$$

где $a(\mathbf{r})$ и $\Phi(\mathbf{r})$ — вещественные функции координат. Волновое число в вакууме $k_0 = \omega/c = 2\pi/\lambda$ введено для удобства, как большой размерный параметр. Подставим выражение (6.5) в уравнение (6.4) и отделим вещественную часть от мнимой. В результате получим два уравнения:

$$(\operatorname{grad}\Phi)^2 = n^2 + \frac{\Delta a}{k_0^2 a}, \tag{6.6}$$

$$a\Delta\Phi + 2\operatorname{grad} a \operatorname{grad}\Phi = 0. \tag{6.7}$$

Допустим теперь, что длина волны мала, а амплитуда a меняется в пространстве не очень быстро, так что соблюдается неравенство

$$\left|\frac{\Delta a}{k_0^2 a}\right| \equiv \frac{\lambda^2}{4\pi^2}\left|\frac{\Delta a}{a}\right| \ll n^2. \tag{6.8}$$

[1]) Предельный переход к геометрической оптике на основе векторных уравнений Максвелла подробно исследован в книге: *Сивухин Д.В.* Лекции по физической оптике, ч. II. Ротапринтное издание. — Новосибирск, 1969.

Для этого достаточно, чтобы выполнялись условия

$$\left|\lambda\frac{\partial a}{\partial x}\right| \ll a, \quad \left|\lambda\frac{\partial^2 a}{\partial x^2}\right| \ll \left|\frac{\partial a}{\partial x}\right| \tag{6.9}$$

для любого направления оси X. Действительно, тогда

$$\left|\lambda^2\frac{\partial^2 a}{\partial x^2}\right| \ll \left|\lambda\frac{\partial a}{\partial x}\right| \ll a,$$

что совпадает с (6.8), так как $|\Delta a| \sim |\partial^2 a/\partial x^2|$. Пренебрегая в (6.6) последним членом, получим

$$(\operatorname{grad}\Phi)^2 = n^2. \tag{6.10}$$

Уравнения (6.10) и (6.7) и составляют *систему уравнений геометрической оптики.* Из их вывода ясно, что *условием применимости геометрической оптики является малость изменения амплитуды волны и ее первых пространственных производных на протяжении длины волны.* В противном случае могут возникать заметные отступления от геометрической оптики. Это происходит, например, в следующих случаях: 1) на границе геометрической тени; 2) вблизи фокуса, т.е. геометрической точки схождения лучей; 3) при распространении света в среде с резко меняющимся показателем преломления (например, в мутной среде); 4) при распространении света в сильно поглощающих средах (например, металлах).

3. Величину Ф Клаузиус (1822–1888) назвал *эйконалом*, а уравнение (6.10) — *уравнением эйконала.* Его можно записать в векторной форме:

$$\operatorname{grad}\Phi = n\mathbf{s}, \tag{6.11}$$

где $\mathbf{s}$ — единичный вектор нормали к фронту волны

$$\omega t - k_0\Phi = \text{const}, \tag{6.12}$$

проведенный в сторону ее распространения.

Уравнение эйконала определяет скорость распространения волнового фронта в направлении нормали $\mathbf{s}$. Действительно, на основании определения градиента можно записать (6.11) в виде $\partial\Phi/\partial s = n$. С другой стороны, дифференцирование уравнения распространения волнового фронта (6.12) дает $\omega\,dt = k_0\,d\Phi$, или

$$\omega\,dt = \frac{\omega}{c}\frac{\partial\Phi}{\partial s}\,ds = \frac{\omega}{c}\,n\,ds = \frac{\omega}{v}\,ds.$$

Отсюда для нормальной скорости волнового фронта находим

$$\frac{ds}{dt} = v, \tag{6.13}$$

т.е. эта скорость такая же, как у плоской волны. Этого и следовало ожидать, так как малый участок волнового фронта в малых объемах пространства должен вести себя как плоский.

Полученный результат позволяет построить волновой фронт F_2 в момент времени $t + dt$, если известно его положение F_1 в момент t. Для этого из каждой точки исходного волнового фронта F_1 (рис. 19) следует отложить в направлении нормали отрезок

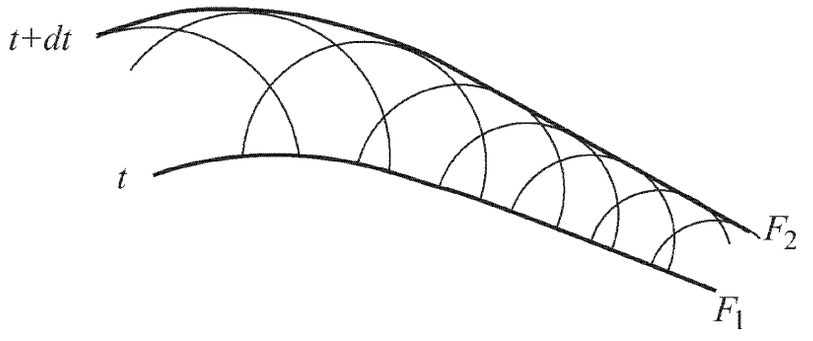


Рис. 19

длиной $v\,dt$. Соединив концы всех таких отрезков, мы и получим волновой фронт F_2 в момент $t + dt$. Вместо этого можно из каждой точки волнового фронта F_1, как из центра, описать сферы радиусом $v\,dt$. Огибающая таких сфер и будет волновым фронтом F_2. Оба построения совершенно эквивалентны. Тем самым построение Гюйгенса (см. § 3, п. 4) распространяется и на волны в неоднородных средах.

4. Второе уравнение геометрической оптики (6.7) теперь можно записать в виде

$$a\Delta\Phi + 2n\frac{\partial a}{\partial s} = 0. \tag{6.14}$$

Если определить *луч как ортогональную траекторию к семейству волновых фронтов, или семейству равных фаз* (6.12), то взятие производной по s можно понимать в смысле дифференцирования по длине луча s. Интегрируя уравнение (6.14) вдоль луча, придем к соотношению

$$a = a_0 \exp\left(-\int\limits_0^s \frac{\Delta\Phi}{2n}\,ds\right), \tag{6.15}$$

где a_0 — амплитуда в «начальной точке» луча, от которой отсчитывается длина s. Формула (6.15) показывает, что для определения волнового поля во всех точках луча достаточно знать его значение в какой-либо одной точке того же луча. Но уравнения геометрической оптики ничего не могут сказать относительно изменения амплитуды поля при переходе от одного луча к соседнему. Они допускают любые изменения амплитуды от луча к лучу. Необходима только достаточная медленность такого изменения, чтобы волна, формально удовлетворяющая уравнениям геометрической оптики, могла быть реализована в действительности. Таким образом, *в приближении геометрической оптики*

световое поле на всяком луче совершенно не зависит от полей других лучей.

Отсюда следует основное представление геометрической оптики о *распространении световой энергии вдоль лучей*, точнее — вдоль «лучевых трубок», образованных лучами. Отсюда также следует, что при нахождении волнового фронта построением Гюйгенса новый волновой фронт не выходит за пределы огибающей вторичных сферических волн Гюйгенса. Тем самым дано полное обоснование гипотезы Гюйгенса об огибающей и показано, что эта гипотеза справедлива только в приближении геометрической оптики.

К представлению о распространении света вдоль лучей можно прийти и другим путем. Умножая уравнение (6.7) на a и замечая, что $\Delta\Phi = \operatorname{div}\operatorname{grad}\Phi = \operatorname{div}(n\mathbf{s})$, перепишем это уравнение в виде

$$a^2\operatorname{div}(n\mathbf{s}) + 2a\operatorname{grad}a(n\mathbf{s}) = 0,$$

или

$$\operatorname{div}(na^2\mathbf{s}) = 0. \tag{6.16}$$

Полученное соотношение по форме совпадает с уравнением непрерывности $\operatorname{div}\mathbf{j} = 0$ для стационарного течения несжимаемой жидкости. Роль линий тока играют световые лучи, а плотности потока жидкости — вектор $\mathbf{j} = na^2\mathbf{s}$, пропорциональный плотности потока световой энергии. Свет как бы течет вдоль узких «световых трубок», т.е. трубок, боковые стенки которых образованы лучами. Через эти боковые стенки свет не проникает. Если σ — поперечное сечение трубки, то вдоль нее величина $na^2\sigma$ сохраняется неизменной, как это видно из уравнения (6.16).

Можно выделить отдельную световую трубку, или *физический световой луч*, поставив на пути распространяющейся волны (6.5) узкую диафрагму. Только диафрагма не должна быть особенно узкой, а световая трубка слишком длинной. Дело в том, что на краях диафрагмы и вблизи боковых границ трубки амплитуда поля меняется резко, т.е. условия применимости геометрической оптики не выполняются. Возникает дифракция света, приводящая к уширению светового пучка. Однако если диафрагма не слишком мала, а световая трубка не слишком длинна, эти эффекты малосущественны. Но они всегда скажутся на больших расстояниях от диафрагмы. В теории дифракции будет показано, что необходимым условием, при выполнении которого можно говорить о физическом световом луче, является неравенство

$$l \ll D^2/\lambda, \tag{6.17}$$

где D — минимальный линейный размер диафрагмы, а l — расстояние от диафрагмы, измеренное вдоль луча.

Оптической длиной линии в однородной среде называется произведение геометрической длины этой линии l на показатель преломления n. Если среда неоднородна, то оптическая длина определяется интегралом $\int n\,dl$, взятым вдоль рассматриваемой линии. Если ABC — геометрическая длина линии, то оптическая длина ее обозначается через (ABC), т.е. заключением в круглые скобки геометрической длины. Из построения Гюйгенса, изложенного в п. 3 этого параграфа, следует, что оптические длины всех лучей между двумя положениями волнового фронта равны между собой.

5. Нетрудно получить из волновых представлений и выражение для кривизны луча в однородной среде. Для этого через бесконечно малый отрезок луча AC проведем соприкасающуюся плоскость (рис. 20). Она пересечет волновые фронты, проходящие через концы этого отрезка, вдоль кривых AB и CD. Пусть BD — бесконечно близкий луч, лежащий в той же плоскости. Так как лучи перпендикулярны к волновым фронтам, то все углы бесконечно малого криволинейного четырехугольника $ABDC$ прямые. А так как оптические длины лучей между любыми двумя положениями волнового фронта одинаковы, то $nl = (n+dn)(l+dl)$, где l и $l+dl$ — длины отрезков AC и BD, а n и $n+dn$ — соответствующие им показатели преломления. С точностью до бесконечно малых высшего порядка отсюда получаем $l\,dn + n\,dl = 0$. По определению радиуса кривизны $l = R\varphi$, где φ — угол между касательными к лучу AC в точках A и C, равный углу между касательными к отрезкам AB и CD в тех же точках. Очевидно, $dl = -a\varphi$, где a — длина отрезка AB. Приращение показателя преломления dn происходит вдоль главной нормали $\mathbf{N}$, так что $dn = (\partial n/\partial N)a$. Учитывая все это, получаем: $R\varphi(\partial n/\partial N)a - na\varphi = 0$, откуда

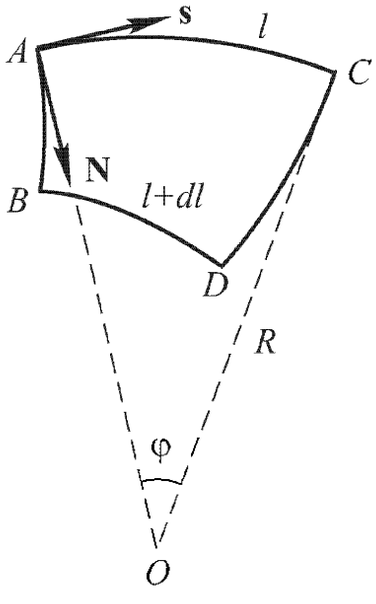


Рис. 20

$$\frac{1}{R} = \frac{1}{n}\frac{\partial n}{\partial N}, \tag{6.18}$$

что совпадает с формулой (4.1).

§ 7. Принцип Ферма

1. Пьер Ферма (1601–1675) выдвинул принцип, согласно которому *свет при распространении из одной точки в другую выбирает путь, которому соответствует наименьшее время распространения.* Ферма руководствовался телеологическими соображениями, согласно которым природа действует целенаправленно: она не может быть расточительной и должна достигать своих целей с наименьшей затратой средств. Подобные

соображения, конечно, чужды науке и не могут служить обоснованием принципа Ферма. Но сам принцип (после введения некоторых уточнений) верен и может оказаться полезным при решении отдельных вопросов геометрической оптики. Это было продемонстрировано уже самим Ферма, который с помощью своего принципа вывел закон преломления Снеллиуса и получил такое же выражение для показателя преломления, что и в волновой теории света. В частности, он пришел к заключению, что скорость света в более преломляющей среде меньше, чем в менее преломляющей.

2. Для доказательства принципа Ферма допустим сначала, что показатель преломления среды меняется в пространстве непрерывно и достаточно медленно, так что условия применимости геометрической оптики выполнены. Пусть в среде распространяется волна вида (6.5), например порожденная точечным источником. Ей соответствует система лучей, представленная на рис. 21. Если эйконал Φ — однозначная функция координат, то из уравнения (6.11) следует, что циркуляция вектора $n\mathbf{s}$ по любому замкнутому контуру равна нулю, т.е.

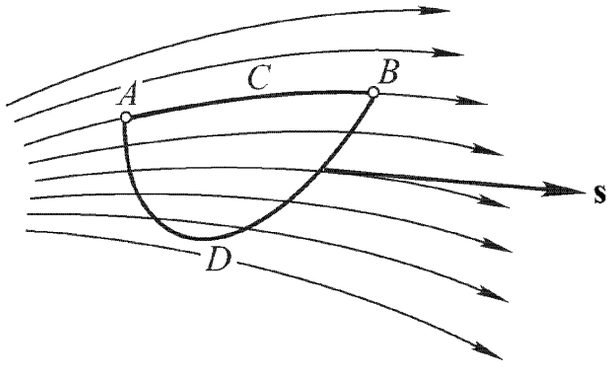


Рис. 21

$$\oint n(\mathbf{s}\,d\mathbf{l}) = 0, \qquad (7.1)$$

где $d\mathbf{l}$ — вектор элементарного смещения вдоль этого контура. Возьмем две произвольные точки A и B, лежащие на одном из лучей. Соединим их произвольной линией ADB. В силу (7.1)

$$\int\limits_{ACB} n(\mathbf{s}\,d\mathbf{l}) = \int\limits_{ADB} n(\mathbf{s}\,d\mathbf{l}).$$

На луче ACB векторы $\mathbf{s}$ и $d\mathbf{l}$ направлены одинаково, следовательно, $(\mathbf{s}\,d\mathbf{l}) = dl$. На линии же ADB $(\mathbf{s}\,d\mathbf{l}) = dl\cos(\mathbf{s}, d\mathbf{l}) \leqslant dl$. Поэтому

$$\int\limits_{ACB} n\,dl \leqslant \int\limits_{ADB} n\,dl. \qquad (7.2)$$

Знак равенства относится только к случаю, когда кривая ADB сама является лучом. Таким образом, *если показатель преломления меняется в пространстве непрерывно, то оптическая длина луча между любыми двумя точками меньше оптической длины всякой другой линии, соединяющей те же точки.* Но это есть другая формулировка принципа Ферма, так как оп-

тическая длина луча пропорциональна времени распространения света вдоль него.

Приведенная формулировка принципа Ферма нуждается в уточнении. В некоторых случаях она может оказаться неверной. Рассмотрим, например, среду со сферически симметричным распределением показателя преломления вокруг центра O (рис. 22). Примером такой среды может служить планетная атмосфера. Предположим, что показатель преломления меняется в пространстве так, что световой луч, выйдя из какой-либо точки перпендикулярно к радиусу, описывает окружность с центром в точке O. Пусть свет попадает из точки A в точку B по большой дуге ACB этой окружности. Но он может пройти из A в B и по дуге ADB той же окружности, затрачивая на распространение меньшее время. Меньшее время потребовалось бы и в том случае, если бы свет избрал какой-либо другой путь, бесконечно близкий к дуге ADB. Все это противоречит принципу Ферма в приведенной выше формулировке.

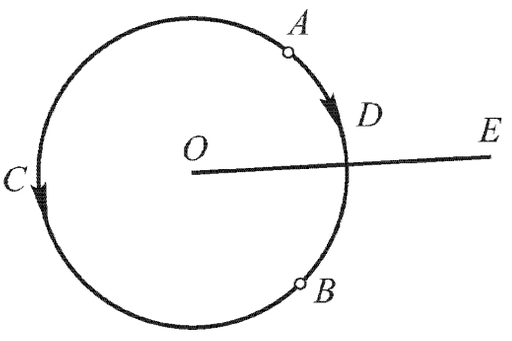


Рис. 22

Причина противоречия состоит в том, что в приведенном примере эйконал Ф не есть однозначная функция координат, как это предполагалось при выводе. Действительно, если луч описывает окружность вокруг центра O, то он вернется в исходную точку с новым значением эйконала: эйконал Ф получит приращение nl, где l — длина описанной окружности. Если окружность описывается m раз, то приращение эйконала будет $2mnl$. Это и значит, что функция Ф неоднозначна. Для справедливости принципа Ферма необходимо наложить на выбор воображаемых путей распространения света такие ограничения, чтобы эйконал Ф вел себя как *однозначная функция координат*. В приведенном примере этого можно достигнуть, поставив перегородку вдоль меридиональной полуплоскости ODE и ограничиваясь только такими путями, которые не пересекают эту перегородку.

Подобным приемом можно воспользоваться и во всех остальных случаях, в которых эйконал Ф оказывается неоднозначным. Впрочем, в применениях принципа Ферма достаточно ограничиться только такими путями, которые проходят бесконечно близко от действительного пути света. В этом случае надобность во введении перегородок отпадает.

3. При наличии поверхностей раздела сред, на которых лучи могут испытывать отражение или преломление, в формулировку и доказательство принципа Ферма надо ввести дополнения. Пусть луч, выйдя из точки A (рис. 23), после отражений или

преломлений в точках C, D, E, ... попадает в точку B. Назовем *виртуальным путем света* любую линию $AC'D'E'B$ между крайними точками A и B, которая получается из $ACDEB$ в результате бесконечно малого бокового смещения ее и отличается от нее бесконечно мало по направлению. Принцип Ферма утверждает, что *оптическая длина действительного светового пути (или пропорциональное ей время распространения) стационарна.*

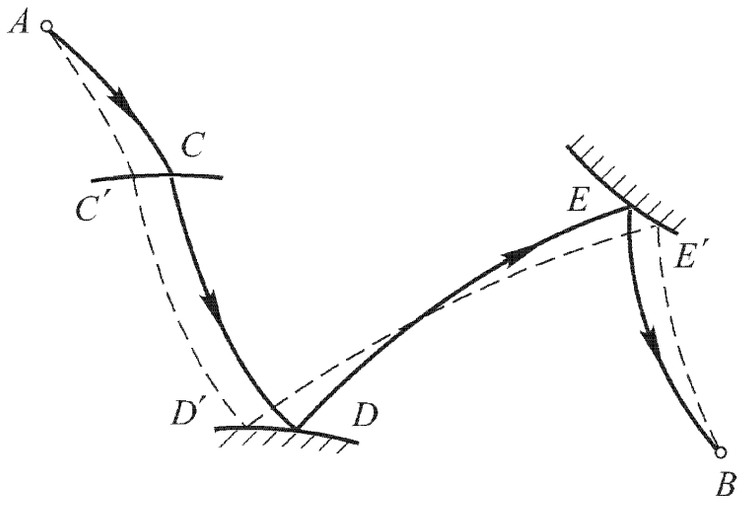


Рис. 23

Это значит, что разность оптических длин действительного и виртуального путей света есть величина *более высокого порядка малости*, чем боковое смещение виртуального пути относительно действительного. Только эта стационарность, а не минимальность оптической длины луча и существенна в приложениях.

При доказательстве достаточно ограничиться преломлением на одной границе. Случай отражения исследуется так же. Пусть MN — граница раздела сред *1* и *2*, а ACB — действительный луч, соединяющий точку A с точкой B (рис. 24). Вообразим два бесконечно узких пучка лучей: один в первой среде, исходящий из точки A, другой во второй среде, сходящийся в точке B. За положительные направления лучей примем направления от A к B. Выберем в этих пучках два луча AC' и $C'B$, пересекающихся на границе раздела в точке C'. Кривую $AC'B$ можно рассматривать как виртуальный путь света, так как луч $C'B$ в общем случае, отнюдь не возникает в результате преломления луча AC'. Обозначим через Φ_1 и Φ_2 эйконалы рассматриваемых пучков лучей, отсчитываемые от точек A и B соответственно. Тогда

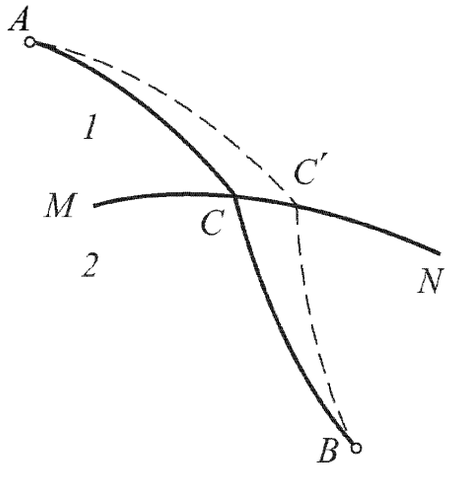


Рис. 24

$$\int\limits_{ACB} n\,ds = \int\limits_{AC} n\,ds + \int\limits_{CB} n\,ds = \int\limits_{AC} n\,ds - \int\limits_{BC} n\,ds = \Phi_1(C) - \Phi_2(C).$$

Вариация интеграла $\int n\,ds$ при смещении точки C в произвольную бесконечно близкую точку C' границы раздела будет

$$\delta \int n\,ds = \delta\Phi_1 - \delta\Phi_2.$$

Если $\delta\mathbf{r} \equiv \overrightarrow{CC'}$ — вектор смещения, то $\delta\Phi_1 = (\operatorname{grad}\Phi_1\delta\mathbf{r}) = n_1(\mathbf{s}_1\delta\mathbf{r})$ и аналогично $\delta\Phi_2 = n_2(\mathbf{s}_2\delta\mathbf{r})$, так что

$$\delta \int n\,ds = (n_1\mathbf{s}_1 - n_2\mathbf{s}_2)\delta\mathbf{r}.$$

В силу закона преломления Снеллиуса вектор $(n_1\mathbf{s}_1 - n_2\mathbf{s}_2)$ перпендикулярен к границе раздела сред в точке падения, а потому и к бесконечно малому смещению вдоль границы $\delta\mathbf{r}$. Таким образом, в первом порядке по δr вариация оптической длины луча ACB обращается в нуль. При доказательстве предполагалось, что виртуальный путь состоит из отрезков лучей AC' и $C'B$. Однако результат не изменится, если эти отрезки заменить произвольными бесконечно близкими к ним линиями, соединяющими те же точки A и C', C' и B. В самом деле, поскольку AC' и $C'B$ — действительные лучи в первой и второй средах, их оптические длины по доказанному выше минимальны. По этой причине замена действительных лучей AC' и $C'B$ бесконечно близкими к ним линиями, соединяющими те же крайние точки, не меняет в первом порядке оптические длины соответствующих путей. Следовательно, вариация оптической длины луча ACB останется равной нулю, каков бы ни был виртуальный путь света. А к этому в рассматриваемом случае и сводится содержание принципа Ферма.

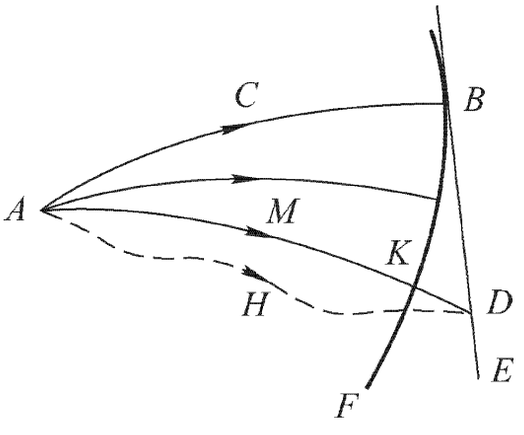


Рис. 25

4. В применениях иногда удобна следующая теорема, являющаяся непосредственным следствием принципа Ферма. *Пусть A и B — произвольные точки луча ACB* (рис. 25). *Проведем через точку B произвольную гладкую поверхность BE, ортогональную к лучу ACB в точке B. Пусть BD — бесконечно малое смещение вдоль этой поверхности. Соединим начальную точку луча A с точкой D произвольной линией AHD, бесконечно мало отличающейся по направлению от луча ACB. Тогда вариация оптической длины при переходе от истинного пути света ACB к виртуальному AHD будет равна нулю.*

Для доказательства возьмем пучок лучей, исходящих из точки A. Все эти лучи ортогональны к волновому фронту BF, а их

оптические длины от точки A до волнового фронта одинаковы. В частности, $(ACB) = (AMK)$. Но по принципу Ферма с точностью до бесконечно малых высшего порядка $(AMK) = (AHK)$. Далее, поскольку поверхности BDE и BKF касаются друг друга в точке B, длина луча KD будет бесконечно малой высшего порядка по сравнению с BD. Поэтому оптическая длина AHD будет отличаться от оптической длины ACB также на величину высшего порядка малости по сравнению с боковым смещением BD. Это и требовалось доказать.

5. Если свет распространяется в однородных средах, граничащих между собой, то в каждой среде путь света будет прямолинеен. В этом случае задача сводится только к нахождению точек на поверхностях раздела сред, в которых происходит отражение и преломление светового луча. Поэтому нет необходимости вводить криволинейные виртуальные пути света. Достаточно ограничиться ломаными виртуальными путями, состоящими из отрезков прямых линий, причем изломы таких путей должны происходить на границах раздела рассматриваемых сред. Даже при таких ограничениях оптическая длина действительного светового пути может быть не только минимальной, но и максимальной или стационарной.

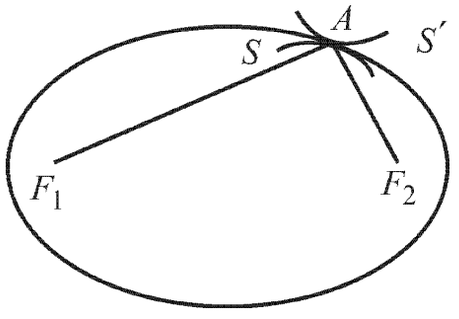


Рис. 26

Чтобы показать это в случае отражения света, возьмем эллипсоидальное зеркало, получающееся от вращения эллипса вокруг его большой оси F_1F_2 (рис. 26).

Пусть F_1 и F_2 — фокусы эллипсоида. Если A — точка на его поверхности, то $F_1A + F_2A = 2a$, где $2a$ — длина большой оси эллипсоида. Поверхность зеркала делит все пространство на две части: *внутреннюю*, сумма расстояний каждой точки которой от фокусов F_1 и F_2 меньше $2a$, и *внешнюю*, для которой эта сумма больше $2a$. Пусть световой луч выходит из фокуса F_1. Тогда после отражения от эллипсоидального зеркала в точке A он пройдет через второй фокус F_2, так как по известному свойству эллипса прямые F_1A и F_2A образуют одинаковые углы с нормалью к поверхности зеркала. При смещении вдоль поверхности зеркала сумма $F_1A + F_2A$, а с ней и время распространения света из F_1 в F_2 не изменяются. Вариация времени распространения при таком смещении равна нулю. Однако это время ни минимально, ни максимально — оно *постоянно*. Именно по этой причине любой луч, вышедший из F_1, обязательно пройдет через F_2, в какой бы точке зеркала он ни отразился. Убедиться в этом можно с помощью таких же рассуждений, какие были приведены в п. 3.

Вообразим теперь зеркало S, касающееся эллипсоида в точке A, обращенное вогнутостью в ту же сторону, что и эллипсоид, но имеющее бóльшую кривизну. Световой луч F_1A после отражения от этого зеркала снова попадает в точку F_2. Однако при смещении точки A по поверхности зеркала S длина ломаной F_1AF_2 уменьшается. Следовательно, время распространения света из F_1 в F_2 вдоль действительного пути *максимально*. Наоборот, если взять зеркало S', имеющее в точке касания меньшую кривизну, чем эллипсоид, или обращенное вогнутостью в противоположную сторону, то время распространения света вдоль действительного пути будет *минимально*. В частности, оно минимально при отражении от *плоского зеркала*. Допустим, наконец, что зеркало SAS' имеет в A *точку перегиба*. Тогда при смещении точки падения луча по поверхности этого зеркала время распространения либо увеличится, либо уменьшится, либо останется неизменным, в зависимости от направления смещения.

6. Чтобы разобрать случай преломления, введем понятие *анаберрационной поверхности*. Пусть точка P находится в однородной среде с показателем преломления n, а точка P' — в однородной среде с показателем преломления n' (рис. 27). Поверхность AA', вдоль которой среды граничат друг с другом, называется *анаберрационной*, если любая точка A этой поверхности удовлетворяет условию

$$n\cdot PA+n'\cdot AP' = C = \text{const}.$$

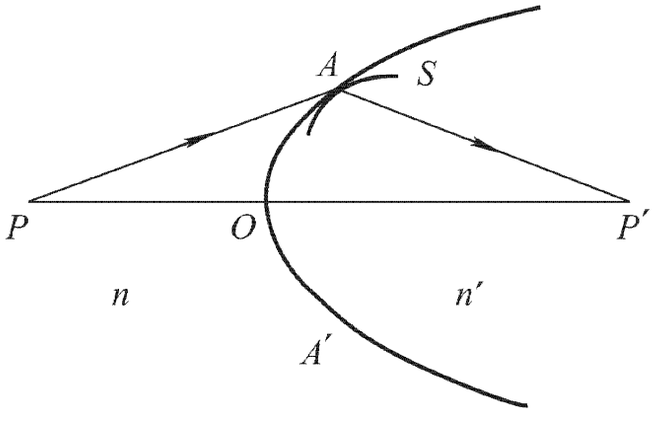


Рис. 27

Для случая преломления анаберрационная поверхность имеет форму так называемого *картезианского овала* (см. задачу 2 к § 9). Он обращен вогнутостью в сторону более преломляющей среды ($n' > n$). Анаберрационная поверхность делит пространство на две части, обладающие следующим свойством. Если точка M расположена в менее преломляющей среде, то сумма $n \cdot PM + n' \cdot MP'$ больше C; если же она лежит в более преломляющей среде, то эта сумма меньше C.

Докажем следующую теорему. *Луч света, вышедший из точки P, после преломления на анаберрационной поверхности обязательно пройдет через точку P'.* Действительно, пусть PA — падающий луч, а $\mathbf{s}$ — единичный вектор, направленный вдоль него. Соединим точку A с точкой P' и обозначим через $\mathbf{s}'$ единичный вектор, направленный вдоль прямой AP'. По определению анаберрационной поверхности вариация оптической дли-

ны ломаной PAP' при смещении точки A по анаберрационной поверхности будет равна нулю. Поэтому, применяя такие же рассуждения, какие были проведены в п. 2, найдем, что вектор $n\mathbf{s} - n'\mathbf{s}'$ перпендикулярен к анаберрационной поверхности в точке A. Отсюда следует, что AP' дает направление преломленного луча.

Доказанной теореме можно дать также следующую формулировку. *Если AA' — анаберрационная поверхность относительно пары точек P и P', то каждая из этих точек будет оптическим изображением другой при преломлении лучей на этой анаберрационной поверхности.* При этом на угловую ширину пучка лучей не накладывается никаких ограничений.

Вернемся к исследованию характера экстремума оптической длины луча при преломлении. Наши рассуждения ничем не будут отличаться от рассуждений, проведенных выше для эллипсоидального зеркала. Допустим, например, что среды граничат друг с другом вдоль поверхности S (рис. 27), касающейся анаберрационной поверхности в точке A. Тогда падающий луч после преломления в точке A опять пройдет через точку P'. Пусть поверхность S обращена вогнутостью в ту же сторону, что и анаберрационная поверхность, и имеет в точке касания бóльшую кривизну. Тогда при смещении точки падения вдоль S она окажется в менее преломляющей среде. Следовательно, смещенный путь будет иметь *меньшую оптическую длину*, чем действительный: время распространения света вдоль действительного пути *максимально*. Напротив, когда кривизна поверхности S в точке касания A меньше кривизны анаберрационной поверхности, а также тогда, когда поверхность S обращена вогнутостью в противоположную сторону, то время распространения вдоль действительного пути *минимально*. В частности, оно минимально при преломлении на *плоской поверхности*.

ЗАДАЧА

Система лучей называется *ортотомной*, если все лучи этой системы ортогональны к одной и той же поверхности. Пользуясь принципом Ферма, доказать *теорему Малюса: ортотомная система лучей остается ортотомной после произвольного числа отражений и преломлений.*

Р е ш е н и е. Пусть все лучи перпендикулярны к поверхности F (рис. 28). Проведем через каждую точку этой поверхности луч и отложим на нем отрезок постоянной (но произвольной) оптической длины L. Геометрическим местом концов таких отрезков будет какая-то поверхность F'. Докажем, что все лучи рассматриваемой системы перпендикулярны к поверхности F', каково бы ни было значение величины L.

Малые отрезки одного из лучей AC и $C'A'$ могут считаться прямолинейными. Возьмем соседний бесконечно близкий луч и притом такой, что длины AB и $A'B'$ бесконечно малы по сравнению с AC и $C'A'$. Соединим B с

C и C' с B прямолинейными отрезками. По принципу Ферма с точностью до бесконечно малых второго или высшего порядков $(BEB') = (BCC'B')$, а по построению $(BEB') = (ACC'A')$. Таким образом, $(BCC'B') = (ACC'A')$. Вычитая отсюда общую часть (CC'), получим: $(AC) + (C'A') = (BC) + (C'B')$. Так как по условию отрезок AC перпендикулярен к AB, то с точностью до второго порядка малости $AC = BC$, а следовательно, $(AC) = (BC)$. Поэтому с той же точностью $(C'A') = (C'B')$, или $C'A' = C'B'$, откуда следует, что $C'A' \perp A'B'$.

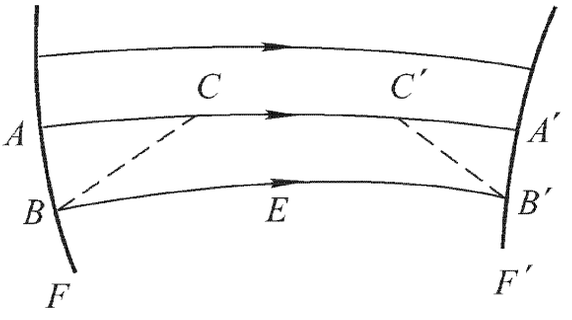


Рис. 28

С точки зрения волновой теории теорема Малюса почти самоочевидна. Действительно, для ортотомной системы лучей поверхность F есть одна из *поверхностей равной фазы* (волновой фронт). Распространяясь по законам геометрической оптики, она продолжает оставаться поверхностью равной фазы, а совокупность лучей — *ортогональной системой*. Конечно, ортогональность может и не соблюдаться. Например, волны вида (6.5) при соблюдении принципа суперпозиции распространяются независимо друг от друга. Каждой из таких волн соответствует ортотомная система лучей. Однако совокупность лучей, соответствующих *всем волнам*, ортотомную систему, вообще говоря, не образует.

§ 8. Групповая скорость

1. До сих пор при рассмотрении скорости распространения волн мы предполагали, что соблюдается принцип суперпозиции и отсутствует дисперсия. При несоблюдении принципа суперпозиции и наличии дисперсии вопрос о скорости распространения волн становится очень сложным. *Ниже предполагается, что принцип суперпозиции соблюдается, но имеется дисперсия.* Сначала рассмотрим плоские волны, распространяющиеся в одном направлении, принимаемом за направление оси X.

Бегущую плоскую монохроматическую волну запишем в виде

$$E = E_0 \cos(\omega t - kx + \delta), \tag{8.1}$$

где E_0 и δ — постоянные. При рассмотрении таких волн дисперсия не играет роли, поскольку частота ω имеет единственное значение. Для выяснения смысла скорости распространения рассмотрим уравнение

$$\omega t - kx + \delta = \text{const}. \tag{8.2}$$

Это есть уравнение плоскости, перпендикулярной к оси X, на которой постоянна фаза волны. Дифференцируя его, получим:

$\omega\,dt - k\,dx = 0$, откуда

$$\frac{dx}{dt} = \frac{\omega}{k}. \tag{8.3}$$

Таким образом, ω/k есть *скорость распространения поверхности постоянной фазы*, ранее обозначавшаяся через v. Она называется *фазовой скоростью волны*. С такой скоростью распространяется синусоидальная волна типа (8.1) без изменения своей формы.

Если бы среда не обладала дисперсией, то говорить о какой-либо другой скорости распространения волны не было бы необходимости. Действительно, произвольное плоское возмущение, распространяющееся в направлении оси X, согласно теореме Фурье можно представить в виде суперпозиции монохроматических волн вида (8.1). При отсутствии дисперсии все эти волны имели бы одну и ту же фазовую скорость, так что форма возмущения все время оставалась бы *одной и той же*. Возмущение в целом бежало бы вперед без изменения вида со скоростью, равной той же фазовой скорости. Не то будет при наличии дисперсии. Тогда монохроматические волны разных частот будут распространяться вперед с разными скоростями, вследствие чего форма всего возмущения будет *непрерывно деформироваться* (исключение составляют только монохроматические волны, распространяющиеся по-прежнему без изменения формы). В этих условиях понятие скорости распространения утрачивает тот ясный смысл, какой оно имело при отсутствии дисперсии. При определенных условиях, однако, можно сохранить представление о скорости распространения немонохроматических волн и в средах, обладающих дисперсией. Важнейшей после фазовой скорости является так называемая *групповая скорость*.

2. Рассмотрим сначала в действительности не существующую непоглощающую среду, в которой фазовая скорость v выражается линейной функцией длины волны λ:

$$v = a + b\lambda, \tag{8.4}$$

а следовательно, частота ω — линейной функцией волнового числа k:

$$\omega = ak + 2\pi b. \tag{8.5}$$

Произвольное плоское возмущение, распространяющееся в среде, разложим на монохроматические волны. Их число, вообще говоря, будет бесконечно велико. Однако можно ограничиться случаем, когда оно равно трем. Это, как будет видно из дальнейшего, не отразится на общности рассуждений и результатов. На рис. 29 представлены эти три синусоиды в какой-то момент времени. Форма результирующего возмущения зависит от их взаимного расположения. Не нарушая общности, можно

принять, что в этот момент какие-то три гребня синусоид A, B, C пространственно совпадают друг с другом.

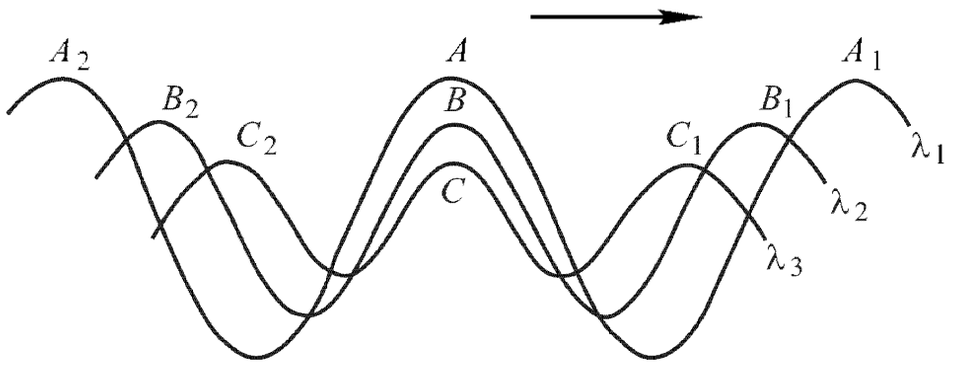


Рис. 29

Допустим ради определенности, что фазовая скорость v возрастает с возрастанием длины волны λ (если предположить противоположное, то рассуждения и окончательный результат не изменятся). Таким образом, если $\lambda_1 > \lambda_2 > \lambda_3$, а v_1, v_2, v_3 — соответствующие фазовые скорости, то $v_1 > v_2 > v_3$. При распространении, например, вправо более длинные волны будут обгонять более короткие, что приведет к непрерывному изменению формы результирующего возмущения. Гребни A, B, C начнут расходиться, гребни A_1, B_1, C_1 разойдутся еще больше, а гребни A_2, B_2, C_2 будут сближаться. Если начало координат поместить в точку, в которой находились гребни A, B, C в начальный момент, то координаты гребней A_2, B_2, C_2 в произвольный момент времени t представятся выражениями

$$x_{A_2}(t) = v_1 t - \lambda_1, \quad x_{B_2}(t) = v_2 t - \lambda_2, \quad x_{C_2}(t) = v_3 t - \lambda_3.$$

В момент времени τ, определяемый условиями

$$v_1\tau - \lambda_1 = v_2\tau - \lambda_2 = v_3\tau - \lambda_3,$$

т.е.

$$\tau = \frac{\lambda_1 - \lambda_2}{v_1 - v_2} = \frac{\lambda_2 - \lambda_3}{v_2 - v_3} = \frac{\lambda_1 - \lambda_3}{v_1 - v_3} = \frac{d\lambda}{dv} = \frac{1}{b},$$

гребни A_2, B_2, C_2 пространственно совпадут, так что первоначальное взаимное расположение синусоид, а с ним и форма всего возмущения восстановятся. Только роль гребней A, B, C перейдет к гребням A_2, B_2, C_2, что, очевидно, никак не может отразиться на форме результирующего возмущения. Полученный результат при линейном законе дисперсии (8.4) имеет общее значение, т.е. справедлив для произвольного числа синусоид и произвольного начального расположения их.

Таким образом, *по истечении времени* $\tau = d\lambda/dv$, *называемого временем восстановления, происходит периодическое восстановление формы возмущения.* Например, в рассмотренном

нами случае трех синусоид в начале координат, где накладывались гребни A, B, C, в начальный момент времени был максимум возмущения. В точности такой же максимум появился через время τ в другом месте пространства, где наложились гребни A_2, B_2, C_2. Распространение возмущения носит как бы прыгающий характер, причем от прыжка к прыжку проходит время τ. Естественно определить скорость возмущения как отношение расстояния, проходимого возмущением за один «прыжок», к промежутку времени между последовательными прыжками. Так определенная величина называется *групповой скоростью* возмущения. В разобранном примере это будет отношение расстояния между двумя последовательными положениями максимума возмущения к времени восстановления τ. В момент времени $t = 0$ координата максимума $x_{\text{макс}}(0) = 0$. В момент τ координата такого же максимума будет

$$x_{\text{макс}}(\tau) = v_1\tau - \lambda_1 = v_2\tau - \lambda_2 = v_3\tau - \lambda_3 = v\tau - \lambda.$$

За время τ максимум проходит путь

$$x_{\text{макс}}(\tau) - x_{\text{макс}}(0) = v\tau - \lambda.$$

Следовательно, групповая скорость будет $u = v - \lambda/\tau$, или

$$u = v - \lambda \frac{dv}{d\lambda}. \tag{8.6}$$

Эта формула впервые была получена Рэлеем (1842–1919) и носит его имя. На рис. 30 приведена графическая интерпретация этой формулы, принадлежащая П.С. Эренфесту (1880–1933). На нем в координатах λ, v представлен график AB прямой (8.4). Так как $u = v - \lambda\, dv/d\lambda = a$, то эта прямая отсекает на оси ординат отрезок BO, длина которого равна групповой скорости u. Формулу (8.6) можно записать в виде

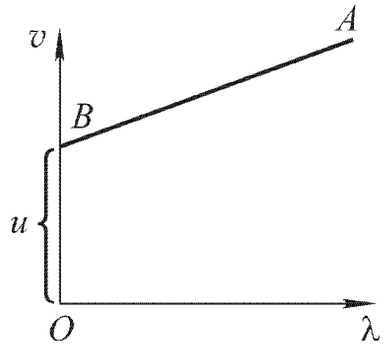


Рис. 30

$$u = v + \frac{1}{\lambda}\frac{dv}{d(1/\lambda)} = \frac{d}{d(1/\lambda)}\left(\frac{v}{\lambda}\right),$$

или

$$u = \frac{d\omega}{dk}. \tag{8.7}$$

Легко также преобразовать (8.6) к виду

$$u = \frac{c}{n}\left(1 + \frac{\lambda}{n}\frac{dn}{d\lambda}\right). \tag{8.8}$$

3. Полученные результаты строго справедливы при линейном законе дисперсии (8.4) (или (8.5)). Однако если возмущение занимает небольшую спектральную область, то эти результаты остаются приближенно верными и в диспергирующих непоглощающих средах. Возмущение такого типа называется *группой*

волн. Точнее, *группой волн называется волновое образование, занимающее столь узкую спектральную область, что в пределах этой области приращение фазовой скорости v с достаточной точностью может считаться пропорциональным соответствующему приращению длины волны λ, а следовательно, приращение частоты ω — пропорциональным соответствующему приращению волнового числа k.* Это значит, что в пределах рассматриваемой спектральной области обе зависимости $v = v(\lambda)$ и $\omega = \omega(k)$ могут быть аппроксимированы линейными функциями λ и k, а именно

$$v = v(\lambda_0) + \left(\frac{dv}{d\lambda}\right)_{\lambda=\lambda_0} (\lambda - \lambda_0), \tag{8.9}$$

$$\omega = \omega(k_0) + \left(\frac{d\omega}{dk}\right)_{k=k_0} (k - k_0), \tag{8.10}$$

где λ_0 — какая-то длина волны, лежащая внутри спектральной области, занимаемой группой, а $k_0 = 2\pi/\lambda_0$ — соответствующее ей волновое число. При допустимости такой аппроксимации можно говорить и о времени приближенного восстановления формы возмущения $\tau = d\lambda/dv$, и о распространении возмущения с групповой скоростью u, определяемой выражением (8.6) или (8.7). На диаграмме Эренфеста в случае группы волн играет роль только малый участок кривой $u = v(\lambda)$, который можно приближенно считать прямолинейным и заменить соответствующим отрезком касательной. Групповая скорость u представится длиной отрезка OB, отсекаемого этой касательной на оси ординат (рис. 31).

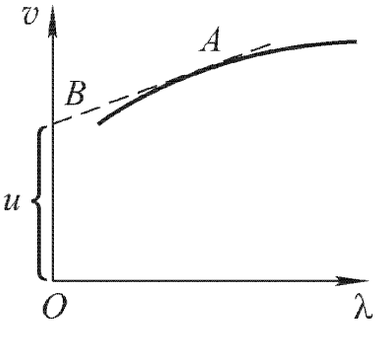


Рис. 31

4. Учет высших членов в разложении (8.9) или (8.10) приводит к следующему характеру распространения возмущения. Возмущение идет вперед, но его форма непрерывно изменяется. Однако по истечении времени $\tau = d\lambda/dv$ возмущение принимает форму, почти совпадающую с исходной, причем за это время оно продвигается вперед на расстояние $x = u\tau$. Можно сказать, что происходит передача энергии возмущения с групповой скоростью u. По истечении последующего промежутка времени той же длительности произойдет то же самое, и т.д. Вообще, в любой момент $t + \tau$ возмущение воспроизводит с малыми искажениями свою форму, какую оно имело в момент t, перемещаясь за время τ на расстояние $u\tau$. Но если возмущение распространяется достаточно долго, то малые изменения, претерпеваемые им за следующие друг за другом равные промежутки времени длительностью τ, будут накапливаться и могут настолько силь-

но исказить само возмущение, что его форма потеряет всякое сходство с исходной.

Чтобы оценить требующееся для этого время, дополним разложение (8.10) членом второй степени по $(k - k_0)$. С учетом формулы (8.7) можно написать:

$$\omega = \omega_0 + u_0(k - k_0) + \frac{1}{2}\left(\frac{du}{dk}\right)_0 (k - k_0)^2,$$

где нулем обозначены значения соответствующих величин при $k = k_0$. Пусть δk — максимальное значение разности $k - k_0$. Тогда соответствующее максимальное изменение фазы, обусловленное наличием квадратичного члена, будет $\frac{t}{2}\left|\frac{du}{dk}\right|(\delta k)^2$. Если это изменение мало по сравнению с величиной порядка π, то оно мало скажется на относительной разности фаз между синусоидами, входящими в группу, и тогда форма возмущения будет мало искажаться наличием квадратичного члена. Таким образом, чтобы на интервале времени t происходило периодическое восстановление исходной формы возмущения, необходимо выполнение условия

$$t \ll \frac{2\pi}{|du/dk|(\delta k)^2}. \tag{8.11}$$

Если перейти к длинам волн, то это условие преобразуется к виду

$$t \ll \frac{\lambda^2}{|du/d\lambda|(\delta\lambda)^2}. \tag{8.12}$$

Если же интервал времени t порядка или больше правой части этого неравенства, то о восстановлении исходной формы возмущения говорить не приходится.

5. Выше предполагалось, что плоские монохроматические волны, входящие в волновое образование, распространяются *в одном и том же направлении*. Рассмотрим теперь случай, когда такие волны занимают по-прежнему узкую область частот, но распространяются в разных направлениях, лежащих в пределах узкого конуса. Соответствующее волновое образование называется *волновым пакетом*. Его можно представить тройным интегралом

$$\mathbf{E}(\mathbf{r}, t) = \int\limits_{k_{0x}-\Delta k_x}^{k_{0x}+\Delta k_x} \int\limits_{k_{0y}-\Delta k_y}^{k_{0y}+\Delta k_y} \int\limits_{k_{0z}-\Delta k_z}^{k_{0z}+\Delta k_z} \mathbf{a}(\mathbf{k}) e^{i(\omega t - \mathbf{kr})}\, dk_x\, dk_y\, dk_z,$$

или сокращенно

$$\mathbf{E}(\mathbf{r}, t) = \int \mathbf{a}(\mathbf{k}) e^{i(\omega t - \mathbf{kr})}\, d\mathbf{k}. \tag{8.13}$$

Частоту ω следует рассматривать как функцию волнового вектора $\mathbf{k}$. Видом этой функции определяется *закон дисперсии*

волн. Если среда *изотропна*, то функция $\omega(\mathbf{k})$ может зависеть только от *длины вектора* $\mathbf{k}$, но не от его направления. Но в *анизотропных средах*, например кристаллах, необходимо учитывать и зависимость ω *от направления* $\mathbf{k}$. Поэтому ниже вид функции $\omega(\mathbf{k})$ не конкретизируется, а рассуждения проводятся в общем виде. Полагая $\omega = \omega_0 + \Delta\omega$, $\mathbf{k} = \mathbf{k}_0 + \Delta\mathbf{k}$, аппроксимируем $\Delta\omega$ линейным выражением

$$\Delta\omega = \frac{\partial\omega}{\partial k_x}\Delta k_x + \frac{\partial\omega}{\partial k_y}\Delta k_y + \frac{\partial\omega}{\partial k_z}\Delta k_z = (\mathbf{u}\,\Delta\mathbf{k}), \tag{8.14}$$

где через $\mathbf{u}$ обозначен вектор с компонентами

$$u_x = \frac{\partial\omega}{\partial k_x}, \quad u_y = \frac{\partial\omega}{\partial k_y}, \quad u_z = \frac{\partial\omega}{\partial k_z}. \tag{8.15}$$

Сокращенно его записывают в символической форме:

$$\mathbf{u} = \frac{\partial\omega}{\partial\mathbf{k}}. \tag{8.16}$$

После подстановки соответствующих значений в выражение (8.13) оно преобразуется в

$$\mathbf{E} = \mathbf{A}(\mathbf{r}, t)e^{i(\omega_0 t - \mathbf{k}_0\mathbf{r})}, \tag{8.17}$$

где введено обозначение

$$\mathbf{A}(\mathbf{r}, t) = \int \mathbf{a}(\mathbf{k})e^{i(\Delta\omega t - \Delta\mathbf{k}\mathbf{r})}\,d\mathbf{k} = \int \mathbf{a}(\mathbf{k})e^{i(\mathbf{u}t - \mathbf{r})\Delta\mathbf{k}}\,d\mathbf{k}.$$

Отсюда видно, что в точке M, движущейся со скоростью $\mathbf{u}$ по закону $\mathbf{r} = \mathbf{u}t + \mathbf{r}_0$ ($\mathbf{r}_0 = \text{const}$), амплитуда $\mathbf{A}$ остается постоянной. В такой точке совершаются гармонические колебания

$$\mathbf{E} = \mathbf{A}e^{i(\omega_0 t - \mathbf{k}_0\mathbf{r})} = \mathbf{A}e^{i[(\omega_0 - \mathbf{k}_0\mathbf{u})t - \mathbf{k}_0\mathbf{r}_0]}$$

с частотой $\omega_0 - \mathbf{k}_0\mathbf{u}$. По истечении времени τ фаза этих колебаний изменяется на $(\omega_0 - \mathbf{k}_0\mathbf{u})\tau$. Если это изменение равно 2π, т.е.

$$\tau = \frac{2\pi}{\omega_0 - \mathbf{k}_0\mathbf{u}} \approx \frac{2\pi}{\omega - \mathbf{k}\mathbf{u}}, \tag{8.18}$$

то вектор $\mathbf{E}$ в движущейся точке M в любой момент времени t примет то же значение, какое он имел в более ранний момент $t - \tau$. Так как это справедливо при любом значении параметра $\mathbf{r}_0$, то происходит периодическое восстановление формы возмущения с периодом τ, причем за это время возмущение перемещается вперед на расстояние $\mathbf{u}\tau$. В результате мы снова приходим к представлению о распространении возмущения с групповой скоростью $\mathbf{u}$, определяемой выражением (8.16).

В изотропных средах векторы $\mathbf{u}$ и $\mathbf{k}$ параллельны. В этом случае (8.18) легко преобразовать к прежнему виду $\tau = d\lambda/dv$. Однако в анизотропных средах векторы $\mathbf{u}$ и $\mathbf{k}$, вообще говоря,

не параллельны, и надо пользоваться более общими выражениями (8.16) и (8.18). Конечно, и здесь из-за наличия членов высших степеней, отброшенных в разложении (8.14), за время τ происходит не точное, а лишь *приближенное* восстановление формы возмущения. За это время возмущение претерпевает малые, может быть едва заметные, искажения. Но на больших интервалах времени эти искажения накапливаются и исходная форма возмущения сможет претерпеть существенные изменения.

6. Изложенные соображения позволяют легко решить вопрос о *скорости движения энергии* или *светового сигнала* в диспергирующих средах в тех областях спектра, в которых применимо понятие групповой скорости (т.е. вдали от полос поглощения). Прежде всего заметим, что фазовая скорость не имеет ничего общего со скоростью движения энергии. Фазовой скоростью устанавливается только связь *между фазами колебаний* в различных точках пространства. Связь такого типа в принципе может существовать и без передачи энергии, как это видно из следующего примера.

Вообразим длинную цепочку спортсменов, расположенных вдоль прямой линии на равных расстояниях друг от друга. Пусть они выполняют одно и то же гимнастическое упражнение, например периодическое движение руками, и притом так, что каждый впереди стоящий спортсмен начинает движение с некоторым запаздыванием по отношению к спортсмену, стоящему за ним. Пусть время запаздывания одно и то же для всех спортсменов. При наблюдении со стороны будет казаться, что по цепочке бежит волна с определенной фазовой скоростью, значение которой зависит от расстояния между соседними спортсменами и от времени запаздывания, о котором говорилось выше. Наличие такой волны, конечно, не означает, что каждый спортсмен приводит в движение впереди стоящего спортсмена. Так и возможность распространения в среде плоской монохроматической волны еще не дает оснований для заключения о переносе энергии с фазовой скоростью.

Строго плоская монохроматическая волна непригодна для наблюдения передачи энергии, поскольку она не имеет ни начала, ни конца во времени и в пространстве. Сама постановка вопроса о передаче энергии требует отказа от такой идеализации. Необходимо перейти к волновому возмущению, *ограниченному в пространстве* по крайней мере с одного конца, т.е. имеющему передовой фронт, перед которым возмущение отсутствует. Подходящим волновым образованием может служить группа волн. Если выполнено условие (8.12), то средняя скорость энергии, переносимой группой, совпадает с групповой скоростью. Действительно, форма группы, какую она имела в момент t, восстанавливается без заметного искажения в более поздний момент

времени $t + \tau$. При этом группа вместе с локализованной в ней энергией за время τ перемещается вперед на расстояние $x = u\tau$. Так как такое восстановление формы имеет место, каков бы ни был момент времени t, то движение энергии с групповой скоростью будет происходить на протяжении как угодно длинного промежутка времени, даже если за это время группа существенно изменит свою форму.

Итак, в области, далекой от области сильного поглощения, скорость движения энергии в группе волн совпадает с групповой скоростью. То же самое приближенно справедливо и для скорости движения энергии в волновом возмущении, занимающем сравнительно широкую спектральную область, если только в пределах этой спектральной области групповая скорость $u = u(\lambda)$ меняется мало. Если ширина спектральной области $\delta\lambda$, занимаемой группой, стремится к нулю, то группа в пределе переходит в монохроматическую волну. Можно поэтому сказать, что средняя скорость переноса энергии в монохроматической волне совпадает с групповой скоростью. Это утверждение следует понимать именно в приведенном смысле, рассматривая монохроматическую волну как предельный случай квазимонохроматической. Нельзя ограничиться идеализированной плоской строго монохроматической волной, отвлекаясь от представления ее как предельного случая квазимонохроматической волны. При такой абстрактной постановке вопроса утрачивается связь с реальными явлениями, а потому с точки зрения физики она бессмысленна.

Прямые измерения скорости света сводятся к измерению расстояния, проходимого световым сигналом за определенный промежуток времени. Из изложенного выше следует, что этот метод практически дает групповую скорость. То же самое, как показывает подробный анализ, относится ко всем известным косвенным методам измерения скорости света. Фазовую скорость, точнее — отношение фазовых скоростей в двух различных средах, можно определить по отношению показателей преломления, используя формулу волновой теории (3.7), в которую входят фазовые скорости света в рассматриваемых средах (см. § 64).

7. Остановимся еще на вопросе о *скорости распространения передового фронта волнового возмущения*. Речь идет о волне, резко ограниченной передовым фронтом, перед которым никакого возмущения нет. Скорость такого фронта точно совпадает со скоростью света в вакууме c. В этом легко убедиться, исходя из основных представлений электронной теории. Согласно этой теории, всякую среду следует рассматривать как вакуум, в который вкраплены молекулы и атомы вещества. Свет распространяется в вакууме между атомами и молекулами вещества, т.е. всегда со скоростью c. Когда световое возмущение дости-

гает какого-либо атома, электроны и атомные ядра приходят в колебания и сами становятся центрами излучения новых электромагнитных волн. Эти вторичные волны накладываются на первичную волну и тем самым определяют все волновое поле в среде. Но из-за инерции электроны и ядра не сразу приходят в колебания. Пока электроны и ядра не пришли в колебания, они не излучают вторичные волны, а потому не оказывают влияния на распространение возмущения. Поэтому ясно, что передовой фронт должен распространяться в среде с той же скоростью, что и в вакууме. Но почему же при измерении скорости света получается не c, а другая величина? Дело в том, что передовой фронт несет слишком малую энергию, а приемники света недостаточно чувствительны, чтобы ее обнаружить. Количественные расчеты, выполненные впервые Зоммерфельдом (1868–1951) и более подробно Л. Бриллюэном (1889–1969), показали, что это действительно так.

ГЛАВА II

ГЕОМЕТРИЧЕСКАЯ ТЕОРИЯ ОПТИЧЕСКИХ ИЗОБРАЖЕНИЙ

§ 9. Понятие оптического изображения

1. Если пучок световых лучей, исходящий из какой-либо точки P, в результате отражений, преломлений или изгибаний в неоднородной среде сходится в точке P', то P' называется *оптическим изображением* или просто *изображением точки* P. Точку P' называют также *фокусом геометрического схождения лучей*. Изображение P' называется *действительным*, если световые лучи действительно пересекаются в точке P'. Если же в P' пересекаются продолжения лучей, проведенные в направлении, обратном распространению света, то изображение называется *мнимым*. При помощи оптических приспособлений мнимые изображения могут быть преобразованы в действительные. Например, в нашем глазу мнимое изображение преобразуется в действительное, получающееся на сетчатке глаза.

Если в некоторый момент времени изменить на противоположное направление магнитного или электрического вектора, то, согласно *принципу обратимости* (см. т. III, § 83, п. 8), форма лучей остается без изменения, но направление распространения света изменится на противоположное. Точка P' будет играть роль *источника света*, а P — *его изображения*. Поэтому P и P' называются *сопряженными* или *взаимно сопряженными точками*. Аналогично, две линии или две поверхности называются *сопряженными*, если одна из них является оптическим изображением другой.

Если желают подчеркнуть, что лучи строго пересекаются в точке P', то изображение называют *стигматическим*. Пучок же лучей, исходящих из одной точки или сходящихся в одной точке, называется *гомоцентрическим*. Примером может служить отражение от *эллипсоидального зеркала*. По свойству эллипсоида

вращения прямые FA и $F'A$ (рис. 32), соединяющие его фокусы F и F' с произвольной точкой A поверхности эллипсоида, наклонены под одинаковыми углами к этой поверхности. Поэтому все лучи, вышедшие из одного фокуса, после отражения от поверхности эллипсоида пересекутся в другом фокусе. Практически более важен случай *параболоидального зеркала*, используемого в астрономических телескопах-рефлекторах. Параболоидальное зеркало есть частный случай эллипсоидального, когда один из фокусов F' удален в бесконечность. В силу известного свойства параболы все лучи, параллельные оси параболоида, после отражения от его вогнутой поверхности пересекутся в фокусе параболоида F (рис. 33). Если же такие лучи отразятся от выпуклой поверхности параболоида, то в фокусе F пересекутся продолжения отраженных лучей (рис. 34). Стигматическое изображение точки P в виде точки P' получается также при преломлении на поверхности, являющейся *анаберрационной* для этой пары точек (см. § 7 и задачи к настоящему параграфу).

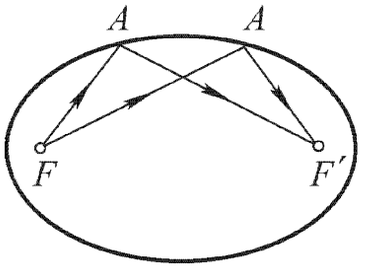


Рис. 32

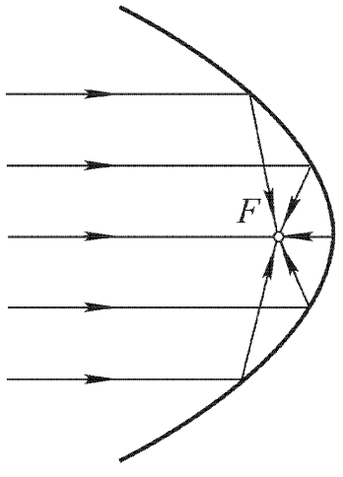


Рис. 33

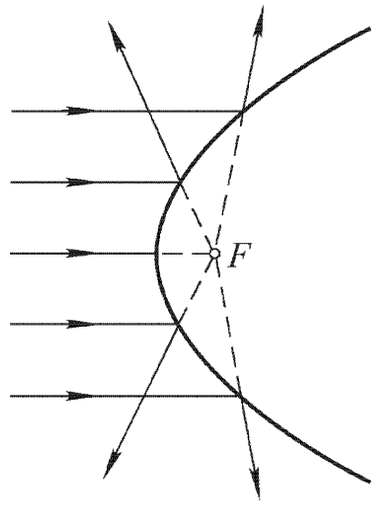


Рис. 34

На практике случаи стигматических изображений, как правило, бывают исключениями. Обычно лучи пересекаются не строго в одной точке, а в некоторой окрестности ее. Изображением светящейся точки на экране будет в этих случаях не математическая точка, а *светлое пятнышко*. Это снижает качество изображения. Однако строго точечное изображение светящейся точки не получается даже в тех случаях, когда по законам геометрической оптики лучи должны точно пересекаться в одной точке. Из-за дифракции света изображение светящейся точки

получается в виде светлого кружка, окруженного темными и светлыми кольцами.

Полная — *физическая* — теория оптических изображений должна учитывать *волновые свойства света*. Но начинать с такой теории, из-за ее сложности, было бы нецелесообразно. Сначала надо уяснить получение изображений с простейшей — *геометрической* — точки зрения, а затем ввести поправки, учитывающие волновую природу света. Геометрическая теория строится на основе одних только законов отражения и преломления света, полностью отвлекаясь от его физической природы. Таким путем, конечно, нельзя установить границы применимости геометрической теории. Это можно сделать только на основе волновой теории. Но значение последней этим не исчерпывается. Более существенно, что волновая теория позволяет определить, *что́ принципиально возможно в оптике*, и указать, как этого достигнуть.

2. Непрерывная совокупность точек, изображаемых оптической системой, называется *пространством предметов*. Непрерывная же совокупность точек, являющихся их изображениями, называется *пространством изображений*. Одна и та же точка может принадлежать к пространству предметов и к пространству изображений, в зависимости от того, рассматривается ли она как *предмет* или как *изображение*. Показатель преломления пространства предметов будем обозначать через n, а пространства изображений — через n'. В оптических приборах световой оптики величины n и n' всегда постоянны, т.е. не меняются от точки к точке. Когда изображение действительное, под n' следует понимать показатель преломления среды *в той точке, где получилось это изображение*. В случае мнимого изображения n' не всегда совпадает с показателем преломления среды в месте нахождения точки P'. Значение n' относится *к той среде, через которую проходят действительные лучи, продолжения которых пересекаются в точке* P'.

Мы будем исследовать и такие случаи, когда показатель преломления среды меняется *непрерывно* от точки к точке, а потому лучи *криволинейны*. Такой случай практически реализуется в *электронной оптике*. Здесь роль линз выполняют *электрические и магнитные поля*, а показателя преломления — *скорость электрона* (см. § 4).

3. С математической точки зрения задача геометрической теории оптических изображений сводится к определению положения изображения при любом заданном положении предмета. При этом общие свойства оптических систем удобно исследовать с помощью следующего положения. *Оптические длины всех лучей, соединяющих сопряженные точки* P *и* P', *одинаковы*. Это непосредственно очевидно, когда изображение P' действитель-

ное, так как тогда сферическая волна, вышедшая из P, превращается в сферическую волну, сходящуюся в P'. *Оптические же длины всех лучей от одного положения волнового фронта до другого одинаковы.* Но это положение можно распространить и на *мнимые изображения.* В этом случае не существует лучей, соединяющих P с P'. Роль луча играет прямолинейное продолжение его в сторону изображения P'. По аналогии с мнимым изображением такое продолжение можно назвать *мнимым лучом.*

Оптическую длину луча следует считать *положительной*, когда он проходится в направлении распространения света, и *отрицательной* в противоположном случае. Чтобы в случае мнимых изображений избежать неопределенности, будем предполагать, что пространство изображений *однородно*, т.е. световые лучи в нем *прямолинейны.* Это не значит, что изображение P' должно обязательно получаться в том месте, где среда однородна. Требуется только, чтобы действительные световые лучи, продолжения которых сходятся в P', были *прямолинейны.*

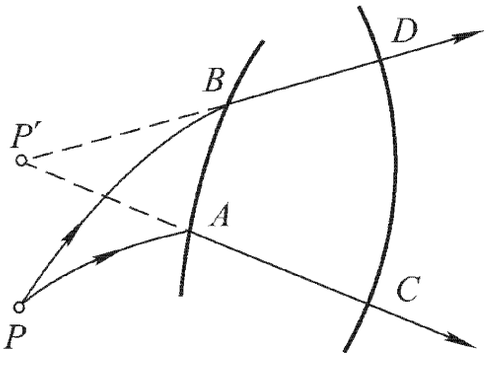


Рис. 35

После этих замечаний обратимся к доказательству нашего утверждения. Пусть лучи PAC и PBD (рис. 35), вышедшие из точки P, на участках AC и BD прямолинейны. Их продолжения пересекаются в точке P', являющейся мнимым изображением точки P. Волновой фронт в однородном пространстве изображений будет иметь форму сферы CD с центром в P'. Очевидно,

$$(PAC) = (PBD), \quad (P'AC) = (P'BD).$$

Почленное вычитание дает

$$(PA) - (P'A) = (PB) - (P'B).$$

Но, согласно нашему правилу знаков,

$$(PA) - (P'A) = (PA) + (AP') = (PAP'),$$

$$(PB) - (P'B) = (PB) + (BP') = (PBP').$$

Следовательно, $(PAP') = (PBP')$, что и требовалось доказать.

Доказанное свойство оптических длин эквивалентно утверждению, что *свет затрачивает одно и то же время, распространяясь вдоль различных лучей от точечного источника до его изображения.* В таком виде это утверждение называется *принципом таутохронизма* (равенства времен распростране-

ния). Принципом таутохронизма мы воспользуемся при изучении явлений интерференции.

Наряду с мнимыми изображениями, следует ввести и *мнимые источники света*, или *мнимые объекты*. Точечный объект называется *мнимым*, если он является точкой пересечения продолжений действительных лучей, проведенных в обратных направлениях. Мнимый объект можно рассматривать как источник *мнимых лучей*. Из множества точечных мнимых объектов составляются мнимые объекты конечных размеров.

Введение мнимых объектов и мнимых лучей освобождает теорию от необходимости раздельного рассмотрения действительных и мнимых изображений. Отпадает также необходимость в раздельном рассмотрении преломления и отражения света, что имеет большое значение в теории оптических систем, содержащих большое количество преломляющих и отражающих поверхностей. Действительно, пусть P' — мнимое изображение точки P, полученное в результате отражения света от зеркала (рис. 36). Согласно принятому нами правилу знаков, оптическая длина мнимого луча AP' отрицательна. Поэтому для оптической длины пути PAP' можно написать:

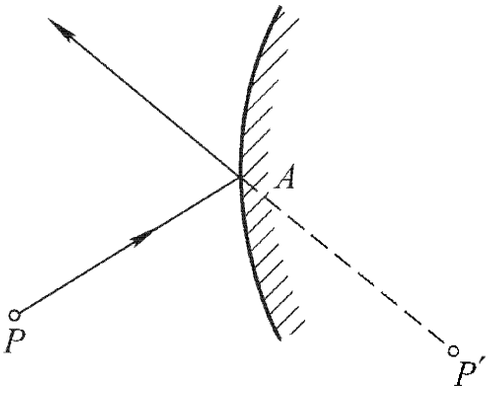


Рис. 36

$$(PAP') = n|PA| - n|AP'| = n|PA| + n'|AP'|,$$

где $n' = -n$. Поэтому *отражение формально математически можно рассматривать как преломление, если только показателю преломления n' приписать отрицательное значение* $(n'/n = -1)$.

ЗАДАЧИ

1. Две однородные среды с показателями преломления n и n' граничат друг с другом вдоль поверхности S (рис. 37), являющейся поверхностью вращения вокруг оси OP' (оптической оси). Найти форму поверхности S, при которой она будет анаберрационной для пары точек P и P', лежащих на оптической оси, из которых точка P удалена в бесконечность, а P' может занимать любое положение на оптической оси.

Р е ш е н и е. Примем оптическую ось за координатную ось X, начало координат поместим в точке пересечения ее с поверхностью S, ось Y направим вверх перпендикулярно к оптической оси. Так как оптические длины лучей от бесконечно удаленной точки P до плоскости OA одинаковы, то условие анаберрационности поверхности S будет $(ABP') = (OP')$, или

$$nx + n'\sqrt{(x-q)^2 + y^2} = n'q,$$

где x и y — текущие координаты точек поверхности S, а q — абсцисса точки P'. Перенеся nx в правую часть и возведя в квадрат, находим уравнение искомой поверхности:

$$(n'^2 - n^2)x^2 + n'^2y^2 - 2n'(n' - n)qx = 0. \tag{9.1}$$

Допустим сначала, что $n'^2 - n^2 > 0$. Тогда уравнение (9.1) представляет эллипсоид вращения с полуосями

$$a = \frac{n'}{n' + n}q, \quad b = \sqrt{\frac{n' - n}{n' + n}}\,q.$$

Эллипсоид вытянут в направлении оси X. Его эксцентриситет равен

$$e = \frac{\sqrt{a^2 - b^2}}{a} = \frac{n}{n'} < 1.$$

Изображение P' действительное.

Пусть теперь $n'^2 - n^2 < 0$. Тогда (9.1) есть уравнение двуполостного гиперболоида вращения с полуосями

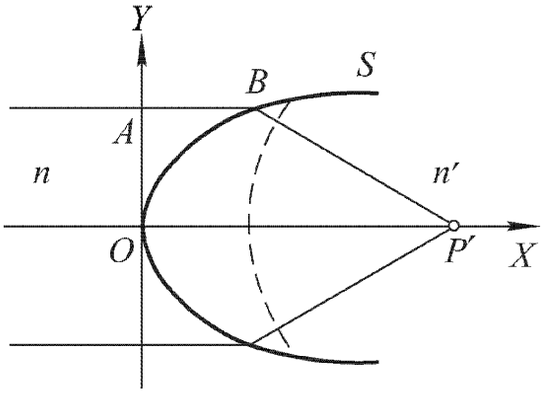


Рис. 37

$$a = \frac{n'}{n' + n}q, \quad b = \sqrt{\frac{n - n'}{n' + n}}\,q$$

и эксцентриситетом

$$e = \frac{\sqrt{a^2 + b^2}}{a} = \frac{n}{n'} > 1.$$

Изображение P' мнимое (рис. 38).

Рассмотрим, наконец, случай, когда $n'^2 - n^2 = 0$. Это может быть либо при $n' - n = 0$, либо при $n' + n = 0$. Первая возможность не представляет интереса, поскольку она соответствует тривиальному случаю, когда обе граничащие среды в оптическом отношении тождественны. Вторая возможность $n' = -n$ может быть реализована при отражении света. В этом случае уравнение (9.1) переходит в

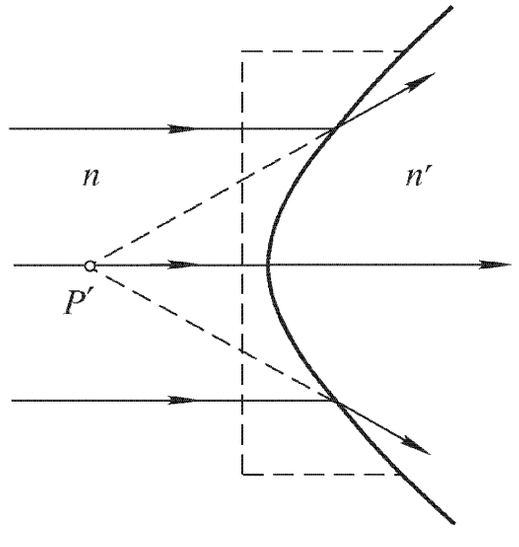


Рис. 38

$$y^2 = 4qx \tag{9.2}$$

и представляет параболоид вращения с параметром $p = 2q$ (параболоидальное зеркало). Если $q > 0$ (см. рис. 34), то фокус P' (на рис. 34 — точка F) мнимый. Если $q < 0$ (см. рис. 33), то он действительный. Эти случаи уже были рассмотрены в тексте.

Результаты решения этой задачи указывают способ построения *идеальной линзы* для пары сопряженных точек, из которых одна бесконечно удаленная. Рассмотрим сначала линзу, ограниченную поверхностью эллипсоида вращения OB и сферической поверх-

ностью с центром в P' (на рис. 37 эта поверхность изображена штриховой линией). Эксцентриситет эллипсоида должен быть равен $1/n$, где n — показатель преломления линзы относительно окружающей среды. Параллельный пучок лучей, падая на поверхность эллипсоида, после преломления на ней превращается в пучок, сходящийся в точке P'. Задняя — сферическая — поверхность линзы не меняет направления лучей, поскольку ее центр находится в точке схождения пучка P'. Таким образом, рассматриваемая линза собирает параллельный пучок лучей строго в одной точке P'. Если точечный источник поместить в P', то после прохождения через линзу пучок лучей станет строго параллельным оптической оси.

Рассмотрим далее линзу, ограниченную плоской поверхностью (на рис. 38 она изображена штриховой линией) и гиперболоидом вращения с эксцентриситетом n. Параллельный пучок лучей, падающих на плоскую поверхность линзы, после прохождения через эту поверхность не изменит направления, а после преломления на поверхности гиперболоида превратится в расходящийся пучок лучей, продолжения которых пересекаются строго в одной точке P'.

2. Найти уравнение картезианского овала (см. § 7, п. 6).

Р е ш е н и е. Пусть $P(q,0)$ и $P'(q',0)$ — сопряженные точки, для которых поверхность, получающаяся от вращения картезианского овала относительно оси симметрии PP', анаберрационна. Поместим начало координат в точку перечения овала с прямой PP'. Тогда по определению анаберрационной поверхности

$$n\sqrt{(x-q)^2+y^2}+n'\sqrt{(x-q')^2+y^2}=n'q'-nq.$$

Освобождаясь от радикалов, получим уравнение картезианского овала:

$$(n^2-n'^2)(x^2+y^2)+4(n^2-n'^2)(n'^2q'-n^2q)(x^2+y^2)x+$$
$$+4nn'(nq-n'q')(nq'-n'q)(x^2+y^2)+4(n'^2q'-n^2q)^2x^2+$$
$$+8nn'(n'-n)(nq-n'q')qq'x=0. \quad (9.3)$$

При определенных значениях параметров n, n', q, q' картезианской овал вырождается в поверхности второго порядка. Тогда получаются, в частности, уже разобранные ранее случаи, изображенные на рисунках 32, 33, 34, 37 и 38.

§ 10. Преломление на сферической поверхности. Сферические зеркала и тонкие линзы

1. Важнейшие из оптических инструментов или их составные части относятся к так называемым *центрированным оптическим системам*. Они представляют собой оптически однородные преломляющие или отражающие среды, отделенные одна от другой сферическими поверхностями, центры кривизны которых расположены на одной прямой, называемой *главной оптической осью* системы. Обычно если это не может привести к недоразумениям, прилагательное «главная» мы будем опускать.

2. Начнем с простейшего случая одной сферической преломляющей поверхности, разграничивающей однородные среды с

показателями преломления n и n'. Можно предполагать (хотя это и не обязательно), что эта поверхность обладает *симметрией вращения* относительно одной из прямых OC, проходящих через *центр кривизны* сферической поверхности (рис. 39). Такая прямая и будет главной оптической осью. Примем ее за координатную ось X. Начало координат поместим в точке O, в которой главная оптическая ось пересекает сферическую поверхность.

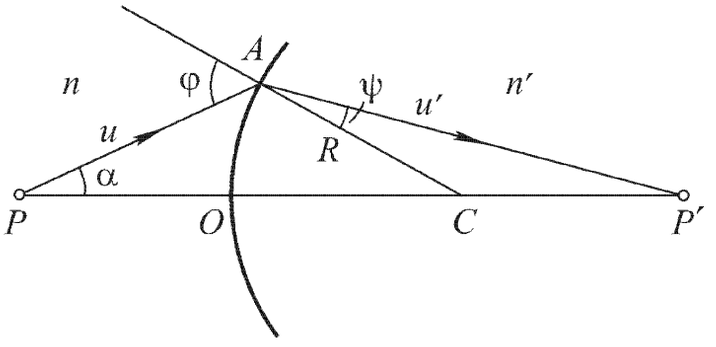


Рис. 39

Ввиду симметрии вращения достаточно ограничиться рассмотрением хода лучей в координатной плоскости XY. Совместим ее с плоскостью рисунка. Абсциссы и ординаты будем отсчитывать от начала координат O. Если направление отсчета совпадает с направлением распространения света вдоль оптической оси, то соответствующая абсцисса считается *положительной*; в противоположном случае она считается *отрицательной*. То же относится и ко всем другим направленным отрезкам. Например, на рис. 39 абсцисса точки P отрицательна, а точка P' положительна. Ордината считается положительной, если соответствующая точка лежит выше оптической оси, и отрицательной, когда она расположена ниже.

Допустим, что точечный источник света P находится на оптической оси системы (рис. 39). Произвольный луч PA после преломления на сферической поверхности пойдет по пути AP'. Обозначим длины AP и AP' через u и u' соответственно. Эти длины отсчитываются от точки A и считаются положительными, если направление отсчета совпадает с направлением распространения света, и отрицательными в противоположном случае. Из рисунка видно, что

$$\text{пл. } PAC + \text{пл. } CAP' = \text{пл. } PAP'.$$

Так как $u < 0$, $u' > 0$, то для этих площадей можно написать:

$$\text{пл. } PAC = \frac{1}{2}\,|PA| \cdot |AC| \sin\varphi = -\frac{1}{2}\,uR\sin\varphi,$$

$$\text{пл.}CAP' = \frac{1}{2}\,u'R\sin\psi,$$

$$\text{пл.}PAP' = -\frac{1}{2}\,uu'\sin(\varphi-\psi) = -\frac{1}{2}\,uu'(\sin\varphi\cos\psi - \sin\psi\cos\varphi).$$

Здесь R — радиус кривизны преломляющей поверхности. Он отсчитывается от сферической поверхности к ее центру (на рис. 39

радиус R положителен). Таким образом,

$$-uR\sin\varphi + u'R\sin\psi = -uu'(\sin\varphi\cos\psi - \sin\psi\cos\varphi).$$

По закону преломления $\sin\varphi/\sin\psi = n'/n$, а потому

$$-u'Rn' + u'Rn = uu'(n\cos\varphi - n'\cos\psi).$$

Отсюда

$$\frac{n}{u} - \frac{n'}{u'} = \frac{n\cos\varphi - n'\cos\psi}{R}. \tag{10.1}$$

Положение точки P' зависит от угла наклона α падающего луча к оптической оси. Ограничимся, однако, малыми углами α и допустим, что углы φ и ψ также малы. Лучи, удовлетворяющие таким условиям, называются *параксиальными* (*приосевыми*) *лучами*. Для них можно принять

$$\cos\varphi = \cos\psi = 1, \quad AP \approx OP = x, \quad AP' \approx OP' = x'.$$

В этом приближении формула (10.1) переходит в

$$\frac{n}{x} - \frac{n'}{x'} = \frac{n - n'}{R}. \tag{10.2}$$

Отсюда видно, что в рассматриваемом приближении положение точки P' не зависит от угла α. Следовательно, все параксиальные лучи, выходящие из одной точки оптической оси, после преломления на сферической поверхности пересекутся приближенно в одной точке, лежащей также на оптической оси. Точка P' будет поэтому *оптическим изображением точки P в параксиальных лучах*. Во всем дальнейшем предполагается, что все лучи, проходящие через центрированные системы, параксиальны.

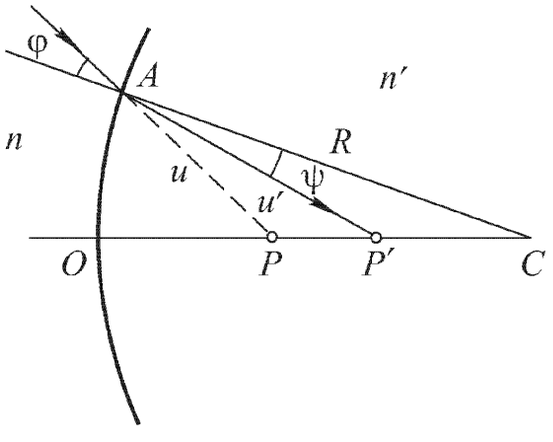


Рис. 40

При выводе предполагалось, что источник света P действительный. Однако все сказанное справедливо и для мнимого источника, т.е. тогда, когда источником служит точка схождения продолжений падающих лучей. В этом легко убедиться, повторив рассуждения применительно к рис. 40.

Полагая формально $n' = -n$, из (10.2) получаем формулу для сферического зеркала:

$$\frac{1}{x} + \frac{1}{x'} = \frac{2}{R}. \tag{10.3}$$

3. Рассмотрим теперь случай, когда точечный источник P не лежит на главной оптической оси (рис. 41). Проведем пря-

мую PC, соединяющую точку P с центром кривизны преломляющей поверхности. Такую прямую можно рассматривать как оптическую ось, сведя тем самым разбираемый случай к предыдущему. Для параксиальных лучей $O'P \approx OQ$, $O'P' \approx OQ'$

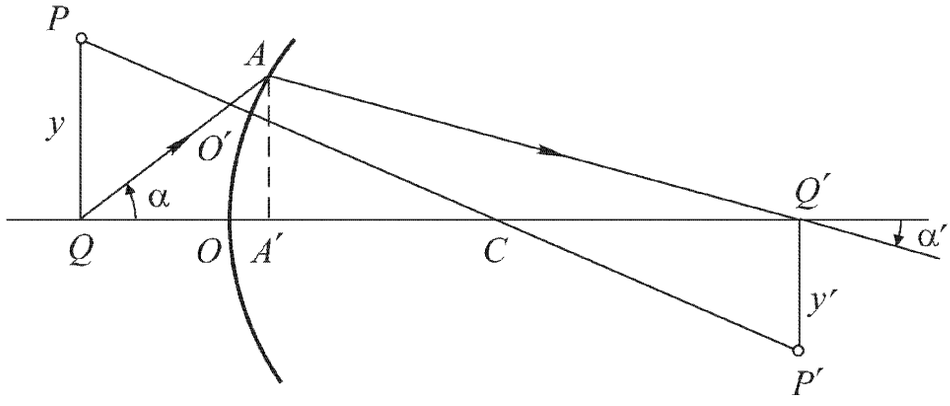


Рис. 41

(здесь Q и Q' — проекции точек P и P' на главную оптическую ось). Абсцисса x' точки P' определится из уравнения (10.2), а ордината y' — из соотношения

$$\frac{y'}{y} = -\left|\frac{CQ'}{CQ}\right| = \frac{x'-R}{x-R},$$

которое следует из подобия треугольников PCQ и $P'CQ'$. В результате находим

$$x' = \frac{n'Rx}{(n'-n)x+nR}, \quad y' = \frac{nRy}{(n'-n)x+nR}. \tag{10.4}$$

Проведем через точку Q произвольный параксиальный луч QA, образующий с главной оптической осью какой-то угол α. Угол наклона преломленного луча AQ' с той же осью обозначим через α'. Отсчет углов производится в направлении от главной оптической оси к лучу. Если это направление противоположно направлению вращения часовой стрелки, то угол считается положительным; в противоположном случае его следует считать отрицательным (на рис. 41 угол α положительный, угол α' отрицательный). Опустим из точки A перпендикуляр AA' на главную оптическую ось. В приближении параксиальной оптики его длину можно представить в виде $|AA'| = -x\alpha = -x'\alpha'$. Следовательно, $x\alpha = x'\alpha'$. Но из формул (10.4) следует

$$\frac{x'}{x} = \frac{n'y'}{ny}, \tag{10.5}$$

а потому

$$ny\alpha = n'y'\alpha'. \tag{10.6}$$

Таким образом, величина $ny\alpha$ не изменяется при преломлении

параксиального луча на сферической поверхности. Эта величина называется *инвариантом Лагранжа–Гельмгольца*, а равенство (10.6) — *теоремой Лагранжа–Гельмгольца.* Теорема, очевидно, справедлива для центрированных систем, состоящих из какого угодно числа преломляющих и отражающих сферических поверхностей.

Формулы (10.4) можно положить в основу геометрической теории любых центрированных систем в параксиальных лучах. Применяя их к первой преломляющей поверхности сложной системы, найдем положение изображения, возникающее от преломления на этой поверхности. Полученное промежуточное изображение играет роль предмета для преломления на второй сферической поверхности. С помощью тех же формул (10.4) можно найти положение второго промежуточного изображения, возникающего от преломления на второй сферической поверхности, и т.д. В конце концов путем последовательного применения формул (10.4) к каждой из преломляющих поверхностей можно найти положение окончательного изображения, даваемого всей системой.

4. В качестве примера рассмотрим тонкую линзу, ограниченную сферическими поверхностями с радиусами кривизны R_1 и R_2. Относительный показатель преломления линзы обозначим через n, приняв за единицу показатель преломления окружающего пространства. Пусть точечный предмет P находится на главной оптической оси линзы. Толщиной линзы будем пренебрегать и поместим начало координат в ее центре. Обозначим через x абсциссу точки P, через x_1 — абсциссу ее промежуточного изображения P_1, возникающего от преломления лучей на первой поверхности линзы. Абсциссу x_1 можно найти с помощью формулы (10.2), если в ней сделать замену $n \to 1$, $n' \to n$, $x' \to x_1$, $R \to R_1$. Это дает

$$\frac{1}{x} - \frac{n}{x_1} = \frac{1-n}{R_1}.$$

Промежуточное изображение P_1 будем рассматривать как предмет при преломлении света на второй сферической поверхности линзы. Изображение точки P, возникающее при таком преломлении, и будет окончательным изображением P', которое дает линза. Абсцисса x' точки P' найдется из формулы (10.2), если в ней сделать замену $n' \to 1$, $x \to x_1$, $R \to R_2$. Таким путем находим

$$\frac{n}{x_1} - \frac{1}{x'} = \frac{n-1}{R_2}.$$

Складывая это равенство с предыдущим, получим

$$\frac{1}{x} - \frac{1}{x'} = -(n-1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right). \tag{10.7}$$

Это хорошо известная формула тонкой линзы

$$\frac{1}{x} - \frac{1}{x'} = -\frac{1}{f}. \tag{10.8}$$

Фокусное расстояние f определяется формулой

$$\frac{1}{f} = (n-1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right). \tag{10.9}$$

Эта формула справедлива для всяких тонких линз: двояковыпуклых, двояковогнутых, плоско-выпуклых и т.д. Надо только придерживаться правила знаков, сформулированного в п. 2 настоящего параграфа. Так, для двояковыпуклой линзы $R_1 > 0$, $R_2 < 0$, а потому фокусное расстояние f положительно. Для двояковогнутой линзы $R_1 < 0$, $R_2 > 0$, и фокусное расстояние f отрицательно. (В обоих случаях предполагается, что $n > 1$.)

§ 11. Общие свойства центрированных оптических систем

1. Свойства центрированных оптических систем в параксиальных лучах были систематически исследованы Гауссом (1777–1855) в 1841 г. Поэтому оптику параксиальных лучей часто называют *гауссовой оптикой*. При изложении относящихся сюда вопросов мы применим *аналитический метод*. Он менее нагляден, чем геометрический метод. Зато аналитический метод отличается большей простотой и систематичностью.

В случае одной преломляющей поверхности координаты x, y точки-предмета связаны с координатами x', y' точки-изображения формулами (10.4). В этих формулах используется одна и та же координатная система в пространствах предметов и изображений. Выберем теперь в этих пространствах *разные системы координат*, получающиеся из исходной системы параллельным переносом вдоль главной оптической оси. Начала координат этих систем лежат на главной оптической оси, но могут и не совпадать друг с другом. При новом выборе координатных систем формулы (10.4) преобразуются в

$$x' = \frac{ax+b}{cx+d}, \quad y' = \frac{e}{cx+d}y,$$

где a, b, c, d, e — постоянные. Их легко вычислить, если известны положения начал координат и параметры преломляющей сферической поверхности. Этими формулами устанавливается соответствие между точками пространства предметов и пространства изображений. Оно называется *коллинеарным со-*

ответствием (точнее — *коллинеарным соответствием с осевой симметрией*).

Допустим теперь, что после прохождения через первую преломляющую поверхность лучи испытывают преломление на второй сферической поверхности. Тогда получится второе изображение — точка P'' с координатами x'', y''. Они связаны с x', y' формулами такого же вида, т.е.

$$x'' = \frac{a'x' + b'}{c'x' + d'}, \quad y'' = \frac{e'}{c'x' + d'}\, y'$$

с новыми коэффициентами a', b', c', d', e'. Исключим из этих и предыдущих соотношений координаты x', y' промежуточного изображения P'. Нетрудно убедиться, что таким путем снова получатся формулы коллинеарного соответствия:

$$x'' = \frac{a''x + b''}{c''x + d''}, \quad y'' = \frac{e''}{c''x + d''}\, y,$$

причем коэффициенты a'', b'', c'', d'', e'' могут быть выражены через десять коэффициентов a, b, …, d', e'. Таким образом, *два или несколько последовательно выполненных коллинеарных соответствий эквивалентны одному коллинеарному соответствию.* Следовательно, *любая центрированная система в параксиальных лучах устанавливает коллинеарное соответствие между точками пространства предметов и пространства изображений.* Его можно выразить формулами

$$x' = \frac{ax + b}{cx + d}, \quad y' = \frac{e}{cx + d}\, y, \quad z' = \frac{e}{cx + d}\, z. \tag{11.1}$$

Здесь добавлена формула для координат z и z'. Ввиду осевой симметрии, она совпадает с соответствующей формулой для y и y'.

Коллинеарное соответствие определяется *четырьмя параметрами*, за которые можно принять отношения четырех из коэффициентов a, b, c, d, e к пятому. Поэтому и произвольная центрированная система характеризуется также *четырьмя параметрами.*

Разрешая уравнения (11.1) относительно x, y, z, получим

$$x = \frac{a'x' + b'}{c'x' + d'}, \quad y = \frac{e'}{c'x' + d'}\, y', \quad z = \frac{e'}{c'x' + d'}\, z', \tag{11.2}$$

причем

$$b' = b, \quad c' = c, \quad a' = -d, \quad d' = -a, \quad e' = \frac{bc - ad}{e}. \tag{11.3}$$

Обратное преобразование, таким образом, выражается также формулами коллинеарного соответствия, что, очевидно, является следствием обратимости светового пути (см. § 10, п. 1).

2. Из формул коллинеарного соответствия вытекают следующие свойства оптических изображений в центрированных системах.

1. *Каждая плоскость пространства предметов изображается в виде плоскости.* Действительно, уравнение плоскости в пространстве предметов

$$Ax + By + Cz + D = 0$$

после замены координат x, y, z их выражениями через x', y', z' по формулам (11.2) переходит в уравнение вида

$$A'x' + B'y' + C'z' + D' = 0,$$

которое представляет плоскость в пространстве изображений.

2. *Каждая прямая пространства предметов изображается в виде прямой*, так как прямую можно рассматривать как линию пересечения двух плоскостей.

3. *Каждая точка пространства предметов изображается в виде точки*, так как всякую точку можно рассматривать как точку пересечения трех плоскостей.

3. Из формул (11.1) следует, что конечным значениям x, y, z соответствуют, вообще говоря, конечные значения x', y', z'. Исключение составляют точки плоскости

$$cx + d = 0. \tag{11.4}$$

Каждая точка такой плоскости изображается бесконечно удаленной точкой. Это означает, что все лучи, вышедшие из одной и той же точки плоскости (11.4), после прохождения через оптическую систему становятся параллельными. Плоскость (11.4) называется *фокальной плоскостью пространства предметов, или передней фокальной плоскостью оптической системы.* Аналогично, плоскость

$$c'x' + d' = 0 \tag{11.5}$$

называется *фокальной плоскостью пространства изображений*, или *задней фокальной плоскостью оптической системы.* Параллельные лучи после прохождения через оптическую систему пересекаются в одной из точек этой плоскости.

Точки пересечения фокальных плоскостей с главной оптической осью называются *фокальными точками*, или *главными фокусами оптической системы.* Главный фокус пространства предметов (*передний главный фокус*) будем обозначать через F, а главный фокус пространства изображений (*задний главный фокус*) — через F'. Согласно формулам (11.4), (11.5) и (11.3), абсциссы главных фокусов системы определяются выражениями:

$$x_F = -\frac{d}{c}, \quad x'_{F'} = -\frac{d'}{c'} = \frac{a}{c}. \tag{11.6}$$

Система может и не иметь фокальных плоскостей. Это будет, когда $c = 0$, и следовательно, $c' = 0$. Такие системы называются *афокальными*, или *телескопическими*. Они являются предельными случаями обычных систем, когда обе фокальные плоскости сдвинуты в бесконечность. После прохождения через афокальную систему всякий параллельный пучок лучей остается параллельным, могут изменяться лишь ширина и направление пучка. Примером афокальной системы может служить зрительная труба (телескоп), установленная на бесконечность. В этом случае задняя фокальная плоскость объектива совмещается с передней фокальной плоскостью окуляра.

Рассмотрим сначала системы с конечными фокусными расстояниями.

4. Отношение y'/y называется *поперечным увеличением*, или просто *увеличением системы*. Согласно формулам (11.1) или (11.2), оно не зависит от y и z. Отсюда следует, что *изображение плоского предмета, перпендикулярного к главной оптической оси, подобно самому предмету*. Если увеличение положительное, то изображение *прямое*. В противоположном случае изображение *обратное*.

Две сопряженные плоскости, отображающиеся друг в друга с поперечным увеличением $y'/y = +1$, называются *главными плоскостями оптической системы*. Уравнения таких плоскостей можно получить, полагая в формулах (11.1) и (11.2) $y' = y$. Это дает

$$cx + d - e = 0, \quad c'x' + d' - e' = 0. \tag{11.7}$$

Первая плоскость называется *главной плоскостью пространства предметов* (*передней главной плоскостью*), вторая — *главной плоскостью пространства изображений* (*задней главной плоскостью*). Точки пересечения главных плоскостей с главной оптической осью называются *главными точками центрированной системы*. Главную точку в пространстве предметов будем обозначать через H, а в пространстве изображений — через H'. Полагая в формулах (11.1) и (11.2) $y = y'$, находим абсциссы главных точек:

$$x_H = \frac{e-d}{c}, \quad x'_{H'} = \frac{e'-d'}{c'} = \frac{bc + a(e-d)}{ec}. \tag{11.8}$$

5. Главные и фокальные точки центрированной системы называются ее *кардинальными точками*. Так как коллинеарное соответствие определяется четырьмя параметрами, то положение четырех кардинальных точек полностью определяет коллинеарное соответствие. Фокальные и главные точки полностью характеризуют оптическую систему в том смысле, что, зная положение этих точек, можно найти изображение любого предмета, даваемое оптической системой.

Допустим сначала, что предмет P точечный и не лежит на главной оптической оси системы (рис. 42). Опустим из точки P

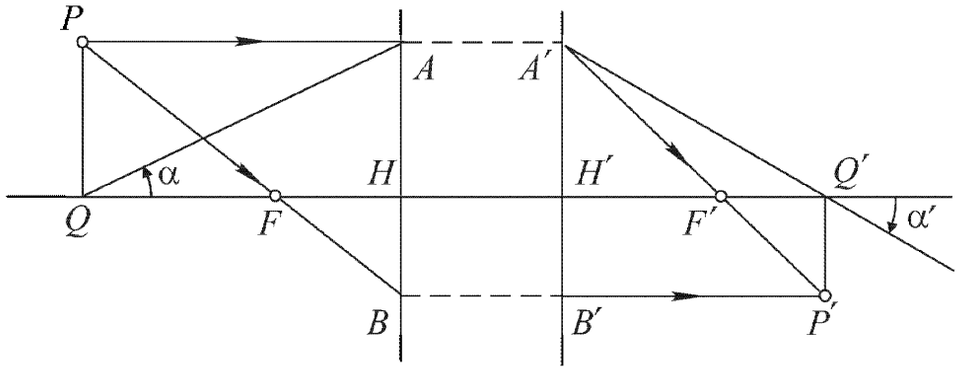


Рис. 42

перпендикуляр PQ на эту ось. Луч PA, параллельный главной оптической оси, или его продолжение в сторону пространства изображений встретит главные плоскости в сопряженных точках A и A'. Поэтому после прохождения через оптическую систему этот луч или его продолжение пройдет через точку A'. Кроме того, он должен пройти через задний фокус F'. Двумя точками A' и F' положение луча в пространстве изображений определяется полностью. Проведем теперь второй луч PF, проходящий через передний фокус F. Он или его продолжение встретит переднюю главную плоскость в точке B. Проведя через B прямую $BB'P'$, параллельную главной оптической оси, найдем положение луча $B'P'$ в пространстве изображений. Точка P', в которой пересекаются лучи $A'P'$ и $B'P'$, и будет изображением точки P. При этом проекция Q точки P на главную оптическую ось изобразится проекцией точки P' на ту же ось. Это дает правило построения изображения точки и в том случае, когда она лежит на главной оптической оси. Отрезок QP изображается отрезком $Q'P'$.

Расстояния главных точек от соответствующих фокальных точек называются *главными фокусными расстояниями* системы. Фокусное расстояние в пространстве предметов будем обозначать через f, в пространстве изображений — через f'. В соответствии с принятым нами правилом знаков фокусное расстояние считается положительным, если падающий свет идет в направлении от главного фокуса к соответствующей главной плоскости. Фокусные расстояния определяются формулами

Надлежащим выбором начал координат можно упростить формулы (11.1) и (11.2). Рассмотрим два важнейших случая.

6. С л у ч а й 1. Начала координат помещены в главные точки H и H'. Абсциссы в этих координатных системах будем обозначать через ξ и ξ', а ординаты — через η и η'. Очевидно, $\eta = y$, $\eta' = y'$, а по самому выбору начал координат $\xi_H = \xi'_{H'} = 0$. Поэтому из (11.8) следует: $e = d$, $b = 0$. Формулы (11.1) преобразуются в

$$\xi' = \frac{a\xi}{c\xi + d}, \quad \eta' = \frac{d}{c\xi + d}\,\eta.$$

Абсциссы фокальных точек, ввиду соотношений (11.6), будут $\xi_F = -d/c$, $\xi'_{F'} = a/c$, а фокусные расстояния $f = d/c$, $f' = -a/c$. Исключая с помощью этих соотношений коэффициенты a, c, d, получим

$$\xi' = \frac{f'\xi}{f - \xi}, \quad \eta' = \frac{f\eta}{f - \xi}. \tag{11.10}$$

Первую из этих формул можно записать также в виде

$$\frac{f}{\xi} + \frac{f'}{\xi'} = -1. \tag{11.11}$$

Как видно из рис. 42, $HA = -\xi\alpha$, $H'A' = -\xi'\alpha'$. Так как $HA = H'A'$, то $\xi\alpha = \xi'\alpha'$. Далее, из уравнений (11.10) находим: $f\eta\xi' = -f'\eta'\xi$, так что

$$f\eta\alpha = -f'\eta'\alpha'. \tag{11.12}$$

Но по теореме Лагранжа–Гельмгольца $n\eta\alpha = n'\eta'\alpha'$, а потому

$$\frac{f}{n} = -\frac{f'}{n'}. \tag{11.13}$$

Такова связь между фокусными расстояниями f и f'.

Допустим, как это обычно бывает, что среда по обе стороны оптической системы одна и та же. Если система содержит четное число отражений (такая система называется *диоптрической*), то $n = n'$, и следовательно, $f = -f'$. В этом случае формула (11.11) переходит в *формулу линзы*:

$$\frac{1}{\xi} - \frac{1}{\xi'} = -\frac{1}{f}. \tag{11.14}$$

Если же система содержит нечетное число отражающих поверхностей (такая система называется *катоптрической*), то $n = -n'$, и следовательно, $f = f'$. Тогда формула (11.11) переходит в *формулу зеркала*:

$$\frac{1}{\xi} + \frac{1}{\xi'} = -\frac{1}{f}. \tag{11.15}$$

В диоптрических системах направления распространения падающего и прошедшего света одинаковы, а в катоптрических противоположны.

7. С л у ч а й 2. Начала координатных систем помещены в главные фокусы F и F'. Абсциссы в этих системах будем обозначать через X и X', а ординаты через Y и Y'. Очевидно, $Y = y = \eta$, $Y' = y' = \eta'$. Кроме того, по самому выбору начал координат $X_F = X'_{F'} = 0$. Следовательно, должно быть $a = d = 0$, $b/c = ff'$, $e/c = f$, как в этом нетрудно убедиться с помощью формул (11.6) и (11.9). В результате формулы (11.1) переходят в

$$XX' = ff', \tag{11.16}$$

$$\frac{Y'}{Y} = \frac{f}{X} = \frac{X'}{f'}. \tag{11.17}$$

Эти простые формулы были известны еще Ньютону, — правда, для частных случаев.

8. К кардинальным точкам, наряду с главными и фокальными, относят еще *узловые точки*. Они естественно появляются при рассмотрении *углового увеличения* системы. Угловое увеличение есть отношение α'/α углов α' и α, образуемых с главной оптической осью проходящим и падающим лучами. *Узловыми точками называются две сопряженные точки K и K', лежащие на главной оптической оси*, которые отображаются друг в друга с угловым увеличением $+1$. Это значит, что всякий луч, проходящий через узловую точку K, после прохождения через оптическую систему остается параллельным своему исходному направлению и проходит через вторую узловую точку системы K'. Для доказательства существования и определения положения узловых точек положим в формуле (10.6) $\alpha = \alpha'$. Это дает $ny = n'y'$, или на основании формулы (11.13) $y'/y = -f/f'$. Сравнивая этот результат с формулами (11.17), находим абсциссы узловых точек относительно фокальных точек:

$$X_K = -f', \quad X'_{K'} = -f, \tag{11.18}$$

или на основании определения фокусных расстояний

$$FK = H'F', \quad F'K' = HF. \tag{11.19}$$

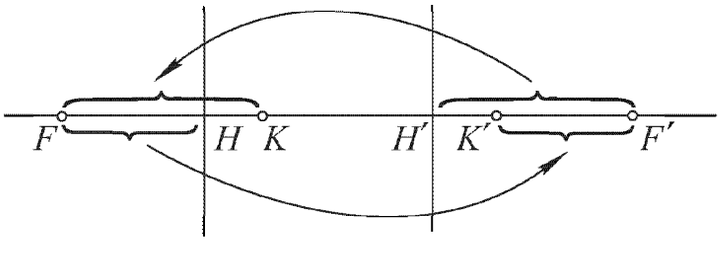


Рис. 43

Это приводит к следующему правилу для нахождения узловых точек. Отрезки HF и $H'F'$ надо перенести параллельно самим себе, чтобы их начальные точки H и H' совпали с фокусами F и F' соответственно. Тогда другие концы этих отрезков F и F' укажут положение узловых точек K' и K (рис. 43). Если $n' = n$, то $f' = -f$.

В этом случае узловые точки совпадают с главными. Узловые точки также могут быть использованы для построения оптических изображений в центрированных системах.

Иногда к числу кардинальных точек относят *обратные главные* и *обратные узловые точки.* Первые характеризуются линейным увеличением, равным минус единице; вторые — угловым увеличением, равным также минус единице. Доказательство существования и определение положения таких пар точек не представляют затруднений.

9. Отношение длины $\delta X'$ изображения бесконечно малого отрезка, параллельного главной оптической оси, к длине δX самого отрезка, называется *осевым* или *продольным увеличением.* Для этого увеличения из формулы (11.16) находим

$$\frac{\delta X'}{\delta X} = -\frac{X'}{X} = -\frac{ff'}{X^2} = -\frac{X'^2}{ff'}. \tag{11.20}$$

Сравнение этих формул с формулами (11.17) показывает, что *осевое увеличение в общем случае не равно поперечному увеличению.* Отсюда следует, что *изображение бесконечно малого объемного предмета в центрированной оптической системе, вообще говоря, не подобно самому предмету. Исключение составляет случай, когда предмет помещен в одной из узловых или в одной из обратных узловых точек системы.* В самом деле, для изображения бесконечно малого объемного предмета с сохранением подобия необходимо и достаточно, чтобы осевое увеличение по абсолютной величине было равно поперечному увеличению, т.е.

$$\frac{ff'}{X^2} = \pm\frac{f}{X}.$$

Отсюда $X = \pm f'$. Формулы (11.17) и (11.13) дают

$$\frac{Y'}{Y} = \pm\frac{f}{f'} = \mp\frac{n}{n'}.$$

После этого из теоремы Лагранжа–Гельмгольца (10.6) получаем $\alpha = \pm\alpha'$, что и требовалось доказать.

Если система диоптрическая (см. п. 6), то знаки f и f' противоположны. В этом случае, как видно из (11.17), δX и $\delta X'$ имеют одинаковые знаки. Отсюда следует, что при перемещении предмета вдоль оптической оси его изображение перемещается *в том же направлении.* Напротив, если система катоптрическая, то при перемещении предмета вдоль оптической оси его изображение перемещается *в противоположном направлении.* А так как в диоптрических системах направления распространения света в пространствах предметов и изображений одинаковы, а в катоптрических противоположны, то для всех систем справедливо такое правило:

Если предмет перемещается вдоль оптической оси в направлении распространения падающего света, то его изображение перемещается в направлении распространения прошедшего света, и наоборот.

10. В случае телескопической системы $c = 0$. Но коэффициент d в нуль обращаться не должен, иначе формулы (11.1) потеряли бы смысл. Не должны обращаться в нуль также коэффициенты a и e, так как в противном случае из формул (11.1) мы получили бы $x' = \text{const}$, $y' = 0$, $z' = 0$, т.е. любая точка изображалась бы всегда одной и той же точкой. Такой случай в оптических системах не встречается, он имеет только формально-математический характер. Итак, для телескопической системы $c = 0$, $d \neq 0$, $a \neq 0$, $e \neq 0$. Следовательно, должно быть также $c' = 0$, $d' \neq 0$, $a' \neq 0$, $e' \neq 0$, как это видно из формул (11.3). Телескопическую систему можно рассматривать как предельный случай фокальной системы, обе фокальные точки которой удалены в бесконечность ($x_F = \pm\infty$, $x'_{F'} = \pm\infty$). Оба фокусных расстояния f и f' телескопической системы бесконечно велики, хотя их отношение $f/f' = -n/n'$ и остается конечным.

Формулы (11.1) для телескопической системы принимают вид

$$x' = Ax + C, \quad y' = By, \tag{11.21}$$

где $A = a/d$, $B = e/d$, $C = b/d$ — постоянные. Отсюда видно, что *всякий параллельный пучок света после прохождения через телескопическую систему остается параллельным.* Постоянные A и B не независимы. Действительно, формулы (11.9) для телескопических систем дают

$$\frac{f}{f'} = -\frac{e^2}{ad} = -\frac{B^2}{A},$$

или на основании (11.13)

$$\frac{B^2}{A} = \frac{n}{n'}. \tag{11.22}$$

Если за начала координатных систем принять любую пару сопряженных точек оптической оси, то формулы (11.21) упростятся и перейдут в

$$x' = Ax, \quad y' = By. \tag{11.23}$$

Таким образом, *поперечное и осевое увеличения телескопической системы постоянны, т.е. не зависят от положения предмета.* Постоянно также и угловое увеличение, так как по формуле (10.6)

$$\frac{\alpha'}{\alpha} = \frac{ny}{n'y'} = \frac{n}{Bn'} = \frac{B}{A}. \tag{11.24}$$

Для зрительных труб и любых систем, у которых $n = n'$,

$$\frac{\alpha'}{\alpha} = \frac{y}{y'}. \tag{11.25}$$

В этом случае угловое увеличение называют просто *увеличением* трубы, опуская прилагательное «угловое». Эта величина показывает, во сколько раз угол, под которым виден бесконечно удаленный малый предмет в трубу, больше угла, под которым он был бы виден невооруженным глазом. Согласно (11.25), угловое увеличение зрительной трубы равно обратному значению ее поперечного увеличения. Отсюда следует, что *увеличение зрительной трубы численно равно отношению ширины падающего пучка лучей к ширине соответствующего выходящего пучка.*

Поясним последнее утверждение на примере *кеплеровой трубы*, состоящей из двух тонких собирающих линз — объектива O и окуляра O' (рис. 44). Задняя фокальная плоскость объектива совпадает с передней фокальной плоскостью окуляра. Пусть в трубу рассматривается бесконечно удаленный предмет. Объектив дает изображение FC этого предмета в общей фокальной плоскости объектива и окуляра. Допустим для простоты, что один из крайних лучей (луч AF) идет вдоль оптической оси трубы, а другой (луч BC) — под углом α к ней. Проходя через оптический центр объектива, луч BC не испытывает преломления. Угол α есть угол, под которым бесконечно удаленный предмет виден невооруженным глазом; угол α' — угол, под которым тот же предмет виден в зрительную трубу. *Угловое увеличение* α'/α, согласно рис. 44, *численно равно отношению фокусного расстояния объектива к фокусному расстоянию окуляра.* Но это отношение в свою очередь равно отношению ширины падающего

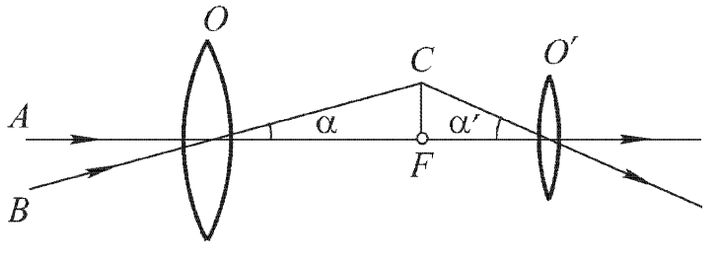


Рис. 44

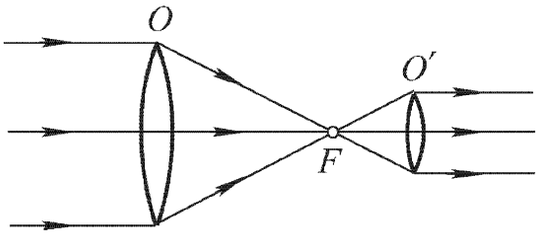


Рис. 45

параллельного пучка лучей к ширине соответствующего выходящего пучка (рис. 45).

11. Более общий подход к вопросу об угловом увеличении телескопической системы дает *принцип Ферма.* Этот подход применим и к телескопическим системам, не обладающим осевой симметрией. *Телескопической системой в общем случае называют любую оптическую систему, при прохождении через которую каждый параллельный пучок света остается параллель-*

ным. Пусть AB (рис. 46) — плоский участок волнового фронта перед телескопической системой. После прохождения через

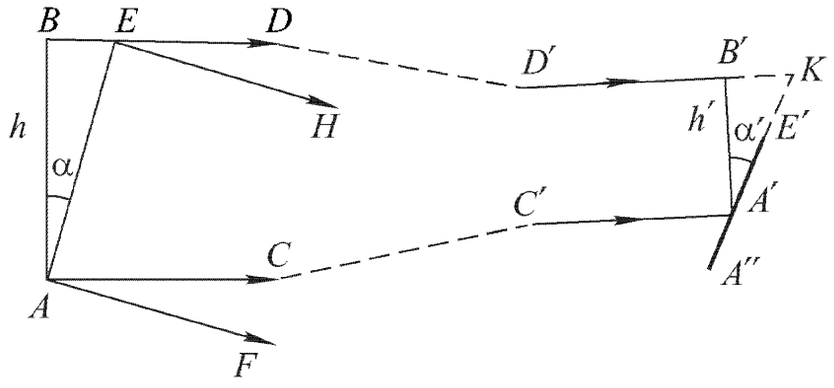


Рис. 46

эту систему (не изображенную на рисунке) он переходит в плоский участок $A'B'$. Продолжением луча AC является луч $C'A'$, а луча BD — луч $D'B'$. Таким образом, $(ACC'A') = (BDD'B')$. Возьмем другой плоский участок волнового фронта AE, наклоненный к AB под бесконечно малым углом α. За оптической системой волновой фронт AE перейдет в волновой фронт $A''E'$, образующий с прежним волновым фронтом $A'B'$ угол α'. Угловое увеличение системы будет $N = \alpha'/\alpha$. Возьмем тот волновой фронт $A''E'$, который проходит через точку A'. На основании следствия принципа Ферма, доказанного в п. 4 § 7, оптические длины лучей AA'' и EE' при смещении вдоль волнового фронта $A''E'$ будут меняться во втором или высшем порядке малости. В этом порядке указанные оптические длины останутся неизменными, если заменить их оптическими длинами $(ACC'A')$ и $(EDD'B'K)$. Таким образом, $(ACC'A') = (EDD'B'K)$. Сравнивая это соотношение с предыдущим, получим $(BEDD'B') = (EDD'B'K)$, откуда $(BE) = (B'K)$, или

$$nh\alpha = n'h'\alpha', \tag{11.26}$$

где n — показатель преломления среды перед телескопической системой, n' — за этой системой, h — поперечное сечение падающего пучка лучей, а h' — выходящего. Если, как обычно бывает, $n' = n$, то

$$N = \frac{\alpha'}{\alpha} = \frac{h}{h'}, \tag{11.27}$$

т.е. *увеличение телескопической системы равно отношению ширин пучков света до и после прохождения через эту систему.* Полученные результаты имеют общий характер и не зависят от конкретного устройства телескопической системы.

Увеличение может и не быть одинаковым по всем направлениям. Рассмотрим, например, призму, преломляющее ребро ко-

торой вертикально. Вертикальные размеры параллельного светового пучка после прохождения через призму не изменяются, тогда как горизонтальные размеры, вообще говоря, меняются. Призма будет давать увеличение в горизонтальном направлении, но не будет увеличивать в вертикальном направлении. Изображение, даваемое призмой, получится вытянутым или сплюснутым в горизонтальном направлении, в зависимости от того, уменьшаются или увеличиваются поперечные размеры светового пучка в этом направлении. Искажения не получится, если держать призму под углом наименьшего отклонения, но в этом случае не будет и увеличения. Комбинацией двух призм, преломляющие ребра которых взаимно перпендикулярны, можно произвольно увеличивать или уменьшать поперечное сечение светового пучка и притом одинаково по всем направлениям. Такая комбинация действует как обыкновенная зрительная труба. Вращая каждую из призм вокруг своего преломляющего ребра, можно получить произвольное увеличение. Недостатками такого оптического прибора являются большие геометрические и хроматические аберрации, вносимые им. Практического значения он не имеет.

§ 12. Сложение центрированных систем. Толстые линзы

1. Пусть две центрированные системы соединены вместе таким образом, что их оптические оси совпадают. Если известны параметры каждой системы, а также их взаимное расположение, то геометрическим построением или аналитическим расчетом можно определить положение всех кардинальных точек сложной оптической системы, состоящей из этих двух систем.

Обозначим фокусные расстояния первой системы через f_1 и f_1', а второй системы — через f_2 и f_2'. Пусть Δ означает расстояние передней фокальной точки F_2 второй системы от задней фокальной точки F_1' первой системы (рис. 47). Это расстояние

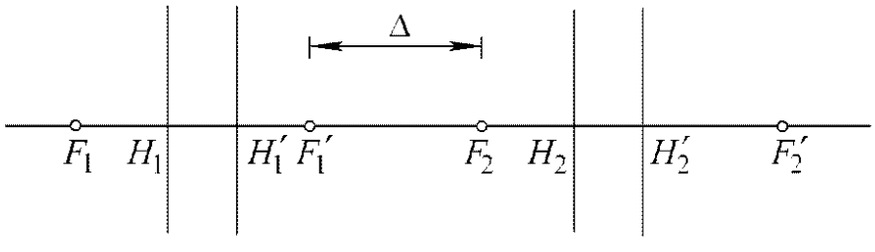


Рис. 47

называется *оптическим интервалом* двух систем и в соответствии с принятым правилом знаков считается положительным,

если падающий свет идет в направлении от фокуса F_1' к фокусу F_2, в противоположном случае оптический интервал считается отрицательным. Заданием оптического интервала полностью определяется взаимное расположение складываемых систем. С целью упрощения вычислений начала координат для каждой из складываемых систем поместим в ее фокальные точки. Фокус F_1 примем за начало координат в пространстре предметов всей сложной системы, а фокус F_2' — за начало координат в пространстве изображений той же системы. Пусть x, y — координаты предмета, а x_1, y_1 — его изображения, даваемого первой из складываемых систем. Тогда

$$xx_1 = f_1 f_1', \quad \frac{y_1}{y} = \frac{f_1}{x}.$$

Примем это промежуточное изображение за «предмет» для второй из складываемых систем. Координаты этого предмета в координатной системе с началом в точке F_2 будут $x_2 = x_1 - \Delta$, $y_2 = y_1$. Если x', y' — координаты изображения, даваемого второй системой (а следовательно, и всей сложной системой) относительно начала F_2', то

$$x_2 x' = f_2 f_2', \quad \frac{y_1}{y_2} = \frac{x'}{f_2'}.$$

Исключая промежуточные координаты x_1, y_1, x_2, y_2, получим

$$x' = \frac{f_2 f_2'}{f_1 f_1' - \Delta \cdot x}\, x, \quad y' = \frac{f_1 f_2}{f_1 f_1' - \Delta \cdot x}\, y. \tag{12.1}$$

Это — формулы коллинеарного соответствия с коэффициентами

$$a = f_2 f_2', \quad b = 0, \quad c = -\Delta, \quad d = f_1 f_1', \quad e = f_1 f_2.$$

Из них по формулам (11.6) и (11.9) находим координаты фокальных точек и фокусные расстояния сложной системы:

$$x_F = \frac{f_1 f_1'}{\Delta}, \quad x'_{F'} = -\frac{f_2 f_2'}{\Delta}, \tag{12.2}$$

$$f = -\frac{f_1 f_2}{\Delta}, \quad f' = \frac{f_1' f_2'}{\Delta}. \tag{12.3}$$

Координаты главных точек определяются выражениями

$$\begin{aligned} x_H &= x_F + f = f_1 \frac{f_1' - f_2}{\Delta}, \\ x'_{H'} &= x'_{F'} + f' = f_2' \frac{f_1' - f_2}{\Delta}. \end{aligned} \tag{12.4}$$

Следовательно,

$$\frac{x_H}{x'_{H'}} = \frac{f_1}{f_2'}. \tag{12.5}$$

Если оптический интервал Δ обращается в нуль, то фокусные расстояния f и f' обращаются в бесконечность, т.е. система будет *телескопической*. (Такой случай осуществляется, например, в зрительной трубе.) В этом случае уравнения (12.1) переходят в (11.23), причем

$$A = \frac{f_2 f_2'}{f_1 f_1'}, \quad B = \frac{f_2}{f_1'}.$$

Угловое увеличение сложной системы будет

$$\frac{\alpha'}{\alpha} = \frac{B}{A} = \frac{f_1}{f_2'}. \tag{12.6}$$

В частности, при $n = n'$

$$\frac{\alpha'}{\alpha} = -\frac{f_1}{f_2}. \tag{12.7}$$

Это согласуется с результатом, полученным выше для кеплеровой трубы.

Если обе складываемые системы телескопические, то составная система также телескопическая, причем ее угловое увеличение равно произведению угловых увеличений складываемых систем. Наконец, *соединение телескопической системы и системы с конечными фокусными расстояниями образует систему с конечными фокусными расстояниями при всякой последовательности расположения обеих систем.*

2. Величина, обратная главному фокусному расстоянию f' пространства изображений, взятая с противоположным знаком, т.е. $-1/f'$, называется *оптической силой системы*. Оптическая сила измеряется *диоптриями*. Диоптрия есть оптическая сила такой системы, фокусное расстояние $|f'|$ которой равно одному метру. Для собирательных тонких линз оптическая сила *положительна*, для рассеивающих *отрицательна*.

Определим оптическую силу сложной системы, зная оптические силы составляющих систем и их взаимное расположение. Будем предполагать, что показатели преломления всех пространств предметов и изображений одинаковы. Обозначим через l_{12} расстояние $H_1'H_2$ передней главной плоскости H_2 второй системы от задней главной плоскости H_1' первой системы. Оптический интервал между рассматриваемыми системами будет

$$\Delta = F_1'F_2 = F_1'H_1' + H_1'H_2 + H_2F_2 = f_1' + l_{12} - f_2 = f_1' + l_{12} + f_2'.$$

Подставляя это значение в формулу (12.3), получим

$$\frac{1}{f'} = \frac{1}{f_1'} + \frac{1}{f_2'} + \frac{l_{12}}{f_1' f_2'}. \tag{12.8}$$

В частности, когда задняя главная плоскость первой системы

совпадает с передней главной плоскостью второй системы, то

$$\frac{1}{f'} = \frac{1}{f'_1} + \frac{1}{f'_2}, \tag{12.9}$$

т.е. *оптическая сила сложной системы равна сумме оптических сил составляющих систем.* Это имеет место, например, для двух тонких линз, прижатых вплотную одна к другой.

Применим полученные результаты к системе двух центрированных тонких линз: собирательной и рассеивающей, поставленных друг за другом. Пусть фокусные расстояния линз по абсолютной величине одинаковы: $f_1 = -f_2$, а потому $f'_1 = -f'_2$. Оптический интервал между линзами $\Delta = f'_1 + l_{12} + f'_2 = l_{12}$, т.е. положителен ($l_{12} > 0$). Из формулы (12.6) получаем

$$f = -f' = -\frac{f'_1 f'_2}{l_{12}} = -\frac{f_1 f_2}{l_{12}} = \frac{f_1^2}{l_{12}} > 0.$$

Далее, по формулам (12.2) и (12.4) находим $x_H = x'_{H'} = 0$, т.е. главные плоскости H и H' системы проходят соответственно через фокусы F_1 и F'_1 и находятся на расстоянии l_{12} друг от друга. Абсцисса главного фокуса F' системы:

$$x'_{F'} = -\frac{f_2 f'_2}{l_{12}} = \frac{f_1^2}{l_{12}},$$

абсцисса линзы L_2:

$$x'_{L_2} = f'_2 = -f'_1 = f_1.$$

Таким образом,

$$x'_{F'} - x'_{L_2} = \frac{f_1^2}{l_{12}} - f_1.$$

Для того чтобы система линз собирала лучи, параллельные главной оптической оси, в действительном фокусе, т.е. была собирающей, необходимо, чтобы эта разность была положительна. Если первая линза рассеивающая, то указанное условие соблюдается всегда, так как в этом случае $f_1 < 0$. Если же первая линза собирательная, то это условие сводится к $l_{12} < f_1$. Эти результаты легче получить непосредственным геометрическим построением, что и рекомендуется сделать читателю.

Системы линз, аналогичные рассмотренной, применяются в современных ускорителях для фокусировки заряженных частиц. На них основан принцип так называемой *жесткой фокусировки.*

3. Каждая центрированная система может рассматриваться как сложная система, состоящая из нескольких подсистем. В качестве подсистем можно взять сферические границы раздела сред, на которых световые лучи испытывают преломление или отражение. Для сферической границы раздела коллинеарное соответствие выражается формулами (10.4). Из них и из формул (11.8) находим прежде всего: $x_H = x'_{H'} = 0$, т.е. обе главные плоскости совпадают между собой и проходят через точку пересечения рассматриваемой преломляющей поверхности с главной оптической осью системы. Для фокусных расстояний f и f' подсистем формулы (10.4) и (11.9) дают

$$f = \frac{Rn}{n' - n}, \quad f' = -\frac{Rn'}{n' - n}. \tag{12.10}$$

Используя эти формулы, а также формулы (12.2), (12.3) и (12.4), можно рассчитать параметры любой центрированной системы.

4. В качестве примера проведем расчет параметров *толстой линзы*. Пусть R_1 и R_2 означают радиусы кривизны преломляющих сферических поверхностей линзы, n_1, n_2, n_3 — показатели преломления первой среды, вещества линзы и второй среды

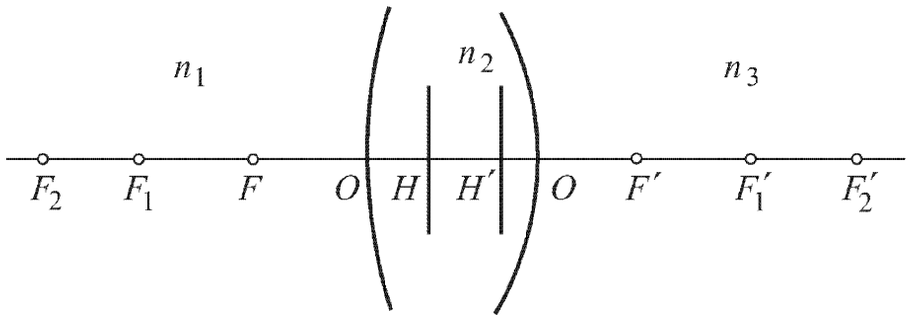


Рис. 48

(рис. 48), f_1 и f_1' — фокусные расстояния при преломлении на передней поверхности линзы, f_2 иf_2' — на задней. В таком случае

$$f_1 = \frac{n_1 R_1}{n_2 - n_1}, \quad f_1' = -\frac{n_2 R_1}{n_2 - n_1},$$
$$f_2 = \frac{n_2 R_2}{n_3 - n_2}, \quad f_2' = -\frac{n_3 R_2}{n_3 - n_2}.$$

Будем рассматривать преломляющие поверхности линзы как центрированные подсистемы, а саму линзу — как сложную систему. Если d — толщина линзы, то $\Delta = f_1' + d - f_2$, или после подстановки значений f_1' и f_2'

$$\Delta = \frac{D}{(n_2 - n_1)(n_3 - n_2)}, \tag{12.11}$$

где введено обозначение

$$D = d(n_1 - n_2)(n_2 - n_3) + n_2[R_1(n_2 - n_3) + R_2(n_1 - n_2)]. \tag{12.12}$$

Для фокусных расстояний f и f' линзы из формул (12.3) получаем

$$f = -n_1 n_2 \frac{R_1 R_2}{D}, \quad f' = n_2 n_3 \frac{R_1 R_2}{D}. \tag{12.13}$$

Координаты фокальных точек линзы F и F' можно вычислить по формулам (12.2), которые дают

$$x_F = -n_1 n_2 \frac{n_2 - n_3}{n_1 - n_2} \frac{R_1^2}{D}, \quad x'_{F'} = n_2 n_3 \frac{n_1 - n_2}{n_2 - n_3} \frac{R_2^2}{D}. \tag{12.14}$$

При этом за начало координат в пространстве предметов линзы принят фокус F_1, а в пространстве изображений — фокус F_2'.

Найдем расстояния h и h' главных плоскостей линзы от точек O и O'. По определению

$$h = OH = OF_1 + F_1F + FH = -f_1 + x_F + f,$$
$$h' = O'H' = O'F_2' + F_2'F' + F'H' = -f_2' + x_{F'}' + f'.$$

Отсюда

$$h = n_1(n_2 - n_3)\frac{R_1 d}{D}, \quad h' = -n_3(n_1 - n_2)\frac{R_2 d}{D}. \tag{12.15}$$

Обычно показатели преломления крайних сред n_1 и n_3 одинаковы. Полагая в этом случае $n_1 = n_3 = 1$, $n_2 = n$, получим

$$f = -f' = -n\frac{R_1 R_2}{D}, \tag{12.16}$$

$$h = (n-1)\frac{R_1 d}{D}, \quad h' = (n-1)\frac{R_2 d}{D}, \tag{12.17}$$

$$D = (n-1)[n(R_1 - R_2) - d(n-1)], \tag{12.18}$$

$$e = HH' = d - h + h'; \tag{12.19}$$

через e здесь обозначено расстояние главной плоскости H' от главной плоскости H (см. рис. 48).

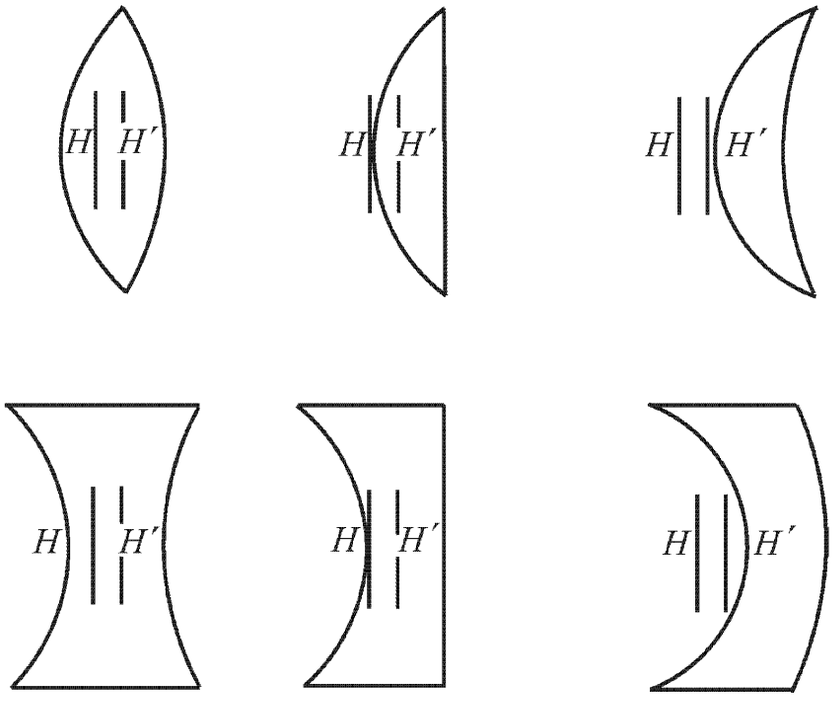


Рис. 49

Если линза не очень толстая, а разность $R_1 - R_2$ не слишком мала, то в выражении (12.18) слагаемым $d(n-1)$ можно

пренебречь. В этом приближении

$$\frac{1}{f} = -\frac{1}{f'} = (n-1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right), \tag{12.20}$$

$$h = \frac{R_1}{n(R_1 - R_2)}d, \quad h' = \frac{R_2}{n(R_1 - R_2)}d, \tag{12.21}$$

$$\frac{h}{h'} = \frac{R_1}{R_2}, \tag{12.22}$$

$$e = \frac{n-1}{n}d. \tag{12.23}$$

На рис. 49 изображены типичные линзы с указанием положений их главных плоскостей ($n = 1{,}5$).

ЗАДАЧИ

1. Для определения углового увеличения зрительной трубы *методом Рамсдена* (1735–1800) трубу устанавливают на бесконечность. Вывернув объектив, устанавливают на его место предмет определенной величины (экран с вырезом). Окуляр трубы дает действительное изображение взятого предмета. Пусть L — величина предмета, а l — величина его изображения. Показать, что угловое увеличение зрительной трубы равно L/l.

Р е ш е н и е. Примем за начала координатных систем фокальные точки окуляра. Тогда в формуле (11.17) следует положить $X = f_1'$, $f = f_2$. Это дает $L/l = f_1'/f_2$, т.е. увеличение трубы (см. § 11, п. 10).

2. Для определения фокусного расстояния собирательной линзы Бессель (1784–1846) предложил следующий метод.

С помощью линзы на экране получается действительное изображение предмета. Пусть A — расстояние от предмета до его изображения. Тогда $A = \xi' + e - \xi$. Исключая с помощью этого соотношения ξ' из (11.14), получим

$$\xi^2 + (A-e)\xi + (A-e)f = 0. \tag{12.24}$$

Если

$$A - e > 4f, \tag{12.25}$$

то уравнение (12.24) имеет два вещественных корня ξ_1 и ξ_2. В этом случае существуют два положения линзы, при которых на экране получаются действительные изображения предмета (при неизменном расстоянии между предметом и экраном). Чтобы перейти от одного изображения к другому, надо сместить линзу на расстояние $a = \xi_1 - \xi_2 = \sqrt{(A-e)^2 - 4f(A-e)}$, откуда

$$f = -f' = \frac{(A-e)^2 - a^2}{4(A-e)}. \tag{12.26}$$

Величины A и a можно измерить. Величина же e — расстояние между главными плоскостями — неизвестна. Для ее определения можно взять другое расстояние A_1 между предметом и экраном и измерить соответствующее смещение линзы a_1. Получится выражение вида (12.26), в котором A и a заменены на A_1 и a_1. Сравнивая эти два выражения, можно вычислить e. Для упрощения расчета можно пренебречь e^2 по сравнению с A^2. Это дает

$$f = -f' = \frac{A^2 - a^2}{4A} - \frac{A^2 + a^2}{4A^2}e. \tag{12.27}$$

§ 13. Ограничение лучей при помощи диафрагм

1. Четкие изображения, как правило, получаются только *в параксиальных лучах*. Непараксиальные лучи на практике устраняются *диафрагмами*. Роль диафрагм могут играть также оправы линз или зеркал.

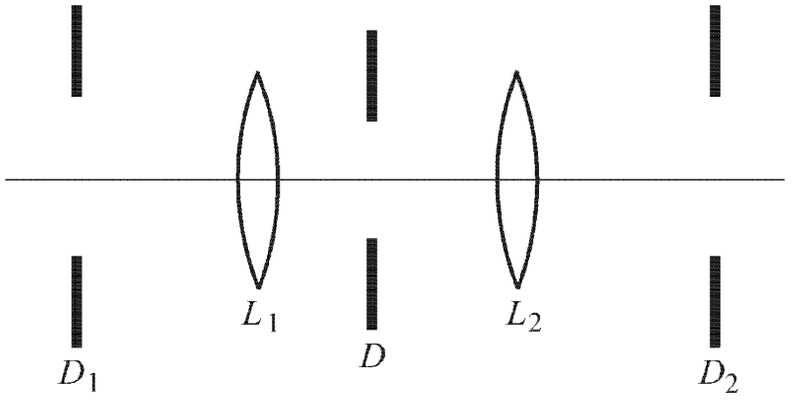


Рис. 50

Пусть D (рис. 50) — какая-либо диафрагма, а D_1 — ее изображение в параксиальных лучах впереди стоящими линзами. Если диафрагму D заменить диафрагмой D_1, то D_1 будет так же ограничивать параксиальные лучи, как и D. В самом деле, в приближении параксиальной оптики всякий луч, проходящий через край отверстия диафрагмы D_1, проходит и через край отверстия диафрагмы D. Следовательно, в этом приближении все реальные диафрагмы можно заменить их *изображениями*, получаемыми с помощью впереди стоящих линз. Тем самым все диафрагмы мысленно как бы переносятся в пространство предметов. Этим достигается то упрощение, что при исследовании действия диафрагм можно *отвлечься от преломления лучей*. Точно так же все реальные диафрагмы можно мысленно перенести в пространство изображений, заменив эти диафрагмы их изображениями, получаемыми с помощью позади стоящих линз.

2. Указанным способом мысленно перенесем все диафрагмы в пространство предметов (рис. 51). Воображаемая диафрагма DD', которая видна под наименьшим углом из точки предмета P, лежащей на главной оптической оси системы, всего более ограничивает лучи, исходящие из P. Она называется *входным зрачком* или *входным отверстием* системы. Реальная диафрагма, изображением которой является входной зрачок, носит название *апертурной* или *действующей диафрагмы*. Если апертурная диафрагма находится перед передней линзой, то она совпадает с входным зрачком. Таким образом, *апертурная диафрагма всего более диафрагмирует лучи, исходящие из точки предмета, лежащей на главной оптической оси системы*. Поэтому от

размеров апертурной диафрагмы зависит *яркость изображения*. При изменении апертурной диафрагмы меняется *светосила прибора*. Угол $2u$, под которым виден входной зрачок из точки предмета P, называется *апертурным углом со стороны предмета*, или *углом раскрытия*. Его называют также *апертурой системы*. Изображение апертурной диафрагмы в параксиальных лучах, получаемое с помощью позади стоящих линз, называется *выходным зрачком* или *выходным отверстием* системы. Очевидно, выходной зрачок есть изображение входного зрачка в параксиальных лучах, даваемое всей системой. Выходной зрачок всего более диафрагмирует лучи, проведенные из точки P', являющейся изображением точки предмета P, которая лежит на главной оптической оси. Угол $2u'$, под которым выходной зрачок виден из P', называется *апертурным углом со стороны изображения*, или *углом проекции* системы (см. рис. 52).

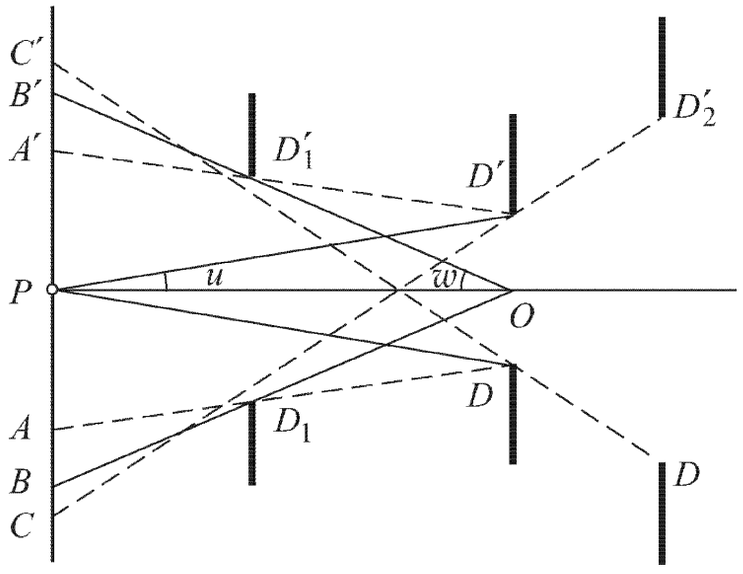


Рис. 51

В зрительных трубах апертурной диафрагмой часто служит край объектива. В таком случае этот край играет роль и входного зрачка, а его параксиальное изображение, даваемое окуляром, будет выходным зрачком. Если держать трубу на некотором расстоянии от глаза против светлого фона, то выходной зрачок будет виден как действительное или мнимое изображение.

В некоторых случаях апертурной диафрагмой может служить *радужная оболочка глаза*. Ее параксиальное изображение, даваемое роговой оболочкой и водянистым телом глаза, называется *зрачком глаза*.

Лучи, проходящие через центр апертурной диафрагмы, называются *главными лучами*. В приближении параксиальной оптики главный луч проходит также через центры входного и выходного зрачков.

Если предмет находится на бесконечности, то входным зрачком служит изображение той диафрагмы в пространстве предметов, которое имеет наименьший диаметр. Аналогично, если изображение получается в бесконечности, то выходным зрачком

служит изображение той диафрагмы в пространстве изображений, которое имеет наименьший диаметр.

3. Если точка предмета лежит на главной оптической оси системы, то всякий луч, проходящий через входной зрачок, пройдет и через оптическую систему. Но если луч исходит из точки предмета, не лежащей на главной оптической оси, то этот луч может и не пройти через оптическую систему даже в том случае, когда он прошел через входной зрачок. Он может быть задержан другими диафрагмами. Чтобы разобрать этот вопрос, рассмотрим сначала только главные лучи.

Та диафрагма, которая всего больше ограничивает главные лучи, называется *диафрагмой поля зрения*. Если бы входной зрачок был бесконечно малым, то все лучи, проходящие через систему, могли бы считаться главными. Поэтому при бесконечно малом входном зрачке поле зрения было бы резко ограничено и определялось размерами диафрагмы поля зрения. Изображение диафрагмы поля зрения, даваемое впереди стоящими линзами, называется *входным окном*, или *входным люком* системы. Изображение диафрагмы поля зрения, даваемое сзади стоящими линзами, называется *выходным окном*, или *выходным люком*. Очевидно, выходной люк есть изображение входного люка, даваемое всей системой. На рис. 51 входной люк изображен диафрагмой D_1D_1'. При бесконечно малом входном зрачке угол поля зрения со стороны предмета равен углу $2w$, под которым входной люк виден из центра входного зрачка. Аналогично определяется угол поля зрения $2w'$ со стороны изображения.

В действительности входной зрачок имеет *конечные размеры*. Это может повести к нарушению резкости границы поля зрения. В самом деле, из рис. 51 видно, что любой луч, исходящий из точек, лежащих между A и A', и прошедший через входной зрачок, пройдет и через оптическую систему. Лучи из точки B' пройдут через оптическую систему, если они направлены в нижнюю половину входного зрачка; лучи, направленные в верхнюю половину входного зрачка, задерживаются диафрагмой поля зрения и через оптическую систему не проходят. Для лучей, выходящих из B', верхняя половина входного зрачка оказывается как бы закрытой диафрагмой поля зрения. Наконец, лучи, исходящие из точек, лежащих дальше C и C', через оптическую систему вообще не пройдут. Таким образом, на краю поля зрения будет наблюдаться непрерывное ослабление освещенности. Оно называется *затенением* или *виньетированием*. При наличии виньетирования поле зрения ограничено не резко. *Чтобы виньетирования не было, необходимо, чтобы входной люк системы лежал в плоскости предмета. Поэтому почти во всех визуальных оптических системах для устранения виньетиро-*

вания применяют диафрагмы поля зрения, помещаемые обычно в передней фокальной плоскости окуляра.

4. В оптических инструментах изображение обычно воспринимается на какую-нибудь поверхность, например матовое стекло, фотографическую пластинку, сетчатку глаза и т.п. Эта поверхность перпендикулярна к главной оптической оси системы и в первом приближении может быть принята за плоскость (*плоскость изображения*). Между тем в большинстве случаев предмет, изображение которого надо получить, бывает *пространственным*, а не представляет собой плоскости или какой-либо простой поверхности. Чему же соответствует картина, получаемая в плоскости изображения оптической системы?

Для ответа на этот вопрос найдем в пространстве предмета плоскость, оптически сопряженную с плоскостью изображения. Она называется *плоскостью установки* или *плоскостью наводки*. Спроецируем предмет из центра входного зрачка на плоскость установки. Эта проекция и будет тем объектом, изображение которого более или менее резко передает оптическая система. Действительно, главный луч, исходящий из любой точки предмета, проходит также через ее проекцию на плоскость установки. Он является центром пучка лучей, исходящих из этой точки. Если точка предмета лежит в плоскости установки, т.е. совпадает со своей проекцией, то ее изображение получится резким. Если же она не лежит в плоскости установки, то ее изображение получится в виде *кружка рассеяния*, центр которого является изображением проекции этой точки на плоскость установки. Чем больше апертурная диафрагма, тем шире пучки, исходящие из точек предмета, а следовательно, тем больше размеры соответствующих кружков рассеяния. Если бы пучки совсем не были ограничены диафрагмами или краями линз, то кружки рассеяния занимали бы всю плоскость изображения и изображение не могло бы вообще быть получено. Отсюда ясно, насколько важную роль играет диафрагмирование при получении изображений объемных предметов.

5. Для резкости изображения необходимо, чтобы диаметр кружка рассеяния d не превосходил известного предела (например, 0,1 мм). Этим определяется *глубина резко изображаемого пространства*. Найдем выражение для этой глубины в приближении параксиальной оптики. Пусть DD' (рис. 52) — входной, а D_1D_1' — выходной зрачки, P — точка предмета на оптической оси, P' — ее изображение, EE' — плоскость изображения. Опишем из P', как из центра, кружок диаметром d. Соединим диаметрально противоположные крайние точки A и A' этого кружка с краями D_1 и D_1' выходного зрачка. В пересечении с оптической осью получатся точки P_1' и P_2'. Пусть P_1 и P_2 — сопря-

женные им точки в пространстве предметов. Расстояние $|\delta X| = P_1P_2$ и определит глубину резко изображаемого пространства. Пусть X и X' — координаты сопряженных точек P и P' отно-

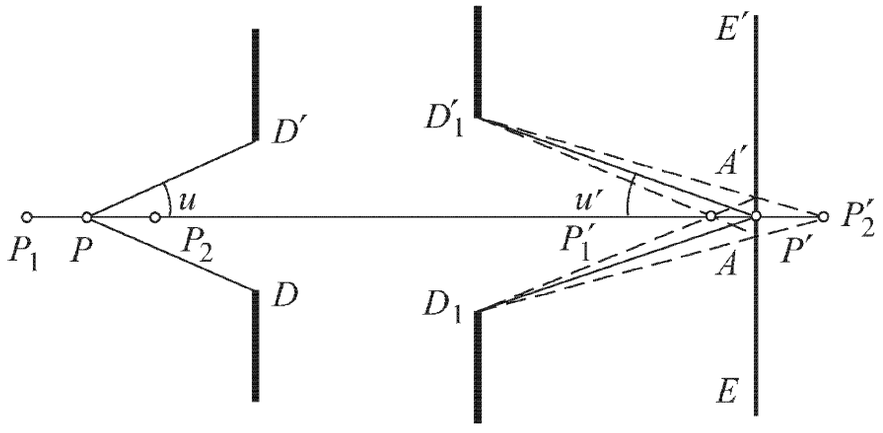


Рис. 52

сительно главных фокусов системы. Из уравнения (11.16) с точностью до членов второго порядка получаем $X'\,\delta X + X\,\delta X' = 0$. С той же точностью, как видно из рис. 52, $\delta X' = d/\operatorname{tg} u' \approx d/u'$. Следовательно,

$$|\delta X| = \left|\frac{X}{X'}\right| \frac{d}{u'}.$$

Комбинация этой формулы с формулами (11.12) и (11.17) приводит к результату

$$|\delta X| = \frac{|X|d}{fu}. \tag{13.1}$$

Если p — расстояние от предмета до входного зрачка, а r — радиус последнего, то $u = r/p$. Для объектива фотоаппарата можно принять $|X| = p$. Тогда

$$|\delta X| = \frac{p^2}{fr} d. \tag{13.2}$$

Каждому фотографу известно, что при недостатке глубины изображения надо либо уменьшить апертурную диафрагму, либо удалиться от фотографируемого объекта.

6. Положение входного зрачка по отношению к плоскости установки определяет *перспективу*, в которой виден предмет. Если входной зрачок расположен за предметом в направлении лучей, то при проецировании предмета из центра входного зрачка на плоскость установки ближние к оптической системе части предмета получат большее увеличение, чем дальние. При таком расположении близкие предметы на изображении будут получаться бóльшими, чем равные им более далекие предметы (*нормальная* или *энтоцентрическая перспектива*). Напротив, если входной зрачок лежит перед предметом, то на изображе-

нии близкие предметы получатся меньшими, чем равные им более далекие предметы (*гиперцентрическая перспектива*). Наконец, в промежуточном случае, когда входной зрачок удален в бесконечность, величина изображения не зависит от расстояния предмета до оптической системы (*телецентрическая перспектива*). Изменение расстояния до предмета сказывается в этом случае *не на величине изображения, а на его резкости.* Поэтому такая перспектива применяется в измерительных микроскопах, так как результат измерения при этом не зависит от точной установки измеряемого предмета. Чтобы ее осуществить, достаточно поместить апертурную диафрагму в задний фокус линзы L_1 (см. рис. 50).

В зрительных трубах, когда предмет практически находится в бесконечности, необходимо, чтобы величина изображения, проецирующегося на шкалу с делениями, не зависела от точной установки шкалы. Этого можно достигнуть, поместив апертурную диафрагму в передний фокус линзы L_2 (рис. 50). Тогда выходной зрачок удалится в бесконечность, главный луч будет параллелен оптической оси и встретит плоскость шкалы на одном и том же удалении от оптической оси. Следовательно, величина изображения на шкале не будет зависеть от ее положения.

§ 14. Астигматические пучки лучей. Каустика

1. При наличии непараксиальных лучей, а также при отсутствии осевой симметрии оптической системы (примером может служить цилиндрическая линза) сферическая волна, исходящая из светящейся точки, после прохождения через оптическую систему перестает быть сферической. В результате светящаяся точка уже не будет изображаться оптической системой в виде точки. Связанные с этим искажения оптических изображений называются *геометрическими* или *лучевыми аберрациями оптических систем.* Помимо лучевых существуют еще *хроматическая аберрация*, т.е. появление окрашенных каемок в изображении, когда оно получается в белом свете, а также *волновые* или *дифракционные аберрации*, обусловленные дифракцией света. Учет последних необходим при рассмотрении разрешающей способности оптических и спектральных приборов, а также других тонких вопросов, связанных с получением оптических изображений. Однако сначала мы отвлечемся от хроматической и дифракционных аберраций, предполагая, что световые пучки, формирующие изображение, — *монохроматические* и к ним применима *геометрическая оптика.*

2. Напомним сначала необходимые сведения из дифференциальной геометрии. Пусть S — какая-то гладкая поверхность,

а O — произвольная точка на ней (рис. 53 *а*). Нормаль к поверхности S в точке O обозначим через $\mathbf{N}$. Проведем через $\mathbf{N}$ плоскость П, пересекающую поверхность S вдоль некоторой кривой L. Если плоскость П вращать вокруг нормали $\mathbf{N}$ в пределах 180°, то кривизна кривой L, вообще говоря, будет изменяться, достигая в каком-то положении L_1 максимума, а в другом положении L_2 — минимума. В дифференциальной геометрии доказывается, что нормальные сечения поверхности S максимальной и минимальной кривизны взаимно перпендикулярны. Эти сечения называются *главными нормальными сечениями поверхности* S, проходящими через точку O. Линии, по которым они пересекают поверхность S в окрестности точки O, называются *линиями кривизны*, а радиусы кривизны последних R_1 и R_2 — *главными радиусами кривизны поверхности* S. Если знаки R_1 и R_2 одинаковы, то линии кривизны обращены вогнутостями *в одну и ту же сторону*; если разные, то *в противоположные стороны*. Если в точке O $R_1 = R_2$, то эта точка называется *точкой закругления поверхности*.

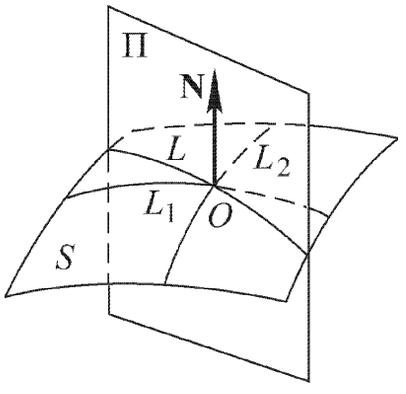


Рис. 53 *а*

3. Пусть теперь поверхность S является волновым фронтом. Возьмем на ней какую-то элементарную площадку dS. Бесконечно узкий пучок лучей волнового фронта, проходящий через эту площадку, называется *астигматическим*, если главные радиусы кривизны R_1 и R_2 не совпадают между собой. Луч, проходящий через центр площадки dS, называется *главным* [1]).

Вообще говоря, две нормали к поверхности S не пересекаются в одной точке. Однако нормали, восстановленные в бесконечно близких точках, лежащих на одной и той же линии кривизны, пересекаются в одной точке (если пренебречь величинами высших порядков малости). Заметив это, возьмем в качестве элемента dS бесконечно малую площадку $ABDC$, имеющую форму криволинейного четырехугольника, сторонами которого служат линии кривизны поверхности S (рис. 53 *б*). Лучи, лежащие в главных сечениях элемента волнового фронта dS, проходящих через его центр, сходятся в различных точках F_1 и F_2, называемых *фокальными точками*; расстояние F_1F_2 называется *астигматической разностью элементарного пучка*. Лучи, проходя-

[1]) При отсутствии симметрии площадки понятие ее центра, строго говоря, теряет смысл. Однако и в этом случае можно указать какую-то, хотя и не совсем четко определенную, точку в середине площадки, которую можно рассматривать как центр последней.

щие через противоположные стороны AB и CD, пересекутся в точках M_1 и N_1 соответственно. Лучи, исходящие из всякой промежуточной линии кривизны между сторонами AB и CD, сойдутся в какой-то промежуточной точке бесконечно малого отрезка M_1N_1, проходящего через фокальную точку F_1. Анало-

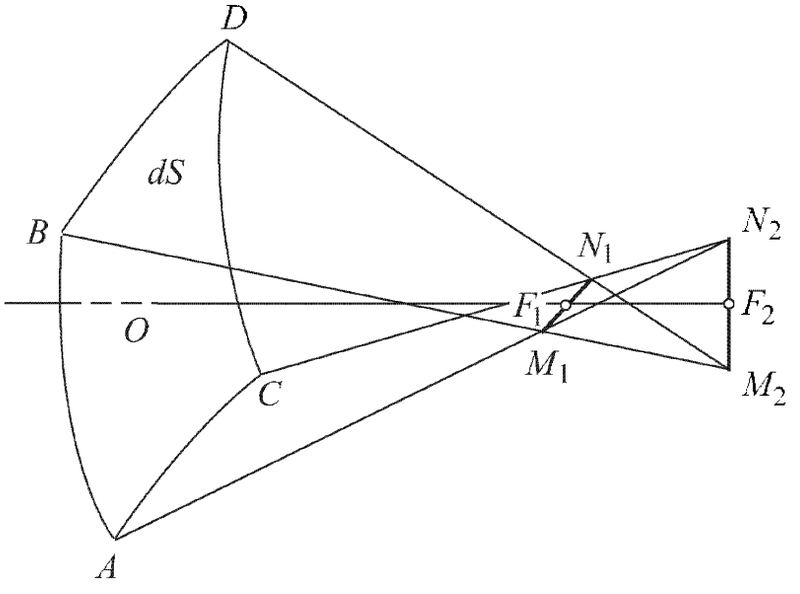


Рис. 53*б*

гично, лучи, исходящие из линий кривизны между сторонами BD и AC, пересекутся в какой-то точке другого бесконечно малого отрезка M_2N_2, проходящего через вторую фокальную точку F_2.

Отрезки M_1N_1 и M_2N_2 называются *фокальными отрезками* элементарного астигматического пучка лучей, исходящих от элемента волнового фронта $ABCD$. Если элементарный астигматический пучок лучей исходит из светящейся точки, то фокальные отрезки M_1N_1 и M_2N_2 принято называть *изображениями* этой точки, даваемыми *астигматическим пучком*, хотя точечного изображения в этом случае и не существует. Если элемент $ABCD$ обладает симметрией прямоугольника, то отрезки M_1N_1 и M_2N_2 будут перпендикулярны между собой, а также к главному лучу пучка. В общем случае это может быть и не так. При бесконечном уменьшении поперечных размеров пучка лучей отрезки M_1N_1 и M_2N_2 стягиваются в фокальные точки F_1 и F_2.

Таким образом, в отличие от гомоцентрического пучка, бесконечно узкий астигматический пучок дает не одно, а *два точечных изображения* F_1 и F_2 светящейся точки. Конечный пучок лучей можно разложить на элементарные астигматические пучки, каждому из которых соответствует пара фокальных точек. Геометрическое место этих точек есть двухлистная поверхность, называемая *каустической поверхностью*, или *каустикой*.

ЗАДАЧА

Точечный источник света P помещен в прозрачной однородной среде, ограниченной плоскостью. Световые лучи, исходящие из P, испытывают преломление на этой плоскости. Найти для них каустическую поверхность.

Р е ш е н и е. Каустика преломленных лучей состоит из двух листов. Один из них есть геометрическое место фокальных точек *меридиональных лучей*, т.е. лучей, лежащих в плоскости падения главного луча элементарного астигматического пучка. Другой — геометрическое место фокальных точек *экваториальных лучей*, т.е. лучей, лежащих в перпендикулярной плоскости, проходящей через главный луч элементарного пучка.

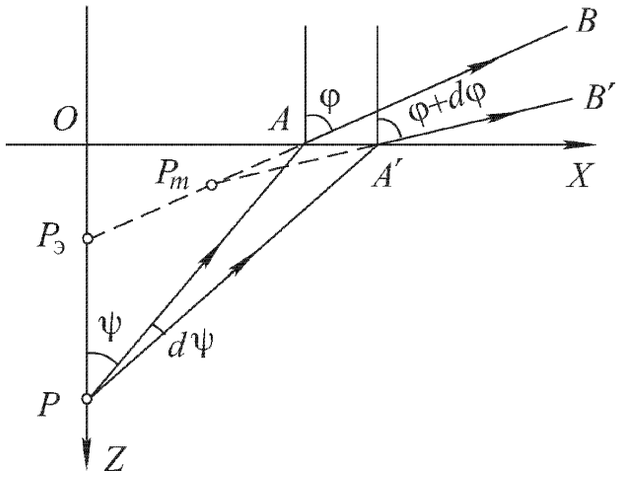


Рис. 54

Пусть n — показатель преломления среды, в которой помещен источник P. Показатель преломления пространства, с которым граничит эта среда, примем за единицу. Введем прямоугольную систему координат с началом O, расположенным на границе среды. За ось Z примем нормаль к поверхности среды, направив эту ось в сторону точки P (рис. 54). Ввиду осевой симметрии достаточно найти сечение каустической поверхности плоскостью, проходящей через ось Z. Прямую, вдоль которой эта плоскость пересекает границу среды, примем за ось X.

Найдем сначала каустику для меридиональных преломленных лучей. Уравнение преломленного луча AB будет

$$z = -\operatorname{ctg}\varphi(x - h\operatorname{tg}\psi),$$

где h — расстояние от точки P до границы среды, ψ — угол падения из среды на эту границу, φ — угол преломления. Для бесконечно близкого луча $PA'B'$ углы φ и ψ получат приращения $d\varphi$ и $d\psi$. Приращение координаты z при одном и том же значении абсциссы x при этом будет равно

$$dz = \frac{x - h\operatorname{tg}\psi}{\sin^2\varphi}\,d\varphi + \frac{h\operatorname{ctg}\varphi}{\cos^3\psi}\,d\psi,$$

или с использованием закона преломления $\sin\varphi = n\sin\psi$:

$$dz = \frac{x - h\operatorname{tg}\psi}{\sin^2\varphi}\,d\varphi + \frac{h\operatorname{ctg}\varphi\cos\varphi}{n\cos^3\psi}\,d\varphi.$$

Координаты x_m и z_m точки P_m, в которой пересекаются продолжения бесконечно близких меридиональных лучей AB и $A'B'$, найдутся отсюда, если приращение dz приравнять нулю. Это дает

$$x_m = h\left(\operatorname{tg}\psi - \frac{\cos^2\varphi\sin\varphi}{n\cos^3\psi}\right), \quad z_m = h\frac{\cos^3\varphi}{n\cos^3\psi}. \tag{14.1}$$

Это и есть *уравнение каустики для меридиональных лучей.* Используя его, нетрудно вывести формулу

$$l_m = l\frac{\cos^2\varphi}{n\cos^2\psi}, \tag{14.2}$$

где l — расстояние от предмета P до точки выхода A преломленного луча, а l_m — расстояние меридионального изображения P_m до той же точки.

Еще проще находится *каустика для экваториальных лучей.* Пусть PAB — один из лучей, исходящих из точки P (см. рис. 54). Если этот луч вращать вокруг перпендикуляра OP к преломляющей поверхности, то получится конус падающих и соответствующий ему конус преломленных лучей. Вершиной второго конуса будет точка $P_э$, в которой продолжение преломленного луча AB пересекает перпендикуляр PO. Бесконечно малые пучки падающих и преломленных лучей, лежащих на поверхностях указанных конусов, для которых луч PAB является главным, будут, очевидно, расположены в плоскостях, перпендикулярных к плоскости падения луча PAB. Значит, лучи этих пучков будут экваториальными, а точка $P_э$ — изображением в этих лучах. Таким образом, все фокальные точки экваториальных лучей расположатся на перпендикуляре PO, т.е. каустика таких лучей выродится в отрезок этого перпендикуляра. Расстояние $l_э$ точки $P_э$ от точки выхода преломленного луча AB будет $P_эA = l\sin\psi/\sin\varphi$, т.е.

$$l_э = \frac{l}{n}. \tag{14.3}$$

Результаты вычислений представлены на рис. 55 для $n = 1{,}5$ (стекло). Каустика экваториальных лучей представляется вертикальным отрезком OP', длина которого равна h/n. Сечение каустики меридиональных лучей

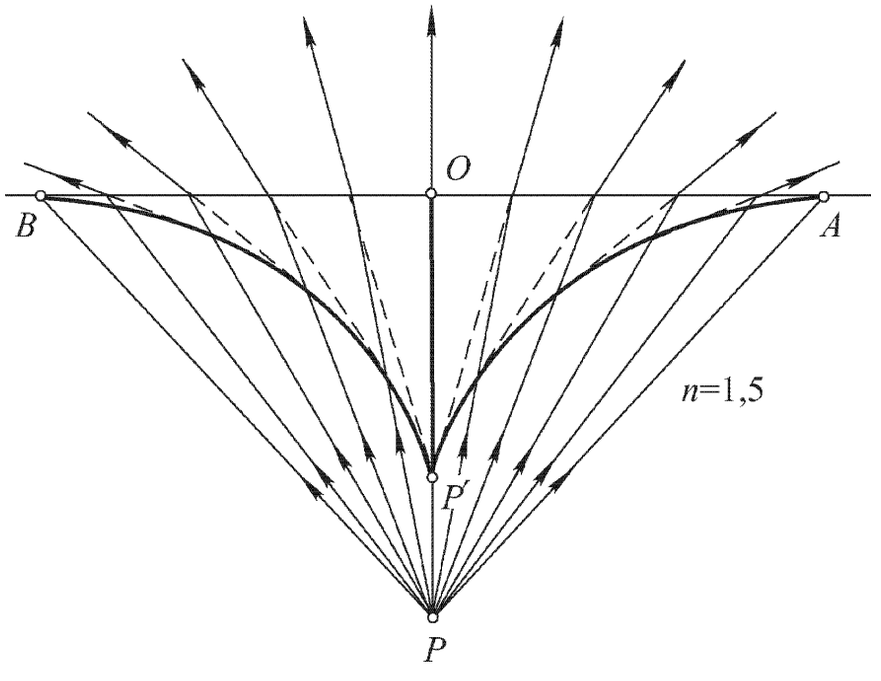


Рис. 55

плоскостью рисунка есть кривая $BP'A$, для которой P' является *точкой возврата.* Продолжения преломленных лучей, изображенные на рисунке штриховыми прямыми, касаются кривой $BP'A$. Таким образом, если начертить еще сечение волнового фронта вышедших преломленных лучей, то

для этого сечения кривая $BP'A$ будет *эволютой*. Сама каустика меридиональных лучей получится от вращения этой эволюты вокруг вертикальной прямой PO. Для параксиальных пучков фокальные точки меридиональных и экваториальных лучей совпадают и получаются в P'. Крайние точки A и B каустики меридиональных лучей получаются от пересечения предельных лучей полного отражения PA и PB с границей среды. Предлагаем читателю начертить каустику меридиональных и экваториальных лучей для случая $n < 1$.

§ 15. Геометрические аберрации центрированных систем

1. Геометрические аберрации в центрированных системах, т.е. отступления от параксиальной оптики, вызываются *непараксиальными пучками лучей*, участвующими в образовании оптических изображений. Дадим классификацию таких аберраций.

Произвольный луч в пространстве предметов можно задать, указав прямоугольные координаты x, z и η, ζ точек его пересечения с предметной плоскостью (т.е. плоскостью, проходящей через изображаемую точку P перпендикулярно к главной оптической оси) и плоскостью входного зрачка. После прохождения через оптическую систему луч пересечет плоскость параксиального изображения в точке с координатами y', z'. Координаты самого параксиального изображения (называемого в дальнейшем также *параксиальным фокусом*) обозначим через y'_0, z'_0. Тогда разности $\Delta y' = y' - y'_0$, $\Delta z' = z' - z'_0$ и можно принять за меру отступлений оптики реальной системы от предельного случая параксиальной оптики.

Координаты y', z' будут функциями аргументов y, z, η, ζ:

$$y' = f_1(y, z, \eta, \zeta), \quad z' = f_2(y, z, \eta, \zeta).$$

Для классификации геометрических аберраций разложим эти функции в степенные ряды по своим аргументам. Линейные члены этих разложений, пропорциональные y и z, соответствуют параксиальной оптике. Линейные члены по η и ζ не войдут, так как в параксиальном приближении y' и z' не зависят от наклона лучей, выходящих из точки P. Не могут войти и члены четных степеней ввиду осевой симметрии оптической системы. Из всего этого следует, что разложения в степенные ряды отклонений $\Delta y' = y' - y'_0$ и $\Delta z' = z' - z'_0$ могут содержать только члены *нечетных степеней* по y, z, η, ζ, причем эти разложения могут начаться с членов, степень которых *не ниже трех*. Считая аргументы y, z, η, ζ малыми, сохраним в разложениях только члены третьей степени. Аберрации, вычисленные в этом приближении, называются *первичными*, или *аберрациями третьего порядка*.

Члены пятой степени вызывают аберрации *пятого порядка*, и т.д. Мы ограничимся только аберрациями третьего порядка.

Переходя к векторным обозначениям, введем три вектора, перпендикулярных к главной оптической оси системы:

$$\mathbf{r} = y\mathbf{j} + z\mathbf{k}, \quad \mathbf{r}' = y'\mathbf{j} + z'\mathbf{k}, \quad \boldsymbol{\sigma} = \eta\mathbf{j} + \zeta\mathbf{k},$$

где $\mathbf{j}$ и $\mathbf{k}$ — единичные векторы, указывающие направления координатных осей Y и Z. Вектор $\mathbf{r}$ определяет положение точки-предмета P в предметной плоскости, вектор $\mathbf{r}'$ — точки пересечения выходящего луча с параксиальной плоскостью изображения, вектор $\boldsymbol{\sigma}$ — точки пересечения падающего луча с плоскостью входного зрачка. Вектор $\Delta\mathbf{r}' = \Delta y'\mathbf{j} + \Delta z'\mathbf{k}$ можно разложить по векторам $\boldsymbol{\sigma}$ и $\mathbf{r}$. Ввиду осевой симметрии коэффициенты этих разложений могут зависеть только от «инвариантов вращения» $\boldsymbol{\sigma}^2$, $(\boldsymbol{\sigma}\mathbf{r})$ и $\mathbf{r}^2$. Поэтому с точностью до членов третьей степени включительно

$$\Delta\mathbf{r}' = [A\boldsymbol{\sigma}^2 + B(\boldsymbol{\sigma}\mathbf{r}) + C\mathbf{r}^2]\boldsymbol{\sigma} + [D\boldsymbol{\sigma}^2 + E(\boldsymbol{\sigma}\mathbf{r}) + F\mathbf{r}^2]\mathbf{r}, \qquad (15.1)$$

где A, B, C, D, E, F — постоянные коэффициенты, зависящие от устройства оптической системы и от положения предметной плоскости.

В дальнейшем будем понимать под σ радиус входного зрачка, выделяя тем самым падающие лучи (или их продолжения), проходящие через точки окружности входного зрачка. Тогда вектор $\Delta\mathbf{r}'$ окажется разложенным по степеням радиуса входного зрачка. Назовем *аберрационной кривой* кривую, по которой плоскость параксиального изображения пересекает пучок лучей, проведенных из точки-объекта P через окружность входного зрачка. Изображением точки P в параксиальной плоскости изображения будет не точка, а какое-то пятнышко, ограниченное аберрационной кривой. Для наглядности можно представить, что в качестве апертурной взята ирисовая диафрагма, радиус которой можно непрерывно менять. Тогда разложение (15.1) определит, как в рассматриваемом приближении будет меняться аберрационная кривая при изменении радиуса этой диафрагмы. Отступления от параксиальной оптики определяются, конечно, суммой (15.1) в целом, а не отдельными слагаемыми, из которых она состоит. Однако при классификации аберраций имеет смысл рассматривать каждое слагаемое в отдельности и рассуждать так, как если бы остальных слагаемых не было совсем. Тогда, в зависимости от степени σ, все аберрации третьего порядка можно разбить на четыре группы, которые мы и рассмотрим.

2. С ф е р и ч е с к а я а б е р р а ц и я. Эта аберрация вызывается членом третьей степени $A\sigma^2\boldsymbol{\sigma}$, так что при наличии одной только сферической аберрации $|\Delta\mathbf{r}'| = A\sigma^3 = \text{const}$.

Следовательно, аберрационной кривой будет *окружность* с центром в параксиальном фокусе и с радиусом $A\sigma^3$. Каждая точка будет изображаться в виде кружка рассеяния, радиус которого пропорционален кубу радиуса входного зрачка и не зависит от положения этой точки. Освещенность кружка рассеяния быстро убывает от центра к краям.

Происхождение сферической аберрации наглядно пояснено на рис. 56. Пусть точечный объект лежит на главной оптической оси системы. Выходящие из него параксиальные лучи встречают плоскость параксиальных изображений в точке G. Лучи, проходящие через окружность выходного зрачка DD', сойдутся на оптической оси в точке M, которая может лежать как ближе, так и дальше G. Лучи, проходящие через какую-либо окружность в плоскости выходного зрачка, концентрическую с окружностью DD', сойдутся на оптической оси между точками M и G. Расстояние MG называется *продольной сферической аберрацией.* Если в плоскости параксиальных изображений AA' поместить экран, то на нем получится светлый кружок радиуса GA. Радиус GA называется *поперечной сферической аберрацией.* С точностью до членов третьего порядка включительно поперечная аберрация пропорциональна кубу апертуры $2u$. Отсюда следует, что продольная аберрация пропорциональна квадрату апертуры.

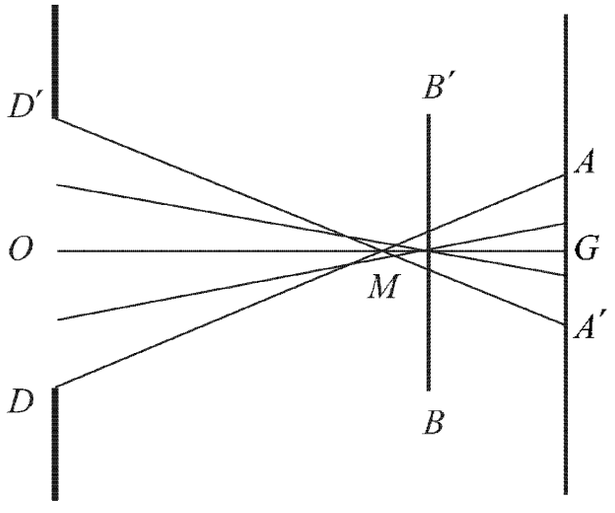


Рис. 56

Если экран перемещать от плоскости AA' по направлению к M, то радиус кружка рассеяния сначала будет уменьшаться, а затем начнет увеличиваться. Нетрудно показать, что наименьший кружок рассеяния получится, когда экран займет положение BB' на расстоянии $^3/_4MG$ от плоскости параксиальных изображений AA'. Однако плоскость BB', строго говоря, не будет плоскостью наилучшей отчетливости изображения. При нахождении последней необходимо учитывать не только размеры кружка рассеяния, но и *распределение освещенности* внутри этого кружка. Исходя из дифракционных соображений, можно показать, что при наличии одной только сферической аберрации плоскость наилучшей отчетливости изображения проходит посередине между точками M и G.

Сферическая аберрация есть единственная из геометрических аберраций, остающаяся и в том случае, когда точка-объект

находится на главной оптической оси системы. Все прочие геометрические аберрации в этом случае исчезают. Это верно в любом приближении, а не только для аберраций третьего порядка.

3. К о м а. Если в разложении (15.1) отличны от нуля только коэффициенты B и D, то соответствующая аберрация называется *комой*. В этом случае

$$\Delta\mathbf{r}' = B(\boldsymbol{\sigma}\mathbf{r})\boldsymbol{\sigma} + D\sigma^2\mathbf{r}.$$

Отсюда легко получить

$$\left[\Delta\mathbf{r}' - \left(D + \frac{B}{2}\right)\sigma^2\mathbf{r}\right]^2 = \left(\frac{B}{2}\,\sigma^2\mathbf{r}\right)^2.$$

Следовательно, аберрационной кривой будет *окружность* радиуса $\frac{1}{2}B\sigma^2 r$, центр которой смещен от параксиального фокуса в направлении вектора $\mathbf{r}$ на расстояние $(D + \frac{1}{2}B)\sigma^2 r$. Теперь легко получить представление о характере изображения точечного объекта при наличии одной только комы в отсутствие других аберраций. Для этого проведем в плоскости входного зрачка произвольную окружность, центр которой совпадает с центром зрачка. Лучи, исходящие из точечного объекта и проходящие через эту окружность, пересекут плоскость параксиального изображения также по окружности. Совокупность таких окружностей и даст изображение рассматриваемого точечного объекта в этой плоскости. Окружности имеют две прямолинейные огибающие, пересекающиеся в параксиальном фокусе

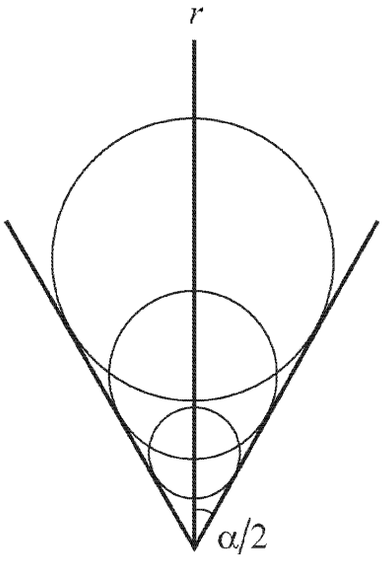


Рис. 57

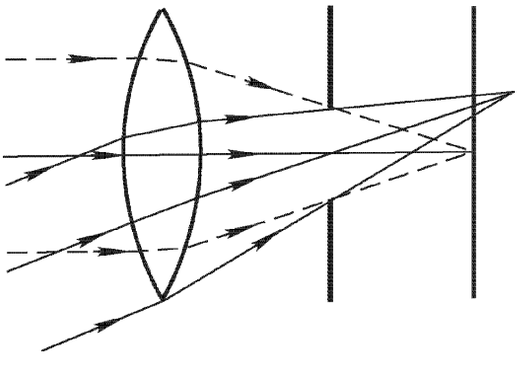


Рис. 58

и составляющие между собой угол α, определяемый соотношением $\sin(\alpha/2) = B/(B + 2D)$. Более подробное исследование, которое мы не приводим, показывает, что $B = 2D$, а потому $\alpha = 60°$. Направление вектора $\mathbf{r}$ является биссектрисой угла между огибающими (рис. 57). Изображение точки, таким образом, напоминает комету. Отсюда и произошло название «кома».

Происхождение комы ясно из рис. 58. Она обусловлена косыми пучками лучей.

4. Астигматизм косых пучков и искривление плоскости изображения. Эти аберрации удобно рассматривать совместно, так как обе они обусловлены членами первой степени по $\boldsymbol{\sigma}$ и второй степени по $\mathbf{r}$. Они возникают, когда оба коэффициента C и E или один из них отличны от нуля. Если все прочие коэффициенты равны нулю, то формула (15.1) переходит в

$$\Delta\mathbf{r}' = Cr^2\boldsymbol{\sigma} + E(\boldsymbol{\sigma}\mathbf{r})\mathbf{r}. \tag{15.2}$$

Для определения формы аберрационной кривой ось Y проведем через точку-объект P. Тогда $\mathbf{r} = y\mathbf{j}$. Уравнение окружности входного зрачка запишем в параметрической форме $\eta = \sigma\cos\varphi$, $\zeta = \sigma\sin\varphi$, где φ — центральный угол, рассматриваемый как параметр. Таким путем из предыдущего соотношения получаем уравнение аберрационной кривой:

$$\begin{aligned} \Delta y' &= (C+E)y^2\eta = (C+E)\sigma y^2\cos\varphi, \\ \Delta z' &= Cy^2\zeta = C\sigma y^2\sin\varphi. \end{aligned} \tag{15.3}$$

Это — эллипс, центр которого находится в параксиальном фокусе, оси параллельны координатным осям Y и Z, а их длины пропорциональны радиусу входного зрачка и квадрату расстояния изображаемой точки от главной оптической оси. Изображением точки будет светлое пятнышко, ограниченное аберрационной кривой. Это указывает на то, что пучок лучей, дающий изображение, — астигматический. При параллельном смещении экрана, на котором получается изображение, вдоль оптической оси оно по-прежнему сохраняет форму эллипса, но форма и размеры эллипса изменяются. При двух положениях экрана эллипс вырождается в прямолинейные отрезки, один из которых параллелен оси Y, а другой — оси Z.

Для доказательства поместим начало координат в центре выходного зрачка и будем характеризовать луч в пространстве изображений точками пересечения его с плоскостью выходного зрачка и с плоскостью параксиального изображения. Координаты первой точки будут $(0, \eta', \zeta')$, второй — $(x', y'+\Delta y', \Delta z')$. Если X, Y, Z — текущие координаты, то уравнение рассматриваемого луча запишется в виде

$$Y = \eta' + \frac{y' + \Delta y' - \eta'}{x'}X, \quad Z = \zeta' + \frac{\Delta z' - \zeta'}{x'}X.$$

Полагая здесь $\eta' = \zeta' = \Delta y' = \Delta z' = 0$, найдем поперечные координаты параксиального фокуса: $Y_0 = y'X/x'$, $Z_0 = 0$ (про-

дольная координата того же фокуса равна x'). Таким образом,

$$\begin{aligned} Y - Y_0 &= \left(1 - \frac{X}{x'}\right)\eta' + \frac{X}{x'}\Delta y', \\ Z - Z_0 &= \left(1 - \frac{X}{x'}\right)\zeta' + \frac{X}{x'}\Delta z'. \end{aligned} \tag{15.4}$$

В последних слагаемых в окрестности параксиального фокуса можно считать $X = x'$, так как это вносит ошибки высшего порядка малости, не учитываемые в рассматриваемом приближении. По той же причине в формулах (15.3) можно положить $y = y'/\beta_1$, $z = z'/\beta_1$, $\eta = \eta'/\beta_2$, $\zeta = \zeta'/\beta_2$, где β_1 и β_2 — линейные поперечные увеличения для объектов, лежащих соответственно в предметной плоскости и в плоскости входного зрачка, как они выражаются в параксиальном приближении. Учтя еще, что y и Y являются лишь различными обозначениями одной и той же ординаты, можем записать (15.3) в виде

$$\Delta y' = \frac{C' + E'}{x'}\,\eta' Y^2, \quad \Delta z' = \frac{C'}{x'}\,\zeta' Y^2,$$

где C' и E' — новые постоянные, а абсцисса x' введена в знаменатели для удобства. В результате всего этого формулы (15.4) в окрестности параксиального фокуса преобразуются в

$$\begin{aligned} Y - Y_0 &= \frac{\eta'}{x'}\,[x' - X + (C' + E')Y^2], \\ Z - Z_0 &= \frac{\zeta'}{x'}\,[x' - X + C'Y^2]. \end{aligned} \tag{15.5}$$

Если

$$X - x' = (C' + E')Y^2, \tag{15.6}$$

то $Y - Y_0 = 0$, т.е. эллипс вырождается в прямолинейный отрезок, параллельный оси Z. Это есть *фокальный отрезок*, образованный *меридиональными лучами*. Аналогично, если

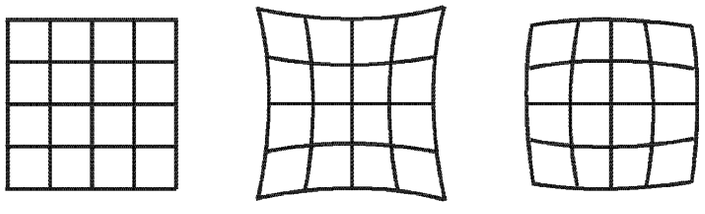
$$X - x' = C'Y^2, \tag{15.7}$$

то $Z - Z_0 = 0$, и эллипс переходит в другой фокальный отрезок, параллельный оси Y. Он образуется *экваториальными лучами*. От вращения кривых (15.6) и (15.7) вокруг главной оптической оси получаются две поверхности, касающиеся друг друга в общей точке пересечения их с главной оптической осью. Эти две поверхности и образуют каустику лучей, прошедших через оптическую систему. Вообще говоря, они имеют только одну общую точку. Каустика и есть та поверхность, в которую переходит плоскость изображения параксиальной оптики. Таким образом, имеет место не только астигматизм, но и *искривление поверхности изображения*.

5. Д и с т о р с и я. Коэффициент F не равен нулю, все прочие коэффициенты равны нулю. Поэтому из (15.1) получаем

$$\Delta r' = Fr^3.$$

Отсюда видно, что при наличии только рассматриваемой аберрации каждая точка изображается резко в виде точки, каковы бы ни были размеры диафрагмы. Однако отклонение изображения точки от соответствующего параксиального фокуса пропорционально кубу ее расстояния r от главной оптической оси. Поэтому происходит *искажение* (*дисторсия*) изображения. Прямые линии, проходящие через главную оптическую ось, изображаются в виде прямых. Все прочие прямые при изображении искривляются. При положительном F изображения точек смещаются относительно соответствующих параксиальных фокусов *наружу*, т.е. от главной оптической оси. Такая дисторсия называется *подушкообразной* (рис. 59 *б*). При отрицательном F смещения происходят внутрь — к главной оптической оси. Соответствующая дисторсия называется *бочкообразной* (рис. 59 *в*).



а *б* *в*

Рис. 59

Все эти выводы сохраняются в силе и для *аберраций высших порядков*. В общем случае дисторсия обусловлена всеми членами разложения $\Delta r'$, не содержащими степеней радиуса входного зрачка σ. В любом приближении при наличии одной только дисторсии точка всегда изображается в виде точки, каковы бы ни были размеры диафрагмы.

6. Устранение (точнее — ослабление) геометрических аберраций на практике достигается комбинацией различных линз (и зеркал), изготовляемых из оптического стекла с различными оптическими характеристиками. Полное устранение всех аберраций невозможно. На практике речь может идти только об определенном оптимуме, зависящем от задач, которые должен решать оптический прибор. Наиболее вредными из всех аберраций третьего порядка являются обычно сферическая аберрация и кома. В большинстве случаев их надо как можно больше ослабить. Уменьшая диафрагму, можно практически полностью устранить обе эти аберрации. После этого подбором линз надо устранить дисторсию, а затем астигматизм и изгиб поверхности изображения. Но уменьшение диафрагмы уменьшает яркость изображения и увеличивает дифракционные ошибки. Дисторсия вредна в фотографических объективах. В астрономических приборах ее можно допускать, так как она не влияет на резкость

изображения, а вызывает только искажение его, которое можно учесть вычислением.

§ 16. Хроматическая аберрация

1. Если используется белый свет, то в изображении возникают дополнительные аберрации. Действительно, показатель преломления зависит от длины волны (дисперсия света). Поэтому оптическая система дает не одно, а множество монохроматических изображений, отличающихся друг от друга по величине и положению. В этом можно убедиться, разложив белый свет на монохроматические составляющие и воспользовавшись принципом суперпозиции. Результирующее изображение, получающееся от наложения таких монохроматических изображений, оказывается нерезким и с окрашенными краями. Это явление называется *хроматической аберрацией*, или *хроматизмом*.

Хроматическая аберрация устраняется путем комбинации линз, изготовленных из стекла с различными оптическими свойствами. Устранение ее для всех лучей спектра невозможно. Обычно совмещают изображения только для каких-либо двух лучей с различными длинами волн. Тогда говорят, что оптическая система *ахроматизована* или *исправлена в хроматическом отношении*. Оставшаяся хроматическая аберрация называется *вторичным спектром*. В большинстве случаев на практике она уже не сказывается существенно на качестве изображения. Выбор лучей, для которых должна быть ахроматизована оптическая система, определяется ее назначением. В случае визуальных приборов эти лучи надо выбирать по разные стороны вблизи желто-зеленой области спектра, к которой наиболее чувствителен человеческий глаз. Обычно в таких приборах ахроматизация производится для фраунгоферовых линий с длинами волн $\lambda_C = 656{,}3$ нм и $\lambda_F = 486{,}1$ нм. В фотографических аппаратах ахроматизация осуществляется для лучей, лежащих ближе к синему концу спектра, так как эти лучи сильнее действуют на фотографическую пластинку. Здесь обычно ахроматизация выполняется для фраунгоферовых линий с длинами волн $\lambda_D = 589{,}3$ нм и $\lambda_G = 434{,}0$ нм.

Для полной ахроматизации, т.е. совмещения изображений двух цветов (например, красных и синих), необходимо, чтобы были равны не только фокусные расстояния для этих цветов, но и совпадали соответствующие им главные плоскости. Однако во многих случаях достаточна уже *частичная ахроматизация*, т.е. либо равенство только фокусных расстояний без точного совмещения главных плоскостей, либо совмещение только главных плоскостей без точного равенства фокусных расстоя-

ний. Это зависит от назначения прибора и определяется тем, что́ в его работе важнее — увеличение изображения или его местоположение. Так, в окулярах зрительных труб и микроскопов главный интерес представляют углы, под которыми глаз видит различные окрашенные изображения предмета. Для равенства этих углов необходимо, чтобы окуляры были ахроматизованы в смысле одинаковости фокусных расстояний (см. § 24, п. 4).

Подробное рассмотрение ахроматизации дается в руководствах по расчету оптических систем. Здесь мы можем остановиться только на простейших принципиальных вопросах и притом только *в параксиальном приближении*. В высших приближениях также возникает хроматическая аберрация, устранение которой требует выполнения особых условий. Но мы не будем их рассматривать.

2. Ахроматизацию фокусного расстояния можно получить с помощью уже одной толстой линзы (см. задачу 2). Однако этот способ не имеет практического значения. Практически более важной является монохроматизация тонкой линзы или системы тонких линз.

Рассмотрим сначала одиночную тонкую линзу или систему тонких линз и подсчитаем, как меняется ее фокусное расстояние при малых изменениях показателя преломления. Взяв логарифмическую производную от выражения (10.9), получим

$$\frac{\delta f}{f} = -\frac{\delta n}{n-1}. \tag{16.1}$$

Так как показатель преломления уменьшается с увеличением длины волны, то из этой формулы следует, что фокусное расстояние f по абсолютной величине больше для красных лучей, чем для синих.

По сравнению с изменением фокусного расстояния положения главных плоскостей линзы меняются ничтожно. Рассмотрим, например, это изменение для главной плоскости пространства предметов. Для не слишком толстой линзы можно воспользоваться формулой (12.21), из которой находим: $\delta h/h = -\delta n/n$, и следовательно,

$$\frac{\delta h}{\delta f} = \frac{h}{f}\frac{n-1}{n} = -\frac{d}{R_2}\left(\frac{n-1}{n}\right)^2.$$

Так как даже для сравнительно толстой линзы, как правило, $d \ll R_2$ и, кроме того, $n-1 < n$, то $\delta h \ll \delta f$. Поэтому тонкая линза будет практически полностью ахроматизована, когда ахроматизовано ее фокусное расстояние.

В практических расчетах принимается, что формула (16.1) приближенно справедлива и для конечных изменений показателя преломления. В случае визуальных приборов полагают $\delta n =$

$= n_C - n_F$ и вводят величину

$$\nu = \frac{n_D - 1}{n_F - n_C}, \tag{16.2}$$

где n_D — показатель преломления для желтой D-линии натрия $\lambda = 589{,}3$ нм. Эта величина называется *коэффициентом дисперсии*, или *числом Аббе*. (Вместо n_D в числителе можно взять показатель преломления для любой другой длины волны из видимой области спектра, так как от этого ν изменится самое большее на 2 %) В результате формула (16.1) преобразуется к виду

$$\frac{\delta f}{f} = \frac{1}{\nu}, \tag{16.3}$$

откуда

$$\delta\frac{1}{f} = -\frac{1}{f\nu}. \tag{16.4}$$

В табл. 2 приведены оптические характеристики некоторых марок оптического стекла. Разность показателей преломления

Таблица 2

Марка стекла	n_D	$n_F - n_C$	$n_F - n_D$	$n_G - n_F$	ν
ЛК 6	1,4704	0,00704	0,00493	0,00339	66,8
К 1	1,4982	0,00765	0,00535	0,00423	65,1
К 3	1,5100	0,00805	0,00565	0,00451	63,4
К 14	1,5147	0,00849	0,00597	0,00477	60,6
БК 4	1,5302	0,00877	0,00617	0,00495	60,5
БК 6	1,5399	0,00905	0,00637	0,00513	59,7
ТК 2	1,5724	0,00996	0,00702	0,00567	57,5
ТК 21	1,6568	0,01285	0,00910	0,00745	51,1
ЛФ 1	1,5406	0,01145	0,00812	0,00673	47,2
ЛФ 7	1,5783	0,01407	0,01002	0,00843	41,1
Ф 6	1,6031	0,01590	0,01135	0,00903	37,9
Ф 8	1,6248	0,01757	0,1260	0,01088	35,6
БФ 1	1,5247	0,00955	0,00674	0,00547	54,9
БФ 28	1,6641	0,01874	0,01342	0,01145	35,4
ТФ 1	1,6475	0,01912	0,01369	0,01172	33,9
ТФ 5	1,7550	0,02743	0,01972	0,01728	27,5

Обозначения: ЛК — легкий крон, К — крон, БК — баритовый крон, ТК — тяжелый крон, ЛФ — легкий флинт, Ф — флинт, БФ — баритовый флинт, ТФ — тяжелый флинт.

$n_F - n_C$ называется *средней дисперсией*, отношения

$$(n_F - n_D)/(n_F - n_C), \quad (n_G - n_F)/(n_F - n_C)$$

и аналогичные им — *относительными* или *частными дисперсиями*.

3. Рассмотрим теперь ахроматизацию тонкой линзы, сложенной из двух тонких линз, прижатых вплотную друг к другу. Обозначим через n_1 и f_1 показатель преломления и фокусное расстояние первой линзы, а через n_2 и f_2 — второй. Фокусное расстояние составной линзы находится по формуле

$$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2}. \tag{16.5}$$

Оно будет одно и то же для длин волн λ_F и λ_C, если выполнено условие

$$\delta \frac{1}{f} = -\left(\frac{1}{f_1\nu_1} + \frac{1}{f_2\nu_2}\right) = 0, \tag{16.6}$$

или

$$f_1\nu_1 + f_2\nu_2 = 0. \tag{16.7}$$

Это и есть *условие ахроматизации составной линзы*. Так как коэффициенты дисперсии для всех оптических стекол положительны, то из равенства (16.6) следует, что фокусные расстояния f_1 и f_2 должны иметь *противоположные знаки*. Из уравнений (16.5) и (16.6) находим

$$\frac{1}{f_1} = \frac{1}{f}\frac{\nu_1}{\nu_1 - \nu_2}, \quad \frac{1}{f_2} = -\frac{1}{f}\frac{\nu_2}{\nu_1 - \nu_2}. \tag{16.8}$$

Если общее фокусное расстояние f положительно, то линза с бóльшим коэффициентом дисперсии будет иметь положительное, а линза с меньшим коэффициентом дисперсии — отрицательное фокусное расстояние. В частности, собирательная линза должна делаться из крона, а рассеивающая — из флинта (см. табл. 2).

Если задано фокусное расстояние f составной линзы, а также указаны сорта стекол, из которых изготовлены первая и вторая линзы, то фокусные расстояния f_1 и f_2 определятся из уравнений (16.8). Тем самым по формуле (10.9) определятся разности $\frac{1}{R_1^{(1)}} - \frac{1}{R_2^{(1)}}$ и $\frac{1}{R_1^{(2)}} - \frac{1}{R_2^{(2)}}$, где $R_1^{(1)}$ и $R_2^{(1)}$ — радиусы сферических поверхностей первой, а $R_1^{(2)}$ и $R_2^{(2)}$ — второй линз. Два из этих радиусов можно выбрать произвольно. Если линзы 1 и 2 должны склеиваться, то надо положить $R_2^{(1)} = R_1^{(2)}$. Остается еще один радиус (или, в общем случае, один свободный параметр). Им целесообразно распорядиться так, чтобы сделать сферическую аберрацию возможно меньше (см. задачу 1).

Ахроматические пары тонких склеенных линз широко применяются в объективах микроскопов. Каждая пара состоит из двояковыпуклой линзы из крона, склеенной с плоско-вогнутой линзой из флинта, плоская поверхность которой обращена в сторону падающего света.

4. Если составные части (подсистемы) сложной оптической системы *полностью ахроматизованы*, то будет *полностью ахроматизована* и система в целом. Действительно, так как первая подсистема полностью ахроматизована, то изображения, даваемые ею, совмещаются в двух цветах. Рассматривая совмещенные изображения как предмет для второй подсистемы, найдем, что изображения, даваемые и этой подсистемой, также совмещаются для тех же двух цветов. Применив это рассуждение для всех подсистем, найдем, что совмещение изображений будет иметь место и для системы в целом.

Обратное утверждение *не справедливо*. Система может быть полностью ахроматизована без того, чтобы были полностью ахроматизованы ее составные части. Однако если система состоит из двух тонких линз, находящихся на некотором расстоянии друг от друга, то справедливо и обратное утверждение. Действительно, пусть x_1, x'_1 и y_1, y'_1 — абсциссы и ординаты точки-предмета и ее промежуточного изображения, даваемого первой линзой, причем за начало координат принят центр этой линзы. Пусть x_2, y_2 и x'_2, y'_2 имеют тот же смысл для второй линзы, причем роль предмета в этом случае играет промежуточное изображение. Очевидно, ординаты промежуточного изображения y'_1 и y_2 одинаковы, а абсциссы x'_1 и x_2 связаны соотношением $x'_1 = x_2 + l$, где l — расстояние между линзами. Поперечные увеличения первой и второй линз равны соответственно

$$\frac{y'_1}{y_1} = \frac{x'_1}{x_1}, \quad \frac{y'_2}{y_2} = \frac{x'_2}{x_2}.$$

Отсюда с учетом соотношения $y'_1 = y_2$ находим

$$\frac{y'_2}{y_1} = \frac{x'_1}{x_1}\frac{x'_2}{x_2}, \tag{16.9}$$

т.е. поперечное увеличение всей системы. По условию сложная система полностью ахроматизована, а потому увеличение должно быть одно и то же для обоих рассматриваемых нами цветов. Значит, вариация левой, а с ней и правой части равенства (16.9) обращается в нуль. Но абсцисса предмета x_1 постоянна, а абсцисса окончательного изображения x'_2 одинакова для обоих цветов. Поэтому, взяв вариацию логарифма от правой части (16.9), получим: $\dfrac{\delta x'_1}{x'_1} - \dfrac{\delta x_2}{x_2} = 0$. Кроме того, из соотношения $x'_1 = x_2 + l$, ввиду постоянства расстояния l между линзами, сле-

дует: $\delta x'_1 = \delta x_2$. Таким образом, $\left(\frac{1}{x'_2} - \frac{1}{x_2}\right)\delta x'_1 = 0$. Если $\frac{1}{x'_1} - \frac{1}{x_2} \neq 0$, т.е. $l \neq 0$, то $\delta x'_1 = 0$, т.е. промежуточные изображения совмещаются в обоих цветах. Но это означает, что первая линза полностью ахроматизована. Учтя свойство обратимости световых лучей, убедимся, что полностью ахроматизована и вторая линза. Исключение составляет случай, когда $\frac{1}{x'_1} - \frac{1}{x_2} = 0$. В этом случае линзы вплотную прижаты друг к другу ($l = 0$), и ахроматизация возможна даже тогда, когда каждая из линз не ахроматизована. Такой случай был разобран выше.

5. Таким образом, *чтобы система из двух тонких линз, расстояние l между которыми не равно нулю, была полностью ахроматизована, необходимо, чтобы были в отдельности полностью ахроматизованы обе составляющие ее линзы.* Однако *если не требовать полной ахроматизации системы, а ограничиться только ахроматизацией ее фокусных расстояний, то такая ахроматизация может быть получена и с неахроматизованными линзами.*

Действительно, фокусное расстояние системы из двух тонких линз определяется формулой (12.8). Из нее находим условие ахроматизации фокусного расстояния:

$$\delta\frac{1}{f} = \left(\frac{1}{f_1} - \frac{l}{f_1 f_2}\right)\frac{\delta n_1}{n_1 - 1} + \left(\frac{1}{f_2} - \frac{l}{f_1 f_2}\right)\frac{\delta n_2}{n_2 - 1} = 0, \qquad (16.10)$$

где n_1 — показатель преломления первой линзы, а n_2 — второй. Если n_1 и n_2 одинаковы, то это условие переходит в

$$l = \tfrac{1}{2}(f_1 + f_2). \qquad (16.11)$$

Таким образом, система из двух тонких линз, изготовленных из одного и того же стекла, будет ахроматизована в отношении фокусного расстояния и притом для всех цветов спектра, если расстояние между линзами равно полусумме их фокусных расстояний. Такой способ ахроматизации применяется в окулярах зрительных труб (см. § 24).

6. Объективы, в которых ахроматизация выполнена для двух цветов спектра, называются *ахроматами.* В некоторых оптических системах (в объективах микроскопов, длиннофокусных астрономических труб и спектральных аппаратов) наличие вторичного спектра (см. п. 1) существенно ухудшает качество изображения. Аббе ввел в микроскопию *апохроматы*, т.е. объективы, в которых ахроматизация выполнена для трех цветов спектра и, кроме того, удовлетворено условие синусов (см. § 18). Остающаяся при этом хроматическая аберрация называется *третичным спектром.*

7. В заключение упомянем о *сложных призмах*, изготовляемых из стекол различных сортов. На рис. 60 изображена ахроматическая призма. Первая призма, из крона, отклоняет и рассеивает лучи. Вторая призма, из флинта, отклоняет их в противоположную сторону. Рассеивающая способность (дисперсия) флинта больше, чем крона. Поэтому вторую призму можно подобрать так, чтобы компенсировать рассеивающую способность первой призмы, т.е. добиться одинаковости направлений каких-либо двух лучей (например, красного и фиолетового) при их выходе из рассматриваемой системы. Однако отклонение лучей, вызываемое первой призмой, компенсируется лишь частично. В результате белый луч проходит через сложную призму без заметной дисперсии, но испытывает отклонение в сторону.

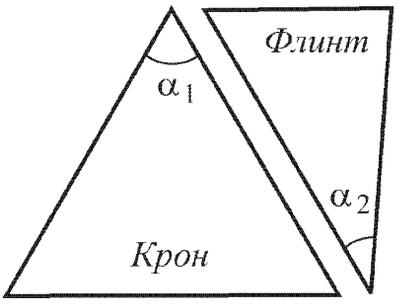
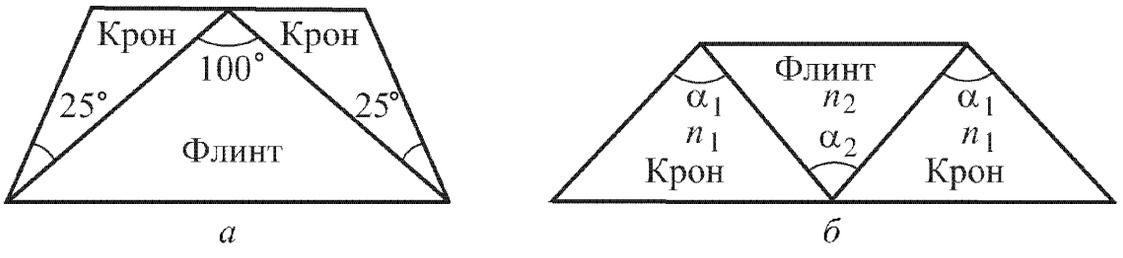


Рис. 60

На рис. 61 *а* изображена *дисперсионная призма*, применяемая в спектрографах. У нее дисперсия очень значительна благодаря большому преломляющему углу внутренней призмы из флинта. Боковые призмы из крона мало влияют на общую дисперсию призмы, но сильно уменьшают отклонение лучей, так как угол между их наружными гранями сравнительно невелик. Кроме того, эти призмы позволяют увеличить преломляющий угол внутренней призмы,



а *б*

Рис. 61

величина которого лимитируется полным отражением. Наконец, на рис. 61 *б* изображена *призма прямого зрения*. Она производит спектральное разложение, но оставляет неотклоненным средний луч пучка света.

ЗАДАЧИ

1. Рассчитать ахроматический объектив с фокусным расстоянием $f =$ $= 50$ см, склеенный из двух линз. Передняя линза изготовлена из крона К 1, задняя — из флинта Ф 8. Задняя линза — выпукло-вогнутая, обращена выпуклой поверхностью наружу, радиус кривизны этой поверхности равен 100 см.

Р е ш е н и е. Из табл. 2 находим для первой линзы: $\nu_1 = 65{,}1$, $n_D = 1{,}4982$; для второй линзы $\nu_2 = 35{,}6$, $n_D = 1{,}6248$. По формулам (16.8) вычисляем фокусные расстояния линз:

$$f_1 = f\frac{\nu_1 - \nu_2}{\nu_1} = 50 \cdot \frac{65{,}1 - 35{,}6}{65{,}1} = 22{,}66 \text{ см},$$

$$f_2 = -f\frac{\nu_1 - \nu_2}{\nu_2} = -50 \cdot \frac{65{,}1 - 35{,}6}{35{,}6} = -41{,}43 \text{ см}.$$

По условию задачи $R_2^{(2)} = -100$ см. По формуле (10.9) находим $R_1^{(2)}$:

$$R_1^{(2)} = \frac{(n-1)f_2 R_2^{(2)}}{(n-1)f_2 + R_2^{(2)}} = \frac{0{,}6248 \cdot 41{,}43 \cdot 100}{0{,}6248 \cdot 41{,}43 - 100} = -34{,}93 \text{ см}.$$

Той же величине равен радиус $R_2^{(1)}$, так что

$$R_1^{(1)} = \frac{(n-1)f_1 R_2^{(1)}}{(n-1)f_1 + R_2^{(1)}} = -\frac{0{,}4982 \cdot 22{,}66 \cdot 20{,}56}{0{,}4982 \cdot 22{,}66 - 20{,}56} = 25{,}04 \text{ см}.$$

Вид рассчитанного объектива изображен на рис. 62.

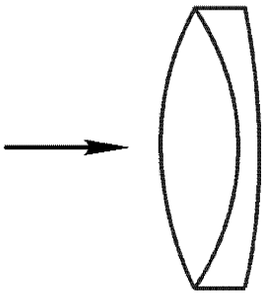


Рис. 62

2. Возможна ли ахроматизация толстой одиночной линзы относительно фокусных расстояний для двух длин волн?

О т в е т. Ахроматизация возможна потому, что в формулу для фокусного расстояния толстой линзы входит n^2. Следовательно, для заданного фокусного расстояния могут существовать два значения показателя преломления n_1 и n_2, которым соответствует это фокусное расстояние. Толщина линзы должна быть равна

$$d = \frac{n^2}{n^2 - 1}(R_1 - R_2),$$

где $n = \sqrt{n_1 n_2}$ — среднее геометрическое из n_1 и n_2. Так как толщина d существенно положительна, то для возможности ахроматизации необходимо, чтобы $R_1 - R_2 > 0$. Этому условию удовлетворяют только двояковыпуклые и плоско-вогнутые линзы. Все прочие толстые линзы не могут быть ахроматизованы.

3. Написать условие ахроматизации преломляющей призмы для двух близких цветов. Призма состоит из двух призм с малыми преломляющими углами, изготовленных из стекол с коэффициентами дисперсии ν_1 и ν_2.

О т в е т. Призмы должны быть обращены преломляющими углами в противоположные стороны. Условие ахроматизации: $\frac{\varphi_1}{\nu_1} - \frac{\varphi_2}{\nu_2} = 0$, где φ_1 — угол отклонения луча, даваемый первой призмой, а φ_2 — второй. Если $\nu_1 \neq \nu_2$, то результирующее отклонение $\varphi_1 - \varphi_2$ отлично от нуля.

§ 17. Условие отсутствия дисторсии

1. В фотографическом объективе особо важное значение имеет устранение изгиба плоскости изображения и дисторсии. Найдем *условие отсутствия дисторсии.* Пусть PP_1P_2

(рис. 63) — плоскость предмета, $P'P'_1P'_2$ — сопряженная ей плоскость изображения, O и O' — центры входного и выходного зрачков. Проведем главные лучи P_1O и P_2O от точек предметной плоскости через центр входного зрачка. Сопряженные им лучи P'_1O' и P'_2O' пройдут через центр O' выходного зрачка. Обозначим через u и u' углы наклона этих лучей к главной оптической оси. Тогда $PP_1 = PO\,\mathrm{tg}\,u_1$, $P'P'_1 = P'O'\,\mathrm{tg}\,u'_1$, $PP_2 = PO\,\mathrm{tg}\,u_2$, $P'P'_2 = P'O'\,\mathrm{tg}\,u'_2$, и следовательно,

$$\frac{P'P'_1}{PP_1} = \frac{\mathrm{tg}\,u'_1}{\mathrm{tg}\,u_1}, \quad \frac{P'P'_2}{PP_2} = \frac{\mathrm{tg}\,u'_2}{\mathrm{tg}\,u_2}.$$

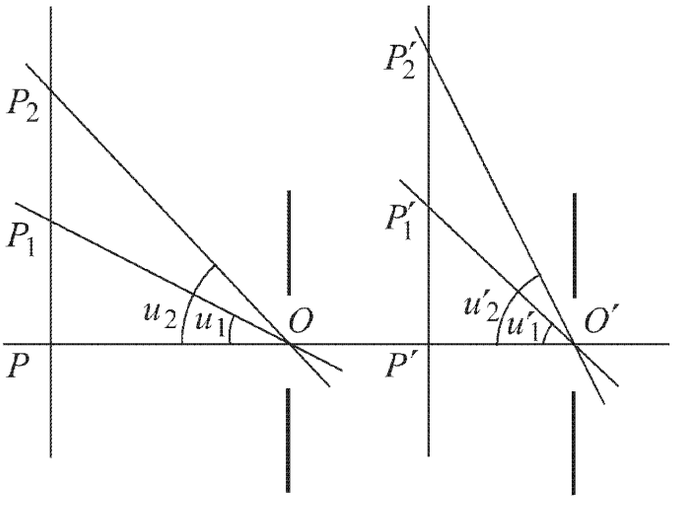


Рис. 63

Отсутствие дисторсии означает, что поперечные увеличения $P'P'_1/PP_1$ и $P'P'_2/PP_2$ одинаковы при любых положениях точек P_1 и P_2. Следовательно, для устранения дисторсии необходимо и достаточно, чтобы соблюдалось равенство

$$\frac{\mathrm{tg}\,u'_1}{\mathrm{tg}\,u_1} = \frac{\mathrm{tg}\,u'_2}{\mathrm{tg}\,u_2}, \tag{17.1}$$

каковы бы ни были значения углов u_1 и u_2. Это равенство называется *условием тангенсов*, или *условием ортоскопии*. Точки O и O' главной оптической оси, удовлетворяющие этому условию, называются *ортоскопическими точками*. Центры входного и выходного зрачков являются центрами перспективы предмета и его изображения. Таким образом, условие отсутствия дисторсии сводится к требованию, чтобы эти центры перспективы были ортоскопическими точками.

2. Простейшим примером, где выполнено условие ортоскопии, может служить симметричный фотографический объектив, изображенный на рис. 64. Он состоит из двух совершенно одинаковых ахроматизованных двойных линз, обращенных друг к другу вогнутыми поверхностями. Апертурная диафрагма помещена посередине между этими линзами. Так как она расположена близко к обеим линзам, то оба изображения ее мнимые, прямые и лежат обычно внутри объектива недалеко от диафрагмы A_1A_2. Изображение D_1D_2, даваемое передней линзой, лежит справа от диафрагмы A_1A_2 и служит входным зрачком системы; изображение $D'_1D'_2$, даваемое задней линзой, лежит левее диафрагмы A_1A_2 и служит выходным зрачком. Выходной зрачок

$D'_1D'_2$ есть изображение входного зрачка D_1D_2, даваемое всей оптической системой.

Как видно из рис. 64, это изображение производится с поперечным увеличением $+1$. Поэтому плоскости входного и выходного зрачков являются главными плоскостями, а центры этих зрачков — главными точками объектива. Возьмем теперь любой падающий луч, продолжение которого проходит через центр входного зрачка. Тогда, ввиду симметрии, продолжение выходящего луча пройдет через центр выходного зрачка, а самый луч внутри системы — через центр апертурной диафрагмы. Следовательно, углы наклона u и u' этих главных лучей к главной оптической оси будут всегда одинаковы, каковы бы ни были значения самих углов u и u'. Это значит, что в рассматриваемой системе соблюдено условие ортоскопии.

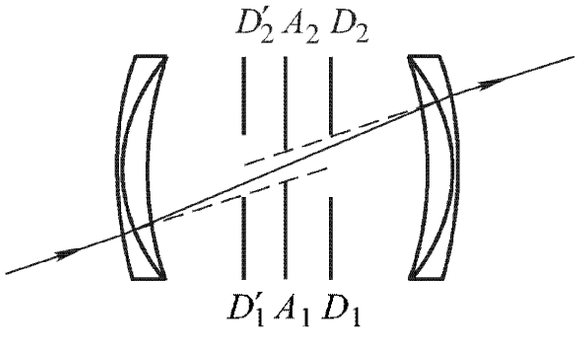


Рис. 64

§ 18. Условие синусов Аббе

1. В микроскопах изучаемые объекты малы и помещаются перед объективом вблизи главной оптической оси. Апертуры лучей, формирующих изображение в микроскопе, должны быть как можно шире (в лучших объективах микроскопов апертуры практически доведены до своего теоретического предела 180°). От этого увеличивается яркость изображения, а главное повышается разрешающая сила объектива, т.е. его способность различать мелкие детали (см. § 56). Поэтому сопряженные точки главной оптической оси, в первой из которых помещается объект, а во второй получается его изображение, даваемое объективом, должны быть *анаберрационными*, т.е. точка P должна изображаться в виде точки P' широкими пучками лучей (см. § 9). Однако этого условия еще недостаточно. Необходимо, чтобы без аберраций изображались все точки малого участка предметной плоскости, проходящей через точку P перпендикулярно к главной оптической оси. Практически речь идет об устранении сферической аберрации и комы. Анаберрационные точки главной оптической оси P и P', удовлетворяющие этому условию, называются *апланатическими*.

Значение апланатизма было выяснено немецким физиком Аббе (1840–1905). Аббе был оптиком-теоретиком. Он не только глубоко понимал теорию оптических приборов, но и указал

пути их совершенствования. Особое внимание он уделил теории и совершенствованию микроскопа. Его многолетнее плодотворное сотрудничество с известной оптической фирмой Цейсса (1816–1888) в Иене (Германия) привело к созданию и массовому производству лучших в мире оптических приборов, в которых практически было достигнуто то, что принципиально возможно с точки зрения волновой теории света. Аббе вывел необходимое и достаточное условие апланатизма, получившее название *условия синусов*. К выводу этого условия мы и перейдем.

2. Ввиду осевой симметрии оптической системы, при выводе достаточно ограничиться рассмотрением изображений не плоских площадок, а *прямолинейных отрезков* малой, но произвольной длины, перпендикулярных к главной оптической оси. Пусть один из таких отрезков, PQ, изображается широкими пучками лучей в виде отрезка $P'Q'$, также перпендикулярного к главной оптической оси (рис. 65). Оптические длины всех лучей, соединяющих сопряженные точки P и P', одинаковы. То же справедливо для лучей, соединяющих сопряженные точки Q и Q'.

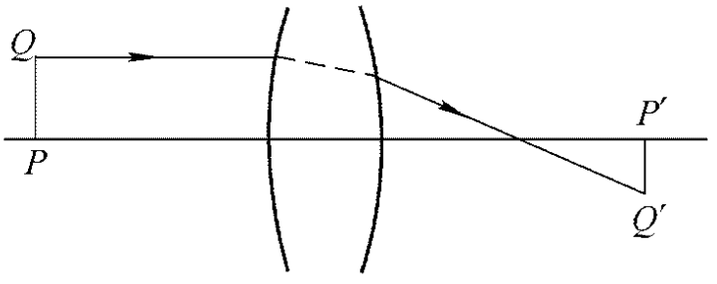


Рис. 65

Докажем, что оптическая длина (PP') любого луча, соединяющего апланатические точки P и P', равна оптической длине (QQ') любого луча, соединяющего сопряженные точки Q и Q'. Иными словами, оптические длины лучей между сопряженными точками отрезка PQ и его изображения $P'Q'$ одинаковы для всех пар сопряженных точек. Ввиду отмеченного выше свойства оптических длин лучей, соединяющих сопряженные точки, достаточно доказать это утверждение для параксиальных лучей, выходящих из точек объекта PQ параллельно главной оптической оси. Волновой фронт, соответствующий таким лучам, после прохождения через объектив становится сферическим, малый участок которого можно считать плоским. Таким образом, объект PQ целиком лежит в плоскости падающего, а его изображение $P'Q'$ — в плоскости прошедшего волновых фронтов. Но оптические длины всех лучей между двумя положениями волнового фронта одинаковы. Отсюда и следует доказываемое утверждение.

Проведем теперь через точки отрезка PQ пучок параллельных лучей под произвольным углом u к главной оптической оси (рис. 66). Так как лучи лежат в одной плоскости, то соответству-

ющий им волновой фронт представится прямолинейным отрезком QA, перпендикулярным к лучам. После прохождения через объектив параллельность лучей нарушится. Однако ввиду малости объекта PQ будет мал и угол расхождения прошедших лучей. Пренебрегая им, можно считать, что все прошедшие лучи наклонены к оптической оси практически под одним и тем же

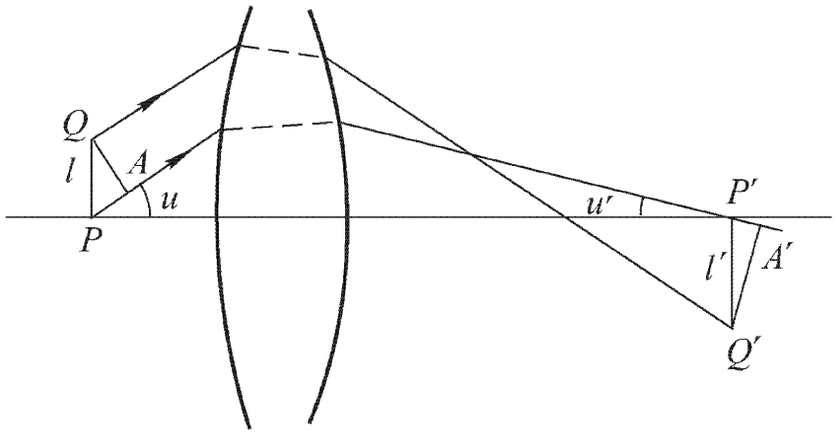


Рис. 66

углом u'. Соответствующий участок волнового фронта представится прямолинейным отрезком $Q'A'$. Ввиду одинаковости оптических длин всех лучей между волновыми фронтами $(AP'A') = (QQ')$. По доказанной только что теореме $(PAP') = (QQ')$, и предыдущее соотношение переходит в $(PAP') = (AP'A')$. Вычитая из обеих частей этого равенства общую часть (AP'), получим: $(PA) = (P'A')$, или

$$nl \sin u = n'l' \sin u', \tag{18.1}$$

где l и l' — длины отрезков PQ и $P'Q'$, а n и n' — показатели преломления пространств предметов и изображений. Равенство (18.1) и выражает условие синусов Аббе. Оно должно выполняться для любых значений угла u и для любых малых значений длины отрезка l. Для параксиальных лучей условие синусов (18.1) переходит в теорему Лагранжа–Гельмгольца (10.6) и, следовательно, выполняется для любой пары сопряженных точек на главной оптической оси.

В следующем параграфе приводится более строгий вывод условия синусов, основанный на теореме косинусов.

3. К важному примеру апланатических точек приводит построение преломленного луча на поверхности шара, указанное Вейерштрассом (1815–1897). Пусть R — радиус шара, n — его показатель преломления относительно окружающей среды (рис. 67). Построим две концентрические с шаровой поверхностью сферы S и S' с радиусами nR и R/n. Продолжим падаю-

щий луч AB до пересечения в точке P со сферой S и соединим точку P с центром шара O. Прямая OP пересечет сферу S' в точке P'. Прямая, соединяющая точку падения B с точкой P', дает направление проломленного луча. Действительно, из подобия треугольников OBP и OBP' следует, что угол $OP'B$ равен углу падения φ, а потому

$$\frac{\sin\varphi}{\sin\psi} = \frac{OB}{OP'} = n,$$

что и доказывает наше утверждение.

Из построения видно, что падающий пучок лучей, сходящийся в точке P, после преломления на шаровой поверхности будет сходиться в точке P'. Обратно, если точечный источник поместить в P', то после преломления получится пучок лучей, расходящихся из P. Следовательно, P и P' являются сопряженными анаберрационными точками и притом апланатическими.

Действительно, ввиду шаровой симметрии не только точки P и P', но и сферы S и S' отображаются друг в друга широкими пучками лучей. Так как обе сферы нормальны к прямой PP', то точки P и P' должны удовлетворять условию синусов, в чем легко убедиться и непосредственно. Точки Q и Q', очевидно, также апланатические. Наконец, центр сферы O можно рассматривать как пару совпадающих апланатических точек, являющихся одновременно узловыми точками системы.

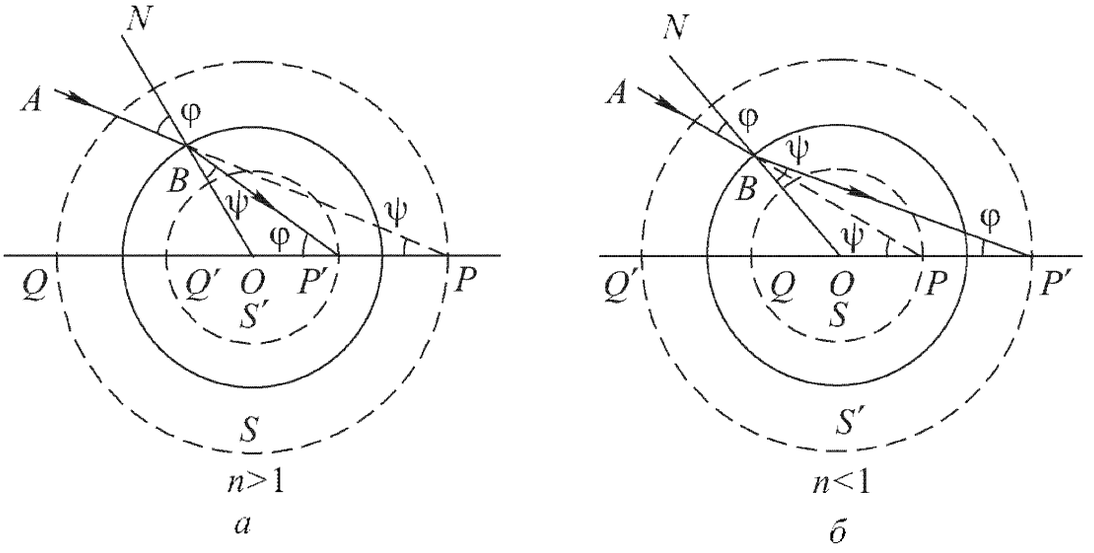


Рис. 67

Итак, на оптической оси QP имеется три пары изолированных апланатических точек: P и P', Q и Q' и двойная апланатическая точка O.

4. Апланатические точки шара используются для построения *иммерсионных объективов* микроскопов с очень большими увеличениями. Иммерсионным называют такой объектив, когда

между покровным стеклом и фронтальной линзой объектива находится слой жидкости (*иммерсия*). Иммерсия была предложена Амичи (1786–1863) в 1840 г. С целью уменьшения вредного для качества изображения преломления на поверхности покровного стекла он заполнял водой промежуток между покровным стеклом и фронтальной линзой объектива. Аббе в 1878 г. стал применять *однородную иммерсию* с показателем преломления, почти равным показателю преломления покровного стекла и фронтальной линзы. В такой системе лучи, исходящие из каждой точки предмета, распространяются практически прямолинейно до выхода из фронтальной линзы объектива. В качестве однородной иммерсии применяется кедровое масло ($n = 1{,}515$). В однородной иммерсионной системе Аббе фронтальная линза объектива состоит из стеклянного полушария, плоская сторона которого обращена к предмету (рис. 68). Предмет помещается на расстоянии $OP = R/n$ от центра (R — радиус полушария). Так как лучи до выхода из фронтальной линзы не испытывают преломления, то P будет апланатической точкой. Изображение предмета получается в сопряженной апланатической точке P' на расстоянии $OP' = Rn$ от центра O. Как видно из построения Вейерштрасса, линейное увеличение равно n^2. Поэтому из (18.1) следует $\sin u' = (\sin u)/n$, так что угол наклона луча к оптической оси уменьшается.

Рис. 68

Амичи для получения большего увеличения и дальнейшего уменьшения углов наклона лучей к оптической оси предложил помещать за фронтальной линзой вогнуто-выпуклую линзу. Точка P' должна находиться в центре кривизны вогнутой поверхности линзы. По отношению к преломлению на этой поверхности P' будет апланатической точкой, совпадающей со своей сопряженной точкой. Точка P' должна одновременно находиться на расстоянии R_1/n_1 от центра кривизны выпуклой поверхности второй линзы (n_1 — показатель преломления этой линзы, R_1 — радиус кривизны ее выпуклой поверхности). Тогда по отношению к преломлению на этой поверхности P' будет апланатической точкой; ее изображение получится в сопряженной апланатической точке P''.

Применив метод Амичи несколько раз, можно добиться какого угодно уменьшения углов наклона лучей к оптической оси.

При этом для всей системы в целом будет выполнено условие синусов. Метод Амичи часто применяют при конструкции объективов микроскопов. Однако таким методом конструируются в лучшем случае первые две линзы, так как иначе возникает сильная, ничем не компенсируемая хроматическая аберрация.

5. Аббе принадлежит простой способ испытания объективов на выполнение условия синусов. Допустим, что во второй апланатической точке объектива P' (рис. 69) помещена малая диафрагма, через которую производится наблюдение. Ее сильно уменьшенное изображение, получающееся в сопряженной апланатической точке P, будет входным зрачком системы.

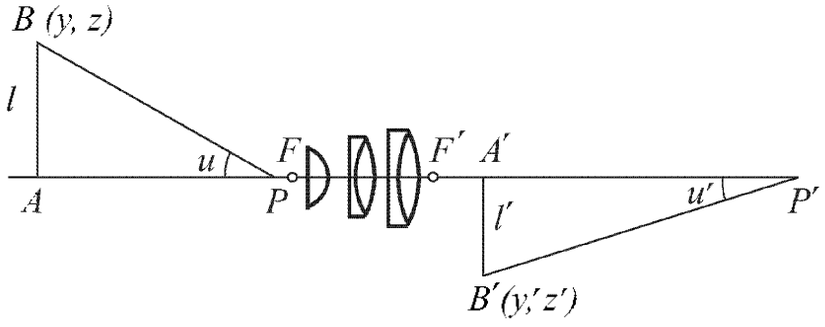


Рис. 69

Рассмотрим изображение конечного участка плоскости AB, даваемое объективом, когда расстояние q этой плоскости от точки P очень велико по сравнению с диаметром входного зрачка. Так как при перемещении объекта влево от апланатической точки P его изображение перемещается также влево (см. § 11, п. 6), то изображение плоскости AB получится левее второй апланатической точки P'. Когда расстояние q достаточно велико, то изображение $A'B'$ плоскости AB можно увидеть невооруженным глазом, если смотреть через диафрагму. Изображение получится сильно искаженным из-за использования косых пучков лучей, наклоненных под большими углами к главной оптической оси. Однако, ввиду исключительной малости входного зрачка, в разложении аберраций в степенные ряды можно пренебречь всеми степенями радиуса входного зрачка σ, сохранив лишь член нулевой степени, не содержащий σ. Иными словами, достаточно ограничиться учетом *одной только дисторсии*. В этом приближении изображение $A'B'$ будет расположено в плоскости параксиальных изображений.

Поставим задачу найти в предметной плоскости AB такое семейство кривых, чтобы его изображение в плоскости $A'B'$ получилось в форме квадратной сетки.

Возьмем в предметной плоскости произвольную точку B с координатами y, z на расстоянии $\sqrt{y^2+z^2}$ от главной оптической оси. Проведем луч BP через центр входного зрачка. Сопряженный луч $B'P'$ однозначно определится условием синусов $\sin u'/\sin u = A$, где $A = nl/(n'l')$, т.е. для рассматриваемого объектива A — величина постоянная, u и u' — углы наклона падающего и прошедшего лучей к главной оптической оси. Положение изображения B' определится точкой пересечения луча $B'P'$ с плоскостью параксиальных изображений. Подставим вместо $\sin u$ и $\sin u'$ их значения

$$\sin u = \sqrt{\frac{y^2+z^2}{q^2+y^2+z^2}}, \quad \sin u' = \sqrt{\frac{y'^2+z'^2}{q'^2+y'^2+z'^2}},$$

где q' — расстояние плоскости $A'B'$ от апланатической точки P'. Учтем при этом, что точки B и B' лежат в одной плоскости, проходящей через главную оптическую ось, так что $z'/y' = z/y$. Тогда получим

$$[A^2q'^2 - (1 - A^2)y'^2]y^2 - y'^2(1 - A^2)z^2 = q^2y'^2.$$

Возьмем теперь в плоскости $A'B'$ семейство прямых, параллельных оси Z: $y' = C$, где C — параметр, который может принимать произвольные значения. Уравнение кривых в предметной плоскости AB, изображением которых является это семейство, найдется из предыдущего соотношения, если в нем y' заменить на C. Таким путем получим

$$\frac{y^2}{a^2} - \frac{z^2}{b^2} = 1, \tag{18.2}$$

где

$$a^2 = \frac{q^2\omega^2}{1 - (1 - A^2)\omega^2}, \quad b^2 = \frac{q^2}{1 - A^2}, \tag{18.3}$$

а ω — новый параметр, связанный с параметром C соотношением

$$\omega = C/Aq'. \tag{18.4}$$

Для объективов микроскопов всегда $A^2 < 1$, так что при $\omega^2 < 1/(1-A^2)$ величина a существенно положительна. В этом случае (18.2) представляет семейство гипербол, осью которого является координатная ось Y. При $\omega = 0$ уравнение (18.2) переходит в $y^2 = 0$ и изображает ось Z. Значения $\omega^2 > 1/(1 - A^2)$ надо исключить, так как в этом случае уравнение (18.2) представляет семейство мнимых эллипсов и никакого оптического изображения не получается. Аналогично, семейство прямых $z' = C$ при тех же значениях параметра ω является оптическим изображением семейства гипербол

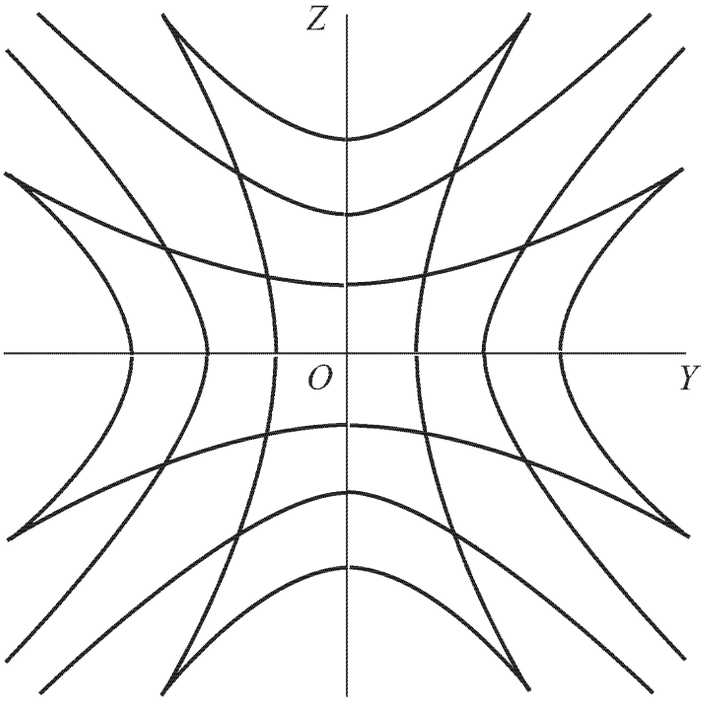


Рис. 70

$$\frac{z^2}{a^2} - \frac{y^2}{b^2} = 1, \tag{18.5}$$

получающегося из семейства (18.2) поворотом на 90° вокруг начала координат O. Оба семейства гипербол (18.2) и (18.5) пересекаются между собой, образуя криволинейную сетку, представленную на рис. 70. При переходе от одной гиперболы к соседней параметр ω должен получать одинаковые приращения. Тогда и приращения параметра C, как это видно из (18.4), также будут одинаковыми. Следовательно, если рис. 70 поместить на расстоянии q перед передней апланатической точкой P, то гиперболы изобразятся в виде квадратной сетки прямых. При $\omega^2 = (1 - A^2)/2$ уравнение (18.2) переходит в $y^2 - z^2 = b^2$, а уравнение (18.5) — в $z^2 - y^2 = b^2$.

Обе эти гиперболы имеют асимптотами биссектрисы координатных углов. Следовательно, b есть расстояние от начала координат до вершины той из гипербол, асимптоты которой совпадают с биссектрисами координатных углов. Для объективов микроскопов A — величина малая, квадратом которой по сравнению с единицей можно пренебречь. В этом приближении $b = q$. Это дает простой способ определения расстояния q, на котором следует поместить рис. 70, чтобы его изображение получилось в виде квадратной сетки.

Теперь становится понятным метод Аббе испытания объективов микроскопов на выполнение условия синусов. Аббе пользовался шаблоном, воспроизведенным на рис. 71. Удалив окуляр микроскопа, следует поместить такой шаблон на расстоянии q перед передней апланатической точкой. Глаз наблюдателя помещается во вторую апланатическую точку. Зрачок глаза играет роль выходного, а его изображение, даваемое объективом, — входного зрачка системы. Если изображение шаблона получается в виде сетки квадратов, то объектив удовлетворяет условию синусов. Если изображение получается слишком мелким, то можно применить вспомогательный микроскоп небольшого увеличения, перед объективом которого помещена малая диафрагма, расположенная во второй апланатической точке P'. При испытании описанным способом большого числа объективов микроскопов, ранее изготовленных лучшими мастерами без всяких расчетов путем последовательных проб и подбора линз, Аббе обнаружил, что все хорошие объективы всегда удовлетворяли условию синусов.

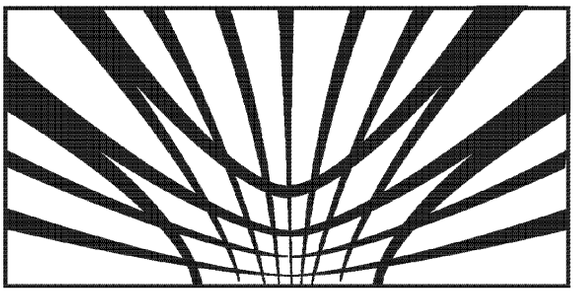


Рис. 71

§ 19. Теорема косинусов. Стигматические изображения широкими пучками лучей

1. Обобщим результаты предыдущего параграфа на случай произвольных оптических систем. Среды, в которых распространяются световые лучи, здесь предполагаются изотропными, но могут быть неоднородными. Таким образом, в общем случае световые лучи будут криволинейными. Пусть P и P' — две точки с радиусами-векторами $\mathbf{r}$ и $\mathbf{r}'$, лежащие на одном луче. Оптическая длина луча, соединяющего эти точки, рассматриваемая как функция их координат, называется *точечным эйконалом*, или *характеристической функцией оптической системы*. Она была введена Гамильтоном (1805–1865) и оказалась весьма полезной при исследовании оптических изображений. Характеристическую функцию будем обозначать через $H = H(\mathbf{r}, \mathbf{r}')$.

Фиксировав положение начальной точки P, будем перемещать конечную точку P'. Вместе с точкой P' будет перемещаться и луч PP', соединяющий ее с точкой P. В результате получится пучок лучей, исходящих в различных направлениях из точки P, как если бы она была источником света (рис. 72). Пусть $F'F'$ — волновой фронт, соответствующий этому пучку, проходящий через начальное положение точки P'. При смещении P' в точку Q', лежащую на луче $PA'Q'$, функция H получает приращение

$\Delta H = (PA'Q') - (PP')$, или, ввиду равенства оптических длин PP' и PA',

$$\Delta H = (PA'Q') - (PA') = (A'Q').$$

Если смещение $d\mathbf{r}' = \overrightarrow{P'Q'}$ бесконечно мало, то ΔH переходит в

$$dH = n'|d\mathbf{r}'|\cos\alpha' = n'(\mathbf{s}'\,d\mathbf{r}'),$$

где α' — угол между вектором $d\mathbf{r}'$ и единичным вектором луча $\mathbf{s}'$ в точке P', а n' — показатель преломления среды в этой точке. Аналогично найдется бесконечно малое приращение функции H, когда конечная точка P' остается неподвижной, а начальная точка P перемещается на $d\mathbf{r}$. В этом случае при вычислении надо воспользоваться пучком лучей, сходящихся в точке P' (рис. 73). В результате получится $dH = -n(\mathbf{s}\,d\mathbf{r})$, где n, $\mathbf{s}$ и $d\mathbf{r}$ имеют такой же смысл, что и в предыдущем случае, но относятся к начальной точке P. При изменении обоих аргументов $\mathbf{r}$ и $\mathbf{r}'$ получим

$$dH = n'(\mathbf{s}'\,d\mathbf{r}') - n(\mathbf{s}\,d\mathbf{r}). \tag{19.1}$$

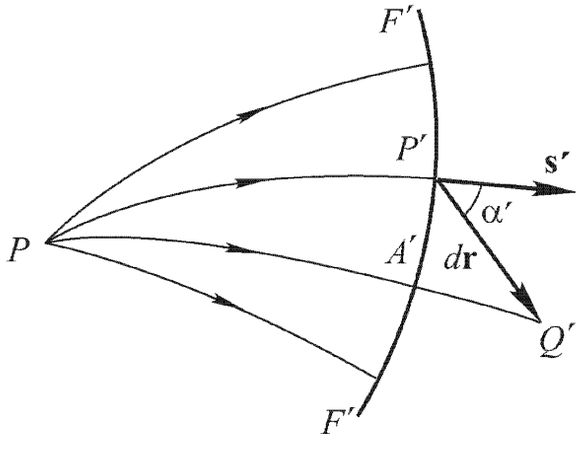


Рис. 72

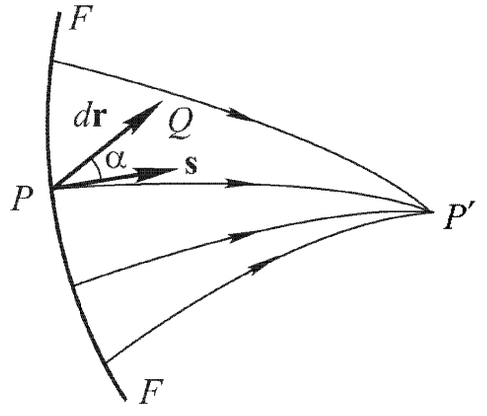


Рис. 73

Докажем теперь следующую теорему, называемую *теоремой косинусов*:

Пусть точка P' (рис. 74) *является стигматическим изображением точки P. Соединим эти точки произвольным лучом, направления которого в P и P' определяются единичными векторами $\mathbf{s}$ и $\mathbf{s}'$. Пусть Q и Q' — две точки, бесконечно близкие к P и P'. Для того чтобы точка Q' была стигматическим изображением точки Q в широких пучках лучей, необходимо и достаточно, чтобы разность*

$$n\mathbf{s}\,d\mathbf{l} - n'\mathbf{s}'\,d\mathbf{l}', \tag{19.2}$$

где $d\mathbf{l} = \overrightarrow{PQ}$, $d\mathbf{l}' = \overrightarrow{P'Q'}$, не зависела от направления луча, соединяющего P с P'.

Для доказательства необходимости теоремы допустим, что Q' является стигматическим изображением точки Q. Обозначим через H оптическую длину какого-либо луча, соединяющего сопряженные точки P и P', а через H' — оптическую длину луча, соединяющего сопряженные точки Q и Q'. В силу известного свойства сопряженных точек величины H и H', а с ними и разность $dH = H' - H = n'\mathbf{s}'\,d\mathbf{l}' - n\mathbf{s}\,d\mathbf{l}$ не зависят от направлений лучей, соединяющих P с P', а Q с Q'. Таким образом, выражение (19.2) не может зависеть от $\mathbf{s}$, и необходимость теоремы доказана.

Докажем теперь ее достаточность. Соединим точки Q и Q' лучом QQ'. Кроме того, через Q проведем произвольный луч QQ'' и отметим на нем такую точку Q'', чтобы оптическая длина (QQ'') была равна оптической длине (QQ'). Так как по условию теоремы P' является стигматическим изображением точки P, то из соображений непрерывности следует, что вектор $\overrightarrow{P'Q''} = d\mathbf{l}''$ бесконечно мал. В таком случае по формуле (19.1)

$$(QQ'') - (PP') = n'\mathbf{s}'\,d\mathbf{l}'' - n\mathbf{s}\,d\mathbf{l}.$$

По той же формуле

$$(QQ') - (PP') = n'\mathbf{s}'\,d\mathbf{l}' - n\mathbf{s}\,d\mathbf{l}.$$

Так как по построению $(QQ') = (QQ'')$, то $\mathbf{s}'\,d\mathbf{l}'' = \mathbf{s}'\,d\mathbf{l}'$, причем по условию теоремы это соотношение должно выполняться для всех направлений вектора $\mathbf{s}'$. Это возможно тогда и только тогда, когда $d\mathbf{l}'' = d\mathbf{l}'$, т.е. когда точка Q'' совпадает с Q'. Таким образом, любой луч, исходящий из Q, пройдет через Q', что и доказывает достаточность теоремы.

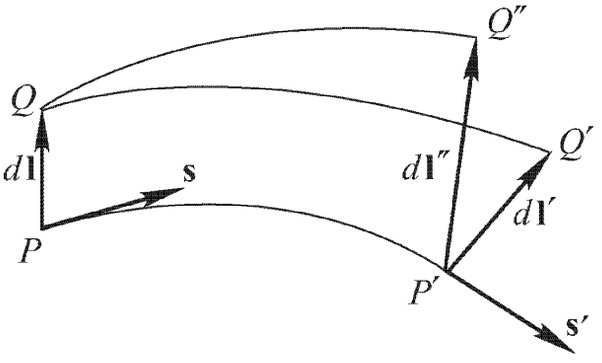


Рис. 74

Чтобы конечная кривая изображалась стигматически широкими пучками лучей, необходимо и достаточно, чтобы условия теоремы косинусов выполнялись для каждой пары сопряженных бесконечно малых отрезков этих кривых.

2. Условие синусов является следствием теоремы косинусов. Действительно, согласно этой теореме разность $n\mathbf{ls} - n'\mathbf{l}'\mathbf{s}' = nl\sin u - n'l'\sin u'$ не должна зависеть от направления луча, соединяющего сопряженные точки P и P' (рис. 75). Но если луч идет вдоль оптической оси, то указанная разность обращается в нуль. В результате снова получается соотношение (18.1), и притом не только как необходимое, но и как *достаточное условие.*

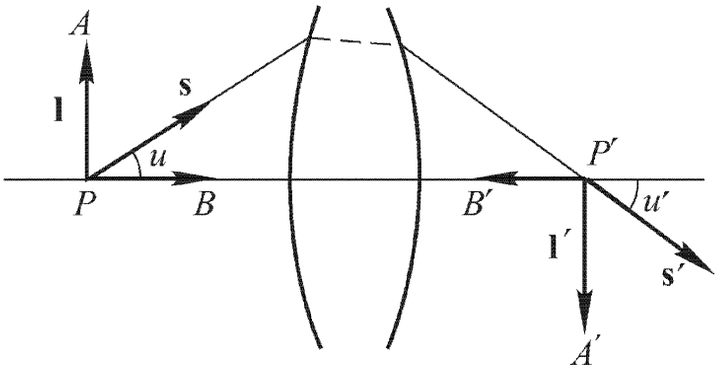


Рис. 75

Допустим теперь, что отрезок $l = PB$ лежит на оптической оси. Найдем условие, при котором он изображается стигматически широкими пучками лучей в виде отрезка $l' = P'B'$, также лежащего на оптической оси. В рассматриваемом случае при $u = 0$

$$n\mathbf{ls} - n'\mathbf{l}'\mathbf{s}' = nl\cos u - n'l'\cos u' = nl - n'l'.$$

Поэтому на основании теоремы косинусов должно быть

$$nl\cos u - n'l'\cos u' = nl - n'l',$$

или

$$nl\sin^2(u/2) = n'l'\sin^2(u'/2), \tag{19.3}$$

каковы бы ни были значения углов u и u'. Это соотношение называется *условием Гершеля.*

Пусть P и P' — сопряженные апланатические точки. Если в бесконечно малых окрестностях этих точек существует другая пара апланатических точек Q и Q' (а следовательно, и бесконечное множество таких пар), то оптическая длина бесконечно малого линейного объекта, лежащего в окрестности точки P, будет равна оптической длине его изображения, получающегося в окрестности точки P'. Действительно, так как P и P' — апланатические точки, то по условию синусов отношение

$$\frac{\sin u}{\sin u'} = \frac{\cos(u/2)}{\cos(u'/2)} \cdot \frac{\sin(u/2)}{\sin(u'/2)}$$

не должно зависеть от u. С другой стороны, так как Q и Q' — также апланатические точки, то бесконечно малый отрезок PQ оптической оси изображается в виде отрезка $P'Q'$ той же оси. Поэтому должно соблюдаться и условие Гершеля: отношение

$$\frac{\sin^2(u/2)}{\sin^2(u'/2)}$$

не должно зависеть от угла u. Оба условия могут быть выполнены одновременно тогда и только тогда, когда $u = \pm u'$, т.е. когда апланатические точки P и P' являются узловыми или обратными узловыми точками системы. Но в таком случае, если отрезок l перпендикулярен к оптической оси, из (18.1) следует: $|nl| = |n'l'|$. Если же отрезок l лежит на оптической оси, то из (19.3) также следует $|nl| = |n'l'|$. Таким образом, теорема доказана для двух частных случаев: когда отрезок l лежит на оптической оси и когда он перпендикулярен к ней. Тем самым она доказана для отрезка l произвольного направления.

Во всех центрированных системах, применяющихся на практике, линейное увеличение, как правило, отлично от n/n'. Поэтому из доказанной теоремы следует, что для таких систем апланатические пары точек, если они существуют, могут быть лишь *изолированными точками* оптической оси. Это значит, что для каждой пары апланатических точек можно указать конечные интервалы оптической оси, содержащие эти точки, внутри которых нет другой пары апланатических точек.

В приближении параксиальной оптики условие синусов и условие Гершеля выполняются всегда. Первое из них в указанном приближении переходит в теорему Лагранжа–Гельмгольца. Для доказательства второго из соотношений (11.20), (11.17), (11.13), (10.6) находим

$$\frac{\delta X'}{\delta X} = \frac{l'}{l} = -\frac{f^2}{X^2}\frac{f'}{f} = \frac{n'}{n}\frac{y'^2}{y^2} = \frac{n}{n'}\frac{u^2}{u'^2},$$

а это при малых углах u и u' совпадает с условием Гершеля (19.3). Таким образом, равенство углов u и u', а следовательно, и равенство оптических длин nl и $n'l'$ не обязательны. Благодаря этому в параксиальной оптике поперечное и продольное увеличения могут принимать любые значения, не обязательно равные n/n'.

3. *Пусть P' является стигматическим изображением точки P. Для того чтобы бесконечно малый элемент плоскости S, проходящий через P, изображался стигматически в виде бесконечно малого элемента плоскости S', проходящей через P', необходимо и достаточно, чтобы выполнялось условие косинусов для двух бесконечно малых непараллельных отрезков, лежащих в плоскости S и проходящих через точку P.*

Необходимость теоремы очевидна. Для доказательства ее достаточности соединим P и P' произвольным лучом. Пусть $d\mathbf{l}_1$ и $d\mathbf{l}_2$ — два бесконечно малых неколлинеарных вектора, проходящих через точку P, для которых удовлетворяется условие теоремы косинусов. Таким образом, по условию разности

$$\begin{aligned} n'\mathbf{s}'\,d\mathbf{l}'_1 - n\mathbf{s}\,d\mathbf{l}_1 &= dH_1, \\ n'\mathbf{s}'\,d\mathbf{l}'_2 - n\mathbf{s}\,d\mathbf{l}_2 &= dH_2 \end{aligned} \tag{19.4}$$

не зависят от направления луча, соединяющего P с P'. Но они могут зависеть от направлений векторов $d\mathbf{l}_1$ и $d\mathbf{l}_2$. Произвольный вектор $d\mathbf{l}$, проходящий через точку P и лежащий в плоскости предмета, можно разложить по векторам $d\mathbf{l}_1$ и $d\mathbf{l}_2$:

$$d\mathbf{l} = a\,d\mathbf{l}_1 + b\,d\mathbf{l}_2,$$

причем коэффициенты a и b не зависят от $\mathbf{s}$. Введем вектор

$$d\mathbf{l}' = a\,d\mathbf{l}'_1 + b\,d\mathbf{l}'_2,$$

умножим соотношения (19.4) на a и b и сложим. Получим

$$n'\mathbf{s}'\,d\mathbf{l}' - n\mathbf{s}\,d\mathbf{l} = dH, \tag{19.5}$$

где $dH = a\,dH_1 + b\,dH_2$. Отсюда видно, что разность (19.5) не зависит от $\mathbf{s}$, т.е. условие теоремы косинусов выполняется для произвольного вектора $d\mathbf{l}$, проходящего через точку P и лежащего в плоскости предмета. Следовательно, предмет изобразится оптической системой стигматически. Вообще говоря, изображение не будет подобно самому предмету.

Направления произвольных неколлинеарных векторов $d\mathbf{l}_1$ и $d\mathbf{l}_2$ можно принять за координатные оси Y и Z в плоскости предмета, направления оптически сопряженных с ними отрезков $d\mathbf{l}'_1$ и $d\mathbf{l}'_2$ — за координатные оси Y' и Z' в плоскости изображения, а сами точки P и P' — за начала координат соответствующих координатных систем. Тогда в окрестностях точек P и P' координатные оси Y и Z изобразятся оптической системой координатными осями Y' и Z'. Координаты сопряженных точек, бесконечно близких к P и P', будут связаны линейными соотношениями

$$y' = Ay, \quad z' = Bz. \tag{19.6}$$

Координатная система $Y'Z'$, вообще говоря, будет косоугольной, даже в том случае, когда система YZ прямоугольная. Однако если в плоскостях предмета и его изображения ввести прямоугольные системы координат, то из соотношений (19.6) и из формул преобразования координат непосредственно следует, что прямоугольные координаты сопряженных точек будут связаны формулами линейного преобразования

$$y' = A_{11}y + A_{12}z, \quad z' = A_{21}y + A_{22}z. \tag{19.7}$$

Из свойств линейного преобразования следует, что в плоскости предмета существует пара взаимно перпендикулярных прямых, которой в плоскости изображения соответствует пара также взаимно перпендикулярных прямых. Если эти четыре прямые принять за координатные оси, то формулы преобразования снова примут вид (19.6), с тем отличием, что теперь обе координатные системы прямоугольны. Вообще говоря, $A \neq B$. Поэтому изображение бесконечно малой площадки происходит с нарушением подобия: бесконечно малый круг изображается в виде эллипса. Только в частном случае, когда $A = B$, система дает подобные изображения бесконечно малых площадок, находящихся в окрестности точки P.

4. Следуя Каратеодори (1873–1950), говорят, что *световой луч лежит в поле инструмента, если он действительно проходит через диафрагмы из пространства предметов в пространство изображений.* Говорят также, что *отрезок кривой лежит тангенциально в поле инструмента, если все лучи, касающиеся этого отрезка, лежат в поле инструмента.*

Пусть бесконечно малая площадка изображается оптической системой стигматически. Пусть, далее, $d\mathbf{l}_1$ *и* $d\mathbf{l}_2$ *— бесконечно малые не параллельные отрезки, пересекающиеся в пределах площадки и лежащие в ее плоскости. Если эти отрезки лежат тангенциально в поле инструмента, то рассматриваемая площадка изображается оптической системой с сохранением подобия. При этом оптическая длина любого отрезка, лежащего на площадке, равна оптической длине сопряженного с ним отрезка.*

По условию теоремы лучи, выходящие из P (рис. 76) в направлениях $d\mathbf{l}_1$ и $d\mathbf{l}_2$, лежат в поле инструмента. В пространстве изображений они пройдут

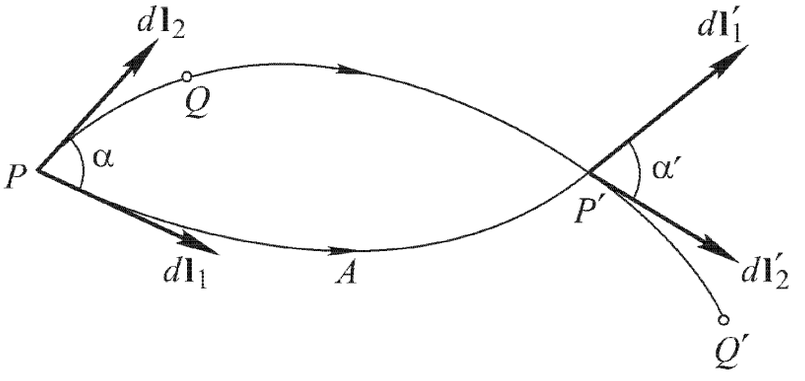


Рис. 76

в направлениях оптически сопряженных отрезков $d\mathbf{l}'_1$ и $d\mathbf{l}'_2$. Рассмотрим сначала изображение отрезка $d\mathbf{l}_1$. Возьмем два луча, исходящих из P в направлениях $d\mathbf{l}_2$ и $d\mathbf{l}_1$. На основании теоремы косинусов

$$n\,dl_1\cos\alpha - n'\,dl'_1\cos\alpha' = n\,dl_1 - n'\,dl'_1. \tag{19.8}$$

Аналогично, рассматривая изображение отрезка $d\mathbf{l}_2$,

$$n\,dl_2\cos\alpha - n'\,dl'_2\cos\alpha' = n\,dl_2 - n'\,dl'_2. \tag{19.9}$$

Допустим, что $dl_1 = dl_2$. Докажем, что тогда $dl'_1 = dl'_2$. В самом деле, почленное вычитание (19.9) из (19.8) дает

$$n'\,dl'_1(1-\cos\alpha') = n'\,dl'_2(1-\cos\alpha'). \tag{19.10}$$

Разность $1-\cos\alpha'$ не может обращаться в нуль. Действительно, по предположению угол α не равен нулю. Но заданием направления в какой-либо точке луч определяется однозначно. Если бы $\alpha' = 0$, т.е. направления обоих рассматриваемых нами лучей в точке P' совпадали, то они совпадали бы и во всех других точках, в частности в начальной точке P. Значит, было бы $\alpha = 0$, вопреки предположению. Поэтому, сокращая на $n'(1-\cos\alpha')$, из (19.10) находим $dl'_1 = dl'_2$, что и требовалось доказать.

Из доказанного следует, что отрезки $d\mathbf{l}_1$ и $d\mathbf{l}_2$ изображаются оптической системой с одинаковым увеличением. Следовательно, в рассматриваемом случае в формулах (19.6) $A = B$, т.е. увеличение любого отрезка в плоскости предмета не зависит от его направления. Отсюда следует, что изображение происходит с сохранением подобия, т.е. является *конформным.*

Но изображение с сохранением подобия характеризуется также сохранением углов. Следовательно, $\alpha = \alpha'$, и формула (19.8) дает $n\,dl_1 = n'\,dl_1'$. Вообще, для всякого отрезка dl, лежащего в плоскости предмета, $n\,dl = n'\,dl'$, $dl'/dl = n/n'$, и вторая часть теоремы доказана.

Итак, *стигматические изображения площадок, тангенциально лежащих в поле инструмента, могут происходить только с вполне определенным увеличением* n/n'. В частности, когда показатели преломления пространств предметов и изображений одинаковы, увеличение равно единице. Это утверждение перестает быть справедливым для площадок, не лежащих тангенциально в поле инструмента.

Примером может служить преломление на сферической поверхности (см. рис. 67). Сфера S отображается на сферу S' стигматически широкими пучками лучей. Однако линейное увеличение, как видно из построения, равно отношению квадратов показателей преломления, а не их первых степеней. Причина этого в том, что ни одна из сфер S и S' не лежит тангенциально в поле инструмента. Напротив, если линейный объект поместить в точку O, то, поскольку последняя является парой совпадающих узловых точек, линейное увеличение будет равно просто отношению показателей преломления в согласии с обсуждаемой нами общей теоремой. Действительно, ввиду шаровой симметрии любой линейный объект, помещенный в центре O, лежит тангенциально в поле инструмента.

Бесконечно малую часть конечной поверхности можно рассматривать как бесконечно малую плоскую площадку. Поэтому *для стигматического изображения конечной поверхности необходимо и достаточно, чтобы стигматически изображались все бесконечно малые площадки, на которые можно разбить эту поверхность.*

5. Рассмотрим, наконец, стигматические изображения объемных объектов широкими пучками лучей. Этот вопрос может быть исследован в точности так же, как и аналогичный вопрос для поверхностных объектов. В частности, может быть доказана следующая теорема:

Пусть точка P' является стигматическим изображением точки P. Для того чтобы бесконечно малый элемент объема в окрестности точки P изображался стигматически, необходимо и достаточно, чтобы выполнялось условие теоремы косинусов для трех бесконечно малых отрезков, проходящих через точку P и не лежащих в одной плоскости.

Новым по сравнению с изображениями элементов поверхностей является то, что в случае стигматического изображения элементов объема всегда существуют три отрезка dl_1, dl_2, dl_3, не находящиеся в одной плоскости и лежащие тангенциально в поле инструмента. Поэтому, повторяя рассуждения, приведенные применительно к изображениям элементов поверхности, приходим к заключению, что эти три отрезка изображаются с одним и тем же увеличением. Как следствие этого, получаем следующую теорему:

Стигматическое изображение элементов объема, если оно осуществляется широкими пучками лучей, всегда конформно, т.е. происходит с сохранением подобия. При этом линейное увеличение равно n/n', так что оптическая длина предмета всегда равна оптической длине изображения.

§ 20. Об абсолютных оптических инструментах

1. С точки зрения геометрической оптики идеалом был бы оптический инструмент, изображающий стигматически широкими пучками лучей каж-

дую точку пространства предметов в виде точки пространства изображений. Такой инструмент называется *абсолютным.* Обсудим вопрос о принципиальной возможности абсолютных оптических инструментов и исследуем их свойства.

Как доказано в предыдущем параграфе, изображения бесконечно малых объектов, даваемые абсолютным оптическим инструментом, всегда *конформны.* При этом оптическая длина любой линии равна оптической длине ее изображения. Отсюда следует, что *в абсолютном оптическом инструменте оптическая длина луча между двумя сопряженными точками одинакова для всех пар сопряженных точек.* Это положение называется *теоремой Каратеодори.*

Для доказательства возьмем две пары сопряженных точек: P, P' и Q, Q' (рис. 76). Через точку P проведем пучок лучей во всевозможных направлениях. Все эти лучи пересекутся в точке P'. Один из них пройдет через точку Q. Возьмем другой произвольный луч PAP'. В силу свойства сопряженных точек $(PAP') = (PQP')$. Луч PQP', поскольку он проходит через точку Q, должен пройти и через сопряженную ей точку Q'. При этом кривая $P'Q'$ будет оптическим изображением кривой PQ, а потому $(PQ) = (P'Q')$. Отсюда следует $(QP'Q') = (PQP')$. Сравнивая это равенство с предыдущим, получаем: $(PAP') = (QP'Q')$, что и требовалось доказать.

2. Допустим теперь, что пространства предметов и изображений *однородны*, т.е. показатели преломления n и n' не зависят от координат. В этом случае *любая прямая пространства предметов изобразится абсолютным инструментом в виде прямой пространства изображений.* В самом деле, проведем через произвольную точку O пространства предметов три луча, лежащих в поле инструмента и не находящихся в одной плоскости. В пространстве изображений сопряженные с ними лучи пересекутся в сопряженной точке O'. Примем эти лучи за координатные оси в пространствах предметов и изображений. Ясно, что координатные оси одной координатной системы будут изображаться координатными осями другой координатной системы. При этом без потери общности можно принять, что одноименные оси являются сопряженными. Так как изображение происходит с постоянным увеличением n/n', то при надлежащем выборе положительных направлений координатных осей координаты сопряженных точек будут связаны соотношениями

$$x' = \frac{n}{n'}\,x, \quad y' = \frac{n}{n'}\,y, \quad z' = \frac{n}{n'}\,z, \tag{20.1}$$

справедливыми для всех точек пространства предметов и пространства изображений, а не только для окрестностей сопряженных точек O и O'. Из них непосредственно следует, что любая прямая изображается прямой. Таким образом, *при постоянных n и n' абсолютный инструмент является телескопической системой.*

Докажем теперь, что *при постоянных n и n' абсолютный оптический инструмент возможен только при $n = n'$.* Для доказательства заметим, что линейное увеличение n/n' не зависит от положения объекта. Чтобы определить это увеличение, поместим бесконечно малый объект тангенциально на границе раздела сред. Тогда его изображение, возникающее как при преломлении, так и при отражении, совместится с самим объектом. Отсюда следует, что $n/n' = 1$, и наше утверждение доказано. Таким образом, показатели преломления сред должны быть одинаковы, и никакого преломления не будет.

Остается исследовать только отражение. По теореме Каратеодори оптическая длина луча между сопряженными точками не зависит от поло-

жения объекта. Поместим точечный объект на границе раздела сред. Его изображение получится в той же точке. Поэтому оптическая длина луча между объектом и его изображением должна равняться нулю, независимо от положения объекта. При преломлении, когда оптические длины лучей на всех участках сохраняют знаки, это было бы возможно только тогда, когда изображение совпадает с самим объектом. Но в этом случае, как мы видели, никакого преломления нет, и об изображении можно говорить лишь условно. Однако при отражении, если изображение получается мнимым, оптические длины лучей на разных участках могут иметь противоположные знаки. Тогда возможен и нетривиальный случай, когда объект и его изображение находятся в разных местах, хотя оптические длины лучей между ними и обращаются в нуль. При этом формулы (20.1) переходят в

$$x' = x, \quad y' = y, \quad z' = z. \tag{20.2}$$

Такой случай осуществляется в плоском зеркале или в системе плоских зеркал. Это единственный абсолютный оптический инструмент, возможный при постоянных n и n'.

3. *Нетривиальные абсолютные оптические инструменты, дающие изображения с увеличением, отличным от единицы, возможны только с использованием неоднородных сред и криволинейных лучей.* Первый пример такого инструмента был приведен Максвеллом. В дальнейшем были приведены и другие примеры. Максвелл назвал свой инструмент *«рыбьим глазом»*, хотя никакого сходства с глазом рыбы у него нет. Более того, «рыбий глаз» Максвелла вообще трудно назвать инструментом в обычном смысле этого слова. Действительно, он представляет собой неограниченную неоднородную среду, показатель преломления которой меняется в пространстве таким образом, что любой луч в этой среде имеет форму окружности. К идее «рыбьего глаза» естественнее всего прийти с помощью геометрического построения, называемого *стереографической проекцией.*

Такая проекция осуществляется следующим образом. Возьмем сферу S, через центр которой проведем координатную плоскость XY (рис. 77). Пусть

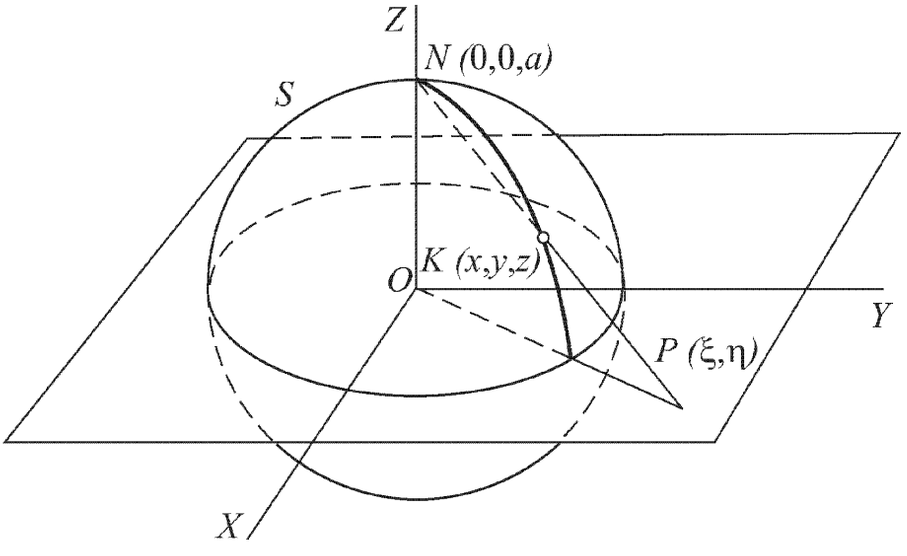


Рис. 77

N — точка пересечения этой сферы с осью Z. Соединим точку N с точкой K сферы прямой NK и продолжим эту прямую до пересечения с плоскостью XY в точке P. Точка P и называется стереографической проекцией точки K на плоскость XY. Легко показать, что координаты x, y, z точки K связаны

с координатами ξ, η соотношениями

$$\xi = a\frac{x}{a-z}, \quad \eta = a\frac{y}{a-z}, \tag{20.3}$$

$$x = 2a^2\frac{\xi}{\xi^2+\eta^2+a^2}, \quad y = 2a^2\frac{\eta}{\xi^2+\eta^2+a^2}, \quad z = a\frac{\xi^2+\eta^2-a^2}{\xi^2+\eta^2+a^2}, \tag{20.4}$$

где a — радиус сферы S.

Каждая окружность сферы S при стереографической проекции преобразуется в окружность плоскости XY и обратно. В самом деле, плоскость $Ax+By+Cz+D=0$ пересекает сферу S по окружности. Если в уравнение этой плоскости подставить значения координат (20.4), то получится уравнение второй степени, в котором будет отсутствовать член с произведением $\xi\eta$, а квадраты ξ^2 и η^2 войдут только в комбинации $\xi^2+\eta^2$. Такое уравнение есть уравнение окружности.

Пусть $d\sigma = \sqrt{dx^2+dy^2+dz^2}$ — элемент длины на сфере S. При стереографической проекции он отобразится элементом длины $ds = \sqrt{d\xi^2+d\eta^2}$ на плоскости XY. Пользуясь формулами (20.4), нетрудно показать, что эти элементы связаны соотношением

$$d\sigma = \frac{2}{1+(r/a)^2}\,ds, \tag{20.5}$$

где $r^2 = \xi^2+\eta^2$. Кривая кратчайшей длины, проведенная на какой-либо поверхности между двумя точками, называется *геодезической линией* этой поверхности. Для сферы геодезическими линиями являются дуги больших кругов. Возьмем на сфере S любые диаметрально противоположные точки $K(x_0, y_0, z_0)$ и $K'(-x_0, -y_0, -z_0)$. Пусть P и P' — их стереографические проекции на плоскость XY. Через прямую, соединяющую точки K и K', можно провести бесконечное множество плоскостей, пересекающих сферу S вдоль больших кругов, проходящих через эти точки. При стереографической проекции эти большие круги изобразятся окружностями, соединяющими точки P и P' (рис. 78). Вдоль этих окружностей интеграл $\int ds/(1+r^2/a^2)$ принимает одно и то же значение, как это непосредственно видно из соотношения (20.5).

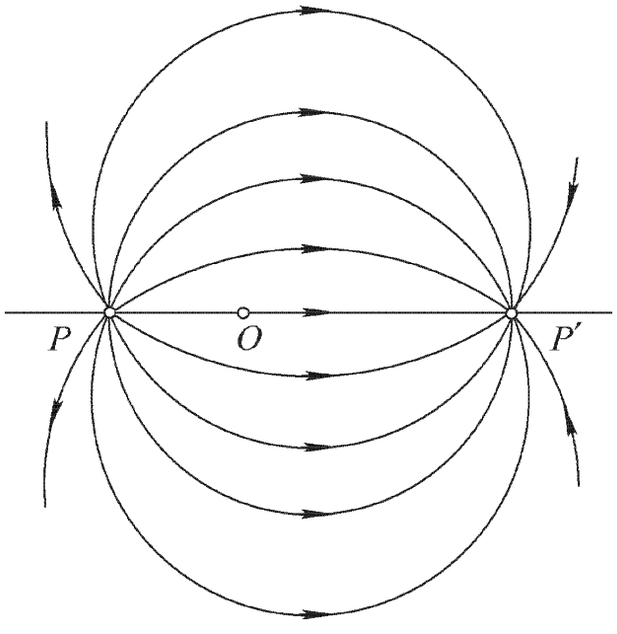


Рис. 78

Вообразим теперь сферически симметричную среду с центром в начале координат O, показатель преломления которой определяется выражением

$$n = \frac{n_0}{1+r^2/a^2}, \tag{20.6}$$

где n_0 — постоянная, а r — расстояние от начала координат. Тогда оптические длины всех окружностей между точками P и P', изображенных на рис. 78, будут одинаковыми. Отсюда следует, что все эти окружности будут совпадать со световыми лучами, исходящими из точки P и сходящимися в

точке P'. Одна из этих окружностей вырождается в прямую, соединяющую точки P и P'. Она получается в результате стереографической проекции большого круга, лежащего в плоскости NKK'. Начало координат O лежит в той же плоскости, а следовательно, и на прямой PP'. Вообще говоря, его положение не совпадает с серединой отрезка PP'. Исключение составляет только случай, когда диаметрально противоположные точки K и K' сферы S лежат в координатной плоскости XY. В этом случае большой круг, лежащий в той же плоскости, совпадает со своей стереографической проекцией, а потому радиус кривизны светового луча будет равен радиусу сферы S, т.е. a.

Ввиду шаровой симметрии все плоскости, проходящие через центр O, эквивалентны. Поэтому луч, исходящий из точки P, пройдет через сопряженную точку P', независимо от того, лежит он в координатной плоскости XY или не лежит. Кроме того, приведенные нами рассуждения применимы для любой пары диаметрально противоположных точек K и K' и соответствующей им пары сопряженных точек P и P'.

Таким образом, получается следующий результат. Все лучи, выходящие из любой точки среды P, описав окружности, собираются в другой точке P', т.е. точка P' является стигматическим изображением точки P в широких пучках лучей. Иными словами, среда с показателем преломления (20.6) обладает свойствами абсолютного оптического инструмента. Она и была названа Максвеллом «рыбьим глазом».

§ 21. Глаз и зрение

1. На рис. 79 приведен горизонтальный разрез правого глаза человека (вид сверху). Глазное яблоко заключено в плотную белую непрозрачную оболочку ww, называемую *склерой*. Через заднюю стенку склеры проходит глазной нерв e и кровеносные сосуды. Передняя часть склеры переходит в прозрачную *роговую оболочку*, или *роговицу* h, более выпуклую, чем склера. Толщина склеры от 0,4 до 1,1 мм, роговицы — около 0,5 мм. За роговицей следуют *передняя глазная камера* a, *хрусталик* l, *задняя глазная камера* gl. Эти части вместе с роговицей составляют *оптическую систему глаза*, дающую при преломлении лучей оптические изображения предметов. Толщина передней глазной камеры, а также хрусталика около 3,6 мм. Передняя глазная камера заполнена прозрачной жидкостью, называемой *водянистой влагой*, задняя — прозрачным студенистым веществом, называемым *стекловидным телом*. Показатели преломления обеих этих сред практически одинаковы и равны 1,336. Показатель преломления роговой оболочки равен 1,376.

Внутренняя поверхность склеры покрыта *сосудистой оболочкой* (толщиной около 0,4 мм), содержащей разветвления кровеносных сосудов, питающих глаз. Спереди она переходит в утолщенную часть st, содержащую кольцевидную ресничную мышцу uu, и далее в *радужную оболочку* ii, не одинаково окрашенную у различных людей и определяющую «цвет глаз». По-

середине радужной оболочки имеется круглое отверстие, называемое *зрачком глаза*, которое выполняет роль апертурной диафрагмы. Изображение этого отверстия, даваемое расположенными перед ним водянистой влагой и роговицей, играет роль входного, а изображение, даваемое хрусталиком и стекловидным телом, — выходного зрачков глаза. Оба они практически совпадают с настоящим зрачком. В зависимости от яркости света зрачок может рефлекторно сужаться и расширяться посредством мышцы *ии*, с которой связана радужная оболочка. Диаметр его может меняться примерно от 2 до 8 мм. Этим достигается ограничение поперечных размеров пучков света, поступающих в глаз.

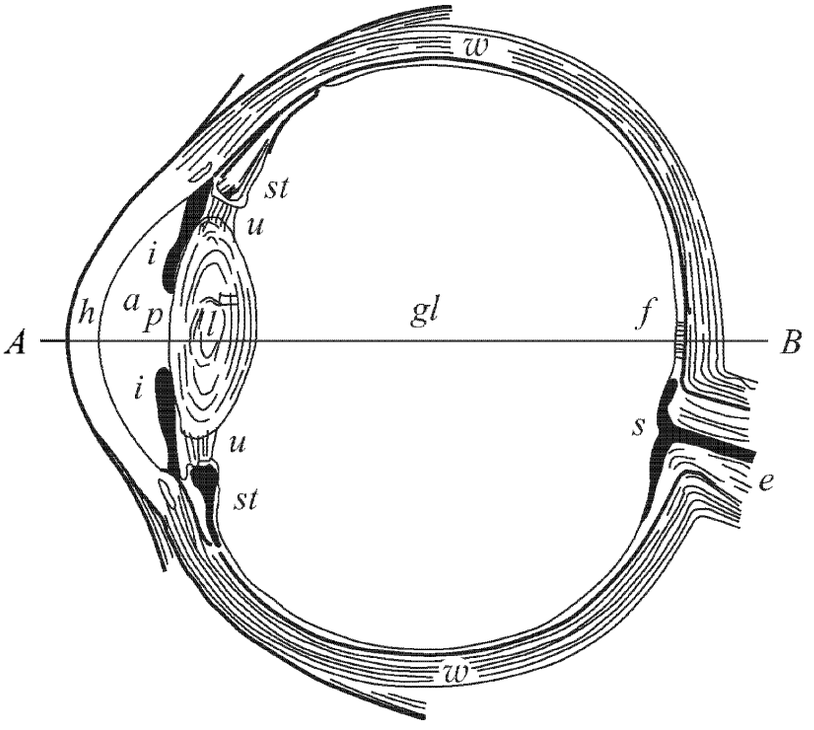


Рис. 79

2. На внутренней поверхности сосудистой оболочки расположена светочувствительная *сетчатая оболочка*, или *сетчатка* (*ретина*), состоящая главным образом из радиально расходящихся разветвлений зрительного нерва e, входящего во внутреннюю полость глаза несколько сбоку от линии прямого зрения AB, ближе к носовой полости. Около точки s толщина сетчатки составляет 0,2–0,22 мм; к периферии она непрерывно уменьшается. Оптическая система глаза дает изображение предметов на сетчатке, где оно вызывает зрительные раздражения, передаваемые в мозг по нервным волокнам.

Недалеко от точки s, где начинаются разветвления зрительного нерва, со стороны виска, находится так называемое *желтое пятно* f (горизонтальный поперечник 1–3 мм, вертикальный 0,8 мм). В его центральной части (размером $\approx 0{,}3$ мм по горизонтали и $\approx 0{,}2$ мм по вертикали) имеется углубление, на-

зываемое *центральной ямкой*. Здесь толщина сетчатки не превышает 0,08–0,1 мм. Желтое пятно и в особенности центральная ямка являются наиболее чувствительными местами сетчатки (при дневном зрении). Изображение рассматриваемой точки пространства всегда приводится к середине центральной ямки. Прямая, проходящая через эту точку и заднюю узловую точку глаза, вместе с параллельной ей прямой, проходящей через переднюю узловую точку, определяет так называемую *линию прямого зрения*. Она не совсем совпадает с *осью симметрии глаза*, так как преломляющие поверхности глаза не вполне симметричны и центрированы. Ось симметрии глаза можно провести в некотором смысле только условно. Обычно ее проводят через крайнюю, наиболее выступающую точку выпуклой поверхности роговицы (несколько выше точки A на рис. 79) и через центр зрачка. Она встречает заднюю стенку глаза где-то между точками f и s, образуя с линией прямого зрения AB угол около 5°.

3. Хрусталик l представляет собой прозрачное бесцветное тело, напоминающее двояковыпуклую линзу, передняя поверхность которой менее выпукла, чем задняя. Он состоит из слоев различной плотности, имеющих волокнистое строение. Наружный слой мягкий и почти студенистый; ядро же более твердое и вполне упругое. Показатель преломления наружного слоя хрусталика 1,405, средних слоев — около 1,429, ядра — 1,454. Мышца $ии$, рефлекторно напрягаясь или расслабляясь, может менять кривизну его поверхностей, главным образом передней. Этим осуществляется *аккомодация*, т.е. изменение оптической силы глаза, позволяющее фокусировать изображения на сетчатке. Так как деформация хрусталика может происходить только в определенных пределах, то для всякого глаза существуют определенные границы, в пределах которых глаз может отчетливо видеть предметы. Эти границы определяют так называемую *область аккомодации глаза*. Наиболее отдаленная граница, которую глаз может видеть отчетливо при вполне ослабленной мышце, называется *дальней*, а ближайшая граница, которую он способен отчетливо видеть при максимальном напряжении мышцы, — *ближней точкой ясного видения*. В ненапряженном состоянии нормальный глаз аккомодирован на рассмотрение бесконечно удаленных предметов, т.е. он собирает параллельные лучи в точке сетчатки. Таким образом, дальняя точка ясного видения для нормального глаза находится в бесконечности. В возрасте до десяти лет ближняя точка нормального глаза лежит впереди на расстоянии 7–8 см от глаза. К 30 годам это расстояние увеличивается примерно до 15 см, к 40–45 годам — до 25 см.

Приближая рассматриваемый предмет к глазу, мы увеличиваем *угол зрения*, а с ним и размеры изображения на сет-

чатке. Это позволяет рассмотреть более мелкие детали. Однако при максимально возможном приближении усиливается напряжение мышцы, деформирующей хрусталик. Работа глаза становится утомительной. В случае *нормального глаза* оптимальное расстояние для чтения и письма составляет около 25 см. Это расстояние для нормального глаза и принимается условно за *расстояние ясного зрения*. При необходимости рассмотреть (в течение ограниченного времени) более мелкие детали его можно уменьшить, оставаясь в пределах области аккомодации глаза.

4. У *близорукого глаза* оба предела ясного видения находятся ближе, а ближний предел *дальнозоркого глаза* — дальше, чем у нормального глаза. Дальний предел дальнозоркого глаза всегда отрицателен, т.е. дальняя точка ясного видения находится не спереди, а позади глаза. Такой глаз в ненапряженном состоянии может собирать на сетчатке только сходящиеся пучки лучей. При сильной дальнозоркости и ближний предел ясного зрения может сделаться отрицательным. У близорукого глаза в ненапряженном состоянии параллельные лучи сходятся перед, а у дальнозоркого — за сетчаткой.

Близорукость может быть обусловлена большей, а дальнозоркость — меньшей длиной глаза по сравнению с длиной нормального глаза. Отступления длины глаза от нормальной даже в пределах одного миллиметра уже могут приводить к заметной близорукости или дальнозоркости. Эти, а также другие недостатки глаза могут возникать из-за неправильных значений кривизны преломляющих поверхностей хрусталика и роговицы, несимметричности этих поверхностей, неправильного положения хрусталика и пр.

Термин «дальнозоркий» очень неудачен, так как с ним ассоциируется ошибочное представление, что дальнозоркость якобы способствует отчетливо видеть далекие предметы. Никакими преимуществами перед нормальным глазом дальнозоркий глаз не обладает, так как вся область перед глазом, отчетливо видимая дальнозорким глазом, отчетливо видна и нормальным глазом. Но область между ближними точками ясного видения дальнозоркого и нормального глаза недоступна для отчетливого рассматривания дальнозорким глазом. Близорукому глазу, напротив, недоступна область, лежащая за дальней точкой ясного видения нормального глаза [1]).

[1]) У близорукого глаза есть по крайней мере то преимущество, что он может выявить более мелкие детали, приближая рассматриваемый предмет на меньшие расстояния, чем это возможно для нормального глаза. У дальнозоркого глаза аналогичное преимущество появилось бы, если бы возникла необходимость рассматривать мнимые изображения, образующиеся с задней стороны глаза.

Как уже отмечено выше, с возрастом (в основном из-за уплотнения хрусталика, теряющего способность достаточно сжиматься) ближняя точка ясного видения удаляется от глаза. Для нормального глаза это удаление к 50 годам составляет около 50 см. Читать на таком расстоянии уже трудно. В дальнейшем указанная точка уходит еще дальше, удаляется в бесконечность и к 60–65 годам перескакивает на противоположную сторону глаза. Затем она начинает приближаться к нему, но уже с противоположной стороны. Это явление называется *старческой дальнозоркостью.*

Дальнозоркость устраняется применением очков с *положительными* (т.е. собирающими), а близорукость — с *отрицательными* (т.е. рассеивающими) стеклами. Положительные стекла приближают, а отрицательные отдаляют ближнюю точку ясного видения.

5. Систематические измерения привели к установлению средних значений всех параметров нормального глаза. Совокупность этих значений определяет так называемый *схематический глаз.*

В табл. 3 приведены эти параметры, как они были определены Гульстрандом. Так как показатели преломления пространств предметов ($n = 1$) и изображений ($n' = 1{,}336$) различны, то переднее f и заднее f' фокусные расстояния не совпадают по величине ($f'/f = -1{,}336$). Задняя узловая точка K' расположена внутри глаза на расстоянии f от заднего фокуса F', т.е. от сетчатой оболочки (см. рис. 43). Если линейный размер предмета равен l, то он виден из точки нахождения глаза под углом зрения $\alpha = l/L$. Поэтому линейный размер изображения на сетчатке будет $l' = f\alpha = fl/L$, или при рассматривании удаленных предметов $\approx 17\alpha$ мм.

6. Сетчатка глаза имеет весьма сложное строение. Схематический разрез ее приведен на рис. 80 *а* (предполагается, что свет падает сверху). Сетчатка состоит из нескольких слоев нервных клеток различного типа и назначения. В последнем слое находятся *светочувствительные рецепторные клетки — палочки и колбочки* (рис. 80 *б*). В них свет вызывает *первичные раздражения,* превращающиеся в электрические импульсы, которые по волокнам зрительного нерва передаются в кору головного мозга. Общая длина палочки 0,06–0,08 мм, колбочки — около 0,035 мм. Число колбочек в глазу достигает 7 миллионов, палочек — до 130 миллионов. Периферия сетчатки заполнена почти исключительно палочками. Число колбочек на единицу площади возрастает по мере приближения к желтому пятну, достигая максимума в его середине. Число колбочек на площади желтого пятна 13–15 тысяч. В его центральной части, в пределах кру-

Таблица 3

Схематический нормальный глаз по Гульстранду

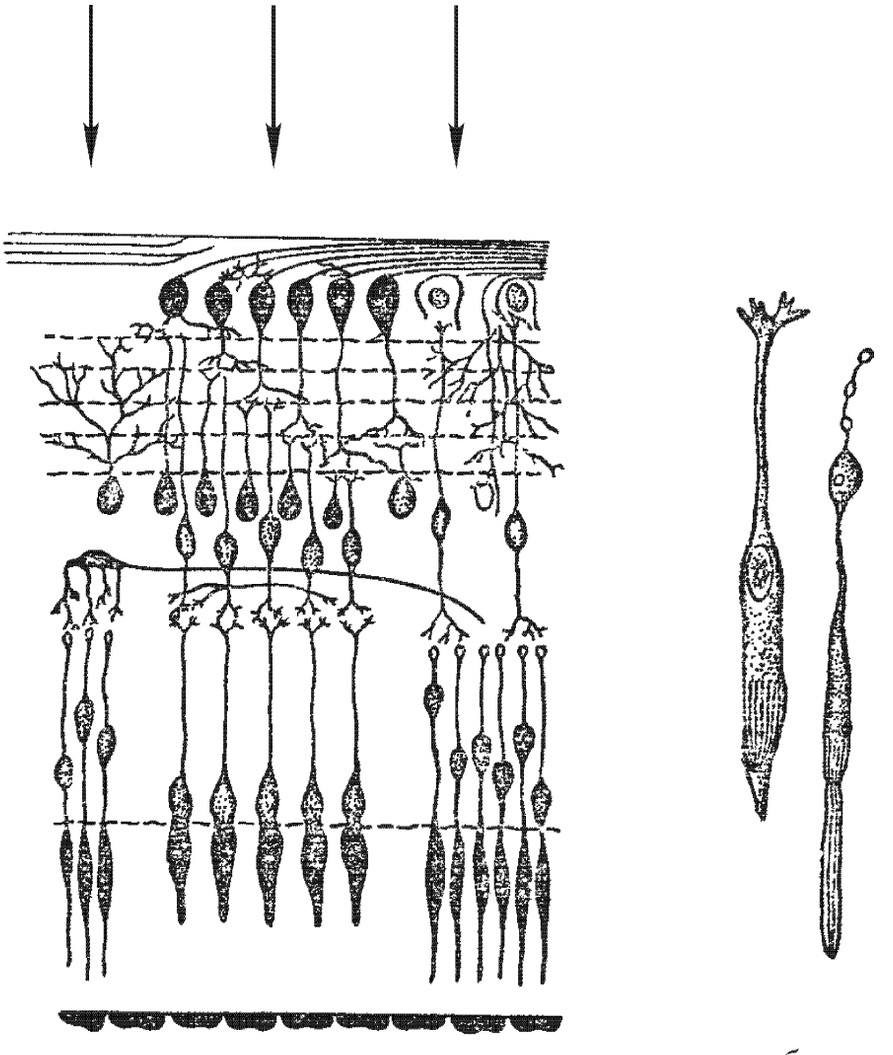
(Все длины — в миллиметрах, все расстояния — от вершины роговой оболочки)

	При расслабленной глазной мышце	При максимальном напряжении мышцы
Расстояние до задней поверхности роговицы	0,5	
Расстояние до передней поверхности хрусталика	3,6	3,2
Расстояние до задней поверхности хрусталика	7,2	
Радиус передней поверхности роговицы	7,7	
Радиус задней поверхности роговицы	6,8	
Радиус передней поверхности хрусталика	10	5,33
Радиус задней поверхности хрусталика	−6	−5,33
Оптическая сила глаза в диоптриях	58,64	70,57
Расстояние до передней главной плоскости	1,348	1,772
Расстояние до задней главной плоскости	1,602	2,086
Расстояние до переднего фокуса	−15,707	−12,397
Расстояние до заднего фокуса	24,387	21,016
Переднее фокусное расстояние	17,055	14,169
Заднее фокусное расстояние	−22,785	−18,930
Расстояние до центральной ямки	24	
Расстояние до входного зрачка	3,047	2,668
Расстояние до выходного зрачка	3,667	3,212
Увеличение в зрачках	0,909	0,941
Расстояние до ближней точки		102,3

га радиуса $\approx 0{,}2$ мм, палочек нет, есть одни только колбочки. Число их здесь достигает 180 тысяч на квадратном миллиметре, так что среднее расстояние между центрами соседних колбочек $\approx 0{,}0025$ мм.

В том месте, где внутрь глаза входит глазной нерв, т.е. вокруг точки s на рис. 79, на сетчатке нет ни палочек, ни колбочек. Это место называется *слепым пятном*. Его легко обнаружить, закрыв левый глаз и смотря правым на крестик рис. 81. Если

глаз приближать к рисунку, то на расстоянии 15–17 см кружок становится невидимым. Это получается тогда, когда изображение кружка попадает на слепое пятно.



а *б*

Рис. 80

Палочки обладают значительно большей чувствительностью к свету, чем колбочки. Ночью и в сумерки (при освещенности меньше 0,01 лк) зрительное ощущение вызывается за счет раздражения одних только палочек. Зато колбочки способны *различать цвета*, а палочки этой способностью не обладают. Практически цветное зрение начинается с освещенностей $\approx$1 лк. При освещенностях выше 10^2 – 10^3 лк зрение осуществляется почти исключительно колбочками.

Рис. 81

Детали предмета различаются наиболее отчетливо, когда его изображение получается на желтом пятне и в особенности на центральной ямке сетчатки. Это — так называемое *фовеальное зрение*, в отличие от *периферического зрения*, при котором изображение предмета получается вдали от центральной ямки. При периферическом зрении человек видит предметы значительно менее четко, различая в них мало деталей. Благодаря наличию на периферии в небольшом количестве колбочек, при периферическом зрении ощущение цветов имеет место, но только при условии, что яркость света достаточно велика, чтобы его воспринимали колбочки.

7. Две близкие светящиеся точки глаз воспринимает раздельно, а не как одну слившуюся точку только при условии, что угловое расстояние между этими точками не меньше определенного предела. Этот предел называется *разрешаемым угловым расстоянием*, а его обратная величина — *разрешающей способностью* или *остротой зрения глаза*. За единицу остроты зрения врачи принимают остроту зрения такого глаза, который разрешает угол в одну угловую минуту. Это — в среднем острота зрения нормального глаза.

Острота зрения максимальна, когда изображение приходится на центральную ямку сетчатки. Здесь плотность расположения колбочек максимальна, а каждая колбочка соединена с отдельным волокном зрительного нерва (число этих волокон — несколько миллионов). Мозг реагирует на раздражение каждой колбочки. Когда изображения двух близких светящихся точек приходятся на одну и ту же колбочку, эти точки действуют как одна светящаяся точка. В этом случае разрешения не получается. Для разрешения необходимо, чтобы изображения этих двух точек приходились на *разные колбочки*. На периферии плотность распределения светочувствительных рецепторов — колбочек и палочек — меньше, а каждое волокно зрительного нерва соединено с большим числом этих рецепторов. Вот почему максимальная разрешающая способность глаза получается при фовеальном зрении, когда изображение получается на центральной ямке сетчатки. Зрение же при помощи палочек предназначено не для повышения разрешающей способности глаза, а для *увеличения его чувствительности* в условиях работы при слабой освещенности. Для этого выгодно, как это и есть на самом деле, чтобы с каждым волокном зрительного нерва была соединена не одна палочка, а *большая группа* их. От этого усилится сигнал, передаваемый по этому волокну.

С изложенной точки зрения разрешаемое угловое расстояние глаза $\delta\vartheta$ равно углу, под которым видно из задней узловой точки глаза среднее расстояние δx между двумя соседними колбочками на центральной ямке, т.е. $\delta\vartheta = \delta x/f$, где f — переднее

фокусное расстояние. Подставив числа, приведенные выше, получим $\delta\vartheta \approx 30''$. И действительно, наблюдения показали, что для нормального глаза $\delta\vartheta$ несколько меньше одной угловой минуты. Приведенная оценка совпадает с той, какая получается на основе *дифракционных соображений* (см. § 56).

Поле зрения, соответствующее желтому пятну, невелико. На это пятно одновременно может проектироваться картина с угловыми размерами около 6° по горизонтальному направлению и около 4° по вертикальному. Поле зрения центральной ямки еще меньше — около 1° по горизонтали и вертикали. Такая ограниченность поля ясного зрения компенсируется, однако, тем, что живой глаз обладает способностью быстро поворачиваться в глазной впадине, за очень короткое время обегая все точки видимой поверхности рассматриваемого большого предмета. Благодаря этому поле ясного зрения глаза расширяется приблизительно до 150° по горизонтали и 120° по вертикали. Вместе с тем отмеченная подвижность глаза позволяет быстро концентрировать внимание на наиболее важных деталях предмета.

8. Глазу присущи все аберрации обычных оптических систем: и геометрические, и хроматические, и дифракционные. Однако геометрические и хроматические аберрации очень мало заметны, так как глаз использует особые средства снижения их вредного действия. Теперь выяснено, что колбочки имеют форму волноводов. Устройство каждого волновода таково, что лучи, проходящие через периферийную зону зрачка, воздействуют на колбочку слабее центральных лучей. В глазу используются асферические преломляющие поверхности, а хрусталик представляет собой линзу, показатель преломления которой непрерывно возрастает к центру. Это приводит к концентрации света преимущественно вблизи центров кружков рассеяния. Поэтому при не очень ярком освещении края кружков рассеяния мало интенсивны. При ярком же освещении зрачок глаза сильно сужен, а от этого кружок рассеяния уменьшается еще сильнее. Все это ослабляет сферическую аберрацию и кому. Астигматизм косых пучков и дисторсия почти незаметны, так как при подвижности глаза изображение каждой точки бессознательно приводится в наиболее выгодное место сетчатки — центральную ямку. Искривление поверхности изображения компенсируется сферической формой сетчатки. Хроматическая аберрация почти незаметна, ввиду чувствительности сетчатки только к сравнительно узкой части спектра.

9. Остановимся теперь схематически на теории *цветного зрения* Юнга–Гельмгольца (1821–1894), которая лучше других теорий согласуется с наблюдаемыми фактами. Она исходит из того экспериментально установленного факта, что ощущение любого цвета можно получить смешением спектрально чистых излуче-

ний красного, зеленого и сине-голубого цветов. На этом основании теория предполагает, что в глазу есть только *три* типа светочувствительных приемников. Они отличаются друг от друга областями спектральной чувствительности. Красный свет воздействует преимущественно на приемники первого типа, зеленый — второго, сине-голубой — третьего. Сложением излучений таких трех цветов в различных количествах можно получить любую комбинацию возбуждений всех трех приемников, а это физиологически эквивалентно получению любого цветового ощущения. Еще не установлено, имеются ли приемники всех трех типов в каждой колбочке или существуют три различных типа колбочек.

Встречаются люди (более 1% среди мужчин и около 0,1% среди женщин), зрение которых характеризуется отсутствием приемников одного из трех указанных типов. Они называются *дихроматами*. Дихромат не различает цвета всех излучений, которые для людей с нормальным зрением различаются по степени возбуждения приемника, недостающего у дихромата. Еще реже (примерно раз на миллион людей) встречаются *монохроматы*, у которых есть приемники только одного типа. Такие люди совсем не различают цвета.

10. Чувствительность глаза к излучениям различных длин волн характеризуется *кривой видности*. На этой кривой по оси абсцисс откладывается длина волны, а по оси ординат — *видность* V_λ, т.е. величина, обратная энергетической мощности излучений, которые при оценке глазом воспринимаются как одинаково яркие. Визуальное сравнение яркостей излучений далеких друг от друга длин волн затруднительно. Поэтому для построения кривой видности обычно применяют *метод малых ступеней*, т.е. сравнивают попарно по видимой яркости излучения столь близких длин волн, что разница в цвете не затрудняет такое сравнение. Несмотря на субъективность этого метода, воспроизводимость результатов достаточно хороша, а кривые видности для различных людей не сильно отличаются друг от друга.

Кривая видности среднего нормального глаза при дневном зрении, утвержденная Международной осветительной комиссией, приведена на рис. 82. Она имеет максимум в желто-зеленой части спектра при $\lambda = 555$ нм, условно принимаемый за единицу. При сумеречном зрении, когда работает только палочковый аппарат, кривая видности сохраняет свой общий вид, но смещается в сторону коротких волн с максимумом около 510 нм. При этом область максимальной чувствительности сетчатки смещается на 10–20° в сторону от центральной ямки.

Еще в 1825 г. Пуркинье (1787–1869) наблюдал, что излучения различного цвета, воспринимаемые глазом как одинаково яркие,

меняют свою видимую яркость не одинаково, если их ослаблять в одно и то же число раз. Яркость излучений с большей длиной волны уменьшается быстрее, чем с более короткой длиной волны. Если ограничить поле зрения, чтобы оно не превосходило 1,5°, и сконцентрировать сравниваемые излучения в пределах центральной ямки, то свет будет восприниматься только колбочками. Исследования показали, что в этом случае явление

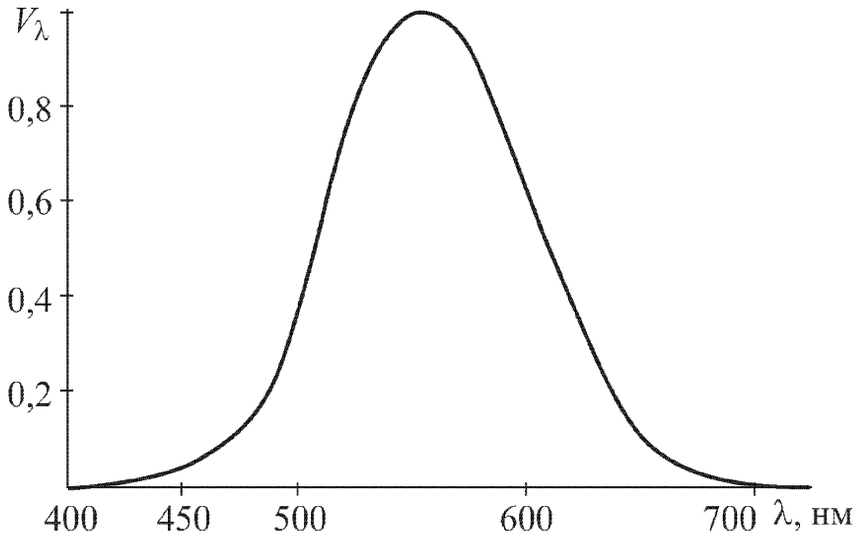


Рис. 82

Пуркинье не наблюдается. Все это хорошо согласуется с двойным механизмом восприятия света: посредством колбочек и посредством палочек, которым соответствуют различные кривые видности.

11. Область, доступная зрительному восприятию глаза, конечно, не обрывается резко на длинах волн 400 и 760 нм. При $\lambda = 400$ нм видность V_λ примерно в 2500 раз, а при $\lambda = 760$ нм — в 20 000 раз меньше, чем в максимуме. В условиях темновой адаптации глаз может видеть в очень слабой степени интенсивные инфракрасные лучи с длинами волн до 950, а ультрафиолетовые — до 300 нм.

Границы видимой области, а также сама форма кривой видности человеческого глаза не случайны. Глаз сформировался в процессе длительной эволюции, приспособившись к условиям освещения земных предметов солнечным светом, а также к условиям сумеречного и ночного освещения. На эту мысль наводит уже то обстоятельство, что на видимую область спектра приходится более 40 % энергии излучения Солнца, хотя она и занимает интервал менее одной октавы. Далее, солнечный спектр вблизи поверхности Земли практически обрывается на длине волны 290 нм. Более короткие волны задерживаются слоем озона в атмосфере. Было бы биологически нецелесообразно, если бы глаз обладал способностью принимать излучение с длинами волн короче 290 нм.

Более того, глаз должен защитить себя от ультрафиолетовых лучей. Они в большинстве случаев химически разлагают органические вещества и могут убивать живые микроорганизмы и клетки. Особенно вредно попадание ультрафиолетовых лучей на сетчатку глаза. Чувствительность сетчатки к ультрафиолетовым лучам довольно велика и, как показал С.И. Вавилов, имеет резкий максимум при $\lambda = 380$ нм. Однако от длинноволновых ультрафиолетовых лучей ($290 < \lambda < 400$ нм), пропускаемых земной атмосферой, глаз защищен собственными средствами. Такие лучи сильно поглощаются внутри глаза, особенно в хрусталике, и лишь ничтожная доля их доходит до сетчатки. С этим и связан сильный спад кривой чувствительности глаза на границе и в области ультрафиолета. При хирургическом удалении хрусталика, если только глаз продолжает функционировать, его чувствительность к ультрафиолетовым лучам заметно повышается [1]).

Высокая чувствительность к инфракрасному излучению, если бы глаз обладал таковой, была бы не только нецелесообразна, но и сделала бы невозможной работу глаза в условиях солнечного освещения. Причина этого — в тепловом излучении глаза. При низкой температуре глаза (310 К) все его тепловое излучение приходится на инфракрасную область. Мощность, излучаемую 1 см^2 стенки глаза внутрь глазной полости, нетрудно оценить, пользуясь законом Стефана–Больцмана (см. § 115). Она составляет около 0,105 Вт/см^2. Общая внутренняя поверхность глазной полости ≈ 17 см^2, так что полная мощность излучения глаза внутрь глазной полости будет $\approx 1{,}7$ Вт. Оценим теперь мощность прямого солнечного излучения, попадающего через зрачок глаза, пользуясь значением солнечной постоянной (0,139 Вт/см^2), приведенным в задаче к § 5. Получим $0{,}139\pi r^2 \approx$ $\approx 0{,}017$ Вт (радиус r зрачка глаза принят равным 2 мм). Эта величина примерно в 100 раз меньше мощности собственного теплового излучения глаза. Если бы чувствительность глаза в инфракрасной области спектра была столь же велика, что и в желто-зеленой, то собственное тепловое инфракрасное излучение глаза затмило бы не только рассеянное излучение неба, но и прямой солнечный свет. Работа глаза как органа зрения стала

[1]) Ультрафиолетовые лучи, попадая в глаз, вызывают его флуоресценцию, преобразуясь в синеватый свет, воспринимаемый сетчаткой. Такое косвенное воздействие ультрафиолета проявляется, например, в следующем опыте. Если посмотреть на мощную ртутную кварцевую лампу, излучающую ультрафиолет, через специальное темное стекло, задерживающее видимые, но пропускающее ультрафиолетовые лучи, то благодаря флуоресценции внутри глаза окружающий воздух покажется наполненным синеватым туманом, напоминающим табачный дым.

бы абсолютно невозможной. В этом, по мнению С.И. Вавилова (1891–1951), одна из причин, почему нет зрения в инфракрасных лучах. Другая причина заключается в малости энергии инфракрасных квантов света. Процесс зрения должен начинаться либо химическими, либо фотоэлектрическими действиями, а они вызываются отдельными квантами света и не могут происходить, если энергия кванта меньше некоторого минимального предела.

12. Для возбуждения светочувствительных рецепторов необходимо, чтобы они поглощали свет. Как правило, чувствительность рецептора к свету определенной длины волны должна быть тем выше, чем сильнее поглощает он свет той же длины волны. Это обстоятельство уже давно побудило исследователей искать в сетчатке глаза *светочувствительные пигменты*. Такой пигмент (*родопсин*, или *зрительный пурпур*) был обнаружен в палочках человеческого глаза. Это — вещество розоватого цвета, разлагающееся (выцветающее) на свету и снова восстанавливающееся в темноте. Спектральная кривая поглощения родопсина близка к сумеречной кривой видности. Извлечь светочувствительные пигменты из колбочек сетчатки человека пока не удалось, хотя наблюдения живого глаза через зрачок, по-видимому, подтверждают, что в центральной части сетчатки имеются пигменты, выцветающие на свету.

Глаз человека обладает способностью приспосабливаться к освещенностям, меняющимся в необычайно широких пределах. Прямые солнечные лучи создают освещенности $\sim 10^5$ лк, а в полной темноте глаз способен отличать от темноты предметы с освещенностью 10^{-6} лк. Глаз способен воспринимать световые потоки в интервале 10^{-17}–10^{-5} Вт. Процесс приспособления глаза к тому или иному уровню яркости света называется *адаптацией*. При повышении яркости происходит *световая*, при понижении — *темновая адаптация*. При переходе от яркости ~ 1000 кд/м^2 к темноте чувствительность глаза возрастает в течение часа примерно в 10 миллионов раз. Сначала чувствительность возрастает очень быстро, затем ее рост замедляется и после часа пребывания в темноте уровень чувствительности почти не меняется. Световая адаптация происходит много быстрее. При средних яркостях она продолжается 1–3 минуты. Изменение чувствительности к световому восприятию в столь широких пределах свойственно палочкам. Темновая адаптация колбочек происходит значительно быстрее, причем чувствительность колбочек возрастает всего в 10–100 раз. В состоянии максимальной световой адаптации глаз может без вреда переносить сравнительно большие яркости (например, яркости белых матовых поверхностей, освещаемых прямым солнечным светом). При больших яркостях необходима искусственная защита гла-

за. Так, наблюдение солнечного затмения можно вести только через закопченное стекло или другой сильно ослабляющий светофильтр. При пребывании на ледниках и в горах на большой высоте необходимо пользоваться темными или цветными очками. Здесь очки необходимы также для защиты от ультрафиолетовых лучей, достигающих на больших высотах значительных интенсивностей и вредно действующих на глаз.

Адаптация глаза происходит автоматически, но механизм ее выяснен не совсем. При резком увеличении яркости света почти мгновенно сужается зрачок. При уменьшении яркости до прежнего уровня зрачок вновь расширяется. Происходит также изменение чувствительности самих рецепторов. Долгое время считалось, что такое изменение чувствительности связано с изменением количества родопсина в палочках. Когда яркость света повышается, родопсин выцветает, и чувствительность глаза понижается. При уменьшении яркости происходит восстановление родопсина и связанное с этим повышение чувствительности. Однако работы последнего времени поставили под сомнение эту точку зрения. Наблюдения показали, что чувствительность глаза к свету меняется всего сильнее, когда уменьшение родопсина еще очень невелико. Наоборот, когда концентрация родопсина резко уменьшается, чувствительность уменьшается незначительно. С изложенной точки зрения это понять трудно. Возможно, играет роль перестройка корковых центров головного мозга, воспринимающих свет, т.е. повышение и понижение их чувствительности.

§ 22. Фотометрические понятия и единицы

1. Излучение в пространстве или в прозрачной однородной среде можно характеризовать *интенсивностью*, *спектральным составом* и *поляризацией*. При этом надо иметь в виду, что пучки строго параллельных лучей являются идеализацией и никогда не встречаются в действительности. Конечной энергией могут обладать лучи, направления которых заполняют *конечные телесные углы*, хотя величина этих углов и может быть очень малой.

Заметив это, возьмем в поле излучения произвольную малую площадку ds. Она должна быть настолько мала, чтобы характеристики излучения практически не изменялись при переходе от одной точки площадки к другой. В то же время линейные размеры площадки ds должны быть велики по сравнению с длинами волн излучения, чтобы к излучению можно было применять понятия и законы геометрической оптики. Проведем через площадку ds лучи, заполняющие какой-то телесный угол Ω.

Энергия, переносимая этими лучами в единицу времени, называется *лучистым потоком* Φ, проходящим через площадку ds в телесный угол Ω. Если телесный угол $d\Omega$ бесконечно мал, а площадка ds перпендикулярна к его оси, то лучистый поток можно представить в виде $d\Phi = I\,ds\,d\Omega$. Величина I есть лучистый поток, отнесенный к единичной площадке, перпендикулярной к излучению, и к единице телесного угла. Она называется *интенсивностью лучистого потока*, или *лучистостью излучения* в направлении оси телесного угла $d\Omega$. Если нормаль к площадке ds образует с направлением излучения угол ϑ, то

$$d\Phi = I\,ds\cos\vartheta\,d\Omega,$$

так как лучистые потоки через площадку ds и ее проекцию $ds_\perp = ds\cos\vartheta$, перпендикулярную к излучению, одинаковы. Для краткости проекцию $ds\cos\vartheta$ называют *видимой величиной площадки* ds, если ее рассматривать под углом ϑ к нормали. При выполнении принципа суперпозиции лучи, проходящие в данный момент времени через определенную точку среды, совершенно независимы друг от друга. Поэтому для полной характеристики состояния излучения необходимо указать интенсивность лучистого потока по любому из бесчисленного множества направлений, проходящих через каждую точку пространства. При этом надо учитывать отдельно каждое из двух прямо противоположных направлений, поскольку независимость лучей имеет место и для этих направлений.

Энергия, проходящая за время dt через площадку ds в телесный угол $d\Omega$, представится выражением $I\,ds\,d\Omega\,dt\cos\vartheta$. При этом время dt, хотя оно и входит в виде дифференциала, должно быть все же велико по сравнению с периодами колебаний волн, входящих в излучение. Иначе значение мощности, например, монохроматического излучения, как энергии, отнесенной к единице времени, при малом интервале dt, в течение которого измеряется энергия, зависело бы от фазы колебаний в момент начала измерения. Независимость имела бы место только тогда, когда время dt случайно содержало бы целое число колебаний. Если же dt велико по сравнению с периодами колебаний любых волн, входящих в излучение, то измеренная мощность излучения практически не будет зависеть от выбора dt.

2. *Объемной плотностью лучистой энергии* называется энергия, содержащаяся в единице объема пространства. Чтобы выразить ее через интенсивность излучения I, возьмем сначала малый конус лучей с телесным углом $d\Omega$ при вершине (рис. 83). Через площадку ds, перпендикулярную к излучению, за время dt проходит энергия $I\,ds\,d\Omega\,dt$. Если v — скорость распространения излучения, то за время dt эта энергия распростра-

нится на расстояние $v\,dt$ и заполнит объем $ds\,v\,dt$, заштрихованный на рис. 83. (Предполагается, что дисперсии нет, так что скорость распространения энергии совпадает с фазовой скоростью v.) Разделив предыдущее выражение на этот объем, найдем $(I/v)\,d\Omega$ — плотность лучистой энергии, создаваемую рассматриваемым элементарным пучком лучей. Полная плотность лучистой энергии u найдется интегрированием этого выражения по всем телесным углам:

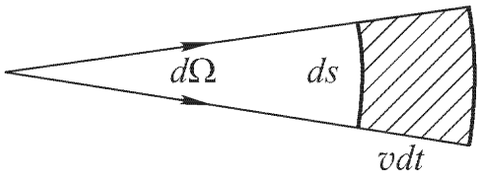


Рис. 83

$$u = \frac{1}{v}\int I\,d\Omega, \tag{22.1}$$

причем величина I может зависеть от направления лучей. Когда I одинакова по всем направлениям, то

$$u = 4\pi I/v. \tag{22.2}$$

3. Величины Φ, I, u можно подвергнуть *спектральному разложению* по частотам или длинам волн. Строго монохроматическое излучение, как и параллельные пучки лучей, никогда не реализуется в действительности. Каждое излучение, обладающее конечной энергией, *занимает конечный интервал частот или длин волн.* Например, величину u можно представить в виде

$$u = \int_0^\infty u_\nu(\nu)\,d\nu = \int_0^\infty u_\omega(\omega)\,d\omega = \int_0^\infty u_\lambda(\lambda)\,d\lambda. \tag{22.3}$$

Подынтегральные выражения $u_\nu\,d\nu$, $u_\omega\,d\omega$, $u_\lambda\,d\lambda$ имеют смысл объемной плотности лучистой энергии в спектральных интервалах $(\nu, \nu+d\nu)$, $(\omega, \omega+d\omega)$, $(\lambda, \lambda+d\lambda)$ соответственно. Ради краткости эти термины обычно применяют к самим функциям $u_\nu(\nu)$, $u_\omega(\omega)$, $u_\lambda(\lambda)$, опуская дифференциалы $d\nu$, $d\omega$, $d\lambda$, хотя это и не совсем логично. При установлении связи между этими функциями надо, конечно, приравнивать друг другу дифференциалы $u_\nu\,d\nu$, $u_\omega\,d\omega$, $u_\lambda\,d\lambda$, поскольку они представляют в различных формах *одну и ту же величину*, если только соответствие между λ, ν и ω устанавливается соотношениями $\lambda = c/\nu = 2\pi c/\omega$. Таким путем находим

$$u_\nu = (\lambda^2/c)u_\lambda. \tag{22.4}$$

Говоря о спектральном распределении, надо указывать, идет ли речь о функции u_ν (распределение по частотам) или о функции u_λ (распределение по длинам волн). Так, в спектре излучения Солнца функция u_ν имеет максимум в инфракрасной области приблизительно при $\lambda = 880$ нм, а функция u_λ — в желто-зеленой части приблизительно при $\lambda = 500$ нм.

Далее, интенсивность излучения I в каждой точке пространства можно представить в виде суммы интенсивностей двух излучений, линейно поляризованных во взаимно перпендикулярных плоскостях. Однако для наших целей в этом пока нет необходимости.

4. Все приведенные энергетические характеристики излучения измеряются в *механических единицах*, например по производимому ими тепловому действию. Так, в системе СИ лучистый поток измеряется в *ваттах* (Вт), интенсивность излучения — в *ваттах на стерадиан-квадратный метр* (Вт/ср·м^2), объемная плотность лучистой энергии — в *джоулях на кубический метр* (Дж/м^3). Такие единицы применяются, например, в теории теплового излучения. Однако в видимой области спектра представляет интерес характеризовать излучение по *зрительному* или *световому ощущению*, оцениваемому по действию света на глаз человека. Соответствующие характеристики и их единицы называются *световыми*, или *фотометрическими*, в отличие от *энергетических* величин и единиц, о которых говорилось выше.

Световые измерения, конечно, содержат немалый элемент субъективизма, поскольку световые впечатления, получаемые различными людьми от одного и того же светового потока, не совсем совпадают между собой. Когда указывают количественные соотношения световых единиц с энергетическими, то при этом имеют в виду не индивидуальный, а какой-то *средний человеческий глаз*. Сами названия величин, характеризующих излучение, при световых измерениях несколько изменяют. Вместо лучистой энергии говорят о *световой энергии*, вместо лучистого потока — о *световом потоке*, вместо интенсивности излучения — об *интенсивности света* и т.п. Вообще, вместо энергетических величин и единиц вводят *световые* или *фотометрические* величины и единицы. Мы не будем вводить для энергетических и фотометрических величин различные обозначения, так как в каждом отдельном случае будет ясно, в каком смысле следует понимать ту или иную величину.

5. *Силой света источника* $\mathcal{J}$ в заданном направлении называют световой поток, посылаемый им в этом направлении и отнесенный к единице телесного угла. (Когда пользуются энергетическими единицами, то говорят об *энергетической силе источника*, измеряя эту величину в Вт/ср.) Обычно это понятие относят к *точечному источнику света*, т.е. характеризуют им источник на расстояниях, больших по сравнению с его линейными размерами. Вообще говоря, величина $\mathcal{J}$ зависит от направления излучения. Световой поток, посылаемый точечным источником в телесный угол $d\Omega$, определяется выражением $d\Phi = \mathcal{J}\, d\Omega$.

Полный световой поток, исходящий от источника,

$$\Phi = \int \mathcal{J}\, d\Omega,$$

причем интеграл берется по полному телесному углу 4π. Величина $\mathcal{J}_0 = \Phi/(4\pi)$ называется *средней сферической силой света источника.*

Единицей силы света источника в системе СИ служит *кандела* (старое название — *свеча*). Это — *основная фотометрическая единица.* Она осуществляется с помощью светового эталона в виде абсолютно черного тела при температуре затвердевания чистой платины (2046,6 К) и давлении 101 325 Па. Устройство эталона показано на рис. 84 (*1* — платина, *2* — трубочка из плавленой окиси тория, *3* — сосуд из плавленой окиси тория, *4* — засыпка из окиси тория, *5* — сосуд из кварца). Нагрев и расплавление платины производятся токами высокой частоты. Излучателем света служит трубочка *2*, стенки которой по всей длине имеют одну и ту же температуру, равную температуре окружающей нагретой платины. *Кандела* (кд) *есть сила света, излучаемого в направлении нормали с* 1/60 см^2 *излучающей поверхности указанного светового эталона.* До введения этого эталона (он был введен с 1 января 1948 г.) единицей силы света источника служила *международная свеча* (м. св.), равная 1,005 кд. Она осуществлялась специально изготовленными электрическими лампочками.

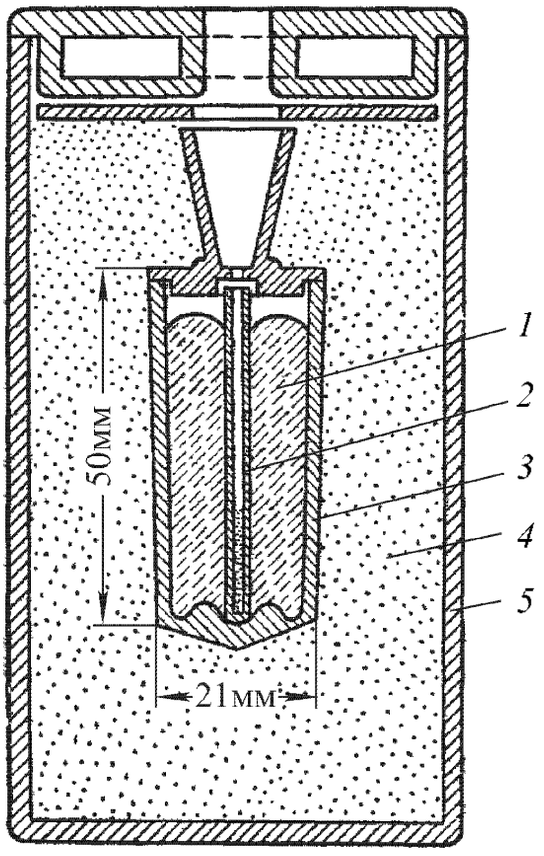


Рис. 84

Единица светового потока есть *люмен* (лм) — световой поток, посылаемый источником в 1 канделу внутрь телесного угла в 1 стерадиан.

Лучистый поток, измеренный в ваттах, можно рассматривать и как световой поток, измеряя его в люменах. *Световая эффективность лучистого потока*, или просто *световая эффективность*, есть число люменов, соответствующее мощности в

один ватт (лм/Вт). Обратная величина (Вт/лм) называется *механическим эквивалентом света*. Из-за различной чувствительности глаза к различным участкам спектра обе эти величины зависят от длины волны λ. Принято приводить их значения для $\lambda = 555$ нм, где чувствительность глаза максимальна. Для такой длины волны эти величины по новейшим измерениям равны соответственно:

$$625 \text{ лм/Вт и } 0{,}00160 \text{ Вт/лм}.$$

Пользуясь кривой видности среднего человеческого глаза, приведенной на рис. 82, нетрудно найти значения этих величин для любой длины волны видимого спектра излучения.

6. Из закона сохранения энергии следует, что полный световой поток, посылаемый источником, не может быть увеличен никакими отражающими и преломляющими устройствами, по крайней мере пока они остаются неподвижными. Такие устройства могут только *перераспределять световые потоки* по различным направлениям, что и осуществляется, например, прожекторами. Если точечный источник света помещен в прозрачной однородной среде, то на любых расстояниях от него остается постоянным не только полный поток Φ, испущенный источником в какой-либо момент времени, но и световой поток $d\Phi = \mathcal{J}\, d\Omega$ в пределах любого телесного угла $d\Omega$, исходящего из источника. Ввиду того, что телесный угол $d\Omega$ никак не связан с расстоянием r до источника, не будет зависеть от r и сила света источника $\mathcal{J}$. Интенсивность света I на расстоянии r найдется делением $d\Phi$ на площадь $ds = r^2\, d\Omega$ перпендикулярного сечения рассматриваемого элементарного пучка лучей. Это дает

$$I = \mathcal{J}/r^2, \tag{22.5}$$

т.е. *интенсивность света обратно пропорциональна квадрату расстояния до точечного источника.*

7. Световой поток, приходящийся на единицу площади освещаемой поверхности, называется *освещенностью* E этой поверхности. (При энергетических измерениях вместо этого термина пользуются термином *энергетическая освещенность*, или *облученность*, единицей которой служит Вт/м^2.) Пусть источник точечный, а лучи падают под углом ϑ к нормали к освещаемой поверхности. Тогда $d\Phi = \mathcal{J}\, d\Omega = \mathcal{J}\, ds \cos\vartheta / r^2$. Разделив на площадь поверхности ds, получим

$$E = \frac{\mathcal{J}}{r^2} \cos\vartheta. \tag{22.6}$$

Таким образом, *освещенность, создаваемая точечным источником в отсутствие поглощения, обратно пропорциональна квадрату расстояния до него и прямо пропорциональна косинусу угла между направлением падающих лучей и нормалью к*

освещаемой поверхности. Первая часть этого утверждения, а также формула (22.5) называются *законом обратных квадратов.* Единица освещенности есть *люкс* (лк) — освещенность, создаваемая световым потоком в 1 люмен, равномерно распределенным по площади в 1 м^2. Освещенность в 1 лк создается точечным источником силою в 1 кд на внутренней поверхности шара радиуса 1 м, если он помещен в центре этого шара и излучает равномерно по всем направлениям.

Чтобы составить конкретное представление о величине люкса, приведем некоторые цифры. Освещенность от Солнца вне земной атмосферы на среднем расстоянии Земли от Солнца $1{,}35 \cdot 10^5$ лк. Освещенность в одну-две десятых люкса создает ночью при безоблачном небе полная Луна. Освещенность, создаваемая молодой Луной или Луной на ущербе, порядка нескольких сотых люкса. Безоблачное звездное небо создает ночью освещенность в тысячные доли люкса. Освещенность в десятитысячные доли люкса позволяет с трудом ориентироваться ночью. При освещенности порядка одного люкса можно с трудом читать. Скорость чтения быстро нарастает при увеличении освещенности до 50 лк. При дальнейшем увеличении освещенности до 100–150 лк она растет более медленно, а дальше этого предела возрастание скорости чтения становится малоощутимым. Освещенность 50 лк уже удовлетворительна для чтения и письма. Инструкциями по охране труда установлены определенные нормы минимальной освещенности рабочих помещений. Освещенность рабочей поверхности (стола) ни для каких видов работ не должна быть меньше 10 лк. При очень тонкой работе, связанной с различением мелких деталей, черточек, букв, рисунков (угол зрения меньше $2'$), требуется освещенность не менее 200 лк. В классах и аудиториях на столах учащихся и черных досках освещенность должна быть не менее 75 лк.

8. Для протяженных (не точечных) источников света вводится понятие *поверхностной яркости*, или просто *яркости* B. (При энергетических измерениях вместо этого термина употребляют термин *энергетическая яркость* и для нее вводят единицу Вт/(ср·м^2).) Понятие поверхностной яркости неприменимо для точечных источников, т.е. источников, угловые размеры которых лежат за пределами разрешающей способности глаза или оптического инструмента, которым он вооружен. Возьмем на поверхности излучающего источника малую площадку ds и проведем от нее световой луч под углом ϑ к ее нормали. При рассмотрении излучения в этом направлении более существенна не сама площадка ds, а ее *видимая величина*, т.е. $ds\cos\vartheta$ (см. п. 1). Яркостью поверхности в рассматриваемом направлении называется световой поток $d\Phi$, исходящий из площадки ds в этом направлении, отнесенный к единице телесного угла и к единице

ее видимой величины:

$$B_\vartheta = \frac{d\Phi}{d\Omega\, ds \cos\vartheta} = \frac{d\mathcal{J}}{ds\cos\vartheta}, \qquad (22.7)$$

где $d\mathcal{J} = d\Phi/d\Omega$ — сила света площадки ds в том же направлении (рис. 85). Буква B снабжена индексом ϑ, так как яркость, вообще говоря, зависит от угла ϑ, под которым рассматривается площадка ds.

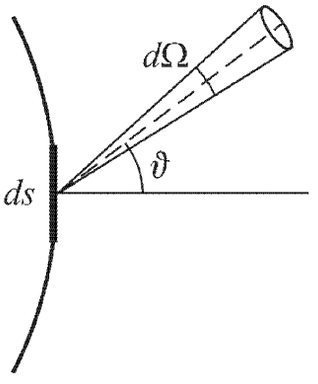


Рис. 85

Источники света, поверхностная яркость которых B не зависит от направления излучения, называются источниками, подчиняющимися *закону Ламберта* (1728–1777). Для таких источников, как видно из (22.7), сила света $d\mathcal{J}$ элементарной площадки ds пропорциональна $\cos\vartheta$. Однородный светящийся шар, подчиняющийся закону Ламберта, кажется одинаково ярким в середине и по краям. Такие наблюдения и привели Ламберта к формулировке своего закона. В действительности от закона Ламберта наблюдаются большие отступления.

В § 113 будет показано, что при температурном излучении поверхность непрозрачного тела излучала бы по закону Ламберта, если бы коэффициент отражения света от этой поверхности для каждой длины волны не зависел от угла падения. Для гладких поверхностей, отражающих зеркально, это условие не выполняется (см. § 65). Но для матовых поверхностей, отражающих диффузно, оно может выполняться с той или иной степенью приближения. Для таких поверхностей при температурном излучении приближенно соблюдается закон Ламберта. Он строго справедлив при температурном излучении абсолютно черного тела. Матовые поверхности, например освещенная белая поверхность тела, покрытая окисью магния, или наружная поверхность колпака из хорошего молочного стекла, освещенного изнутри, являются источниками, довольно хорошо подчиняющимися закону Ламберта. Однако к этим случаям вывод закона Ламберта, приводимый в § 113, неприменим, так как в них речь идет не о самосветящихся телах и температурном излучении, а о телах, рассеивающих свет от посторонних источников.

Единицей яркости является *кандела на квадратный метр* (кд/м^2). Это — яркость плоской поверхности, сила света которой в перпендикулярном направлении составляет одну канделу с каждого квадратного метра. Если при тех же условиях сила света равна одной канделе с каждого квадратного сантиметра, то соответствующая единица называется *стильб* (сб). Очевидно, 1 сб = 1 кд/см^2 = 10^4 кд/м^2. В табл. 4 приведены значения яркости некоторых светящихся поверхностей.

Таблица 4

Источник света	Яркость, кд/м^2	Источник света	Яркость, кд/м^2
Ночное безлунное небо	Около $1 \cdot 10^{-4}$	Кратер обычной электрической угольной дуги	$1{,}5 \cdot 10^8$
Полная Луна, видимая сквозь атмосферу	$1 \cdot 10^3$	Шаровая ртутная лампа сверхвысокого давления (СВДШ)	$1{,}2 \cdot 10^9$
Пламя обычной стеариновой свечи	$2{,}5 \cdot 10^3$	Солнце	$1{,}5 \cdot 10^9$
Ясное дневное небо	$1{,}5 \cdot 10^4$	Импульсная стробоскопическая лампа (ИСШ)	$1 \cdot 10^{11}$
Газосветная лампа	$5 \cdot 10^4$		
Спираль газонаполненной лампы накаливания	$5 \cdot 10^6$		

9. *Светимостью* K называется полный световой поток, посылаемый единицей светящейся поверхности в одну сторону, т.е. в телесный угол $\Omega = 2\pi$. Ее единица такая же, что и единица освещенности, т.е. лм/м^2. (При энергетическом рассмотрении вместо этого термина употребляется *энергетическая светимость*, или *излучательность*, а единицей служит Вт/м^2.) Так как световой поток с единицы поверхности в телесный угол $d\Omega$ равен $d\Phi = B_\vartheta \cos\vartheta\, d\Omega$, то

$$K = \int B_\vartheta \cos\vartheta\, d\Omega = 2\pi \int\limits_0^{\pi/2} B_\vartheta \cos\vartheta \sin\vartheta\, d\vartheta. \tag{22.8}$$

Для поверхностей, излучающих по закону Ламберта, яркость $B_\vartheta = B$ не зависит от угла ϑ, а потому в этом случае

$$K = \pi B. \tag{22.9}$$

10. Рассчитаем теперь освещенность, создаваемую протяженным источником света с известной поверхностной яркостью B_v. Возьмем на поверхности источника бесконечно малую площадку ds (рис. 86). Пусть ds' — освещаемая площадка, r — расстояние между площадками, ϑ и ϑ' — углы, составляемые нормалями к площадкам с прямой MN, соединяющей их. Телесный угол, под которым из площадки ds видна площадка ds', равен $ds' \cos\vartheta'/r^2$. В него площадка ds посылает световой поток $d\Phi = B_\vartheta\, ds \cos\vartheta\, ds' \cos\vartheta'/r^2$, или $d\Phi = B_\vartheta\, d\Omega \cos\vartheta'\, ds'$, где $d\Omega = ds \cos\vartheta/r^2$ — телесный угол, под которым из освещаемой площадки ds' видна излучающая площадка ds. Разделив на ds', найдем освещенность dE площадки ds', создаваемую потоком $d\Phi$.

Полная освещенность E найдется интегрированием полученного выражения по всей видимой поверхности источника:

$$E = \int B_\vartheta \cos\vartheta' \, d\Omega. \qquad (22.10)$$

В качестве примера возьмем источник в виде равномерно светящегося диска с поверхностной яркостью $B_\vartheta = B$, не зависящей от угла ϑ. Освещаемую площадку ds' расположим на геометри-

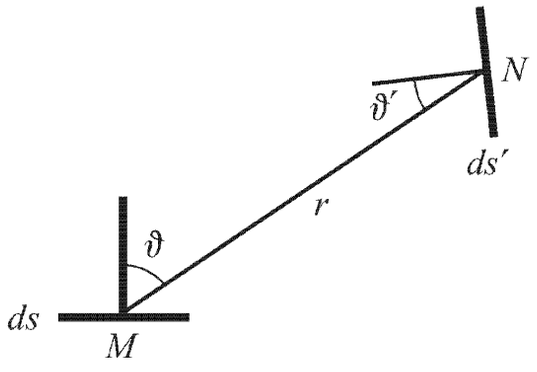


Рис. 86

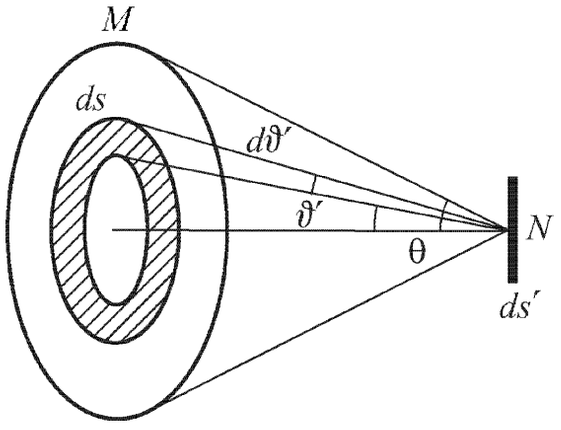


Рис. 87

ческой оси диска перпендикулярно к ней (рис. 87). Полагая в (22.10) $d\Omega = 2\pi \sin\vartheta' \, d\vartheta'$ и интегрируя, получим

$$E = \pi B \sin^2\theta, \qquad (22.10)$$

где θ — угол, составляемый крайним лучом MN с геометрической осью диска. Формула (22.11) справедлива и для шара, а также для источника произвольной формы, видимая граница которого имеет форму окружности.

11. В нашу задачу не входит изложение методов измерения и описание приборов, применяемых в фотометрии. Ограничимся только схематическим описанием одного из наиболее совершенных фотометров, построенных Луммером (1860–1925) и Бродхуном. Наиболее существенной частью этого фотометра является стеклянный кубик, состоящий из двух прямоугольных стеклянных призм A и B (рис. 88). Гипотенузная плоскость призмы A сошлифовывается под шаровую поверхность, так что от этой плоскости остается только резко очерченный участок в виде плоского круга. Затем призмы гипотенузными плоскостями плотно прикладываются друг к другу (на оптический контакт). Между сравниваемыми источниками света L_1 и L_2 ставится белый экран S, обе поверхности которого одинаково диффузно рассеивают свет. Рассеянный свет отражается двумя одинаковыми зеркалами S_1 и S_2 на кубик фотометра. На призму A попадает свет от источника L_1, на призму B — от источника L_2. Из призмы A свет может попасть в зрительную трубу толь-

ко через плоский круг на ее гипотенузной плоскости. Наоборот, свет из призмы B может попасть в ту же трубу только после полного отражения от гипотенузной поверхности этой призмы, внешней по отношению к тому же кругу. Труба сфокусирована для наблюдения плоскости указанного круга. Наблюдатель видит два поля зрения различной яркости, граничащие друг с другом вдоль окружности этого круга. Передвигая один из источников L_1 или L_2, добиваются исчезновения видимой границы между обоими полями, чтобы все поле зрения получилось равномерно освещенным. Это произойдет тогда, когда освещенности обеих сторон экрана S станут одинаковыми. Если r_1 и r_2 — расстояния (точечных) источников L_1 и L_2 до экрана в этом положении, то для отношения сил света $\mathcal{J}_1$ и $\mathcal{J}_2$ этих источников по закону обратных квадратов можно написать

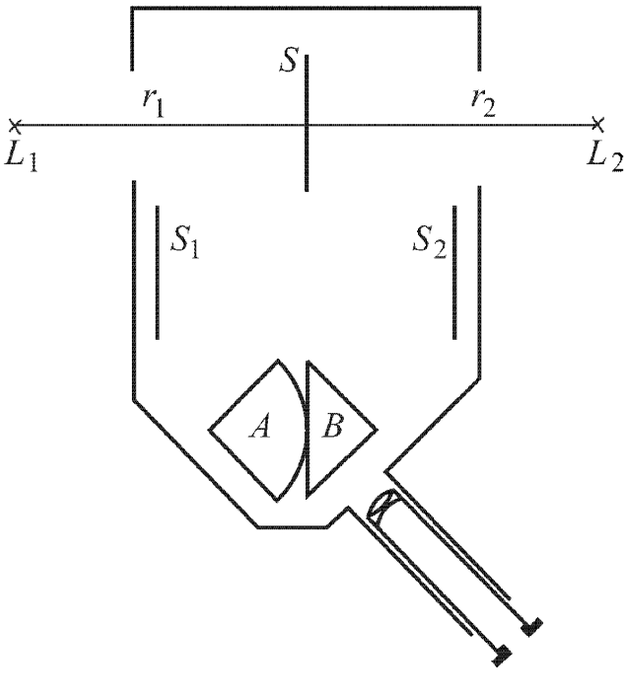


Рис. 88

$$\frac{\mathcal{J}_1}{\mathcal{J}_2} = \frac{r_1^2}{r_2^2}.$$

ЗАДАЧИ

1. Найти освещенность E площадки, создаваемую бесконечной плоскостью, излучающей свет по закону Ламберта с постоянной поверхностной яркостью B. Площадка и плоскость параллельны между собой.

Ответ. $E = \pi B$.

2. Вычислить освещенность E на горизонтальной площадке, освещаемой небесной полусферой, в предположении, что яркость неба B постоянна.

Ответ. $E = \pi B$.

3. Найти освещенность, создаваемую однородным светящимся шаром радиуса a на расстоянии R от его центра, если освещаемая площадка перпендикулярна к радиусу, а поверхность шара излучает по закону Ламберта с поверхностной яркостью B. Показать, что в этих условиях на любых расстояниях от центра шара строго выполняется закон обратных квадратов, т.е. освещенность площадки меняется обратно пропорционально квадрату R.

Ответ. $E = \pi B a^2/R^2$.

4. Найти освещенность, создаваемую равномерно светящимся диском радиуса a на его геометрической оси, перпендикулярной к плоскости диска, если освещаемая поверхность перпендикулярна к этой оси и находится

на расстоянии R от центра диска. Диск излучает по закону Ламберта с поверхностной яркостью B. С какой относительной точностью будет выполняться закон обратных квадратов, если диск рассматривать как точечный источник света, помещенный в его центре?

О т в е т. $E = \pi B a^2/(R^2 + a^2)$. Если $a/R \ll 1$, то

$$E \approx \frac{\pi B a^2}{R^2}\left(1 - \frac{1}{2}\frac{a^2}{R^2}\right).$$

Закон обратных квадратов выполняется с относительной точностью $a^2/(2R^2)$. При $a/R \sim 1/10$ эта точность будет около 1 %.

5. Предполагая, что излучение Солнца подчиняется закону Ламберта, определить его поверхностную яркость B при наблюдении с поверхности Земли (т.е. с учетом поглощения и рассеяния света земной атмосферой), если освещенность поверхности Земли, создаваемая солнечными лучами при перпендикулярном падении в тех же условиях, — $E \approx 10^5$ лк. Средний телесный угол, под которым солнечный диск виден с Земли, $\Omega = 6{,}8 \cdot 10^{-5}$ ср.

О т в е т. $B = E/\Omega \approx 1{,}5 \cdot 10^9$ кд/м^2.

§ 23. Яркость и освещенность оптического изображения. Нормальное увеличение

1. Во всех дальнейших расчетах, если нет специальной оговорки, не будем учитывать потери света при отражении на границах линз, призм и других отражающих поверхностях. Будем также пренебрегать поглощением и рассеянием света. Потери света на отражение довольно значительны. Так, для призматического бинокля они составляют около 50 % (см. § 67). С учетом всех этих потерь яркость и освещенность изображения получились бы несколько меньше вычисленных. Сначала рассмотрим случай, когда размеры предмета заметно превосходят предел разрешения оптической системы. Тогда система дает геометрически подобные изображения, и можно ограничиться точностью геометрической оптики.

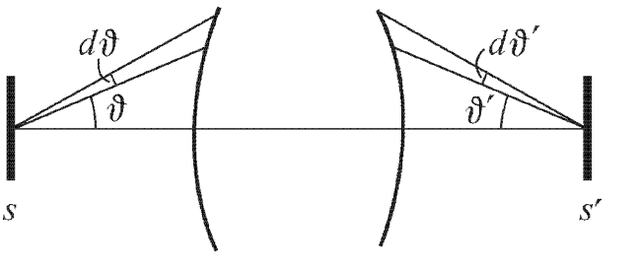


Рис. 89 *а*

Допустим, что предмет помещен на главной оптической оси центрированной системы и имеет форму малой площадки s, перпендикулярной к этой оси (рис. 89 *а*). Изображением площадки s будет какая-то другая площадка s'. Обозначим через B_ϑ яркость площадки s под углом ϑ к ее нормали, или (что то же) к главной оптической оси системы. Соответствующую яркость изображения обозначим через $B'_{\vartheta'}$. Соот-

ветствие означает, что угол ϑ, составляемый падающим лучом с оптической осью системы, после выхода из нее переходит в ϑ'. Будем предполагать, что в образовании изображения участвуют не только параксиальные, но и широкие пучки лучей, так что должно выполняться условие синусов (18.1). Из него следует

$$sn^2 \sin^2 \vartheta = s'n'^2 \sin^2 \vartheta'. \tag{23.1}$$

Если условие синусов не выполняется, то все результаты, получаемые ниже, также сохраняют силу, но только для параксиальных лучей. Возьмем теперь два конуса с общей вершиной на площадке s, образующие которых составляют с главной оптической осью углы ϑ и $\vartheta + d\vartheta$. (На рис. 89 *а* указаны только сечения этих конусов плоскостью рисунка.) Величина телесного угла, заключенного между этими конусами, будет $d\Omega = 2\pi \sin \vartheta \, d\vartheta$. Световой поток, посылаемый площадкой s внутрь телесного угла $d\Omega$, равен

$$d\Phi = B_\vartheta s \cos \vartheta \, d\Omega = 2\pi B_\vartheta s \sin \vartheta \cos \vartheta \, d\vartheta.$$

В пространстве изображений световой поток $d\Phi$ перейдет в поток

$$d\Phi' = 2\pi B'_{\vartheta'} s' \sin \vartheta' \cos \vartheta' \, d\vartheta',$$

пронизывающий площадку s'. При отсутствии потерь света $d\Phi = d\Phi'$, т.е.

$$B_\vartheta s \sin \vartheta \cos \vartheta \, d\vartheta = B'_{\vartheta'} s' \sin \vartheta' \cos \vartheta' \, d\vartheta'.$$

Учтем теперь соотношение (23.1), дифференцирование которого дает

$$sn^2 \sin \vartheta \cos \vartheta \, d\vartheta = s'n'^2 \sin \vartheta' \cos \vartheta' \, d\vartheta'.$$

В результате получим

$$\frac{B_\vartheta}{n^2} = \frac{B'_{\vartheta'}}{n'^2}, \tag{23.2}$$

чем и определяется яркость изображения. Если показатели преломления пространств предметов и изображений одинаковы ($n = n'$), то $B'_{\vartheta'} = B_\vartheta$, т.е. яркость изображения равна яркости предмета. Как видно из вывода, одинаковость яркостей обусловлена тем, что увеличение площади изображения сопровождается уменьшением в такое же число раз телесного угла, в котором распространяется световой поток. Для источников света, подчиняющихся закону Ламберта, B_ϑ не зависит от угла ϑ. В этом случае нет смысла сохранять индексы ϑ и ϑ', так что можно написать

$$B' = (n'/n)^2 B. \tag{23.3}$$

2. Иначе ведет себя освещенность изображения, проецирующегося на экран. В этом случае величина телесного угла, в котором распространяются лучи после выхода из оптического

прибора, не играет роли. Существенны только площади изображения и полный световой поток, концентрирующийся на этой площади. При неизменном световом потоке освещенность будет тем больше, чем меньше площадь изображения.

При количественном исследовании этого вопроса будем предполагать, что изображаемая площадка s имеет форму диска, плоскость которого перпендикулярна к главной оптической оси, и излучает свет по закону Ламберта. Если бы оптического прибора не было, то освещенность экрана, согласно формуле (22.11), была бы равна $E_0 = \pi B \sin^2\theta$, где 2θ — угол зрения, т.е. угол между лучами, исходящими из диаметрально противоположных точек площадки s и направленными к центру площадки s'. При наличии оптического прибора угол схождения крайних лучей с вершиной на площадке s' меняется и становится равным $2\theta'$. Это есть угол между крайними лучами, идущими к центру площадки s' от диаметрально противоположных точек выходного зрачка прибора. Полный световой поток через площадку s', как было показано выше, представляется интегралом

$$\Phi' = 2\pi B' s' \int \sin\vartheta' \cos\vartheta' \, d\vartheta' = \pi B' s' \sin^2\theta',$$

а освещенность площадки s' будет Φ'/s', т.е.

$$E = \pi B' \sin^2\theta' = \pi B \left(\frac{n'}{n}\right)^2 \sin^2\theta'. \tag{23.4}$$

Если $n = n'$, то

$$\frac{E}{E_0} = \left(\frac{\sin\theta'}{\sin\theta}\right)^2. \tag{23.5}$$

Эта формула показывает, что влияние оптического прибора на освещенность изображения сводится к изменению угла между крайними лучами, исходящими из противоположных точек предмета. Например, действие лупы как зажигательного стекла (в отсутствие потерь света) эквивалентно приближению источника (Солнца) на такое расстояние, чтобы он был виден невооруженным глазом под тем же углом зрения, как и входной зрачок (свободное отверстие) лупы из ее главного фокуса.

Если изображение получается в главной фокальной плоскости объектива (как в фотоаппарате), то в параксиальном приближении $\sin\theta' = D'/(2f')$, где D' — диаметр выходного зрачка, а f' — заднее фокусное расстояние объектива. Освещенность изображения будет пропорциональна квадрату этого отношения, а с ним и отношению D/f', где D — диаметр входного зрачка. Отношение D/f' называется *относительным отверстием*, а его квадрат $(D/f')^2$, определяющий освещенность изображения, — *светосилой объектива*.

В иммерсионных микроскопах (см. § 18, п. 4) свет от конденсора, помещенного перед объективом, попадает в пространство, заполненное жидкостью (иммерсией) с показателем преломления n, в которой помещается предмет. Пусть $2\theta_0$ — угол (апертура) между крайними направлениями лучей в конденсоре, которые попадают в микроскоп. Так как сам конденсор находится в воздухе, то по закону преломления $\sin\theta_0 = n\sin\theta$. Световой поток пропорционален $\sin^2\theta_0$, т.е. $(n\sin\theta)^2$. При одной и той же апертуре 2θ световой поток, попадающий в объектив, а с ним и освещенность изображения пропорциональны квадрату показателя преломления n. Величину $n\sin\theta$, следуя Аббе, называют *числовой апертурой*. Таким образом, применение иммерсии увеличивает освещенность изображения. Так как, однако, окончательное изображение получается в воздухе, то и при наличии иммерсии его яркость может быть самое большее равна яркости источника света.

3. Перейдем теперь к зрительной оценке поверхностной яркости предметов, т.е. оценке при наблюдении их глазом.

Отдельные элементы сетчатки — колбочки и группы палочек, соединенные с одним и тем же волокном зрительного нерва, — реагируют на световое раздражение независимо друг от друга. Увеличение освещаемой поверхности сетчатки при сохранении неизменной ее освещенности не усиливает интенсивности светового раздражения отдельного элемента, а увеличивает лишь число освещенных элементов. Если отвлечься от трудно поддающихся учету различных физиологических и психологических факторов, с которыми связаны, например, многочисленные обманы зрения, то следует сказать, что глаз судит о яркости предмета не по общему световому потоку, попадающему в него, а прежде всего по потоку, приходящемуся на единицу площади сетчатки, т.е. по *освещенности* последней. Поэтому, если исключить все субъективные факторы, то вопрос о зрительной оценке яркости сведется к вопросу об освещенности сетчатки.

Освещенность сетчатки невооруженного глаза называется *естественной освещенностью*. Она найдется по формуле (23.4), в которой следует положить $n = 1$, если, как это обычно и бывает, рассматриваемый предмет находится в воздухе. Таким образом,

$$E = \pi B(n'\sin\theta')^2. \tag{23.6}$$

Угол θ' есть угол, под которым из заднего фокуса глаза виден диаметр его выходного зрачка. В формулу (23.6) совсем не входит расстояние до предмета. *При неизменном зрачке глаза освещенность сетчатки E не зависит от этого расстояния, а определяется только поверхностной яркостью предмета.* Вот почему одинаковые фонари, находящиеся на разных расстояни-

ях от нас, кажутся нам одинаково яркими (если, конечно, пренебрежимо малы потери света при распространении его от фонаря до глаза).

Допустим теперь, что глаз вооружен, т.е. наблюдение ведется через зрительную трубу (телескоп) или микроскоп. В этом случае оптический прибор и глаз следует рассматривать как единую оптическую систему, к которой применимы формулы (23.5) и (23.6). Однако величина угла θ' может измениться. Смотреть в оптический прибор следует так, чтобы плоскость выходного зрачка прибора совмещалась с плоскостью входного зрачка глаза. Только тогда поле зрения будет резко очерчено. Угол θ' определяется тем из зрачков, который сильнее диафрагмирует световые лучи. Если входной зрачок глаза меньше выходного зрачка прибора, то угол θ' определяется зрачком глаза. В этом случае получается естественная освещенность сетчатки, т.е. сетчатка будет освещена так же, как и при отсутствии прибора (если, конечно, пренебречь дополнительными потерями света, вносимыми оптическим прибором). Если же входной зрачок глаза больше выходного зрачка прибора, то угол θ' определится зрачком прибора. В этом случае освещенность сетчатки будет меньше естественной. Будем непрерывно увеличивать выходной зрачок прибора, начиная с очень малых размеров. Сначала освещенность сетчатки будет возрастать пропорционально площади выходного (а, следовательно, и входного) зрачка (в параксиальном приближении). Когда выходной зрачок прибора сделается равным входному зрачку глаза, освещенность E достигнет максимума, а затем будет оставаться постоянной.

4. Изложенными соображениями руководствуются при выборе рационального увеличения зрительных труб, телескопов, биноклей и микроскопов. *Нормальным* или *равнозрачковым* увеличением таких приборов называют такое увеличение, когда выходной зрачок прибора равен входному зрачку глаза.

В телескопе, как и во всякой зрительной трубе, апертурной диафрагмой и входным зрачком служит свободное отверстие объектива. Объектив дает изображение в своей задней фокальной плоскости, которая одновременно является передней фокальной плоскостью окуляра. При этом условии изображение в телескопе получается на бесконечности, и его отчетливо будет видеть нормальный глаз в ненапряженном состоянии. Все параллельные пучки лучей после прохождения через телескоп остаются параллельными, т.е. телескоп является телескопической оптической системой. Меняется только ширина пучков. Если падающие лучи параллельны главной оптической оси, то ширина пучка равна диаметру объектива D. Увеличение телескопа N есть отношение угла, под которым виден малый предмет в

телескоп, к углу, под которым он был бы виден, если бы рассматривался невооруженным глазом. Как было показано в § 11 (п. 10), для телескопических систем такое увеличение равно отношению ширины падающего пучка параллельных лучей к ширине выходящего пучка. В телескопе ширина выходящего пучка равна диаметру выходного зрачка D'. При нормальном увеличении $D' = d$, где d — диаметр входного зрачка глаза. Таким образом, нормальное увеличение телескопа определяется выражением

$$N_{\text{норм}} = D/d. \tag{23.7}$$

Диаметр d зрачка глаза не остается постоянным, а зависит от яркости. Увеличение, равнозрачковое в дневных условиях, оказывается слишком большим при наблюдении в сумерки и ночью. В последних случаях требуются трубы с большим выходным зрачком, т.е. либо с малым увеличением, либо, что гораздо лучше, с большим объективом.

Освещенность изображения не зависит от увеличения, если оно меньше нормального. Увеличения меньше нормального нецелесообразны, так как при этом используется только центральная часть объектива; периферийные лучи, достаточно удаленные от оптической оси, задерживаются радужной оболочкой и не попадают в зрачок глаза. Нормальное увеличение есть максимальное увеличение, при котором освещенность изображения еще максимальна. При таком увеличении объектив используется полностью. Поэтому окуляры к зрительным трубам подбирают так, чтобы они давали нормальное увеличение.

В некоторых случаях имеет смысл применять увеличения, превосходящие нормальное в 2–4 раза, несмотря на то, что это связано с значительным уменьшением освещенности изображения (соответственно в 4–16 раз). Дело в том, что разрешаемый предел глаза 30–$60''$ соответствует случаю, когда изображение получается на центральной ямке желтого пятна. На периферии разрешаемый предел больше. Он увеличивается также при длительной работе и усталости глаза. В этих условиях выгоднее пойти на некоторое уменьшение освещенности, чтобы получить большие увеличения и возможность различать более мелкие детали предметов. Однако далеко идти в этом направлении нецелесообразно из-за уменьшения освещенности и поля зрения.

При N-кратном увеличении угловые размеры поля зрения на местности уменьшаются в N, а его площадь в N^2 раз. Поэтому в биноклях предпочитают увеличение меньше нормального. Этим достигается большее поле зрения и облегчается нахождение нужного объекта на местности. Бинокли рассчитывают для сравнительно широких выходных зрачков (в пределах 3–5 мм). Так, полевые бинокли с диаметром объектива 3 см имеют 6- или

8-кратные увеличения (диаметр выходного зрачка соответственно 5 и 3,75 мм).

5. Зрительные трубы, как обнаружил еще М.В. Ломоносов (1711–1765), улучшают видимость изображений предметов при слабых освещенностях, например ночью. Конечно, освещенность изображений при этом не только не увеличивается, но даже уменьшается из-за неизбежных потерь света на отражение от поверхностей линз и поглощение в стекле. Однако при слабых освещенностях сильно уменьшается разрешающая способность глаза. Ночью, когда освещенность падает до десятитысячных долей люкса, угол, разрешаемый глазом, возрастает с $\sim 1'$ до $\sim 1°$. Зрительные трубы увеличивают угол зрения, а потому их применение способствует различению контуров и крупных деталей предметов, даже если они неразличимы невооруженным глазом.

В трубах, предназначенных для ночных наблюдений, число линз должно быть минимальным, а сами линзы должны быть простыми, чтобы максимально уменьшить потери света на отражение. Увеличение трубы должно быть как можно ближе к нормальному, чтобы использовать весь световой поток, падающий на объектив. Конечно, нормальное увеличение определяется размерами зрачка глаза в ночных условиях (6–8 мм).

6. Найдем теперь нормальное увеличение микроскопа. Увеличением микроскопа N называется отношение угла φ', под которым виден малый объект в микроскоп, к углу φ, под которым он был бы виден невооруженным глазом, если бы был помещен от него на расстоянии ясного зрения L (см. § 21, п. 3). Если y — линейный размер рассматриваемого объекта, то $\varphi = y/L$. Поместим этот объект $y = AB$ в передней апланатической точке

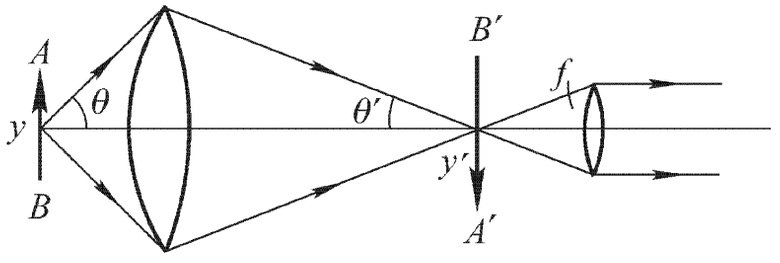


Рис. 89 *б*

объектива (рис. 89 *б*); для упрощения рисунка сложные объектив и окуляр изображены тонкими линзами, это не отражается на общности выводов и результатов). Его изображение $A'B' = y'$ получится в задней апланатической точке. Так как угол θ' всегда мал, то его синус можно заменить самим углом, записав условие синусов в виде $ny\sin\theta = y'\theta'$. Так же, как и в телескопе, изображение y' должно получаться в передней фокальной плос-

кости окуляра. Если f — фокусное расстояние окуляра, то $2f\theta'$ будет шириной выходящего пучка лучей. При нормальном увеличении эта ширина должна быть равна диаметру d входного зрачка глаза, т.е.

$$d = 2f\theta' = 2\frac{fny\sin\theta}{y'}.$$

Но $y = L\varphi$, $y' = f\varphi'$, так что $y/y' = L\varphi/(f\varphi') = L/(Nf)$. Подставляя это в предыдущее соотношение, получим [1])

$$N_{\text{норм}} = 2Ln\frac{\sin\theta}{d}. \tag{23.8}$$

Нормальное увеличение для микроскопа еще в большей степени есть величина субъективная, чем для телескопа, так как оно зависит от двух параметров глаза: диаметра зрачка d и расстояния ясного зрения L. Характеризуя увеличение микроскопа, обычно полагают $d = 2$ мм, $L = 250$ мм. Для сухих систем ($n = 1$) верхний предел числовой апертуры $n\sin\theta$ составляет 1, для иммерсионных 1,5. Соответствующие нормальные увеличения будут $N = 250$ и $N = 375$. По тем же соображениям, что и для телескопа, от этих значений можно несколько отступать в бóльшую сторону, но не более чем в 2–4 раза. Таким образом, в лучших безиммерсионных микроскопах верхний предел увеличения можно принять равным 1000, а в иммерсионных — 1500. Бóльшие увеличения не только бесполезны, но и вредны. К этому вопросу, а также к вопросу об увеличении телескопа мы вернемся в § 56 и разберем его с новой точки зрения — в связи с разрешающей способностью этих приборов, определяемой волновой природой света.

7. В заключение рассмотрим случай, когда размеры объекта меньше предела разрешения оптической системы. Такой случай типичен при наблюдении звезд невооруженным глазом и в телескоп. В этом случае оптическое изображение не передает формы предмета, а представляет собой дифракционный кружок, окруженный темными и светлыми кольцами (см. рис. 180, с. 316), причем почти весь свет сконцентрирован в центральном кружке. Назначение телескопа при наблюдении звезд состоит не в том, чтобы различать детали рассматриваемого объекта (они находятся за пределами его разрешающей способности), а в том, чтобы *увеличить световой поток*, попадающий в глаз наблю-

[1]) Очень часто при выводе формулы (23.8) вводят ненужное предположение, что изображение в микроскопе получается на расстоянии ясного зрения L от глаза. На самом деле это может быть и не так. Справедливость формулы (23.8) никак не связана с тем, на каком расстоянии от глаза получается изображение в микроскопе. Величина L входит в формулу (23.8) совсем по другой причине, как это видно из нашего вывода.

дателя, и тем самым обнаруживать все более и более слабые звезды.

Применять увеличение меньше нормального не имеет смысла, так как при этом будет использоваться не весь объектив. Поэтому в дальнейшем достаточно ограничиться случаем, когда диаметр выходного зрачка телескопа D' меньше или равен диаметру входного зрачка глаза d ($D' \leqslant d$). Тогда размер дифракционного кружка будет определяться величиной D'. Как будет показано в теории дифракции (см. § 55), диаметр дифракционного кружка будет обратно пропорционален D', а его площадь обратно пропорциональна D'^2. Световой поток, попадающий в глаз, $\Phi \sim D^2$, где D — диаметр объектива, а освещенность дифракционного кружка на сетчатке

$$E \sim \frac{D^2}{1/D'^2} = D^2 D'^2.$$

Вводя увеличение телескопа $N = D/D'$, получим

$$E \sim D^4/N^2.$$

Видимость звезды определяется не только освещенностью дифракционного кружка на сетчатке, но и освещенностью окружающего фона. Освещенность фона $E_{\text{фон}}$ определяется яркостью B неба. Ее можно найти по формуле (23.6), в которой под θ' теперь следует понимать угол, под которым из заднего фокуса глаза виден диаметр D' выходного зрачка телескопа. Таким образом, $E_{\text{фон}} \sim BD'^2 = BD^2/N^2$, и следовательно,

$$E/E_{\text{фон}} \sim BD^2. \tag{23.9}$$

Эта формула объясняет, почему в телескоп звезды видны и днем.

ЗАДАЧА

С помощью собирательной линзы можно зажечь папиросу, фокусируя на ней прямой солнечный свет. Можно ли, взяв линзу достаточных размеров (например, объектив большого астрономического телескопа), сделать то же самое, используя свет полной Луны?

Р е ш е н и е. Для простоты будем считать, что излучение Солнца подчиняется закону Ламберта. Прямые солнечные лучи, падающие на Землю перпендикулярно, создают на ней освещенность $E_{\text{C}} = \pi B \sin^2 \vartheta_{\text{C}}$, где B — поверхностная яркость Солнца, а ϑ_{C} — угловой размер его. Из повседневного опыта известно, что такой освещенности для воспламенения папиросы недостаточно. Средняя освещенность поверхности Земли будет в π раз меньше, т.е. $\overline{E}_{\text{C}} = B \sin^2 \vartheta_{\text{C}}$. Если пренебречь поглощением и рассеянием света, то действие линзы сводится просто к увеличению видимых угловых размеров Солнца. Однако максимальная освещенность не может быть больше πB.

Так как диаметр лунной орбиты пренебрежимо мал по сравнению с расстоянием до Солнца, то средняя освещенность поверхности Луны будет такой же, как и у Земли, т.е. $\overline{E}_{\text{С}} = B\sin^2\vartheta_{\text{С}}$. Если Луна как источник света подчиняется закону Ламберта, то ее поверхностная яркость будет $B\sin^2\vartheta_{\text{С}}$. Лучи от Луны, падая на поверхность Земли перпендикулярно, создают освещенность $B\sin^2\vartheta_{\text{С}}\sin^2\vartheta_{\text{Л}}$, где $\vartheta_{\text{Л}}$ — угловой радиус Луны. Эта освещенность много меньше, чем $B\sin^2\vartheta_{\text{С}}$. Поэтому, каковы бы ни были размеры линзы, с помощью лунного света воспламенить папиросу нельзя.

§ 24. Оптические инструменты

1. Л у п а. Лупа, как и микроскоп, дает увеличенные изображения предметов, находящихся на небольших расстояниях от глаза. *Увеличение N лупы (и микроскопа) определяется как отношение угла, под которым виден (малый) предмет через лупу, к углу, под которым он был бы виден невооруженным глазом, если бы был помещен от глаза на расстоянии ясного зрения L* (см. § 23, п. 6). Можно также сказать, что увеличение лупы (и микроскопа) есть отношение линейных размеров изображения предмета на сетчатке при рассматривании его в лупу к линейным размерам изображения того же предмета на сетчатке, когда он рассматривается невооруженным глазом с расстояния ясного зрения. При увеличениях, не превосходящих 5, в качестве лупы обычно применяется простая собирательная линза, при больших увеличениях — система, состоящая из таких линз.

При вычислении увеличения лупы будем для общности рассматривать ее как сложную центрированную систему и воспользуемся уравнениями такой системы в форме Ньютона (11.16) и (11.17). Предмет помещается между передним главным фокусом и передней главной плоскостью лупы. Лупа дает прямое мнимое изображение его. Пусть X, Y — координаты точки предмета относительно переднего главного фокуса, X', Y' — координаты соответствующей точки изображения относительно заднего главного фокуса, a — координата глаза (точнее — его передней узловой точки) относительно того же заднего главного фокуса. Расстояние глаза от изображения будет $a - X'$. Глаз через лупу видит отрезок Y' под углом $\varphi' = Y'/(a - X')$. Соответствующий отрезок предмета Y был бы виден невооруженным глазом с расстояния ясного зрения под углом $\varphi = Y/L$. Отсюда

$$N = \frac{\varphi'}{\varphi} = \frac{Y'}{Y}\,\frac{L}{a - X'},$$

или на основании (11.17)

$$N = \frac{f}{X}\,\frac{L}{a - X'} = \frac{X'}{f'}\,\frac{L}{a - X'}.$$

Для лупы $n = n'$, а потому $f' = -f$. Следовательно, с учетом (11.16)

$$N = \frac{L}{f + aX/f}. \tag{24.1}$$

Если глаз помещен в заднем главном фокусе лупы ($a = 0$), то $N = L/f$. То же самое увеличение получится, если предмет поместить в переднем главном фокусе лупы. В этом случае глаз должен быть аккомодирован на бесконечность, т.е. для нормального глаза аккомодирующая мышца должна находиться в расслабленном состоянии. Однако многие наблюдатели ведут наблюдение с напряжением аккомодирующей мышцы, чтобы изображение получалось на привычном для них расстоянии ясного зрения. В этом случае $a - X' = L$, так что

$$N = \frac{X'}{f'} = \frac{a - L}{f'} = \frac{L - a}{f}.$$

Если лупой является тонкая линза, вплотную придвинутая к глазу, то $a = -f$, так что $N = (L/f) + 1$. Рассматриваемый предмет всегда помещается вблизи переднего главного фокуса, т.е. $a/f \ll 1$. Величина aX/f также меньше единицы. Поэтому, независимо от способа наблюдения, увеличение лупы практически всегда определяется формулой

$$N = L/f. \tag{24.2}$$

Подставляя сюда $N = \varphi'/\varphi = (\varphi'/Y)L$, получим $\varphi' = Y/f$. Размер изображения предмета на сетчатке глаза равен $f_{\text{гл}}\varphi' = (f_{\text{гл}}/f)Y$, где $f_{\text{гл}}$ — переднее фокусное расстояние глаза (см. § 21, п. 5). Этот размер определяется практически только фокусным расстоянием лупы f, так как от глаза к глазу величина $f_{\text{гл}}$ меняется незначительно. Таким образом, действие лупы эквивалентно уменьшению расстояния ясного зрения с величины L до f; лупа как бы приближает рассматриваемый предмет к глазу на фокусное расстояние f.

Теоретически увеличение лупы ничем не ограничено. Так, лупа с фокусным расстоянием $f = 0{,}25$ мм имела бы увеличение $N = 1000$. Изготовление луп со столь малыми значениями f, а следовательно и с малыми диаметрами линз, очень затруднительно, а пользование ими крайне неудобно. Кроме того, для таких луп очень велики геометрические и хроматические аберрации. Поэтому увеличение луп, применяемых в действительности, в настоящее время не превосходит 40. Бóльшие увеличения получаются с помощью микроскопа.

Изображение в лупе формируется с помощью сравнительно тонких пучков лучей, выделяемых зрачком глаза. Поэтому требования к лупам в отношении исправления аберраций не очень

жесткие. Хроматическая аберрация положения при малых увеличениях не исправляется вовсе. При более значительных увеличениях и необходимости иметь большое поле зрения надо исправить кому и астигматизм, а также хроматическую разность увеличений. Астигматизм исправляется для главных лучей, проходящих через центр вращения глаза.

Простая плоско-выпуклая линза дает хорошие изображения вплоть до восьмикратного увеличения (т.е. когда ее фокусное расстояние f не менее 3 см). Линза должна быть обращена к глазу плоской стороной. В этом случае, так как предмет помещен вблизи фокуса лупы, преломление лучей происходит практически только на выпуклой поверхности лупы, а не распределяется между обеими поверхностями ее. Это несколько увеличивает сферическую аберрацию. Зато в противоположном случае, когда к глазу обращена выпуклая сторона лупы, значительно увеличиваются аберрации изображений периферийных частей предмета.

2. М и к р о с к о п. Микроскоп состоит из двух собирающих систем линз — *объектива* и *окуляра*, расположенных на большом расстоянии друг от друга. Они помещаются на концах трубки (*тубуса*), которая обычно укрепляется вертикально на штативе микроскопа, объектив ввинчивается в нее снизу, а окуляр вставляется сверху. Предмет помещается на малом расстоянии перед передним фокусом объектива. Объектив дает его действительное обратное увеличенное изображение, которое рассматривается через окуляр, действующий как лупа. Пусть f_1 и f_1' — главные фокусные расстояния объектива, f_2 и f_2' — окуляра. Тогда фокусные расстояния f и f' всей оптической системы микроскопа определятся формулами (12.3), т.е.

$$f = -\frac{f_1 f_2}{\Delta}, \quad f' = \frac{f_1' f_2'}{\Delta}, \tag{24.3}$$

где Δ — оптический интервал между объективом и окуляром. В применении к микроскопу величину Δ называют *оптической длиной тубуса*. Ее следует отличать от *механической длины тубуса* T. Последняя обычно составляет 160 или 190 мм; величина Δ зависит от объектива и окуляра и, как правило, больше T.

Для безиммерсионного микроскопа его увеличение, согласно формуле (24.2), будет

$$N = \frac{L}{f} = -\frac{L}{f_1}\frac{\Delta}{f_2} \tag{24.4}$$

(при наличии иммерсии с показателем преломления n это выражение надо умножить на n). Величину N можно представить в виде $N = N_1 N_2$, где $N_1 = L/f_1$ — увеличение объектива как лупы, $N_2 = \Delta/f_2$ — добавочное увеличение, обусловленное оку-

ляром. При заданном общем увеличении N для устранения ошибок в изображении осевых точек выгоднее делать бóльшим N_1, а для уменьшения ошибок всего поля зрения выгоднее применять бóльшие N_2. По этой причине, а также по соображениям, указанным в п. 1, требуемое общее увеличение N не получают с помощью одних только сильных объективов или одних только сильных окуляров, а идут на компромисс, комбинируя и те и другие.

В микроскопе величина Δ, а также фокусные расстояния f_1 и f_2 положительны. Фокусное расстояние f всей системы, а с ним и увеличение N отрицательны. Это значит, что изображение, получаемое в микроскопе, обратное. Говоря в дальнейшем об увеличении микроскопа, мы обычно будем иметь в виду *абсолютное значение* N.

Величина Δ обычно лежит в пределах 150–200 мм. Фокусное расстояние f_1 объектива при самых больших увеличениях его (100–120) не меньше 1,5 мм, а окуляра — не меньше 8–10 мм. При достаточно больших значениях Δ можно получить увеличение до 2500–3000 раз без чрезмерного уменьшения f_1 и f_2. (Впрочем, столь бóльшие увеличения бесполезны, как это будет подробно разобрано в § 56.)

Штатив микроскопа обычно имеет револьверное приспособление для смены объективов. Такое же приспособление может применяться для смены окуляров. Это позволяет варьировать увеличение микроскопа в широких пределах.

Со стороны предмета в микроскопе велики углы раскрытия, а со стороны изображения — углы наклона лучей. Вследствие этого у объектива особенно велики сферическая аберрация и кома, а у окуляра — астигматизм. Поэтому к объективу и окуляру предъявляются различные требования.

3. Основные требования, предъявляемые к объективу микроскопа: 1) высокая числовая апертура; 2) исправление сферической аберрации для всего очень широкого пучка лучей; 3) выполнение условия синусов для того же пучка; 4) ахроматизация. Ахроматизация состоит в выполнении условия синусов по меньшей мере для двух цветов, а также в получении через объектив ахроматических действительных изображений предмета.

Всем этим требованиям в той или иной мере удается удовлетворить применением довольно сложных конструкций объективов, особенно при больших увеличениях. Наиболее совершенными объективами до настоящего времени являются *апохроматы*, разработанные и рассчитанные Аббе. Впервые апохромат Аббе был выпущен фирмой Цейсса в 1886 г. Это — иммерсионный объектив, состоящий из 10 линз. Он изображен на рис. 90 в увеличенном виде (примерно вдвое). Роль нижней (полусферической) линзы была выяснена в § 18 (п. 4). Поверхность

этой линзы погружена в масло, в котором помещен и предмет. Эта линза обладает весьма сильной хроматической аберрацией, уничтожаемой остальными линзами. Из этих остальных линз рассеивающие изготовлены из различных сортов флинта, собирающие — из крона. Аббе нашел, что, перемещая слегка некоторые из этих линз относительно других, можно достигнуть почти полного уничтожения хроматической аберрации. Апохромат ахроматизован для трех и удовлетворяет условию синусов для двух цветов спектра. Фокусное расстояние апохромата Аббе $f = 2$ мм, числовая апертура $n \sin\theta = 1{,}4$, а геометро-оптические качества таковы, что можно считать фактически достигнутыми границы разрешающей способности микроскопа, определяемые волновой природой света (см. § 56). Апохроматы применяются во всех наиболее ответственных микроскопических и микрофотографических исследованиях.

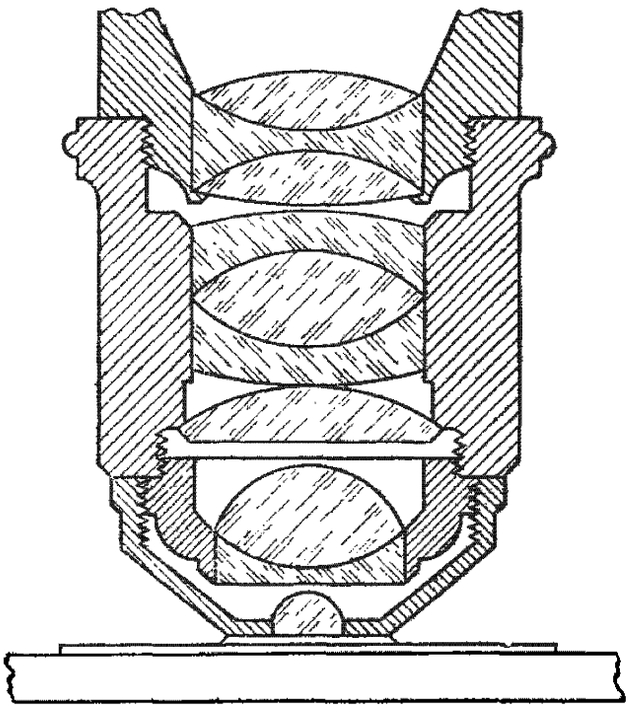


Рис. 90

Менее совершенными, но более распространенными являются *ахроматы*, т.е. объективы, ахроматизованные для двух цветов спектра. Если удовлетвориться частичной ахроматизацией объектива, то надо совместить положения передних фокальных точек для двух цветов (*ахроматизация положения*). Иначе положение изображения предмета, помещенного вблизи передней фокальной точки объектива, очень сильно менялось бы с изменением цвета. Фокусные расстояния при этом остаются неахроматизованными, т.е. неравными. Поэтому хроматическая разность увеличений остается. Она исправляется применением особых окуляров, называемых *компенсационными.* Такие окуляры всегда применяются и в апохроматах, так как у них разность увеличений для лучей различного цвета исправлена не вполне.

Вообще, ахроматизация положения, т.е. совмещение фокусов, или точек схождения лучей различного цвета, должна производиться во всех оптических системах, предназначенных для получения действительных изображений объектов с малыми угловыми размерами. Например, объективы зрительных труб

должны быть ахроматизованы в смысле совмещения задних фокальных точек для лучей разного цвета.

Свет от объекта попадает в объектив, проходя через *покровное стекло*. В сухих объективах толщина покровного стекла влияет на сферическую аберрацию. Поэтому все расчеты объективов производят в предположении, что толщина покровного стекла равна 0,17 мм (0,15–0,20 мм). Все сильные сухие объективы в настоящее время снабжены *коррекционной оправой*, с помощью которой можно немного менять расстояние между верхней и нижней линзами объектива и тем самым уничтожать сферическую аберрацию, вносимую покровным стеклом несоответствующей толщины. В объективах с однородной иммерсией, когда показатели преломления покровного стекла, иммерсионной жидкости и фронтальной линзы объектива одинаковы, толщина покровного стекла не имеет значения, так как ее можно компенсировать слоем иммерсионной жидкости между покровным стеклом и объективом.

4. Основные требования, предъявляемые к окуляру: 1) устранение астигматизма в косых пучках; 2) устранение дисторсии; 3) ахроматизация.

Кривизна поверхности изображения обычно невелика и существенно не мешает визуальным наблюдениям. Если объектив дает изображение, свободное от хроматической аберрации, то достаточна уже частичная ахроматизация окуляра. Однако теперь существенна не ахроматизация положения, а *ахроматизация увеличения*, т.е. равенство фокусных расстояний для лучей различного цвета. Действительно, благодаря значительной длине тубуса микроскопа лучи попадают на окуляр под малыми углами наклона к главной оптической оси микроскопа. В этом случае угловое увеличение, даваемое окуляром, определяется формулой (24.2), независимо от того, как аккомодирован глаз: на бесконечность или на расстояние ясного зрения. Если фокусные расстояния окуляра ахроматизованы, то угловое увеличение будет одинаковым для всех цветов. Поэтому изображения предмета на сетчатке глаза во всех цветах совместятся между собой, т.е. окончательное изображение получится неокрашенным, даже если положения главных фокусов и главных плоскостей окуляра не ахроматизованы. Если же изображение, даваемое объективом, не свободно от хроматической аберрации, то ее можно компенсировать хроматической аберрацией противоположного знака, которую дает окуляр. Так, апохроматы дают хроматическую разность увеличений: синее изображение увеличивается сильнее красного. Аббе комбинирует их с такими компенсированными окулярами, у которых фокусное расстояние для синих лучей больше, а следовательно, увеличение меньше, чем для красных лучей.

В § 16 (п. 5) было показано, что фокусное расстояние системы из двух тонких линз, изготовленных из стекла одного и того же сорта, будет одним и тем же для всех цветов спектра, если расстояние l между линзами равно полусумме ее фокусных расстояний: $l = {}^1\!/_2(f_1 + f_2)$. Составлением окуляров из двух линз одновременно можно достигнуть устранения астигматизма косых пучков, а потому окуляры и конструируются по этому принципу. Линза, обращенная к объективу, называется *коллективом*, или *собирательной линзой*, а обращенная к глазу — *глазной линзой*. Коллектив выполняет ту же роль, что и конденсор проекционного аппарата; он собирает все лучи, участвующие в образовании изображения, направляя их в зрачок глаза наблюдателя. При отсутствии специальной диафрагмы поля зрения роль этой диафрагмы выполняет оправа коллектива. Вот почему коллектив часто называют также *полевой линзой*. Увеличение окуляра определяется главным образом глазной линзой. Простейшими окулярами описанного типа, применяемыми в микроскопах и телескопах, являются *окуляр Гюйгенса* и *окуляр Рамсдена* (1735–1800).

5. Окуляр Гюйгенса состоит из двух плоско-выпуклых или двояковыпуклых линз, обращенных более выпуклыми поверхностями в сторону падающих лучей (рис. 91). Для уменьшения сферической аберрации Гюйгенс выбрал фокусные расстояния и взаимное расположение линз так, чтобы отклонение луча, первоначально направленного параллельно главной оптической оси, при преломлении распределялось поровну между обеими линзами. Нетрудно подсчитать (что рекомендуется сделать читателю), что в параксиальном приближении это будет достигнуто при условии $l = f_1 - f_2$, где f_1 — фокусное расстояние собирательной, а f_2 — глазной линзы. Комбинируя это уравнение с условием $l = {}^1\!/_2(f_1 + f_2)$, находим: $f_1 = 3f_2$, $l = 2f_2$. Таким образом, в окуляре самого Гюйгенса $f_2 : l : f_1 = 1 : 2 : 3$ (в современных конструкциях окуляра Гюйгенса чаще пользуются отношением $2 : 3 : 4$). Оптический интервал между линзами окуляра $\Delta = -2f_2$. По формулам (12.2), (12.3) и (12.4) нетрудно рассчитать

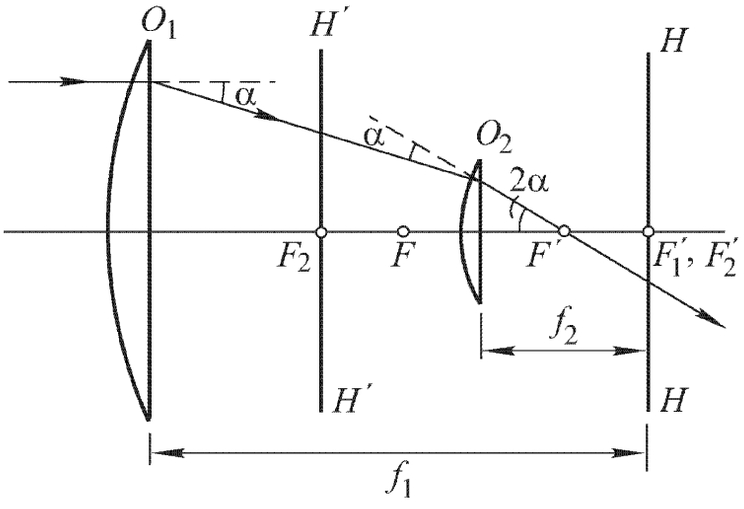


Рис. 91

положение главных фокусов, главных плоскостей и фокусные расстояния окуляра Гюйгенса. Результаты расчета приведены на рис. 91 (положение переднего фокуса F_1 коллектива не показано).

На рис. 92 приведен ход лучей в окуляре Гюйгенса. Предположим, что глаз аккомодирован на бесконечность, а потому окончательное изображение, даваемое микроскопом, должно получаться в бесконечности. Значит, объектив должен давать перевернутое изображение AB предмета в передней фокальной плоскости окуляра. Этого можно добиться, перемещая весь ми-

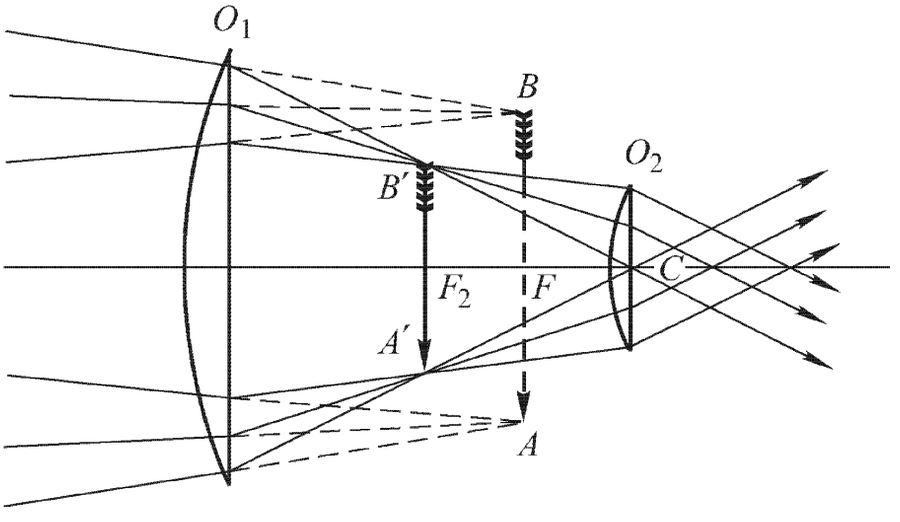


Рис. 92

кроскоп вверх или вниз под предметным столиком. Изображение AB для коллектива O_1 играет роль *мнимого предмета*. Коллектив даст его действительное уменьшенное изображение $A'B'$ в передней фокальной плоскости глазной линзы O_2, так как окончательное изображение должно получиться в бесконечности (в этом можно убедиться и непосредственным расчетом). Изображение $A'B'$ увеличивается глазной линзой O_2, действующей как лупа. Угол $A'CB'$ определяет поле зрения. Его можно изменить, поместив в фокальной плоскости F_2 глазной линзы диафрагму поля зрения.

Окуляр Гюйгенса мало пригоден для точных измерений. Действительно, в этом окуляре шкалу микрометра следовало бы располагать между линзами окуляра. Но тогда шкала микрометра сравнивалась бы с изображением предмета, даваемым объективом микроскопа, которое искажено первой линзой окуляра. Аберрации этой линзы, и прежде всего ее дисторсия, вносили бы в измерения значительные ошибки, не поддающиеся точному учету. Окуляр Гюйгенса не может быть использован и в качестве обыкновенной лупы, предназначенной для непосредственного рассматривания мелких предметов, так как рассматрива-

емый предмет надо было бы опять помещать между линзами окуляра. Окуляр Гюйгенса применяется в качестве компенсационного окуляра в комбинации с объективами-ахроматами. В этом случае, однако, глазную линзу изготовляют из двух склеенных сферических стекол различного сорта.

6. Окуляр Рамсдена состоит из двух плоско-выпуклых линз с одинаковыми фокусными расстояниями ($f_1 = f_2$), обращенных выпуклыми поверхностями одна к другой (рис. 93). В первоначальном варианте окуляра расстояние между линзами $l = f_1 = f_2$, что обеспечивает ахроматизацию фокусного расстояния f всего окуляра. Но в таком варианте собирательная линза находится в передней фокальной плоскости глазной линзы. Поэтому глазная линза отчетливо изображает каждую пылинку, грязь и

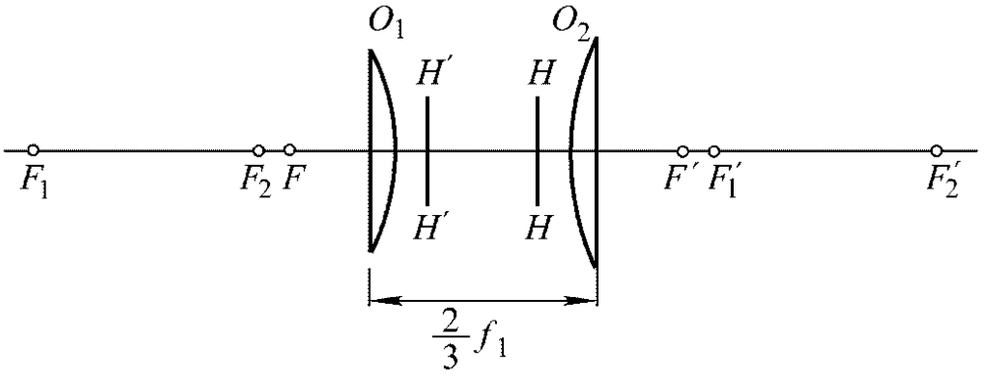


Рис. 93

царапину на поверхности собирательной линзы, что, конечно, портит изображение предмета. Чтобы это устранить, расстояние между линзами уменьшают, выбирая обычно $l = {}^2/_3 f_1$. Это, конечно, нарушает ахроматизацию окуляра. Однако возникающие хроматические явления мало заметны. Их можно устранить, применяя ахроматическую глазную линзу, склеенную из собирательной (крон) и рассеивающей (флинт) линз. Это делается в *окулярах типа Кельнера.* Рассеивающая линза должна быть обращена к глазу. Желательно также ахроматизовать и коллектив окуляра, но этого почти никогда не делают.

Положение фокальных точек и главных плоскостей в окуляре Рамсдена показано на рис. 93, а на рис. 94 приведен ход лучей в этом окуляре. Сравнивая их с рисунками 91 и 92, видим невыгоду окуляра Рамсдена: он удлиняет всю систему микроскопа, так как передняя фокальная плоскость окуляра, в которой объектив должен давать действительное изображение AB, лежит перед передней линзой окуляра на расстоянии ${}^1/_4 f_1$ (а не между линзами окуляра, как в случае окуляра Гюйгенса).

Но в этой особенности окуляра Рамсдена есть и его преимущество. В фокальной плоскости F можно поместить микрометр

(шкалу на стеклянной пластинке или нити, перемещаемые микрометрическим винтом) и таким путем измерить размеры действительного изображения, даваемого объективом микроскопа. При этом микрометр и действительное изображение, даваемое объективом, оба находятся перед окуляром, и аберрации полевой линзы окуляра искажают их одинаково. Тем самым в окуляре Рамсдена устраняется основной недостаток окуляра Гюйгенса как измерительного прибора. Зная линейное увеличение объек-

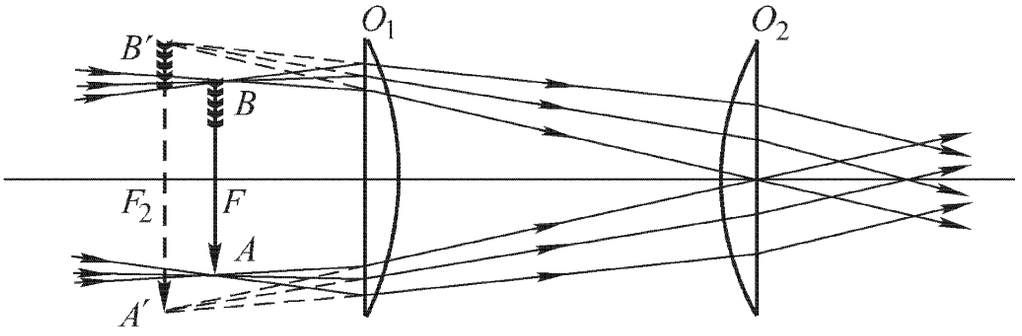


Рис. 94

тива микроскопа, легко вычислить и размеры самого предмета. Поэтому в микроскопах, служащих для измерения длин и углов, применяются окуляры Рамсдена с окулярными микрометрами.

В зависимости от назначения применяются и более сложные окуляры, состоящие не из двух, а из большего количества линз.

7. На рис. 95 изображен ход лучей в микроскопе с окуляром Гюйгенса. Главные лучи, исходящие из крайних точек предмета AB, начерчены толстыми, а крайние лучи, исходящие из тех же точек, — тонкими линиями. Если увеличение меньше нормального, то зрачок глаза наблюдателя служит апертурной диафрагмой, а его изображение, даваемое всем микроскопом, — входным зрачком этого прибора. Если же увеличение больше нормального, то апертурной диафрагмой и входным зрачком является диафрагма или края полушаровой линзы CD. Ее изображение $C'D'$ будет выходным зрачком микроскопа. Диафрагма поля зрения должна быть помещена в том месте, где получается изображение $A'B'$, даваемое объективом и собирательной линзой окуляра. Изображение $A''B''$, которое видит наблюдатель, можно получить на любом расстоянии, что достигается небольшим поднятием или опусканием всего микроскопа. Обычно его получают на расстоянии ясного зрения (25 см для нормального глаза) или в бесконечности. Перемещая окуляр, можно менять оптическую длину тубуса, а с ней и увеличение микроскопа.

Из рисунка видно главное преимущество микроскопа перед лупой. Оно состоит в том, что в микроскопе задача получения изображения распадается на две части, из которых одна реша-

ется объективом, другая окуляром. Объектив дает изображение предмета при максимально возможной числовой апертуре. Окуляр, как и лупа, при большом поле зрения дает изображение точек предмета с помощью узких пучков лучей, так как эти пучки ограничены выходным зрачком микроскопа.

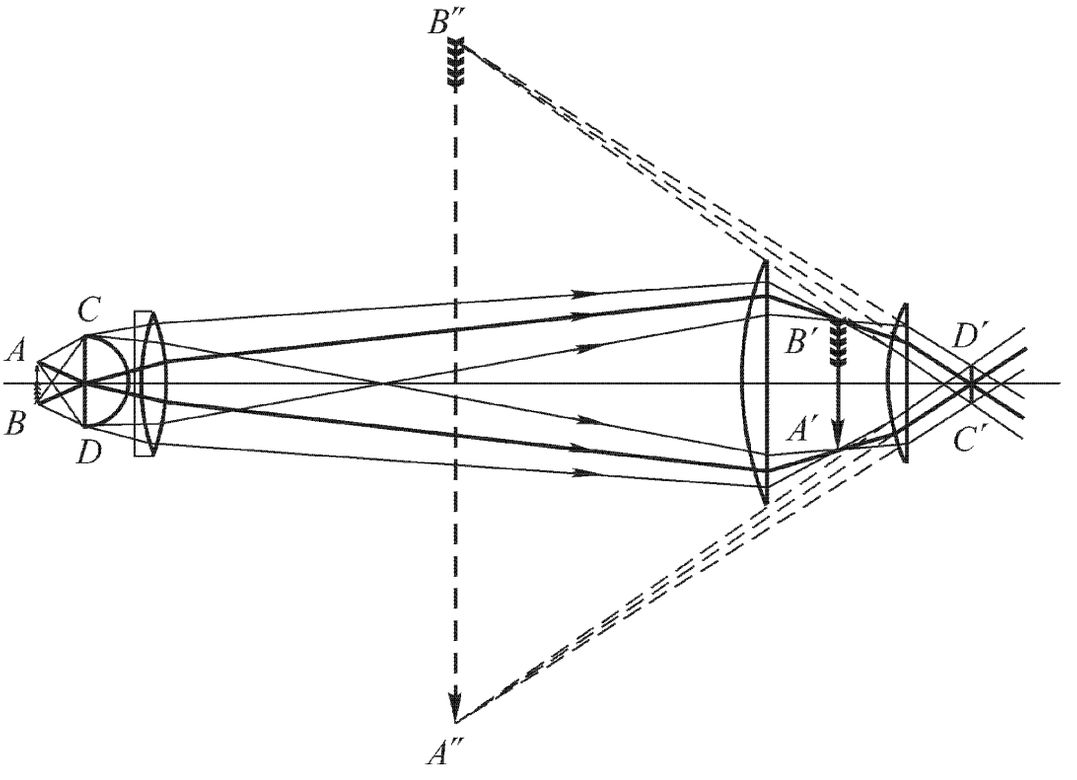


Рис. 95

Объекты, рассматриваемые в микроскоп, обычно не самосветящиеся, а освещаются посторонним источником. Для этой цели употребляется *конденсор*, состоящий из одной или нескольких короткофокусных линз, помещаемый под предметным столиком микроскопа. Чтобы полностью использовать высокую числовую апертуру объектива микроскопа, наклон освещающих лучей к оптической оси должен быть большим. Вот почему для конденсора берут короткофокусные линзы. Конечно, конденсор не может повысить поверхностную яркость источника света, его действие эквивалентно приближению источника к освещаемому объекту.

8. З р и т е л ь н ы е т р у б ы и т е л е с к о п ы. Главными частями зрительной трубы являются *объектив* и *окуляр*. Объектив дает в задней фокальной плоскости обратное уменьшенное изображение AB удаленного предмета, рассматриваемое в окуляр как в лупу (рис. 96). Для нормального глаза, аккомодированного на бесконечность, задний фокус объектива должен быть совмещен с передним фокусом окуляра. Это совмещение нарушается, но незначительно, для близорукого и дальнозорко-

го глаза. Таким образом, параллельный пучок лучей после прохождения через трубу, установленную на бесконечность, остается параллельным, т.е. зрительная труба является *телескопической оптической системой*. Увеличение (угловое) таких систем равно отношению ширины падающего пучка лучей к ширине соответствующего выходящего пучка, или отношению фокусного расстояния объектива f_1 к фокусному расстоянию окуляра f_2 (см.§ 11, п. 10). О нормальном увеличении подробно говорилось в п. 4 предыдущего параграфа.

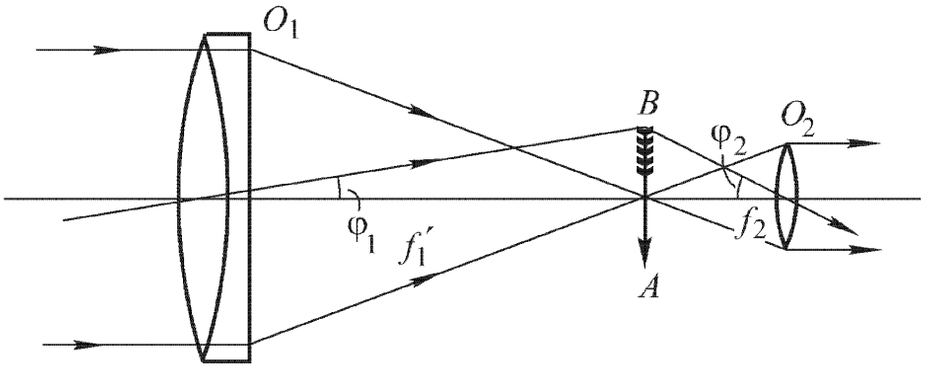


Рис. 96

Лучи, падающие на объектив зрительной трубы, наклонены под малыми углами к главной оптической оси. Поэтому нет надобности в тщательном исправлении аберраций, зависящих от этих углов. Объективом служит ахроматическая двойная линза, исправленная на сферическую аберрацию. Кома, т.е. отступление от условия синусов, может быть исправлена не столь тщательно. Ахроматизация окуляра должна удовлетворять тем же условиям, что и в случае микроскопа: если главный луч падает на окуляр под небольшим углом к главной оптической оси, то можно ограничиться ахроматизацией только фокусного расстояния. Поэтому для микроскопов и зрительных труб можно пользоваться одними и теми же окулярами. Если требуются микрометрические измерения изображения (как в астрономических трубах), то предпочтительнее пользоваться окуляром Рамсдена, а не Гюйгенса.

На рис. 96 изображен ход лучей в зрительной трубе Кеплера (для простоты сложный окуляр заменен простой собирательной линзой). Эта труба дает перевернутое изображение. Чтобы устранить этот недостаток, за изображением AB, даваемым объективом, на двойном фокусном расстоянии помещают добавочную ахроматизованную собирательную линзу, которая переворачивает изображение AB, не меняя его величины. Полученное изображение и рассматривается в окуляр. Окончательное изображение, которое видит глаз, получается прямым.

В трубе Галилея окончательное изображение получается прямым без введения промежуточной линзы. Это достигается тем, что в качестве окуляра применяется простая двояковогнутая линза, передний фокус которой F_2 совпадает с задним фокусом объектива F_1' (рис. 97). Поле зрения галилеевой трубы неве-

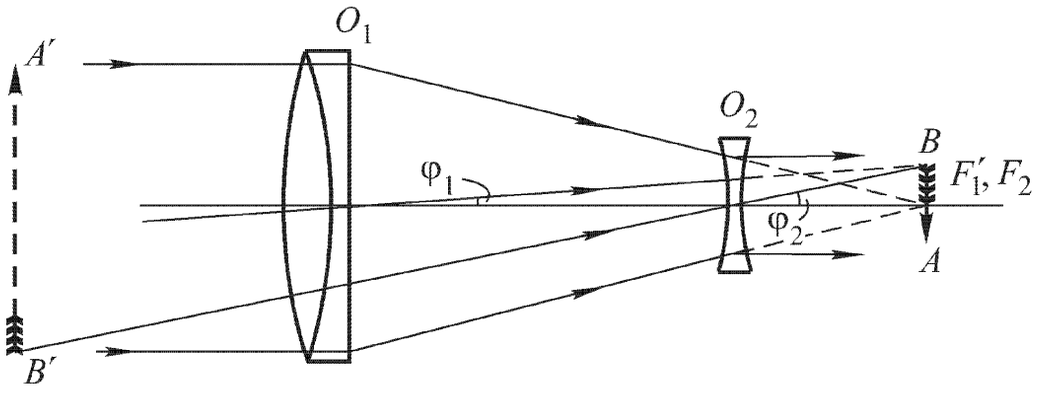


Рис. 97

лико. Зато при том же объективе галилеева труба значительно короче кеплеровой. Поэтому труба с отрицательным окуляром применяется только в театральных биноклях.

В обыкновенном бинокле, состоящем из двух зрительных труб, переворачивание изображения, даваемого объективом, достигается в каждой трубе с помощью четырехкратного полного отражения от гипотенузных поверхностей двух прямоугольных стеклянных призм A и B, расположенных, как показано на рис. 98. Луч выходит из призмы B параллельно падающему лучу, но получает боковое смещение. В остальном устройство зрительной трубы не отличается от устройства трубы Кеплера. Так как луч проходит расстояние между призмами трижды, то можно значительно уменьшить длину трубы, раздвигая призмы A и B друг относительно друга. При несколько иной конструкции можно получить бинокулярную трубу, у которой расстояние между объективами во много раз превышает расстояние между глазами наблюдателя. В такой «стереотрубе» сильно увеличивается *стереоскопический эффект*, обусловленный зрением двумя глазами.

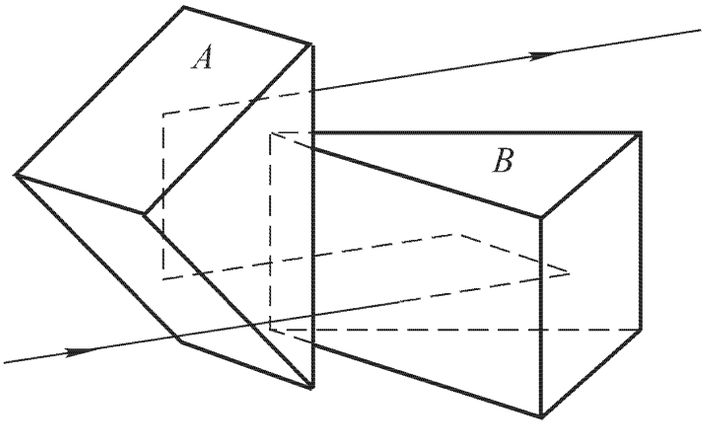


Рис. 98

9. *Астрономические трубы*, или *телескопы*, бывают двух типов: преломляющие (*рефракторы*) и отражающие (*рефлекторы*). В преломляющих объективом служит линза, в отражающих — вогнутое зеркало, называемое *главным зеркалом.*

Изготовление объективов для телескопов-рефракторов представляет большие трудности, быстро возрастающие с увеличением размеров объектива. Главная трудность состоит в получении большого диска стекла с вполне однородным показателем преломления и отсутствием натяжений. Последние снимаются медленным *отжигом* стекла, длящимся месяцами. Необходима большая точность в изготовлении сферических поверхностей, в центровке и сборке линз. Влияние неизбежных небольших случайных отклонений от полной однородности показателя преломления стекла устраняется деформацией точных сферических поверхностей объектива путем *местной полировки — ретуши.* Изготовление объектива для рефракторов больших обсерваторий тянется годами. Самым крупным рефрактором в настоящее время является телескоп Иеркской обсерватории в США. Диаметр его объектива составляет 40 дюймов (1 м). Диаметр объектива рефрактора Главной астрономической обсерватории АН СССР в Пулкове равен 30 дюймов (75 см).

Значительно легче изготовить главное зеркало телескопа, так как свет не проходит сквозь стекло, и последнее может быть менее высокого качества, но должно быть лишь хорошо отожжено. Такое стекло может быть отлито в диски огромного размера. Поэтому крупнейшими являются отражательные телескопы (рефлекторы). Самый крупный телескоп с диаметром главного зеркала 6 м недавно установлен в СССР на северных отрогах Кавказского хребта вблизи станицы Зеленчукской. До этого самым крупным был телескоп в обсерватории Маунт Паломар (Калифорния, США) с диаметром главного зеркала 5 м.

Отражающая поверхность главного зеркала телескопа-рефлектора шлифуется на стекле. Ей придают форму параболоида, чтобы лучи, параллельные главной оптической оси, собирались в фокусе параболоида. При изготовлении поверхности зеркала сначала придают форму сферы, радиус которой R почти в точности равен удвоенному фокусному расстоянию F параболоида. Затем производят параболизацию этой поверхности путем снятия шлифовкой (ретушью) тонкого слоя стекла на центральной части зеркала. В наиболее толстом месте, как нетрудно подсчитать, толщина слоя должна составлять $^1/_8(r/R)^4R$, где r — радиус кривизны зеркала. При $r = 50$ см, $F = 5$ м для этой толщины получаем $7{,}8 \cdot 10^{-4}$ см, т.е. примерно 15 длин волн желто-зеленого света. Отражающая поверхность зеркала покрывается тонким слоем серебра, алюминия или родия. С течением време-

ни под действием воздуха отражательная способность металлического слоя постепенно уменьшается. Серебряный слой должен обновляться примерно каждые полгода. Алюминиевый слой более устойчив и держится годами. Еще лучшие результаты дает покрытие зеркала слоем родия.

Важное преимущество зеркального телескопа перед рефрактором состоит в том, что этот телескоп абсолютно ахроматичен. Это позволяет не только работать со светом любой окраски, но и значительно повысить светосилу телескопа с соответствующим уменьшением его длины.

Свет должен дважды пройти трубу рефлектора, прежде чем он соберется в фокусе. Поэтому требуется небольшое вспомогательное зеркало (или призма), чтобы сместить фокальную плоскость в более удобное положение для наблюдения глазом или помещения фотографической пластинки. В зеркальном телескопе Ньютона плоское вспомогательное зеркало s смещало фокус F вбок, как указано на рис. 99 (на этом и последующих рисунках

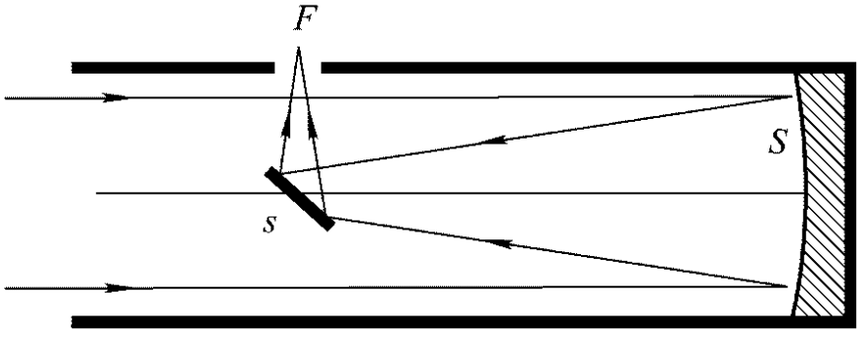


Рис. 99

окуляр не изображен, указано только смещенное положение F фокуса оптической системы).

В телескопе Вильяма Гершеля (рис. 100) вспомогательного зеркала нет, главный фокус F смещается вбок трубы путем небольшого наклона главного зеркала S. За несколько лет до Гершеля такой телескоп был изобретен Ломоносовым, но его

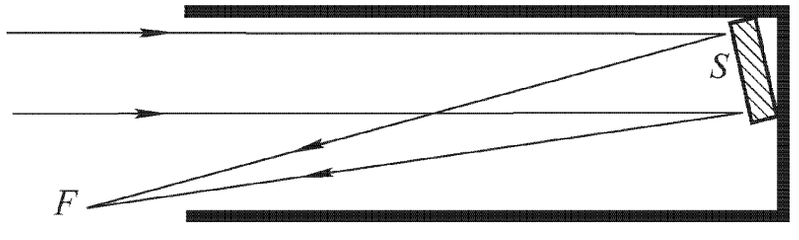


Рис. 100

изобретение оставалось неизвестным вплоть до начала XX века. Недостаток системы Ломоносова–Гершеля состоит в том,

что изображение получается не на главной оптической оси. Это ухудшает качество изображения. Впрочем, сам Гершель пользовался столь длиннофокусными зеркалами, что требовались лишь небольшие углы наклона, и ухудшение изображения было незначительным.

В системе Кассегрена (рис. 101) применяется выпуклое гиперболическое зеркало s, располагающееся на главной оптической оси несколько ближе фокуса главного зеркала S. Фокус гиперболической поверхности вспомогательного зеркала s сов-

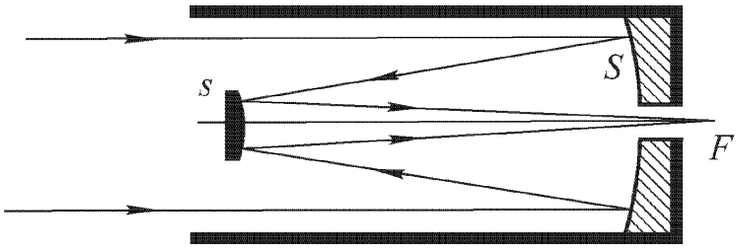


Рис. 101

падает с фокусом главного зеркала S. После отражения лучей от зеркала образуется новый фокус F оптической системы, совпадающий со вторым фокусом гиперболической поверхности зеркала s. Лучи выходят из системы через отверстие в центре главного зеркала S и попадают в окуляр или на фотопластинку.

Сходное устройство имеет телескоп Грегори (рис. 102). Здесь вспомогательным зеркалом s служит вогнутое эллиптическое зеркало, располагающееся на главной оптической оси несколько дальше фокуса главного зеркала S. Фокус зеркала S совмеща-

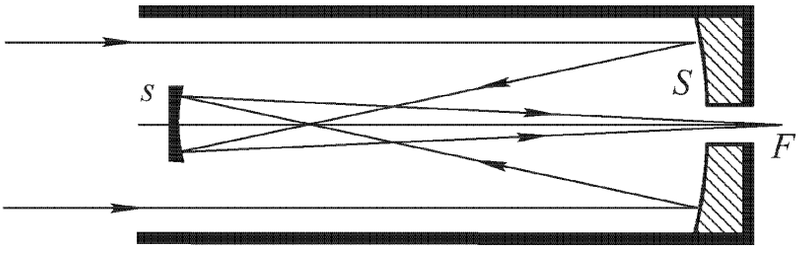


Рис. 102

ется с первым фокусом эллиптического зеркала s. Смещенный фокус F всей оптической системы получается в точке нахождения второго фокуса эллиптической поверхности зеркала s. Лучи также выходят из системы через отверстие в центре главного зеркала S и попадают в окуляр или на фотопластинку. По качеству изображения телескоп Грегори несколько превосходит телескоп Кассегрена. Но для него требуется гораздо более длинная труба. По этой причине система Грегори почти никогда не применяется.

10. К а м е р а Ш м и д т а и м е н и с к о в ы е с и с т е м ы М а к с у т о в а. Придание отражающей поверхности главного зеркала телескопа параболической формы, а также использование в качестве вспомогательных эллиптических и гиперболических зеркал устраняет сферическую аберрацию, но сохраняет все прочие геометрические аберрации, так как геометрические фокусы параболоида, эллипсоида и гиперболоида являются *только анаберрационными, но не апланатическими точками*. Зеркальным объективам телескопов всегда свойственны значительные аберрации комы и астигматизма. Вследствие этого поле зрения, где получаются четкие изображения, у этих приборов невелико и измеряется минутами, а в лучших случаях — десятками минут. В 1930 г. Б. Шмидт, сотрудник Гамбургской обсерватории в Бергедорфе, предложил новый тип телескопа, получивший название камеры Шмидта. Короткофокусная камера Шмидта с относительным отверстием $D/f = 1$ может давать резкие изображения при поле зрения $\approx 25°$. Параболическое же зеркало при таком же относительном отверстии может иметь полезное поле зрения, измеряемое только несколькими дуговыми минутами.

Рисунок 103 поясняет основную идею камеры Шмидта. На этом рисунке S_1S_2 — вогнутое сферическое зеркало с центром в точке C, а D_1D_2 — отверстие диафрагмы с центром в той же точке, введенной только для лучшего уяснения указанной идеи. Пучок лучей A, параллельных главной оптической оси зеркала, после отражения пройдет через кружок с центром в главном фокусе F. Размеры кружка определяются сферической аберрацией. Параллельный пучок A_1, падающий наклонно, пройдет через аналогичный кружок с центром в побочном фокусе F_1. Геометрическим местом всех таких фокусов будет сфера, радиус которой вдвое меньше радиуса зеркала S_1S_2. Она изображена штриховой линией. Таким образом, бесконечно удаленные точки в результате отражения света от сферического зеркала S_1S_2 изобразятся кружками, расположенными на этой штриховой сфере.

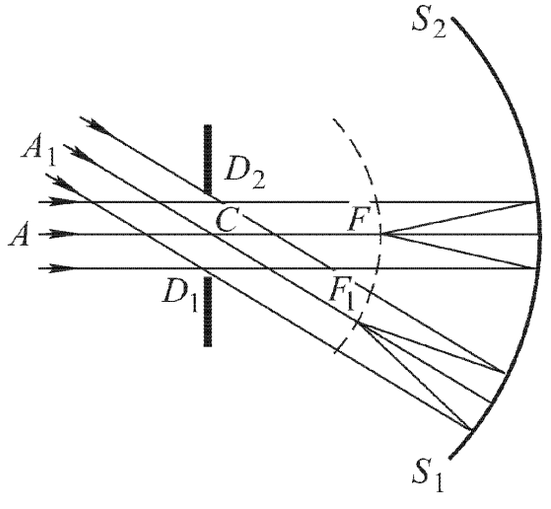


Рис. 103

Размеры кружков почти одинаковы, так как они образуются параллельными пучками лучей практически одинакового поперечного сечения. Наклон пучков влияет лишь на положение кружков на штриховой сфере, но не на их величину. Впрочем,

полная одинаковость кружков несущественна. Существенно, что единственными геометрическими аберрациями будут сферическая аберрация и искривление поверхности изображения. Комы, астигматизма и дисторсии не будет.

В обычных зеркальных телескопах сферическая аберрация устраняется параболизацией отражающей поверхности зеркала. Но такой способ годится только для точек, лежащих на главной оптической оси зеркала. Шмидт предложил устранять сферическую аберрацию коррекционной стеклянной пластинкой, помещаемой перед сферическим зеркалом S_1S_2. Одна поверхность пластинки плоская, вторая — поверхность вращения надлежащей формы, мало отклоняющаяся от плоскости. Это отклонение настолько мало, что оно совершенно незаметно невооруженному глазу.

Сечение коррекционной пластинки P_1P_2 представлено на схематическом рис. 104 в преувеличенном виде. Центральная часть пластинки более толстая — действует как собирательная линза, приближая фокус центральных лучей к зеркалу. Края пластинки более тонкие и, подобно рассеивающей линзе, удаляют от зеркала фокусы наклонных пучков лучей.

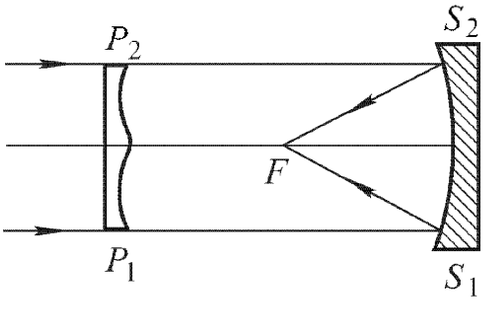


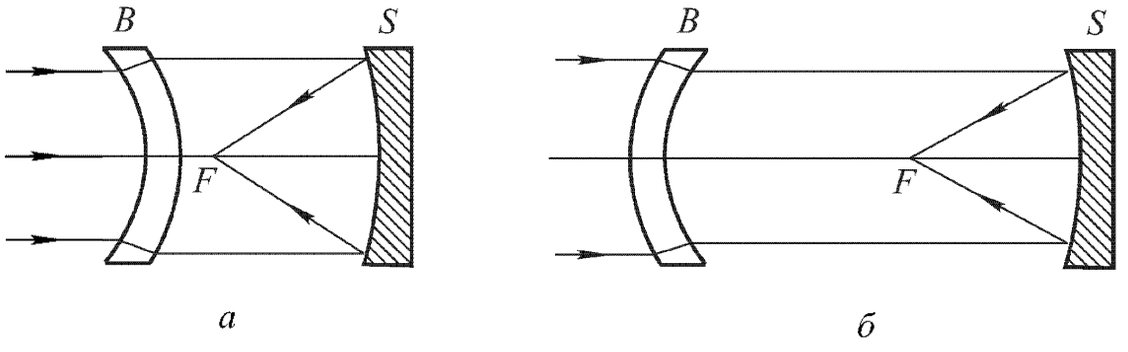
Рис. 104

Помещение коррекционной пластинки вплотную к зеркалу не может дать никаких преимуществ по сравнению с параболизацией его поверхности. Однако коррекционная пластинка, если правильно рассчитать ее форму, помещенная на определенном расстоянии от зеркала, устраняет сферическую аберрацию не только для пучков лучей, параллельных главной оптической оси, но и для пучков, наклоненных к ней. Обычно коррекционную пластинку помещают в центре сферического зеркала и для этого положения рассчитывают ее форму. При этом другие аберрации практически не возникают. Изготовление коррекционных пластинок Шмидта — очень трудная задача.

В 1941 г. Д.Д. Максутов (1896–1964) для исправления аберраций сферического зеркала предложил применять *менискu*, значительно более легкие в изготовлении, чем коррекционные пластинки Шмидта с несферическими поверхностями. Системы Максутова быстро получили широкое распространение. При определенном соотношении между радиусами кривизны сферических поверхностей мениска и его толщиной мениск ахроматизован (для двух требуемых цветов спектра).

На рис. 105 *а* приведена простейшая система Максутова, состоящая из сферического зеркала S и мениска B, обращенно-

го к нему выпуклой стороной. Если мениск ахроматизован, то в фокусе F образуется ахроматическое изображение. Сферическая аберрация мениска подбирается так, чтобы компенсировать сферическую аберрацию зеркала S. При правильно выбранном расстоянии l между мениском и зеркалом оказывается исправленной и кома, т.е. система становится апланатической. Если при том же расстоянии l мениск повернуть к зеркалу вогнутой стороной, то апланатизм нарушается. Его можно восстановить, увеличивая расстояние l (рис. 105 *б*). Первая ориентация менис-



а *б*

Рис. 105

ка более выгодна, так как при ней телескоп получается вдвое короче, чем при второй. Однако для некоторых целей предпочтительнее вторая ориентация.

Астигматизм менисковой системы может быть сделан ничтожно малым, а в частных случаях и вполне устранен. Искривление поверхности изображения остается. Но из всех аберраций эта аберрация является наименее вредной, и существуют способы ее устранения. Соединение камеры Шмидта или менисковых систем Максутова с окуляром производится так же, как в телескопах Ньютона, Кассегрена и Грегори.

§ 25. Электрические и магнитные линзы

1. Допустим, что показатель преломления n меняется в пространстве *непрерывно*. Проведем поверхности равного показателя преломления и притом настолько часто, что показатели преломления между каждыми соседними поверхностями можно будет считать величинами постоянными. Тогда непрерывное изменение величины n заменится скачкообразным, происходящим на границах между слоями. Если среда обладает осевой симметрией, то эти границы будут поверхностями вращения, вершины которых лежат на оси симметрии системы. В малой окрестности вокруг оси симметрии их можно аппроксимировать сферами, центры которых также лежат на той же оси. Таким путем мы приходим к центрированной системе тонких сферических линз, у которой ось симметрии служит главной оптической осью и к которой применимы все результаты оптики параксиальных лучей. Увеличивая число слоев бесконечно и одновременно устремляя к нулю их толщины, мы восстановим в пределе первоначальное непрерывное распределе-

ние показателя преломления. Отсюда следует, что *осесимметричную среду с непрерывно изменяющимся в пространстве показателем преломления можно рассматривать как предельный случай центрированной системы линз и применять к ней законы и методы оптики параксиальных лучей. Такая среда обладает способностью давать оптические изображения.*

Оптические системы с непрерывно изменяющимися показателями преломления принципиально возможны, но из-за трудностей их изготовления в световой оптике они не встречаются. (Исключение составляет хрусталик глаза, показатель преломления которого возрастает от периферии к центру.) Аналогом таких систем являются электронные и ионные приборы (электронный микроскоп, электронный осциллограф, электронно-лучевая трубка в телевидении и пр.), в которых роль лучей играют электроны или ионы, движущиеся в электростатических или магнитных полях, создаваемых заряженными электродами или катушками, по которым текут электрические токи. Эти электроды называются *электрическими*, а катушки — *магнитными линзами*. Получение изображений в таких системах изучается в *электронной и ионной оптике*.

2. Возможность фокусировки заряженных частиц и получения изображений в осесимметричных системах электростатических линз непосредственно следует из формальной аналогии между геометрической оптикой и классической механикой, отмеченной нами в п. 1 § 4. Действительно, согласно этой аналогии, траектория частицы в потенциальном силовом поле совпадает со световым лучом в непрерывной среде, показатель преломления которой пропорционален скорости (а в релятивистской механике — импульсу) частицы, однозначно определяющейся ее положением в пространстве. Например, можно положить

$$n = \sqrt{2(W - eV)}, \tag{25.1}$$

где W — полная энергия частицы, e — ее заряд, V — потенциал электростатического поля.

3. Несколько труднее понять действие магнитных линз и комбинаций таких линз с электростатическими линзами. Для исследования вопроса рассмотрим статическое электромагнитное поле, обладающее симметрией вращения вокруг некоторой оси. Эту ось примем за ось X цилиндрической системы координат. Расстояние до оси X будем обозначать через r, а азимутальный угол — через φ. Ввиду симметрии вращения, электрическое и магнитное поля не могут зависеть от φ. Предположим кроме того, что всякая плоскость, проходящая через ось X, является плоскостью симметрии поля. Тогда не будет φ-составляющих электрического и магнитного полей ($E_\varphi = B_\varphi = 0$). Оставшиеся x- и r-составляющие этих полей будут функциями только координат x и r. Уравнение $\operatorname{div} \mathbf{B} = 0$ после умножения на r запишется в виде

$$\frac{\partial}{\partial x}(rB_x) + \frac{\partial}{\partial r}(rB_r) = 0. \tag{25.2}$$

Введем потенциальную функцию $A(x, r)$, определяемую условием $rB_x = -\partial A/\partial r$. Тогда предыдущее уравнение преобразуется в

$$\frac{\partial}{\partial r}(rB_r) = \frac{\partial}{\partial x}\left(\frac{\partial A}{\partial r}\right) = \frac{\partial}{\partial r}\left(\frac{\partial A}{\partial x}\right),$$

откуда $rB_r = \partial A/\partial x$, причем «постоянную интегрирования» (точнее — функцию x) мы без ущерба для общности положили равной нулю. Таким

образом,

$$rB_x = -\frac{\partial A}{\partial r}, \quad rB_r = \frac{\partial A}{\partial x}. \tag{25.3}$$

Отсюда

$$A(x, r) = \int r(B_r\, dx - B_x\, dr). \tag{25.4}$$

В силу (25.2) входящий сюда интеграл не зависит от пути интегрирования, а только от его начальной и конечной точек. Условимся помещать начальную точку на оси системы. На оси системы в силу симметрии $B_r = 0$. Поэтому не имеет значения, в каком месте этой оси выбрать начальную точку. От этого значение интеграла (25.4) не зависит, и он может быть представлен в виде

$$A(x, r) = -\int_0^r B_x r\, dr. \tag{25.5}$$

Обратимся теперь к исследованию движения частицы с массой m и зарядом e, ограничиваясь ради простоты нерелятивистским приближением. Угловую скорость $\dot{\varphi}$ частицы можно найти из уравнения моментов

$$\frac{d}{dt}(mr^2\dot{\varphi}) = M_x,$$

где M_x — момент сил, действующих на частицу, относительно оси X. Электрическое поле не влияет на момент M_x. Он создается только силой $\mathbf{F} = \frac{e}{c}[\mathbf{vB}]$, действующей со стороны магнитного поля. Вычисляя этот момент, находим

$$M_x = \frac{er}{c}(B_r v_x - B_x v_r) = \frac{e}{c}\left(v_x \frac{\partial A}{\partial x} + v_r \frac{\partial A}{\partial r}\right).$$

Если под x, r и φ понимать координаты движущейся частицы, то $A(x, r)$ станет функцией времени t. Производная этой функции равна

$$\frac{dA}{dt} = \frac{\partial A}{\partial x}\frac{dx}{dt} + \frac{\partial A}{\partial r}\frac{dr}{dt} = v_x \frac{\partial A}{\partial x} + v_r \frac{\partial A}{\partial r}.$$

Следовательно,

$$M_x = \frac{e}{c}\frac{dA}{dt}, \quad \frac{d}{dt}(mr^2\dot{\varphi}) = \frac{e}{c}\frac{dA}{dt}.$$

Отсюда

$$\dot{\varphi} = \frac{e}{mcr^2}A + \frac{C}{mr^2}.$$

Постоянная C должна обращаться в нуль. Иначе (так как на оси системы функция A равна нулю) при $r = 0$ мы получили бы $r\dot{\varphi} = \infty$. А это невозможно, так как $r\dot{\varphi}$ есть линейная скорость частицы в ее вращении вокруг оси X. Таким образом,

$$\dot{\varphi} = \frac{eA}{mcr^2}. \tag{25.6}$$

Возьмем теперь уравнение движения частицы

$$m\mathbf{a} = e\left(\mathbf{E} + \frac{1}{c}[\mathbf{vB}]\right), \tag{25.7}$$

где $\mathbf{a}$ — ускорение. При проецировании его на ось X и направление радиуса r слева получим соответственно $ma_x = m\ddot{x}$ и ma_r. Радиальное ускорение в

цилиндрической системе координат определяется выражением $a_r = \ddot{r} - \omega^2 r$ (см. т. I, § 46). Введем еще электрический потенциал V и учтем соотношения (25.3) и (25.6). Тогда после недлинных преобразований из (25.7) получим

$$m\ddot{x} = -e\frac{\partial}{\partial x}\left(V + \frac{eA^2}{2mc^2r^2}\right), \quad m\ddot{r} = -e\frac{\partial}{\partial r}\left(V + \frac{eA^2}{2mc^2r^2}\right). \tag{25.8}$$

Задача о движении частицы распалась на две *независимые задачи*: определение угловой координаты φ и определение координат x и r. Для решения первой задачи имеется уравнение (25.6), а для второй — уравнения (25.8). Формально вторая задача идентична с задачей определения траектории заряженной частицы в плоском электростатическом поле с потенциалом

$$U = V + \frac{eA^2}{2mc^2r^2}. \tag{25.9}$$

Ее формальным аналогом в оптике служит задача о распространении светового луча в неоднородной изотропной среде с показателем преломления

$$n = \sqrt{2m(W - eV) - \frac{e^2A^2}{c^2r^2}} = \sqrt{n_0^2 - \frac{e^2A^2}{c^2r^2}}, \tag{25.10}$$

где $n_0 = \sqrt{2m(W - eV)}$ — «показатель преломления» в отсутствие магнитного поля. Для получения окончательной формы траектории частицы надо наложить на этот «луч» *дополнительное вращение*, выражаемое формулой (25.6).

Учтем теперь условие параксиальности. Для этого разложим $B_x(x, r)$ в ряд по степеням r. Ввиду осевой симметрии, это разложение может содержать только четные степени r. Оборвем разложение на члене нулевой степени — в этом приближении поле B_x не зависит от r. Вынося B_x из-под знака интеграла (25.5) и выполнив интегрирование, получим

$$A = -\frac{1}{2}B_x r^2, \tag{25.11}$$

а после подстановки в формулу (25.6)

$$\dot{\varphi} = -eB_x/(2mc). \tag{25.12}$$

Отсюда

$$\frac{d\varphi}{dx} = \frac{d\varphi}{dt}\frac{dt}{dx} = -\frac{1}{v_x}\frac{eB_x}{2mc}. \tag{25.13}$$

В параксиальном приближении скорость v_x также не зависит от расстояния r до оси X и равна скорости частицы, движущейся вдоль этой оси. Поэтому производные $d\varphi/dx$ имеют одинаковые значения для всех частиц, независимо от наклона их траекторий к оси X. Если l — расстояние от предметной плоскости до плоскости изображения, то на этом расстоянии в параксиальном приближении все частицы поворачивают вокруг оси симметрии системы на один и тот же угол

$$\varphi = -\frac{e}{2mc}\int_0^l \frac{B_x}{v_x}\,dx. \tag{25.14}$$

Следовательно, *в смысле получения изображений система будет вести себя так, как если бы магнитного поля не было, а электростатическое*

определялось потенциалом (25.9). *Магнитное поле приводит еще к несущественному повороту всего изображения вокруг оси симметрии системы на угол, определяемый формулой* (25.14). В высших приближениях угол поворота зависит от наклона траектории к оси системы. Это ведет к появлению дополнительных аберраций, обусловленных наличием магнитного поля.

4. Отметим еще одно обстоятельство. Если электрического поля нет, а магнитное поле однородно, то частица движется по спирали, вращаясь с циклотронной частотой $\omega = -eB/mc$ (см. т. III, § 86). Между тем формула (25.12) для этого случая дает угловую скорость, вдвое меньшую ω. Недоразумение легко разъясняется, если заметить, что ω есть угловая скорость вращения частицы *вокруг оси спирали*, тогда как формула (25.12) определяет ее вращение *вокруг одной из образующих* той же спирали. Допустим, например, что магнитное поле перпендикулярно к плоскости рисунка (рис. 106), а частица вращается по окружности в той же плоскости. За ось X можно, конечно, принять любую прямую, направленную вдоль поля $\mathbf{B}$. Поэтому можно считать, что частица выходит из какой-то точки M этой оси. Описав окружность, она снова возвращается в исходную точку. Ее вращение вокруг точки M описывается формулой (25.12), тогда как вращение вокруг центра O окружности происходит с циклотронной частотой ω. За время dt частица повернется вокруг центра O на угол $\omega\,dt$, а вокруг точки M — на угол $d\varphi = {}^1\!/_2\omega\,dt$, так что $\dot\varphi = {}^1\!/_2\omega$.

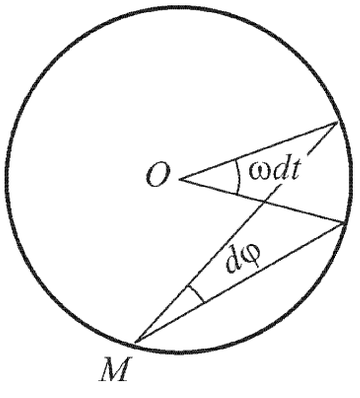


Рис. 106

5. Выше фокусировка частиц в электрических линзах была объяснена с помощью аналогии со световой оптикой. Сделаем теперь то же самое, рассматривая силы, действующие на частицу. На рис. 107 представлена линза, состоящая из трех соосных металлических цилиндров одинакового диамет-

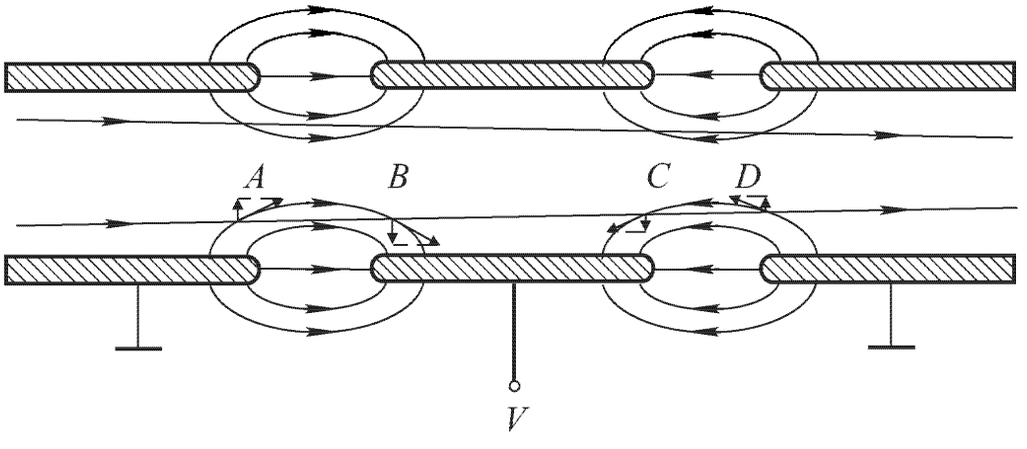


Рис. 107

ра. Крайние цилиндры заземлены, на средний подан положительный или отрицательный потенциал. Линзы такого типа используются в электронно-лучевых трубках и некоторых электронных микроскопах. На рисунке изображены электрические силовые линии с указанием направлений сил, действующих на частицу.

Допустим, что частица влетает в линзу, двигаясь параллельно ее оси. В области A действующая на нее сила имеет составляющую, направленную вверх. Эта сила будет смещать частицу вверх. В области B направление вертикальной составляющей силы изменится на противоположное. Однако, так как под действием электрического поля скорость частицы непрерывно возрастает, на прохождение области B частица затрачивает меньше времени, чем на прохождение области A. Поэтому поперечная скорость, приобретенная частицей в области A, не может быть скомпенсирована скоростью противоположного направления, которую она получает в области B. В результате в областях A и B и по выходе из них частица будет двигаться вверх, приближаясь к оси линзы. Аналогично, в области C на частицу действует сила, стремящаяся удалить ее от оси линзы, а в области D — приблизить. Но в этих областях частица замедляется, а потому проводит в области D большее время, чем в C. Поэтому при прохождении обеих областей C и D вертикальная скорость частицы, направленная вверх, возрастет.

Эти разъяснения объясняют, почему частицы приближаются к оси линзы. Конечно, из них не следует, что все частицы пучка соберутся в одной и той же точке на оси линзы. Для доказательства этого требуется уже количественное рассмотрение, которое и было проведено выше на основе аналогии со световой оптикой.

ЗАДАЧИ

1. Вычислить фокусное расстояние тонкой оптической линзы, показатель преломления которой в пространстве меняется непрерывно.

Р е ш е н и е. Сначала не будем вводить предположение о тонкости линзы, а рассмотрим среду, обладающую симметрией вращения вокруг оси X. Уравнение луча в меридиональной плоскости представится в виде $r = r(x)$. Обозначим через u угол, образуемый касательной к лучу с осью X (рис. 108). В параксиальном приближении квадратом этого угла пренебрегают. В этом приближении кривизна луча определяется выражением $1/R = -du/dx$, причем радиус кривизны R мы считаем положительным, когда луч обращен вогнутостью к главной оптической оси X, и отрицательным в противоположном случае. Воспользуемся формулой (4.1). В пределах точности параксиальной оптики $\partial r/\partial N = -\cos u \approx -1$, $\partial x/\partial N = \sin u \approx u$, так что

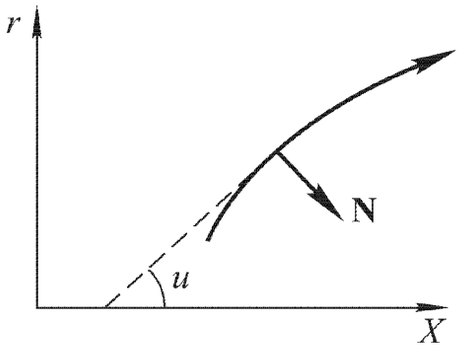


Рис. 108

$$\frac{\partial n}{\partial N} = \frac{\partial n}{\partial r}\frac{\partial r}{\partial N} + \frac{\partial n}{\partial x}\frac{\partial x}{\partial N} \approx -\frac{\partial n}{\partial r} + u\frac{\partial n}{\partial x}.$$

Так как на луче $r = r(x)$, то $n(x, r)$ можно рассматривать как сложную функцию от x, т.е. $n(x) = n[x, r(x)]$. Ее производная по x определяется выражением

$$\frac{dn}{dx} = \frac{\partial n}{\partial x} + \frac{\partial n}{\partial r}\frac{\partial r}{\partial x} = \frac{\partial n}{\partial x} + u\frac{\partial n}{\partial r},$$

т.е. отличается от частной производной $\partial n/\partial x$ на величину первого порядка малости по u. Поэтому замена $\partial n/\partial x$ на dn/dx в предыдущем выражении вносит ошибку второго порядка, которой мы пренебрегаем. Сделав эту за-

мену и воспользовавшись формулой (4.1), получим

$$-\frac{1}{R}=\frac{du}{dx}=-\frac{1}{n}\frac{\partial n}{\partial N}=\frac{1}{n}\frac{\partial n}{\partial r}=-u\frac{dn}{dx},$$

или

$$\frac{d}{dx}(nu)=\frac{\partial n}{\partial r}. \tag{25.15}$$

Разложим $n=n(x,r)$ в ряд по степеням r и оборвем разложение на члене второй степени. Член первой степени должен отсутствовать, ввиду осевой симметрии системы. Таким образом,

$$n=n_0(x)+\frac{1}{2}\left(\frac{\partial^2 n}{\partial r^2}\right)_0 r^2,$$

а потому

$$\frac{\partial n}{\partial r}=r\left(\frac{\partial^2 n}{\partial r^2}\right)_0.$$

Здесь нуль в индексе означает, что величина с таким индексом берется при $r=0$, т.е. на главной оптической оси системы. В результате уравнение (25.15) переходит в

$$\frac{d}{dx}(nu)=r\left(\frac{\partial^2 n}{\partial r^2}\right)_0. \tag{25.16}$$

Допустим теперь, что линза тонкая. Пусть P и P' — оптически сопряженные точки на ее оси (рис. 109). Отрезки соединяющего их луча вне линзы прямолинейны. Проинтегрируем уравнение (25.16) по x в пределах от

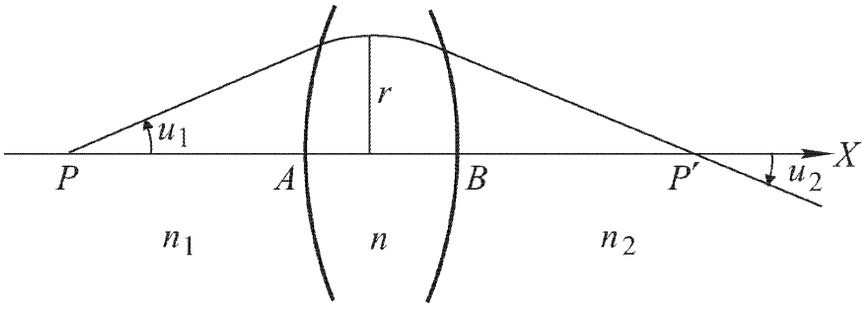


Рис. 109

$-\infty$ до $+\infty$. Фактически это сводится к интегрированию по отрезку AB, так как вне этого отрезка $\partial^2 n/\partial r^2=0$. Внутри линзы, поскольку она тонкая, радиус r можно считать постоянным и вынести его из-под знака интеграла. Это дает

$$n_2u_2-n_1u_1=r\int\limits_{-\infty}^{+\infty}\left(\frac{\partial^2 n}{\partial r^2}\right)dx,$$

где n_1 — показатель преломления в пространстве предметов, а n_2 — в пространстве изображений. Обозначим, как и раньше, через ξ и ξ' расстояния предмета P и его изображения P' от центра линзы. С учетом правила знаков $u_1=-r/\xi$, $u_2=-r/\xi'$, так что предыдущее уравнение переходит в

$$\frac{f}{\xi}+\frac{f'}{\xi'}=-1, \tag{25.17}$$

где

$$\frac{1}{f} = -\frac{1}{n_1}\int\limits_{-\infty}^{+\infty}\left(\frac{\partial^2 n}{\partial r^2}\right)_0 dx, \quad \frac{1}{f'} = \frac{1}{n_2}\int\limits_{-\infty}^{+\infty}\left(\frac{\partial^2 n}{\partial r^2}\right)_0 dx. \tag{25.18}$$

2. Вычислить фокусные расстояния тонкой электростатической линзы (примером такой линзы может служить «одиночная линза», представленная на рис. 110).

Р е ш е н и е. Потенциал поля V определим по формуле $^1/_2 mv^2 = |e|V$. При таком определении V — существенно положительная величина.

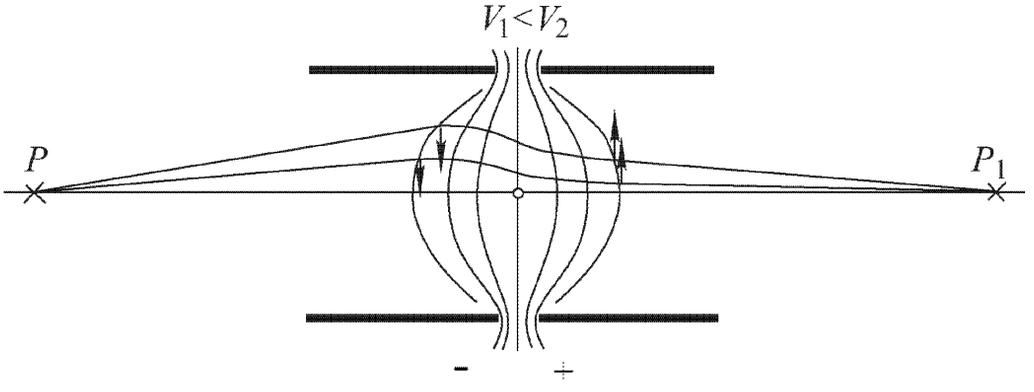


Рис. 110

В формулах (25.18) делаем замену $n \to \sqrt{V}$. Учитывая, что на оси системы $\partial V/\partial r = 0$, получаем

$$\left(\frac{\partial^2 n}{\partial r^2}\right)_0 \to \frac{\partial^2\sqrt{V}}{\partial r^2} = \frac{1}{2\sqrt{V}}\frac{\partial^2 V}{\partial r^2}.$$

При отсутствии свободных зарядов $\nabla^2 V = 0$. На оси $\partial^2 V/\partial y^2 = \partial^2 V/\partial z^2 = \partial^2 V/\partial r^2$, так что $\nabla^2 V = \partial^2 V/\partial x^2 + 2(\partial^2 V/\partial r^2) = 0$. В результате из (25.18) находим

$$\frac{1}{f} = \frac{1}{4\sqrt{V_1}}\int\limits_{-\infty}^{+\infty}\frac{1}{\sqrt{V}}\frac{\partial^2 V}{\partial x^2}\,dx. \tag{25.19}$$

Проинтегрировав по частям и приняв во внимание, что на пределах интеграла $\partial V/\partial x = 0$, получим

$$\frac{1}{f} = \frac{1}{8\sqrt{V_1}}\int\limits_{-\infty}^{+\infty} V^{-3/2}E_x^2\,dx, \quad \frac{1}{f'} = -\frac{1}{8\sqrt{V_2}}\int\limits_{-\infty}^{+\infty} V^{-3/2}E_x^2\,dx, \tag{25.20}$$

где V_1 — потенциал пространства предметов, V_2 — потенциал пространства изображений, $E_x = -\partial V/\partial x$ — напряженность электрического поля на оси. Если $V_1 = V_2$, то $f = -f'$.

Так как $V > 0$, то для тонких электростатических линз $f > 0$, $f' < 0$. Такие линзы всегда будут собирательными. Оптические линзы, даже тонкие, могут быть и собирательными, и рассеивающими. Это различие связано с тем, что функция $n(r)$ может быть какой угодно, тогда как потенциал

$V(r)$ при отсутствии свободных зарядов должен удовлетворять уравнению Лапласа $\nabla^2 V = 0$. Толстые электростатические линзы могут быть и собирательными, и рассеивающими. В этом нетрудно убедиться на примере системы, состоящей из двух тонких электростатических линз. Формулы (12.3) показывают, что такая система будет собирательной, когда оптический интервал Δ отрицателен, и рассеивающей, когда он положителен.

3. Предполагая, что $n_0 \gg \left|\dfrac{eA}{er}\right|$, вычислить фокусное расстояние тонкой магнитной линзы.

Р е ш е н и е. Разлагая (25.10) в ряд и отбрасывая члены высших степеней, получим

$$n = n_0 - \frac{e^2 A^2}{2n_0 c^2 r^2} = n_0 - \frac{e^2 r^2}{8n_0 c^2} B_x^2,$$

откуда

$$\frac{\partial^2 n}{\partial r^2} = -\frac{e^2}{4n_0 c^2} B_x^2.$$

Следовательно,

$$\frac{1}{f} = -\frac{1}{f'} = \frac{e^2}{4n_0^2 c^2} \int\limits_{-\infty}^{+\infty} B_x^2\, dx = \frac{e^2}{8mc^2 W} \int\limits_{-\infty}^{+\infty} B_x^2\, dx. \tag{25.21}$$

Как и электростатическая, тонкая магнитная линза всегда будет собирательной.

4. Вычислить фокусное расстояние кругового витка радиуса a тонкой проволоки, по которой течет ток I.

О т в е т. В гауссовой системе единиц

$$\frac{1}{f} = \frac{e^2 \pi^2 a^4 I^2}{8mc^4 W} \int\limits_{-\infty}^{+\infty} \frac{dx}{(x^2 + a^2)^3} = \frac{3\pi^3 e^2}{16mc^4 W} \frac{I^2}{a}. \tag{25.22}$$

Выражая энергию W в электрон-вольтах, полагаем $W = |e|V_0$ и переходим к практическим единицам (сантиметры, вольты, амперы). Тогда получаем для электронов

$$f = -f' = 98 \frac{V_0 a}{I^2}. \tag{25.23}$$

Для короткой катушки с числом витков N

$$f = -f' = 98 \frac{V_0 a}{N^2 I^2}. \tag{25.24}$$

5. Показать, что формула (25.18) для f в предельном случае тонкой линзы с резкими краями переходит в формулу (10.9). Считать, что показатели преломления n_1 и n_2 по обе стороны линзы одинаковы.

У к а з а н и е. Принять, что показатель преломления n меняется только в бесконечно тонких слоях вблизи каждой сферической границы линзы. На всякой сферической поверхности, проведенной внутри этих слоев параллельно ближайшей поверхности линзы, функция n постоянна. Пользуясь этим, дифференцирование по r можно заменить дифференцированием по x, а затем вычислить интеграл в формулах (25.18).

ГЛАВА III

ИНТЕРФЕРЕНЦИЯ СВЕТА

§ 26. Общие сведения об интерференции

1. В *интерференции* и *дифракции* проявляются *волновые свойства света*. После открытия этих явлений на них смотрели сначала как на доказательство исключительно *волновой природы света*. Такая точка зрения оказалась недостаточной. В XX веке были открыты *корпускулярные свойства света*, а затем *волновые свойства обыкновенных частиц*: электронов, протонов, нейтронов, атомов, молекул и пр. Как это ни парадоксально, природа света и вещества оказалась *двойственной*: корпускулярно-волновой. С открытием этого факта связан коренной пересмотр физических воззрений, приведший к построению *квантовой механики*. От этого, однако, значение интерференционных и дифракционных явлений не уменьшилось. В наши дни интерференция и дифракция света имеют важные практические применения, например в спектроскопии и метрологии.

2. С помощью различных приспособлений, например отражений, преломлений или как-либо иначе, можно наложить один световой пучок на другой. Если выполняется принцип суперпозиции (а это предполагается всюду в настоящей главе), то каждый пучок будет проходить через область перекрытия так, как если бы другого пучка не было совсем. Пусть $\mathbf{E}_1$ — напряженность электрического поля, создаваемая первым пучком в произвольной точке A области перекрытия, а $\mathbf{E}_2$ — вторым. Согласно принципу суперпозиции, результирующая напряженность поля в той же точке A, создаваемая обоими пучками, будет представляться векторной суммой $\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2$. Так же ведут себя любые скалярные или векторные волны, какова бы ни была их физическая природа. Требуется только, чтобы они подчинялись *принципу суперпозиции*. Однако частота световых волн совсем иного порядка, чем частота, скажем, звуковых волн или волн на

поверхности воды. Это ограничивает возможности приемников света.

Всем приемникам света присуща определенная *инерционность*. Ее можно характеризовать временем *установления* или *разрешения* приемника τ. Так, для глаза $\tau \approx 0{,}1$ с. Это есть время, в течение которого глаз сохраняет зрительное впечатление, т.е. «видит свет» даже тогда, когда излучение уже перестало в него попадать. Глаз не замечает, например, быстрых миганий света, если они следуют друг за другом через времена, малые по сравнению с τ (кино, телевидение). У фотоматериалов τ (время экспозиции) обычно порядка 10^{-2}–10^{-4} с. Существуют приемники, для которых время разрешения может быть сделано гораздо меньше. В ячейках Керра это время может быть доведено до 10^{-8}–10^{-9}. Наиболее быстродействующие современные фотоэлектрические приемники имеют время разрешения порядка 10^{-10} с, но оно может быть сделано еще меньше.

Однако даже такие времена очень велики по сравнению с периодами оптических колебаний. Средний период колебаний T электромагнитного поля в видимой области спектра составляет около 10^{-15} с. Поэтому ни один приемник света не позволяет измерить мгновенное значение напряженности электрического или магнитного поля в световой волне. Для этого время разрешения приемника должно было бы быть мало по сравнению с периодом световых колебаний T. Все приемники могут измерять только величины, *квадратичные по полю*, усредненные за времена, не меньшие времени разрешения приемника. Сюда относятся энергетические и фотометрические величины: лучистый и световой поток, яркость, освещенность и пр.

В явлениях интерференции, дифракции и пр. представляют интерес не абсолютные, а только *относительные значения* этих величин. Например, нас может интересовать относительное распределение освещенности на экране, куда попадает свет. При такой постановке вопроса нет смысла точно указывать, о какой именно энергетической или фотометрической величине идет речь в том или ином конкретном случае. Заключения будут относиться к любой усредненной по времени величине, квадратичной по напряженности электрического поля. Эту нечетко определенную величину принято называть *интенсивностью света* или *интенсивностью колебаний*. Ниже она обозначается через I. За I мы будем обычно принимать усредненное по времени значение квадрата напряженности электрического поля, т.е. $I = \overline{\mathbf{E}^2}$.

3. Найдем теперь интенсивность света в какой-либо точке пространства, где перекрываются два пучка света. Возведя равенство $\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2$ в квадрат и произведя усреднение по вре-

мени, получим

$$I = I_1 + I_2 + I_{12}, \tag{26.1}$$

где I_1 — интенсивность света первого пучка, а I_2 — второго. Последнее слагаемое

$$I_{12} = 2(\overline{\mathbf{E}_1\mathbf{E}_2}), \tag{26.2}$$

учитывающее взаимодействие пучков, называется *интерференционным членом.* Если взять независимые источники света, например две электрические лампочки, то повседневный опыт показывает, что $I = I_1 + I_2$, т.е. результирующая интенсивность равна сумме интенсивностей налагающихся пучков, а потому интерференционный член обращается в нуль. Тогда говорят, что пучки *не коррелированы* или *не когерентны* между собой. Однако, если накладывающиеся пучки не независимы, например один получается отражением другого от зеркала, то в некоторых случаях интерференционный член I_{12} не обращается в нуль, а потому $I \neq I_1 + I_2$. В одних точках пространства результирующая интенсивность I больше, в других меньше суммы интенсивностей I_1 и I_2. Это явление называется *интерференцией волн*, или *интерференцией колебаний.* Говорят, что источники света и соответствующие им пучки, дающие интерференцию, *коррелированы* или *когерентны* между собой, т.е. колебательные и волновые процессы в них протекают во времени согласованно. Говорят также, что когерентные пучки *интерферируют*, а некогерентные *не интерферируют* между собой.

Все сказанное о двух пучках относится и к случаю наложения нескольких пучков. Интерференция двух пучков называется *двухлучевой*, многих пучков — *многолучевой.*

4. Начнем с идеализированного случая, когда обе волны *строго монохроматические* и имеют одну и ту же частоту. *Монохроматическая волна* — это строго синусоидальная волна с постоянными во времени частотой ω, амплитудой a и начальной фазой φ. Амплитуда и фаза колебаний могут меняться от одной точки пространства к другой, частота одна и та же для колебательного процесса *во всем пространстве.* Монохроматическое колебание в каждой точке пространства длится *бесконечно долго*, не имея ни начала, ни конца во времени. Поэтому строго монохроматические колебания и волны никогда не могут быть точно реализованы в действительности. Однако эти идеализации играют громадную роль в учении о колебаниях и волнах, в чем мы уже убедились в гл. X третьего тома и еще в большей степени убедимся в дальнейшем.

Допустим сначала, что в рассматриваемой точке наблюдения оба вектора $\mathbf{E}_1$ и $\mathbf{E}_2$ параллельны или антипараллельны. Тогда можно отвлечься от векторного характера колебаний, считая их

скалярными. Представим эти колебания в вещественной форме:

$$E_1 = a_1 \cos(\omega t + \varphi_1), \quad E_2 = a_2 \cos(\omega t + \varphi_2), \tag{26.3}$$

где a_1 и a_2, φ_1 и φ_2 — амплитуды и начальные фазы колебаний. Если ввести комплексные амплитуды $A_1 = a_1 e^{i\varphi_1}$ и $A_2 = a_2 e^{i\varphi_2}$, то в комплексной форме те же колебания представятся так:

$$E_1 = A_1 e^{i\omega t}, \quad E_2 = A_2 e^{i\omega t}. \tag{26.4}$$

Результирующее колебание будет

$$E = E_1 + E_2 = (A_1 + A_2) e^{i\omega t}.$$

Это — также монохроматическое колебание с той же частотой ω и комплексной амплитудой $A = A_1 + A_2$. Чтобы найти вещественную амплитуду a и начальную фазу φ результирующего колебания, запишем последнее соотношение так:

$$a e^{i\varphi} = a_1 e^{i\varphi_1} + a_2 e^{i\varphi_2}.$$

Умножая его на комплексно-сопряженное, получим

$$a^2 = a_1^2 + a_2^2 + 2a_1 a_2 \cos(\varphi_2 - \varphi_1), \tag{26.5}$$

а после отделения вещественной части от мнимой

$$a\cos\varphi = a_1 \cos\varphi_1 + a_2 \cos\varphi_2, \quad a\sin\varphi = a_1 \sin\varphi_1 + a_2 \sin\varphi_2.$$

Отсюда

$$\operatorname{tg}\varphi = \frac{a_1 \sin\varphi_1 + a_2 \sin\varphi_2}{a_1 \cos\varphi_1 + a_2 \cos\varphi_2}. \tag{26.6}$$

На рис. 111 приведена *векторная диаграмма* сложения рассматриваемых колебаний, из которой также нетрудно получить результаты (26.5) и (26.6).

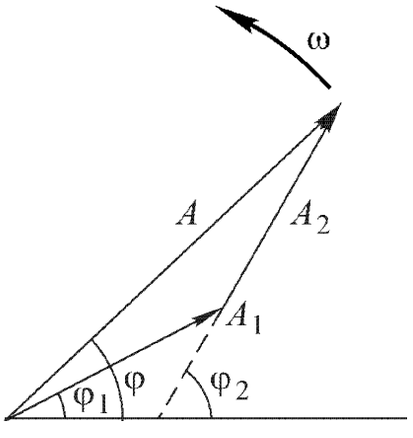


Рис. 111

Вводя интенсивности колебаний, результат (26.5) можно записать в виде

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos(\varphi_2 - \varphi_1). \tag{26.7}$$

Если колебания синфазны, т.е. фазы φ_1 и φ_2 одинаковы или отличаются на четное число π, то интенсивность I максимальна и равна

$$I_{\text{макс}} = \left(\sqrt{I_1} + \sqrt{I_2}\right)^2. \tag{26.8}$$

Если колебания противофазны, т.е. фазы φ_1 и φ_2 отличаются на нечетное число π, то получается минимальная интенсивность:

$$I_{\text{мин}} = \left(\sqrt{I_1} - \sqrt{I_2}\right)^2. \tag{26.9}$$

Если колебания совершаются в квадратуре, т.е. их фазы отличаются на $m\pi \pm \pi/2$ (m — целое число), то $I = I_1 + I_2$. В этом

случае интенсивность результирующего колебания равна сумме интенсивностей складываемых колебаний.

5. Не представляет труда написать интерференционный член и для общего случая, когда складываются *векторные* колебания, причем между декартовыми компонентами каждого вектора могут существовать произвольные разности фаз. Предоставляя это сделать читателю, заметим, что Френель и Араго обнаружили на опыте, что *две световые волны, распространяющиеся в одном направлении, никогда не интерферируют между собой, если они линейно поляризованы во взаимно перпендикулярных плоскостях.* На этом основании Френель пришел к заключению, что световые волны *поперечны.*

Покажем это, не вводя никаких специальных предположений о физической природе «светового вектора» $\mathbf{E}$, совершающего колебания в световой волне. Допустим, что вдоль оси Z распространяются две волны, плоскости колебаний которых взаимно перпендикулярны. В одной волне колебания происходят в координатной плоскости XZ, в другой — в координатной плоскости YZ. Представим световые векторы этих волн в виде $\mathbf{E}_1 = \mathbf{E}_{1x} + \mathbf{E}_{1z}$ и $\mathbf{E}_2 = \mathbf{E}_{2y} + \mathbf{E}_{2z}$, где координатные индексы указывают, какой из координатных осей параллелен соответствующий вектор. Перемножая скалярно и усредняя по времени, находим интерференционный член:

$$I_{12} = 2(\overline{\mathbf{E}_{1z}\mathbf{E}_{2z}}) = 2a_{1z}a_{2z}\cos(\varphi_{2z} - \varphi_{1z}),$$

где a_{1z}, a_{2z} — амплитуды, а φ_{1z}, φ_{2z} — фазы соответствующих продольных колебаний. Опыты Френеля и Араго показали, что интерференционный член обращается в нуль, каковы бы ни были фазы колебаний. Отсюда следует, что по крайней мере одна из амплитуд, например a_{1z}, равна нулю, т.е. первая волна поперечна. Но тогда и вторая волна должна быть поперечной, так как нет никаких оснований предпочесть первую волну второй.

6. Допустим теперь, что перекрываются две плоские волны:

$$E_1 = a_1\cos(\omega t - \mathbf{k}_1\mathbf{r} + \delta_1), \quad E_2 = a_2\cos(\omega t - \mathbf{k}_2\mathbf{r} + \delta_2). \quad (26.10)$$

Снова предположим, что векторы $\mathbf{E}_1$ и $\mathbf{E}_2$ параллельны или антипараллельны, так что от векторного характера колебаний можно отвлечься. Сравнивая эти выражения с (26.3), видим, что в рассматриваемом случае

$$\varphi_2 - \varphi_1 = \mathbf{K}\mathbf{r} + (\delta_2 - \delta_1), \quad (26.11)$$

где введен новый вектор $\mathbf{K} = \mathbf{k}_1 - \mathbf{k}_2$. Он параллелен биссектрисе угла, внешнего по отношению к углу α, который составляют между собой волновые векторы $\mathbf{k}_1$ и $\mathbf{k}_2$ (рис. 112). Поверхности равных разностей фаз $\varphi_2 - \varphi_1 = \text{const}$ суть параллельные плоскости, перпендикулярные к вектору $\mathbf{K}$. Они обозначены на

рис. 112 штриховыми прямыми. Вдоль каждой из этих плоскостей, следовательно, интенсивность результирующего колебания будет постоянна. Она максимальна, когда разность фаз $\varphi_2 - \varphi_1$ содержит π четное число раз, и минимальна, когда π содержится нечетное число раз. В частности, если складываются волны с одной и той же интенсивностью I, то интенсивность в максимуме будет $4I$, а в минимуме — нуль. Расстояние Δx между двумя соседними плоскостями максимальной или минимальной интенсивности найдется из условия $K\,\Delta x = 2\pi$. Так как длины обоих волновых векторов $\mathbf{k}_1$ и $\mathbf{k}_2$ одинаковы и равны $k = 2\pi/\lambda$, то $K = 2k\sin(\alpha/2)$, и следовательно,

Рис. 112

$$\Delta x = \frac{2\pi}{K} = \frac{\pi}{k\sin(\alpha/2)} = \frac{\lambda}{2\sin(\alpha/2)}. \tag{26.12}$$

Для малых углов α

$$\Delta x \approx \lambda/\alpha. \tag{26.13}$$

Если поставить плоский экран, то он пересечет плоскости равной интенсивности вдоль параллельных прямых; на экране появятся светлые и темные «интерференционные полосы». Расстояние между серединами соседних светлых или темных полос называется *шириной интерференционной полосы.* Если плоскость экрана параллельна плоскости $(\mathbf{k}_1, \mathbf{k}_2)$, в которой лежат волновые векторы $\mathbf{k}_1$ и $\mathbf{k}_2$, то ширина интерференционной полосы равна Δx, т.е. определяется выражением (26.12). То же самое получится, если экран установлен в перпендикулярной плоскости перпендикулярно к биссектрисе угла между волновыми векторами $\mathbf{k}_1$ и $\mathbf{k}_2$. Если же, оставляя экран перпендикулярным к плоскости $(\mathbf{k}_1, \mathbf{k}_2)$, повернуть его на угол φ, то ширина интерференционной полосы сделается равной $\Delta_\varphi x = \Delta x/\cos\varphi$.

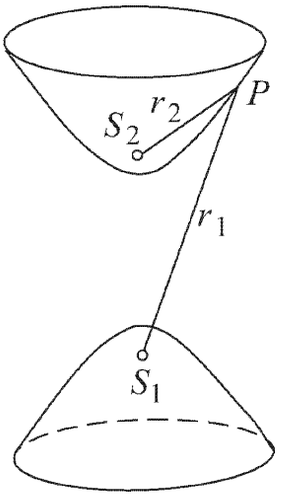


Рис. 113

7. Разберем теперь случай, когда перекрываются *сферические монохроматические волны* от двух точечных источников света S_1 и S_2 (рис. 113). В этом случае амплитуды a_1 и a_2 складываемых колебаний обратно пропорциональны расстояниям r_1 и r_2 от точки наблюдения до источников S_1 и S_2. Поэтому интенсивность света будет меняться вдоль каждой интерференционной полосы. Однако это изменение медленное, и от него можно отвлечься. Основное значение для результи-

рующей интенсивности имеет разность фаз складываемых колебаний. Поверхности равных разностей фаз $r_1 - r_2 = \text{const}$ будут двухполостными гиперболоидами вращения с фокусами S_1 и S_2. На экране, перпендикулярном к линии источников S_1S_2, получатся интерференционные полосы в виде концентрических колец с центром в точке пересечения экрана с указанной линией. Если же экран параллелен линии источников S_1S_2, то интерференционные полосы будут гиперболами с фокусами S_1 и S_2. В последнем случае в небольшой центральной части интерференционная картина практически будет состоять из равноотстоящих параллельных светлых и темных полос.

Если фазы колебаний обоих источников света одинаковы, то разность фаз складываемых колебаний будет

$$\Delta\varphi = k(r_2 - r_1) = (2\pi/\lambda)(r_2 - r_1).$$

Когда она равна $2m\pi$ (m — целое число, положительное или отрицательное), то получается максимум интенсивности колебаний (светлая интерференционная полоса). Если же $\Delta\varphi = 2\pi(m+{}^1\!/_2)$, то интерференционная полоса будет темной. Условия максимума и минимума можно также записать в виде

$$\Delta r \equiv r_2 - r_1 = \begin{cases} m\lambda & \text{(светлая полоса)}, \\ (m + {}^1\!/_2)\lambda & \text{(темная полоса)}. \end{cases} \tag{26.14}$$

Величина $r_2 - r_1$ называется *разностью хода* интерферирующих лучей. Если интерферирующие лучи проходят через среды с различными показателями преломления, то величину $r_2 - r_1$ надо заменить на

$$\Delta = \int n_2\, dl - \int n_1\, dl,$$

т.е. на *оптическую разность хода* интерферирующих лучей. Для светлых интерференционных полос оптическая разность хода составляет целое, для темных — полуцелое число длин волн (в вакууме). Целое число m называется *порядком интерференции*. Порядок интерференции есть округленная до целого числа оптическая разность хода интерферирующих лучей, выраженная в длинах волн (в вакууме).

8. Основные результаты, найденные выше, можно также получить на следующем простом примере. Пусть S_1 и S_2 — близко расположенные точечные монохроматические источники света (рис. 114). Для увеличения интенсивности интерференционных полос вместо них можно взять два коротких линейных источника, например две узкие ярко освещенные щели. Плоскость экрана $Э$ предполагается параллельной плоскости, в которой лежат линейные источники S_1 и S_2. Пусть CO — перпендикуляр к этой плоскости, проходящий через середину между источниками S_1

и S_2, D — длина этого перпендикуляра, d — расстояние между источниками S_1 и S_2. Предполагается, что не только расстояние d, но также длины источников и линейные размеры экрана малы по сравнению с расстоянием D. Тогда интерференционные полосы на экране будут прямолинейны и перпендикулярны к линии, соединяющей источники S_1 и S_2. Начало координат поместим в точке O на экране, ось X направим параллельно линии источников S_1S_2. Если x — абсцисса точки наблюдения A, то

$$r_1^2 = D^2 + (x + d/2)^2, \quad r_2^2 = D^2 + (x - d/2)^2,$$

так что $r_1^2 - r_2^2 = 2xd$, и следовательно, $r_1 - r_2 = 2xd/(r_1 + r_2)$. Так как $x \ll D$, то без существенной ошибки знаменатель $r_1 + r_2$ можно заменить на $2D$. Тогда

$$r_1 - r_2 = \frac{xd}{D} = \alpha x \ ^{1}), \tag{26.15}$$

где $\alpha \approx d/D$ — угол схождения интерферирующих лучей, т.е. угол, под которым из точки O (или в рассматриваемом приближении из любой точки экрана) видно расстояние d между источниками S_1 и S_2. Интенсивность света найдется по формулам (26.7) и (26.11). Если источники S_1 и S_2 одинаковы и синфазны, то эти формулы дают

$$I = 2I_1\left(1 + \cos\frac{2\pi\alpha x}{\lambda}\right). \tag{26.16}$$

Интенсивность I периодически меняется вдоль оси X от нуля до максимального значения $I_{\text{макс}} = 4I_1$. Пространственный период изменения интенсивности $\Delta x = \lambda/\alpha$ есть ширина интерференционной полосы.

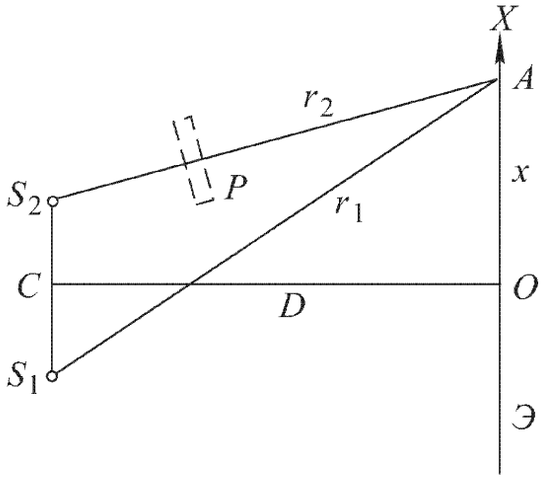


Рис. 114

Если на пути одного из лучей, например S_2A, ввести прозрачную плоскопараллельную пластинку P с толщиной l и показателем преломления n, то оптическая длина этого луча увеличится на $(n-1)l$, а разность хода между лучами S_1A и S_2A уменьшится на такую же величину. Прежнее значение разности хода получится в какой-то другой точке A', отстоящей от S_1 и S_2 на

---

[1]) Если пренебречь членами третьей степени по d, то $r_1 - r_2 = xd/\sqrt{D^2 + x^2}$. Пренебрегая далее пятыми степенями x/D, отсюда получим

$$r_1 - r_2 = \frac{xd}{D}\left(1 - \frac{x^2}{2D^2}\right) \approx \alpha x\left(1 - \frac{x^2}{2D^2}\right). \tag{26.15а}$$

расстояниях r'_1 и r'_2. Положение точки A' найдется из условия

$$r'_1 - [r'_2 + (n-1)l] = r_1 - r_2, \quad \text{или} \quad (r'_1 - r'_2) = (r_1 - r_2) + (n-1)l.$$

Это значит, что произойдет смещение всей интерференционной картины на $N = (n-1)l/\Delta x$ полос в ту сторону, с какой была введена пластинка P. На этом основаны интерференционные методы измерения малых изменений показателя преломления, обладающие высокой чувствительностью. Для таких измерений надо пользоваться белым светом (см. § 30). В монохроматическом свете все полосы одинаковы, и их смещение трудно измерить — для этого толщину пластинки l надо было бы увеличивать непрерывно от нуля до окончательной величины.

9. Наличие интерференционных максимумов и минимумов в волновом поле, конечно, не может привести к противоречию с принципом сохранения энергии. Обсудим этот вопрос для только что исследованного случая двух одинаковых синфазных монохроматических точечных источников S_1 и S_2.

Максимумы интенсивности находятся на гиперболах $r_1 - r_2 = m\lambda$, минимумы — на гиперболах $r_1 - r_2 = (m + {}^1\!/_2)\lambda$. Максимальное значение, которое может принимать разность $r_1 - r_2$, равно расстоянию между источниками d, а максимальное значение m — целой части, содержащейся в числе d/λ. Когда $d \gg \lambda$, число интерференционных полос в волновом поле велико. В этом случае интерференция приводит только к *пространственному перераспределению* лучистой энергии, в результате чего в одних местах пространства (в максимумах) плотность лучистой энергии увеличивается, в других (в минимумах) уменьшается, а общий запас энергии в пространстве остается неизменным.

Полный поток лучистой энергии через любую замкнутую поверхность, окружающую источники S_1 и S_2, равен сумме потоков через ту же поверхность, которые излучались бы каждым из этих источников в отсутствие другого. Однако это равенство *не точное, а только приближенное* и в среднем выполняется тем точнее, чем больше d/λ. Когда $d < \lambda$, в волновом поле уже нет линии нулевой интенсивности. Когда же $d \ll \lambda$, то во всех точках пространства складываются колебания, фазы которых практически одинаковы. В этом случае интенсивность результирующего колебания, а с ним и поток результирующего излучения, исходящий от обоих источников, в четыре раза больше соответствующих величин для одного из источников в отсутствие другого. Таким образом, если расстояние между источниками S_1 и S_2 меньше длины волны λ, то поток лучистой энергии через замкнутую поверхность, окружающую эти источники, *больше суммы потоков*, которые излучали бы те же, но уединенные источники.

Однако принцип сохранения энергии и не требует равенства этих величин. Источники вместе действительно больше излучают энергии, чем в том случае, когда они находятся далеко друг от друга. Но это увеличение излучения происходит не за счет нарушения сохранения энергии, а *за счет работы генератора*, который должен поддерживать постоянными амплитуды колебаний в источниках. Если же колебания в источниках свободные, то увеличение излучения приводит просто к более быстрому затуханию этих колебаний.

Рассмотренный случай трудно реализовать в оптическом диапазоне спектра из-за малости длин световых волн. Но в области радиоволн это не только возможно, но и практически используется для получения *направленных излучений и увеличения их мощности*. Конечно, возможен и такой случай, когда совместное действие двух источников приводит не к увеличению, а *к уменьшению* общего излучения. Это будет, например, когда колебания в источниках происходят в противоположных фазах.

10. Строго монохроматические волны одной и той же частоты, даже от двух различных точечных источников, *всегда интерферируют между собой*, т.е. когерентны. Интерференционная картина от таких источников *устойчива* в том смысле, что распределение интенсивности волнового поля в пространстве остается *неизменным во времени*. Интерференция излучений от *независимых источников света*, например от электрических лампочек, недоступна наблюдению глазом. Глаз не наблюдает интерференцию даже при наложении излучений от двух независимых самых узких спектральных линий, излучаемых разреженными газами. Это служит лучшим доказательством того, что излучения реальных источников света никогда не бывают строго монохроматичными.

Невозможность визуального наблюдения интерференционных полос от независимых источников света можно пояснить на примере идеализированных источников, излучающих *квазимонохроматический свет*. Такой свет представляется колебаниями вида (26.3), в которых, однако, амплитуды a_1, a_2 и фазы φ_1, φ_2 медленно и хаотически меняются во времени, т.е. испытывают заметные изменения за времена, очень большие по сравнению с периодом T самих световых колебаний. Примером может служить излучение изолированного атома. Возбужденный атом испускает ряд или, как принято говорить, *цуг волн* в течение времени $\tau_{\text{изл}}$, характерная длительность которого порядка 10^{-8} с (см. § 89). В таком цуге содержится 10^6–10^8 волн. За время $\tau_{\text{изл}}$ атом «высвечивается» и переходит в невозбужденное состояние. В результате различных процессов, например столк-

новений с другими атомами или ударов электронов, атом может снова вернуться в возбужденное состояние, а затем начать излучать новый цуг волн. Таким образом, получится последовательность цугов, испускаемых атомом через малые и нерегулярно меняющиеся промежутки времени. Пусть теперь на экран попадают излучения от двух независимых атомов. При наложении двух цугов, излучаемых этими атомами, на экране получится какая-то картина интерференционных полос. Положение полос определяется разностью фаз между колебаниями обоих цугов. А такая разность фаз быстро и беспорядочно меняется от одной пары цугов к следующей. В течение секунды десятки и сотни миллионов раз или чаще одна система интерференционных полос будет сменяться другой. Глаз или другой приемник света не в состоянии следить за этой быстрой сменой интерференционных картин и фиксирует только равномерную освещенность экрана.

Смещение интерференционных полос при замене одной пары цугов другой происходить не будет, если у этих цугов начальные фазы δ_1 и δ_2 хотя и меняются нерегулярно во времени, но *одинаково*. Действительно, в этом случае $\delta_2 - \delta_1 = 0$ и, как видно из (26.11), разность фаз складываемых колебаний $\varphi_2 - \varphi_1$ содержит только слагаемое $\mathbf{Kr}$, *регулярно* меняющееся в пространстве. Чтобы этого достигнуть, надо излучение от *одного и того же источника* расщепить на два или несколько пучков и заставить их попадать на экран *различными путями*. Если при этом будут накладываться пучки, полученные из *одного и того же цуга*, то интерференционные полосы на экране будут сохранять свое положение и глаз зафиксирует устойчивую (неподвижную) интерференционную картину. Некоторые опыты, основанные на таком принципе, будут описаны в следующем параграфе.

11. Отметим, что *нет принципиальных запретов* на возможность получения интерференции от двух *независимых* источников света. Такая возможность определяется *уровнем техники генерирования и приема излучения*. Еще в 1947 г. Г.С. Горелик (1906–1957) предлагал осуществить *гетеродинирование*, или *демодуляцию* видимого света. В видимом свете можно выделить дублеты или мультиплеты, т.е. сложные спектральные линии, состоящие из двух или нескольких одиночных спектральных линий с разностями частот $\nu_1 - \nu_2$ порядка 10^9–10^{10} Гц. По идее Горелика такой свет должен падать на катод фотоэлемента (а еще лучше фотоумножителя). Если верно предположение, что мгновенный фототок пропорционален квадрату напряженности результирующего электрического поля, то его можно представить в виде

$$\mathscr{I} = A(\mathbf{E}_1 + \mathbf{E}_2)^2 = A(\mathbf{E}_1^2 + \mathbf{E}_2^2 + 2\mathbf{E}_1\mathbf{E}_2),$$

где $\mathbf{E}_1$ и $\mathbf{E}_2$ — напряженности электрического поля, соответствующие частотам ν_1 и ν_2, а A — постоянная прибора. В этом выражении слагаемое $2A(\mathbf{E}_1\mathbf{E}_2)$ содержит составляющую с разностной частотой $\nu_1 - \nu_2$. Частоте $\nu_1 - \nu_2 \sim 10^{10}$ Гц соответствует длина волны $\lambda \sim 3$ см, лежащая в микроволновой области радиоволн. Ее можно усилить радиотехническими средствами и сделать доступной наблюдению. Тем самым биения интенсивности, возникающие при наложении пучков света от двух источников, превратятся в колебания электрического тока.

Для осуществления опыта в принципе безразлично, испускаются ли волны с частотами ν_1 и ν_2 одним источником или двумя независимыми источниками света. Трудность опыта состоит в том, что на выделяемый сигнал разностной частоты $\nu_1 - \nu_2$ накладывается «шум», интенсивность которого превосходит интенсивность сигнала примерно в 10^4 раз. Эта трудность была преодолена в 1955 г. Форрестором и его сотрудниками. Источником света служила яркая одиночная зеленая линия ртути $\lambda = 546{,}1$ нм, расщеплявшаяся в магнитном поле (эффект Зеемана). Применялся фотоэлемент с сурьмяно-цезиевым катодом. Положительный результат опыта доказывает, что если и существует запаздывание выхода электронов с поверхности фотокатода по сравнению с попаданием на нее фотонов, то оно значительно меньше 10^{-10} с.

С изобретением лазеров техника эксперимента получила источники света столь высокой степени монохроматичности, что наблюдение интерференционных максимумов и минимумов от двух независимых лазеров с помощью надлежащих приемников света (например, тех же фотоэлементов) стало сравнительно легко осуществимой экспериментальной работой.

§ 27. Классические интерференционные опыты

1. О п ы т Ю н г а. Юнг получал полосы интерференции по способу, описанному им в публичных лекциях 1807 г. Яркий пучок света от Солнца падал на экран с малым отверстием или узкой щелью S (рис. 115). Дифрагированный свет шел ко второму экрану с двумя узкими отверстиями или щелями S_1 и S_2. На этих щелях свет также претерпевал дифракцию, в результате чего получались два перекрывающихся расходящихся пучка света с вершинами в S_1 и S_2. Ввиду общности происхождения эти пучки когерентны. На экране в месте перекрытия пучков наблюдались параллельные интерференционные полосы (см. предыдущий параграф, п. 8). Расстояние между щелями S_1 и S_2 должно быть велико по сравнению с шириной каждой щели.

Удобный способ получения нужных щелей был указан Рэлеем. Стеклянные пластинки покрываются тонким слоем серебра и тем самым становятся непрозрачными. На серебряном слое одной из пластинок лезвием бритвы проводится одна, а на другой две близко расположенные параллельные линии, которые и используются в качестве щелей.

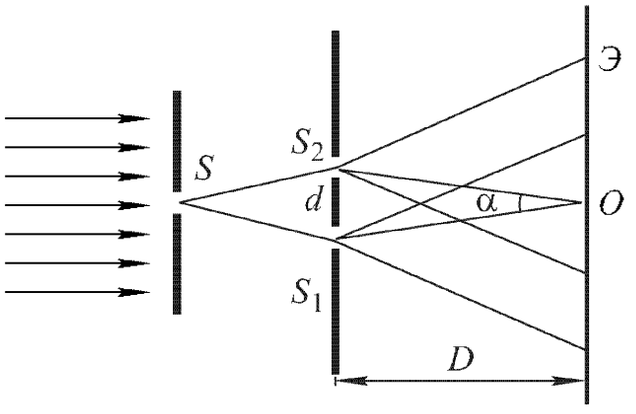


Рис. 115

Оценим по формуле (26.13) ширину Δx интерференционной полосы. Допустим, что расстояние d между щелями S_1 и S_2 составляет 1 мм, а расстояние от щелей до экрана $D = 1$ м. Тогда $\alpha = d/D = 0{,}001$ рад. Для красного света ($\lambda = 600$ нм) получаем $\Delta x = \lambda/\alpha = 6 \cdot 10^5$ нм $= 0{,}6$ мм. В синем свете ширина полосы будет порядка 0,4 мм. Таким путем впервые Юнг измерил длины световых волн, хотя эти измерения и не могли быть точными. В опыте Юнга из-за дифракции свет распространяется от щелей S_1 и S_2 в различных направлениях неравномерно. Влияние этого осложняющего обстоятельства будет выяснено в гл. IV.

Задолго до Юнга, в 1665 г., аналогичный опыт был поставлен Гримальди. Однако в опыте Гримальди свет от Солнца падал непосредственно на щели S_1 и S_2. Дополнительной щели S не было. При такой постановке опыта интерференционные полосы получаться не могли, ввиду значительных угловых размеров Солнца (см. § 28, п. 7).

При использовании лазеров, генерирующих практически параллельные пучки лучей, щель S в опыте Юнга не нужна.

2. З е р к а л а Ф р е н е л я. В 1816 г. Френель осуществил следующий интерференционный опыт. Свет от узкой ярко освещенной щели S (рис. 116) падал на два плоских зеркала CD и CE, наклоненных друг к другу под углом, близким к 180°. Щель S устанавливалась параллельно линии C пересечения плоскостей зеркал. При отражении падающий пучок разделялся на два когерентных пучка, как бы исходивших от прямых S_1 и S_2, являющихся мнимыми изображениями щели S в зеркалах. Прямой свет от источника S загораживался непрозрачной ширмой MN. На экран $Э$ попадали только отраженные пучки. В области AB, где пучки перекрывались, наблюдались параллельные интерференционные полосы. Для расчета ширины интерференционной полосы Δx введем следующие обозначения:

$CO = a$, $CS = b$, $\angle HCD = \varphi$. Так как точки S, S_1 и S_2 лежат на одной окружности радиуса b, то, предполагая угол φ малым, можем написать $d = S_1S_2 = 2b\varphi$. Угол α, под которым из точки O видно расстояние S_1S_2, равен $\alpha = d/(a+b) = 2b\varphi/(a+b)$, а потому

$$\Delta x = \frac{\lambda}{\alpha} = \frac{a+b}{2b\varphi}\lambda. \tag{27.1}$$

Угол α можно просто измерить по шкале зрительной трубы. Для этого трубу следует поместить в точке O и установить ее на отчетливое видение изображений S_1 и S_2 щели S. Тогда можно

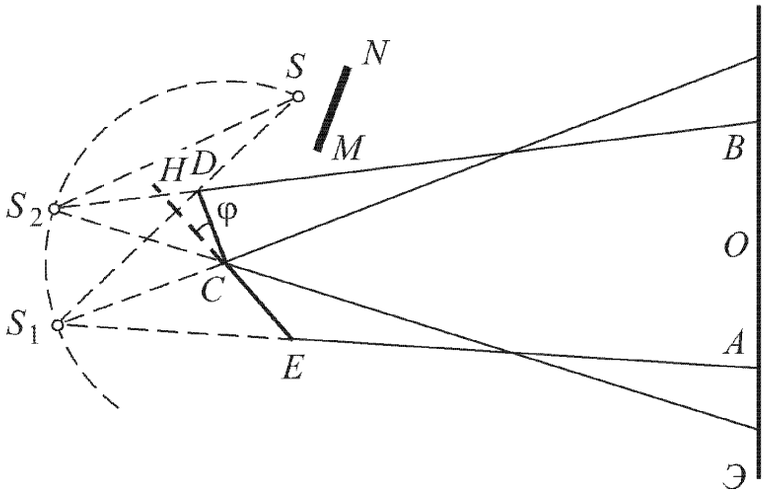


Рис. 116

найти λ по формуле $\lambda = \alpha\,\Delta x$. Ширина области перекрытия $AB = 2a\varphi$, а потому число полос, которые могут наблюдаться на экране, равно

$$N = \frac{2a\varphi}{\Delta x} = \frac{2a\varphi\alpha}{\lambda} = \frac{4ab}{\lambda(a+b)}\varphi^2. \tag{27.2}$$

В опыте Френеля интерференционная картина искажается дифракцией на ребре C, вдоль которого пересекаются плоскости зеркал.

3. Бипризма Френеля. Бипризма Френеля состоит из двух стеклянных призм с малыми преломляющими углами, сложенных своими основаниями (рис. 117). Практически она изготовляется из целого куска стекла. Источником света служит ярко освещенная щель S, устанавливаемая параллельно ребру бипризмы. После преломления в бипризме падающий пучок света разделяется на два когерентных пучка с вершинами в мнимых изображениях S_1 и S_2 щели S. В области AB экрана пучки перекрываются и дают систему параллельных интерференционных полос. Пусть β — преломляющий угол бипризмы, n — ее показатель преломления, a и b — длины отрезков CO и SC. Каждая половина бипризмы отклоняет параксиальный

луч на угол $(n-1)\beta$. Расстояние d между изображениями S_1 и S_2 равно $d = S_1S_2 = 2b(n-1)\beta$, а угловое расстояние между ними $\alpha = d/(a+b)$. Ширина полосы

$$\Delta x = \frac{\lambda}{\alpha} = \frac{\lambda(a+b)}{2b(n-1)\beta}. \qquad (27.3)$$

Искажения интерференционной картины, вносимые дифракцией на ребре C бипризмы, более существенны, чем в опыте с зеркалами Френеля. Однако осуществить опыт с бипризмой значительно легче.

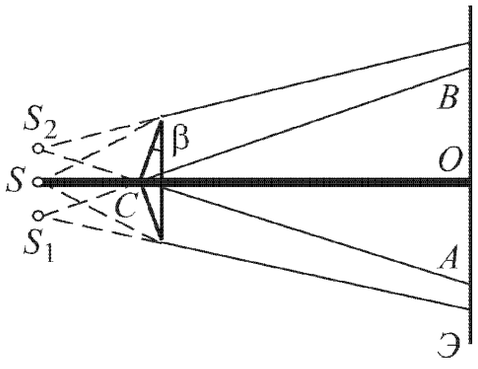


Рис. 117

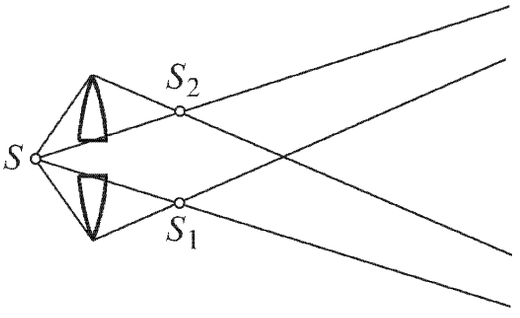


Рис. 118

4. Б и л и н з а Б и й е. Собирательная линза разрезается пополам, и обе половинки ее раздвигаются. Образовавшаяся система называется *билинзой*. На билинзу направляется свет от щели S, параллельной плоскости разреза (рис. 118). В S_1 и S_2 получаются действительные изображения щели S. Пучки света, проходящие через S_1 и S_2, перекрываются. В местах перекрытия наблюдается интерференция. Если щель S поместить между билинзой и ее фокальной плоскостью, то изображения S_1 и S_2 будут мнимыми. В этом случае интерференция невозможна, так как световые пучки не перекрываются. Чтобы получить интерференцию и при таком расположении щели S, надо из центральной части линзы вырезать плоскопараллельный кусок стекла и обе половинки линзы сблизить между собой. Тогда пучки будут перекрываться, и интерференция станет возможной.

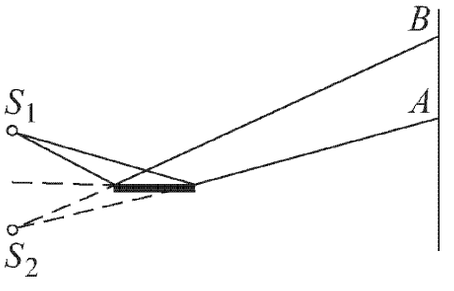


Рис. 119

5. З е р к а л о Л л о й д а (1800–1881). Свет, исходящий из узкой ярко освещенной щели S_1 (рис. 119), отражается от полированной плоской поверхности черного стекла. На участке AB экрана, где перекрываются прямой и отраженный свет, наблюдаются полосы интерференции.

6. Во всех описанных опытах полосы интерференции можно наблюдать на белом матовом экране. Можно пользоваться также матовым стеклом, наблюдая полосы интерференции на его задней стороне через лупу или микроскоп малого увеличения. Такой *объективный* способ наблюдения применял и Френель в первых опытах. Но он заметил, что полосы интерференции не исчезали, а становились более яркими, когда матовое стекло убиралось. Поэтому в дальнейших исследованиях Френель перешел к *субъективному* методу наблюдения интерференционных (и дифракционных) полос. Идея, на которой основан субъективный способ наблюдения полос, весьма проста. Допустим, что интерференционные полосы получены на задней плоскости П матового стекла. Интенсивность света в какой-либо точке P плоскости П определяется разностью фаз интерферирующих лучей, приходящих в эту точку. Отобразим теперь с помощью линзы плоскость П на сопряженную плоскость П′. Лучи, вышедшие из точки P, снова соберутся теперь в точке P'. Но так как оптические длины всех лучей между сопряженными точками одинаковы, то интерферирующие лучи придут в точку P' *с той же разностью фаз*, какой они обладали в точке P. Поэтому в плоскости П′ получится изображение не только плоскости П, но и всей системы интерференционных полос на ней. Линза как бы переносит интерференционную картину с плоскости П на плоскость П′. Так же действует и оптическая система глаза, с помощью которой интерференционные полосы получаются на сетчатке. Если пользоваться зрительной трубой, установленной на бесконечность, то таким путем будет наблюдаться интерференционная картина, получающаяся на бесконечном (т.е. очень далеком) расстоянии от глаза.

7. Интерференционный опыт Поля (1884–1976). Описанные выше установки отличаются малой светосилой, а потому мало пригодны для демонстрации интерференции света в большой аудитории. Опыт Поля свободен от этого недостатка. Здесь источник света S (рис. 120), которым может служить небольшая, но достаточно яркая ртутная лампа, помещается перед пластинкой слюды, толщина которой порядка нескольких сотых долей миллиметра. Свет отражается от передней и задней поверхностей пластинки с образованием двух мнимых изображений S_1 и S_2 источника S. Прямой свет от источника S загораживается ширмой. Получаются два широких когерентных пучка света с вершинами в точках S_1 и S_2. Падая на экран, потолок или

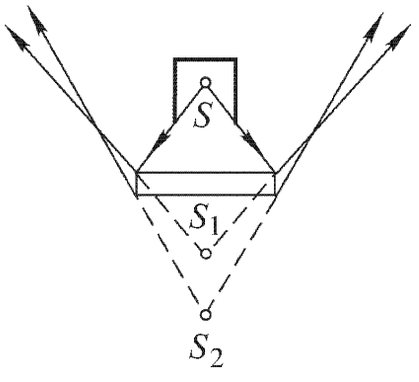


Рис. 120

стены аудитории, они дают при интерференции систему светлых и темных колец (см. п. 6 предыдущего параграфа). Благодаря большому количеству света кольца хорошо видны во всей аудитории. Преимущества метода Поля будут выяснены в § 28.

8. О п ы т М е с л и н а. Этот опыт является видоизменением опыта Бийе. Только в нем половинки разрезанной линзы не раздвигаются перпендикулярно к оптической оси, а смещаются вдоль нее, и притом на значительное расстояние (рис. 121). Половинки билинзы дают действительные изображения точечного источника S в точках S_1 и S_2 на главной оптической оси. Интерференционные полосы наблюдаются в заштрихованной области, где накладываются соответствующие пучки света, т.е. между изображениями S_1 и S_2. В плоскости, перпендикулярной к главной оптической оси, интерференционные полосы имеют вид концентрических полуколец с общим центром на этой оси.

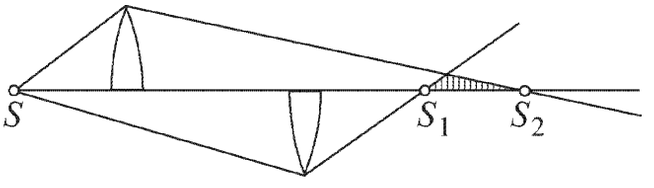


Рис. 121

Интересно, что центр колец *темный*. Явление связано с тем, что один пучок, прежде чем интерферировать со вторым, проходит через свой фокус S_1, тогда как другой пучок проходит через свой фокус S_2 позже. На этом основании можно прийти к заключению, что прохождение волны через фокус сопровождается изменением фазы на π.

Это легко понять на примере скалярной сферической волны, сходящейся к своему центру O и заполняющей полный телесный угол 4π. Представим поле этой волны в виде $E_1 = \frac{1}{r} f_1\left(t + \frac{r}{v}\right)$, где r — расстояние от центра O, а v — скорость распространения волны. Когда волна проходит через центр O, начинает образовываться расходящаяся сферическая волна $E_2 = \frac{1}{r} f_2\left(t - \frac{r}{v}\right)$. Волновое поле в пространстве теперь представится в виде

$$E = \frac{1}{r} f_1\left(t + \frac{r}{v}\right) + \frac{1}{r} f_2\left(t - \frac{r}{v}\right).$$

Но напряженность поля E нигде не может обращаться в бесконечность, в частности и при $r = 0$. Поэтому необходимо, чтобы в любой момент времени выполнялось соотношение $f_2(t) = -f_1(t)$, т.е. функция f_2 может отличаться от функции f_1 только знаком. Если колебания f_1 и f_2 синусоидальны, то в точке O их фазы будут противоположны. А это и значит, что при прохождении через фокус фаза волны меняется на π.

Из приведенного рассуждения следует, что на передовом фронте функция f_1 должна оставаться непрерывной. Если бы здесь она испытывала разрыв, то при подходе передового фронта к центру O напряженность поля E возрастала бы неограниченно. То же самое относится и к заднему фронту волны, а также к функции f_2.

ЗАДАЧИ

1. Как надо поместить собирательную линзу в опыте с бипризмой (см. рис. 117), чтобы ширина интерференционной полосы Δx не зависела от положения экрана? Найти Δx. Найти максимальное число N полос, которое может наблюдаться в этой установке, если оно получается при удалении экрана от бипризмы на расстояние L.

Ответ. $\Delta x = \dfrac{\lambda}{2(n-1)\beta}$; $N = \dfrac{4L(n-1)^2\beta^2}{\lambda}$.

2. При каком расстоянии L экрана от бипризмы в предыдущей задаче число интерференционных полос N будет максимальным, если расстояние между вершинами бипризмы равно l? Найти N. При каком положении экрана полосы пропадут?

Ответ. $N = l(n-1)\beta/\lambda$; $L = \dfrac{l}{4(n-1)\beta}$. Если экран удалить от бипризмы на расстояние L или больше, то полосы пропадут.

3. Найти число полос интерференции N в установке с бипризмой (см. рис. 117).

Ответ. $N = \dfrac{4ab}{a+b}\dfrac{(n-1)^2\beta^2}{\lambda}$.

4. В опыте с бипризмой Френеля экран для наблюдения интерференционных полос расположен перпендикулярно к оси установки. Нулевая полоса получается в центре экрана в вертикальной плоскости, проходящей через ребро бипризмы и ось установки. В какую сторону и на какую величину x сместится нулевая полоса, если освещающую щель немного сместить в сторону на величину h перпендикулярно к оси установки?

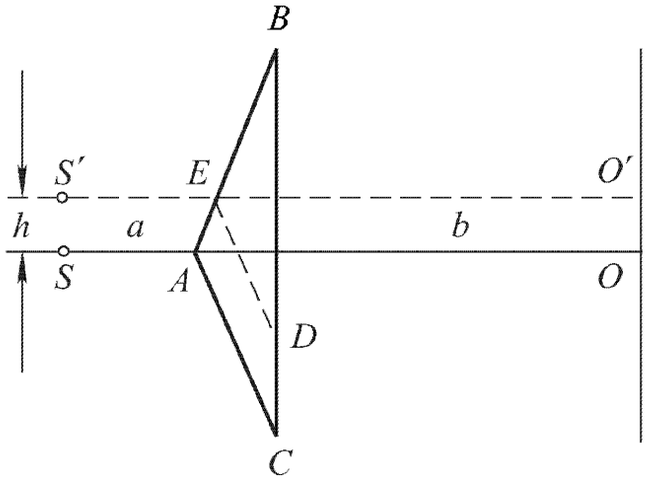


Рис. 122

Решение. При смещении источника из положения S в положение S' (рис. 122) действие бипризмы CAB сведется к действию ее части DEB и плоскопараллельной пластинки $CAED$. Суммируя смещения, вызываемые этими частями в отдельности, найдем, что центр интерференционной картины из прежнего положения O сместится вверх на расстояние $x = hb/a$.

§ 28. Влияние размеров источника света. Пространственная когерентность

1. Увеличение размеров источника света приводит к ухудшению *контрастности* интерференционных полос и даже к их полному исчезновению. Допустим сначала, что источник состоит из двух одинаковых некогерентных светящихся точек A и B, находящихся на малом расстоянии l друг от друга (рис. 123). Как было выяснено в § 26 (п. 10), для получения интерференции от одного источника A надо расщепить свет, исходящий от этого источника, на два пучка, попадающих на экран *различными путями.* То же самое надо сделать со светом от источника B. Конкретный способ расщепления пучка лучей на два в последующих рассуждениях не имеет никакого значения. На рис. 123 один пучок света попадает к месту интерференции P непосредственно от источников A и B, другой — после отражения от зеркала M_2N_2. Последующие рассуждения применимы и к криволинейным лучам. Однако, во избежание ненужных усложнений в написании формул, показатели преломления сред в точках A и B, а также в месте нахождения экрана, на котором наблюдается интерференция, предполагаются равными единице.

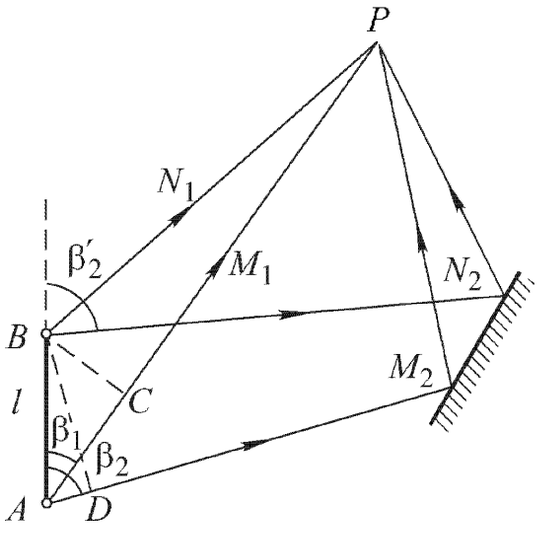


Рис. 123

Пусть AM_1P и AM_2P — два луча, исходящие из точки A под углами β_1 и β_2 к линии AB и сходящиеся в точке P, где они и интерферируют. Аналогичные лучи, исходящие из точки B, пусть будут BN_1P и BN_2P. Если расстояние $l = AB$ достаточно мало, то для разности оптических длин лучей AM_1P и BN_1P можно написать: $(AM_1P) - (BN_1P) = (AC) = l\cos\beta_1$. Аналогично для AM_2P и BN_2P: $(AM_2P) - (BN_2P) = (AD) = l\cos\beta_2$. Следовательно,

$$[(AM_1P) - (AM_2P)] - [(BN_1P) - (BN_2P)] = l(\cos\beta_1 - \cos\beta_2).$$

В первой квадратной скобке стоит оптическая разность хода лучей, исходящих из точки A. Она определяет характер интерференции этих лучей в точке P: максимум, минимум или какой-либо промежуточный случай. Аналогично, во второй квадратной скобке стоит оптическая разность хода лучей, исходящих из точки B, от которой зависит характер интерференции этих

лучей. Разность этих величин

$$\Delta = l|\cos\beta_1 - \cos\beta_2|$$

характеризует фазовый сдвиг интерференционной картины от точечного источника A относительно интерференционной картины от точечного источника B. Она определяет результат наложения одной из этих картин на другую. Если величина Δ равна нулю или мала по сравнению с длиной волны λ, то максимумы одной картины накладываются на максимумы другой, а минимумы — на минимумы. Тогда происходит усиление интерференционных картин, а полосы интерференции наиболее контрастны: наблюдается плавный переход от яркой освещенности в максимумах к полной темноте в минимумах. При возрастании Δ контрастность полос сначала ухудшается. Когда $\Delta = \lambda/2$, т.е.

$$l|\cos\beta_1 - \cos\beta_2| = \lambda/2, \tag{28.1}$$

максимумы одной картины накладываются на минимумы другой, и полосы интерференции исчезают. При дальнейшем возрастании Δ полосы интерференции появляются вновь, причем интерференционная картина периодически повторяется от наибольшей контрастности до их полного исчезновения. При

$$\Delta \equiv l|\cos\beta_1 - \cos\beta_2| = m\lambda \tag{28.2}$$

(m — целое число) полосы столь же контрастны, что и при $\Delta = 0$. При

$$\Delta \equiv l|\cos\beta_1 - \cos\beta_2| = \left(m + \frac{1}{2}\right)\lambda \tag{28.3}$$

полосы пропадают, а экран освещается равномерно. Если Δ не превышает примерно $\lambda/4$, то контрастность полос лишь немного уступает контрастности при $\Delta = 0$. То же самое будет при $\Delta = (m+\alpha)\lambda$, где α — правильная дробь, не превышающая примерно $1/4$. Это несколько произвольное условие в дальнейшем мы будем рассматривать как *условие хорошей контрастности интерференционных полос*. Заметим еще, что в пределах принятой точности расчета угол β_2 можно отождествить с углом β_2', что и делается в дальнейшем без особых оговорок.

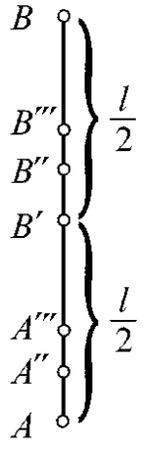


Рис. 124

2. Допустим теперь, что источник света протяженный и имеет форму равномерно светящейся линии AB длины l, все точки которой излучают свет некогерентно (рис. 124). Разобьем ее мысленно на бесконечное множество пар некогерентных точечных источников (A, B'), (A'', B''), ..., находящихся на расстоянии $l/2$ друг от друга. Применим к таким парам точечных источников результаты, полученные выше. Для этого надо расстояние l заменить на $l/2$. Если $\frac{\Delta}{2} = \frac{l}{2}|\cos\beta_1 - \cos\beta_2| = \frac{\lambda}{2}$, то

интерференционных полос от каждой пары точечных источников не получится, они создадут на экране только светлый фон. То же самое будет, если размеры источника $l = AB$ увеличить в 2, 3, ... раз. Итак, если выполнено условие

$$\Delta \equiv l|\cos\beta_1 - \cos\beta_2| = m\lambda, \tag{28.4}$$

то протяженный источник создает на экране только освещенный фон без интерференционных полос (в противоположность тому, что было в случае пары точечных источников, когда при том же условии получалась наилучшая контрастность полос). Если $\Delta = (m + \alpha)\lambda$, где α — правильная дробь, то протяженный источник можно мысленно разбить на две части, длины которых относятся как $m : \alpha$. Меньшая часть источника даст интерференционные полосы на светлом фоне, создаваемом большей частью того же источника. От этого контрастность полос уменьшается. Чем больше m, тем хуже контрастность. При больших m интерференционные полосы практически наблюдаться не будут. Условие хорошей контрастности интерференционных полос в случае протяженного источника света можно ориентировочно представить в виде

$$l|\cos\beta_1 - \cos\beta_2| \leqslant \lambda/2. \tag{28.5}$$

Если крайние интерферирующие лучи выходят из какой-либо точки протяженного источника симметрично по отношению к перпендикуляру к линии AB, проведенному из этой точки, т.е. $\beta_2 = \pi - \beta_1$, то $\cos\beta_2 = -\cos\beta_1$, и предыдущее условие можно представить в виде

$$l\sin(\Omega/2) \leqslant \lambda/4, \tag{28.6}$$

где Ω — угол между этими крайними лучами. Его часто называют *углом интерференции*. Следует, однако, иметь в виду, что условие (28.6) можно применять только при наличии той *симметрии*, которая была использована при его выводе. Задания только одного угла интерференции (вопреки распространенному противоположному утверждению) совершенно недостаточно, чтобы в общем виде записать условие хорошей контрастности интерференционных полос. Для этого надо ввести два угла β_1 и β_2, как это и было сделано выше.

Результаты, полученные выше для точечных источников, без всяких затруднений распространяются на источники света, имеющие форму коротких параллельных прямолинейных отрезков, смещенных один относительно другого в поперечном направлении на расстояние l и параллельных плоскости экрана, в средней части которого наблюдается интерференция. (Под средней частью здесь понимается малая окрестность экрана вблизи линии, вдоль которой экран пересекается с плоскостью, проведенной через середину источников перпендикулярно к их длине.)

При том же расположении справедливы и результаты (28.4) и (28.5) для некогерентных протяженных источников, имеющих форму прямоугольных полосок, если только под l понимать теперь их ширину. Такими источниками в оптике обычно служат прямоугольные щели, ярко освещаемые широкими пучками лучей. Если выполнено условие

$$l|\cos\beta_1 - \cos\beta_2| \ll \lambda,$$

то щель действует как бесконечно тонкий линейный источник света.

3. Таким образом, для получения интерференционных полос от двух источников света недостаточно, чтобы эти источники состояли из попарно когерентных точечных источников. (Примером могут служить источник и его оптическое изображение или два оптических изображения одного и того же источника.) Даже в случае строго монохроматического света необходимо еще, чтобы размеры источников не превосходили определенного предела, зависящего от взаимного расположения и расстояния между ними, а также от положения экрана, предназначенного для наблюдения интерференционных полос.

Два источника, размеры и взаимное расположение которых позволяют наблюдать интерференционные полосы (при необходимой степени монохроматичности света), называются *пространственно когерентными.* Если же интерференционных полос (даже при идеальной монохроматичности света) получить нельзя, то источники называются *пространственно некогерентными* (см. далее п. 8).

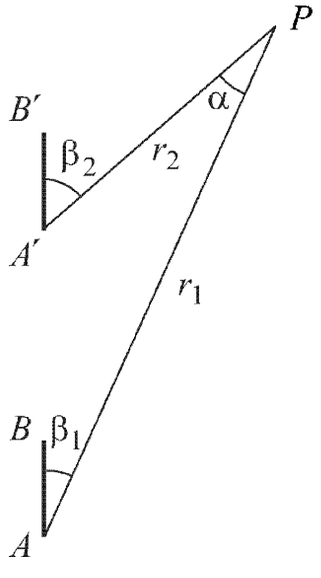


Рис. 125

4. Посмотрим теперь, при каких условиях будет соблюдаться пространственная когерентность в классических интерференционных опытах, описанных в предыдущем параграфе. Во всех опытах источники света линейные (за исключением опыта Меслина, где они точечные, так что вопрос о пространственной когерентности здесь практически не возникает). При этом в опытах с зеркалами и бипризмой Френеля, а также в установках Бийе и Поля взаимное расположение когерентных источников AB и $A'B'$ «прямое», т.е. соответствует рис. 125, а в опыте с зеркалом Ллойда — «обратное», как представлено на рис. 126 (стр. 221). На этих рисунках соответствующие когерентные пары точечных источников обозначены через (A, A'), (B, B') и т.д. При этом в установках Френеля, Ллойда и Бийе, где источниками служат освещаемые щели, последние предполагаются перпендикулярными к плоскости рисунка, а $l = AB$ означает ширину щели. В установке Поля линейные источники AB и $A'B'$ (например, изображение ртутной лампы, имеющей форму небольшого цилиндра) ориентированы вертикально, так что здесь AB означает длину l источника света.

5. Начнем с прямого расположения (см. рис. 125). В этом случае $\beta_2 - \beta_1 = \alpha$. Ввиду малости угла α схождения интерферирующих лучей можем написать

$$\cos\beta_1 - \cos\beta_2 = \sin\beta\Delta\beta = \alpha\sin\beta,$$

где с принятой точностью расчета под β можно понимать любой из углов β_1 и β_2 или любое промежуточное значение между ними. Условие (28.5) хорошей контрастности интерференционных полос принимает вид

$$l \leqslant \lambda/(2\alpha\sin\beta), \tag{28.7}$$

или, ввиду (26.13),

$$l \leqslant \Delta x/(2\sin\beta). \tag{28.8}$$

В установках Френеля и Бийе углы β_1 и β_2 практически не отличаются от $\pi/2$, так что в этом случае $l \leqslant \Delta x/2$, т.е. ширина щели не должна превосходить примерно половину ширины интерференционной полосы (предполагается, что плоскость экрана перпендикулярна к оси установки CO, см. рис. 116).

В интерференционном опыте Поля (см. рис. 120), напротив, углы β составляют 20–50 °, а интерференционные полосы, если они наблюдаются на потолке аудитории, имеют форму колец. Для упрощения вычислений заменим пластинку слюды двумя отражающими параллельными плоскостями, расстояние между которыми равно толщине h этой пластинки. Тогда расстояние между соответствующими точками источников (например, между A и A' или B и B') будет $2h$, так что $2h\sin\beta = r\alpha$, где r в нашем приближении равно r_1 или r_2 либо любой промежуточной величине между r_1 и r_2. Подставляя отсюда значение α в (28.7), получаем условие пространственной когерентности:

$$l \leqslant r\lambda/(4h\sin^2\beta). \tag{28.9}$$

В опыте Поля толщина пластинки слюды h должна быть очень мала, так что мнимые источники AB и $A'B'$ сдвинуты один относительно другого на ничтожную величину $2h$, во много раз меньшую длины l самого источника. Допустим, например, что $\lambda = 500$ нм $= 5\cdot10^{-5}$ см, $h = 0{,}05$ мм, $r = 8$ м, $\beta = 30°$. Тогда, согласно формуле (28.9), должно быть $l \leqslant 8$ см. Таким образом, для получения хорошей контрастности полос источник света может иметь сравнительно большие размеры, а потому его можно взять светосильным. В этом основное преимущество установки Поля при демонстрации явлений интерференции. Другое преимущество — большие апертуры интерферирующих пучков, позволяющие получать полосы интерференции на большой площади.

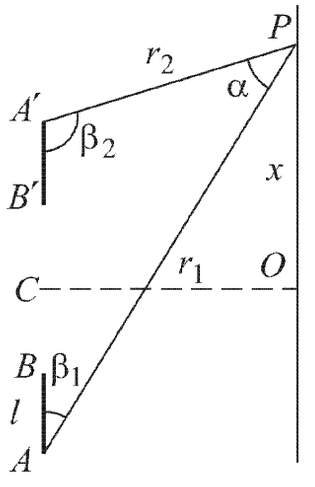


Рис. 126

6. Перейдем теперь к обратному расположению когерентных источников AB и $A'B'$ (рис. 126), реализующемуся в опыте Ллойда. В этом случае в середине экрана O условие (28.5) выполняется при любой ширине щели l. При другом расположении точки наблюдения P

$$\cos\beta_1 = (AC + OP)/r_1, \quad \cos\beta_2 = -(OP - A'C)/r_2.$$

Так как $AC = A'C$, то, пренебрегая в знаменателях различием между r_1 и r_2, отсюда получим: $\cos\beta_1 - \cos\beta_2 = 2x/r$, где $x = OP$ — координата точки P. Поскольку интерферирующие лучи почти перпендикулярны к плоскости экрана, условие хорошей

контрастности интерференционных полос запишется в виде $2lx/r \leqslant \lambda/2$, т.е.

$$x \leqslant r\lambda/(4l). \tag{28.10}$$

На вдвое большем расстоянии интерференционные полосы пропадут. Полное число N полос, которые будут наблюдаться до такого расстояния, найдется делением $2x$ на ширину полосы $\Delta x = \lambda/\alpha$. Это дает

$$N = \frac{r\alpha}{2l} = \frac{d}{l}, \tag{28.11}$$

где $2d = r\alpha$ — расстояние между источниками света. Это выражение дает оценку наивысшего порядка интерференции в монохроматическом свете в установках типа опыта Ллойда.

7. Рассмотрим, наконец, интерференционный опыт Юнга (рис. 127). Источником света служит прямоугольная полоска ширины l (например, ярко освещенная щель). Из каждой точки этой полоски, например, точки A, лучи идут к щелям S_1 и S_2 экрана под углами β_1 и β_2. Если опять пренебречь в знаменателях различием между r_1 и r_2, то можно написать: $\cos\beta_1 - \cos\beta_2 = d/r$, где d — расстояние между серединами щелей S_1 и S_2. Тогда условие (28.5) переходит в $ld/r \leqslant \lambda/2$, или

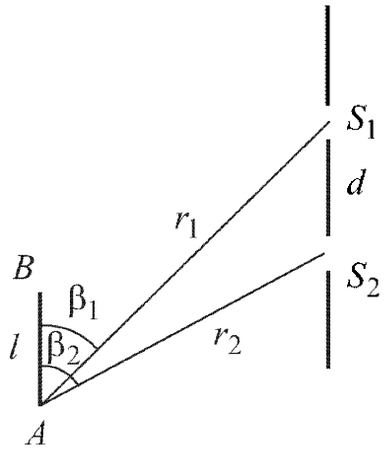


Рис. 127

$$\varphi \leqslant \lambda/(2d), \tag{28.12}$$

где $\varphi = l/r$ — угловой размер источника AB, если его рассматривать с места расположения щелей S_1 и S_2 (так как в опыте Юнга углы β_1 и β_2 почти не отличались от $\pi/2$). Когда $\varphi = \lambda/d$, интерференционные полосы пропадают.

Допустим, что свет направляется к щелям S_1 и S_2 непосредственно от Солнца, как это было в опыте Гримальди (см. § 27, п. 1; в опыте Гримальди источниками S_1 и S_2 служили два рядом расположенных небольших отверстия). Угловой размер Солнца $\varphi \approx 30' = 0{,}0087$ рад, длину волны примем равной $\lambda = 550$ нм. Тогда для получения каких бы то ни было интерференционных полос необходимо выполнение условия $d < \lambda/\varphi \approx$ $\approx 6 \cdot 10^4$ нм $= 0{,}06$ мм. Такое условие в опытах Гримальди, конечно, не могло соблюдаться, а потому не могла наблюдаться и интерференция света.

8. Более важным является следующее обобщение. Допустим, что свет, освещающий отверстия S_1 и S_2, монохроматичен. Будет или не будет в этих условиях наблюдаться интерференция пучков света, прошедших через малые отверстия S_1 и S_2, при заданном расстоянии d между ними, зависит от степени углового расхождения лучей, освещающих эти отверстия. Пусть, например, источник имеет форму диска, плоскость которого параллельна плоскости экрана, на котором сделаны отверстия S_1 и S_2. Если φ — угловой диаметр диска, то интерференция будет наблюдаться, когда отверстия S_1 и S_2 можно покрыть кругом с диаметром $d < \lambda/\varphi$. Если же этого сделать нельзя, то интерференции не будет. В первом случае говорят, что световые пучки, прошедшие через отверстия S_1 и S_2, *пространственно когерентны*, а во втором — *пространственно некогерентны*. Минимальная площадь поперечного сечения пучка

$$\sigma = \pi\lambda^2/(4\varphi^2), \tag{28.13}$$

обладающая этим свойством, называется *площадью пространственной когерентности* освещающего света. Когда свет строго монохроматичен, то все нарушения когерентности носят чисто пространственный характер, т.е. связаны с различием направлений световых лучей. Для строго плоских волн все направления лучей одинаковы ($\varphi = 0$), так что площадь когерентности σ становится бесконечной. По мере удаления от источника угловое расхождение лучей, попадающих в прибор, уменьшается, а их пространственная когерентность повышается. Примером могут служить звезды. Несмотря на громадные линейные размеры звезд, свет от них доходит до нас с высокой степенью пространственной когерентности. Световые лучи от лазеров характеризуются высокой направленностью, недоступной никаким другим земным источникам света. Именно эту особенность лазерного излучения имеют в виду, когда говорят о его высокой пространственной когерентности.

ЗАДАЧИ

1. Для характеристики контрастности интерференционных полос Майкельсон ввел функцию

$$V = \frac{I_{\text{макс}} - I_{\text{мин}}}{I_{\text{макс}} + I_{\text{мин}}}, \tag{28.14}$$

называемую *видностью* (или *видимостью*) их. Определить видность V в случае интерференции двух одинаково интенсивных монохроматических пучков света, попадающих на экран по различным путям от точечного источника света.

Р е ш е н и е. Согласно (26.7), интенсивность света на экране при освещении обоими источниками дается выражением $I = 2I_1(1 + \cos\delta)$, где I_1 — интенсивность, создаваемая одним из них. При смещении точки наблюдения по экрану из-за изменения разности фаз δ между интерферирующими пучками интенсивность I меняется от $I_{\text{макс}} = 4I_1$ до $I_{\text{мин}} = 0$, так что $V = 1$.

2. Решить ту же задачу для источника света, состоящего из двух одинаковых некогерентных светящихся точек A и B, расположенных на расстоянии l друг от друга (см. рис. 123). Интерференционные полосы наблюдаются на удаленном экране.

Р е ш е н и е. Интенсивности света в интерференционных картинах на экране, создаваемые в отдельности светящимися точками A и B, равны соответственно $I_A = 2I_1(1+\cos\delta_1)$, $I_B = 2I_1(1+\cos\delta_2)$, где δ_1, δ_2 и I имеют такой же смысл, что и в предыдущей задаче. Результирующая интенсивность

$$I = I_A + I_B = 4I_1\left(1 + \cos\frac{\delta_1 - \delta_2}{2}\cos\frac{\delta_1 + \delta_2}{2}\right).$$

Величина $\delta_1 - \delta_2$ есть разность фаз между обеими интерференционными картинами. Согласно вычислениям, приведенным в п. 1,

$$\delta \equiv |\delta_1 - \delta_2| = k\Delta = kl|\cos\beta_1 - \cos\beta_2|,$$

где $k = 2\pi/\lambda$ — волновое число. При перемещении по экрану точки наблюдения изменения фаз δ_1 и δ_2 происходят быстро и определяют ширину интерференционных полос. Напротив, разность фаз $\delta \equiv |\delta_1 - \delta_2|$ изменяется медленно, оставаясь практически постоянной на протяжении многих интерференционных полос. Она определяет видность, т.е. контрастность полос.

Максимальные и минимальные интенсивности будут

$$I_{\text{макс}} = 4I_1\left(1+\left|\cos\frac{\delta}{2}\right|\right), \quad I_{\text{мин}} = 4I_1\left(1-\left|\cos\frac{\delta}{2}\right|\right),$$

а потому

$$V = \left|\cos k\frac{\Delta}{2}\right| = \left|\cos\frac{\pi\Delta}{\lambda}\right|. \tag{28.15}$$

При изменении Δ, т.е. расстояния l между светящимися точками A и B, видность V периодически изменяется от 1 до 0 (рис. 128).

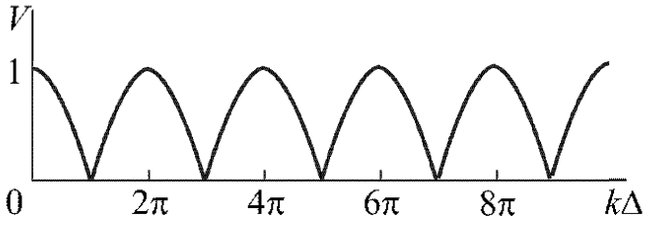


Рис. 128

3. Найти видность интерференционных полос, когда источником света является прямолинейный отрезок длины l, все точки которого излучают некогерентно (см. рис. 124).

Р е ш е н и е. Направим ось Y вдоль излучающего отрезка, поместив начало координат O в его середине. Пусть φ — разность фаз между двумя лучами, исходящими из точки O и попадающими в точку наблюдения P по различным путям. Разность фаз между такими же лучами, исходящими из другой точки отрезка с координатой y и приходящими в ту же точку P, будет

$$\delta = ky(\cos\beta_1 - \cos\beta_2) + \varphi = (k\Delta/l)y + \varphi.$$

Лучи, исходящие по различным путям из отрезка dy, при интерференции на экране создают интенсивность

$$dI = (1+\cos\delta)\,dy$$

(в условных единицах). Так как точки источника излучают некогерентно, то полная интенсивность будет

$$I = \int\limits_{-l/2}^{+l/2}\left[1+\cos\left(\frac{k\Delta}{l}y+\varphi\right)\right]dy = l+\frac{l}{k\Delta}\left[\sin\left(\frac{k\Delta}{2}+\varphi\right)+\sin\left(\frac{k\Delta}{2}-\varphi\right)\right].$$

При смещении вдоль экрана будет меняться фаза φ. Значения фазы φ, при которых интенсивность I экстремальна, найдутся из условия

$$\cos\left(\frac{k\Delta}{2}+\varphi\right) = \cos\left(\frac{k\Delta}{2}-\varphi\right),$$

откуда $\varphi = m\pi$, где m — целое число;

$$I = l + l\,\frac{\sin(k\Delta/2)}{k\Delta/2} \quad \text{при четном } m,$$

$$I = l - l\,\frac{\sin(k\Delta/2)}{k\Delta/2} \quad \text{при нечетном } m.$$

Следовательно,

$$I_{\text{макс}} = l + l\,\frac{\sin(k\Delta/2)}{k\Delta/2}, \quad I_{\text{мин}} = l - l\,\frac{\sin(k\Delta/2)}{k\Delta/2},$$

$$V = \frac{\sin(k\Delta/2)}{k\Delta/2}. \tag{28.16}$$

Кривая видности представлена на рис. 129. Из рисунка видно, как быстро с увеличением длины l источника уменьшаются максимумы на кривой видности. Той же кривой представляется видность интерференционных полос в опытах типа зеркал и бипризмы Френеля (если только точки щелевого источника света излучают некогерентно). В этом случае l означает ширину щели.

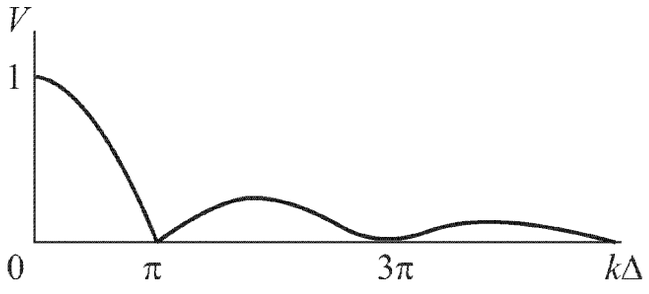


Рис. 129

4. Решить предыдущую задачу в предположении, что все точки источника излучают когерентно и притом в одинаковых фазах.

О т в е т.

$$V = \frac{2\left|\dfrac{\sin(k\Delta/2)}{k\Delta/2}\right|}{1+\left(\dfrac{\sin(k\Delta/2)}{k\Delta/2}\right)^2}. \tag{28.17}$$

§ 29. Спектральное разложение

1. До сих пор интерференция исследовалась только в идеальном случае монохроматического света. Интерференцию в немонохроматическом свете можно исследовать, разлагая свет по теореме Фурье на монохроматические составляющие. Если волновое поле в точке наблюдения описывается периодической функцией $E = E(t)$ с основным периодом τ и основной частотой $\Omega = 2\pi/\tau$, то его можно представить в виде вещественной части *ряда Фурье*:

$$E(t) = \sum_{n=0}^{\infty} a_n e^{in\Omega t}, \tag{29.1}$$

коэффициенты которого определяются выражением

$$a_n = \frac{1}{\tau}\int_{-\tau/2}^{\tau/2} E(t)e^{-in\Omega t}\,dt \tag{29.2}$$

(см. т. III, § 128; конечно, в случае света постоянного слагаемого с коэффициентом a_0 не будет). Средняя за период объемная плотность энергии колебаний (в условных единицах) будет

$$\overline{w} = \frac{1}{\tau}\int_{-\tau/2}^{\tau/2} |E|^2\,dt = \sum_{n=0}^{\infty}|a_n|^2 = \sum_{n=0}^{\infty} w_n. \tag{29.3}$$

Она равна сумме средних плотностей энергии монохроматических колебаний, из которых складывается результирующее колебание. То же заключение справедливо и для интенсивности колебаний, если понимать под интенсивностью усредненную по периоду τ любую энергетическую величину, характеризующую поле излучения в рассматриваемой точке пространства.

Полученный результат остается приближенно верным и для случая, когда функция $E(t)$ не периодична, а представляется суперпозицией монохроматических колебаний, частоты которых распределены по спектру совершенно произвольно. Только в этом случае усреднение надо производить не по периоду τ (которого теперь не существует), а по времени, весьма большому по сравнению с периодами всех монохроматических колебаний, входящих в суперпозицию. Результат приближенно верен и в случае суперпозиции почти гармонических колебаний с произвольными частотами, например для света, состоящего из узких спектральных линий.

2. Если функция $E(t)$ не периодична, то она представляется не рядом, а *интегралом Фурье*. Для возможности такого представления на функцию $E(t)$ приходится накладывать различные (достаточные) ограничения, например требовать, чтобы она была *абсолютно интегрируема* во всем бесконечном интервале $(-\infty, +\infty)$, т.е. чтобы сходился интеграл

$$\int\limits_{-\infty}^{+\infty} |E(t)|\, dt.$$

Это обстоятельство, однако, в физике не создает никаких существенных затруднений, даже в тех случаях, когда вводят функции, не обращающиеся в нуль на бесконечности, т.е. не удовлетворяющие требованию абсолютной интегрируемости.

Действительно, пусть $E(t)$ — такая функция. Разделим t на интервалы времени, достаточно длительные по сравнению с периодами световых колебаний. Световое поле на каждом из таких интервалов $(t_0, t_0+\tau)$ и его воздействие на приемник при любом значении t_0 практически совсем не зависят от полей на соседних интервалах. Поэтому при рассмотрении света только на интервале $(t_0, t_0+\tau)$ функцию $E(t)$ вне рассматриваемого интервала можно заменить любой другой функцией. В частности, ее можно периодически продолжить за пределы интервала $(t_0, t_0+\tau)$ с периодом τ. Но тогда для представления функции $E(t)$ в интервале $(t_0, t_0+\tau)$ можно воспользоваться рядом Фурье (29.1). При этом, ввиду малости частоты $\Omega = 2\pi/\tau$, целесообразно ввести обозначения $\Delta\omega = \Omega$, $\omega_n = n\Omega$. Тогда

$$E(t) = \sum a_n e^{i\omega_n t} = \sum \frac{a_n}{\Omega} e^{i\omega_n t} \Delta\omega.$$

После аппроксимации суммы интегралом получаем

$$E(t) = \int_0^\infty a(\omega) e^{i\omega t}\, d\omega, \tag{29.4}$$

где $a(\omega) = a_n/\Omega$, или с учетом (29.2)

$$a(\omega) = \frac{1}{\tau\Omega} \int_{-\tau/2}^{\tau/2} E(t) e^{-i\omega t}\, dt = \frac{1}{2\pi} \int_{-\tau/2}^{\tau/2} E(t) e^{-i\omega t}\, dt. \tag{29.5}$$

Интеграл Фурье получается из этих формул, если в последнем выражении конечные пределы заменить на бесконечные $-\infty$ и $+\infty$ (см. т. III, § 128). Однако здесь мы не будем делать этого, оставляя время τ неопределенным. При физической постановке задач всегда можно достигнуть необходимой точности, выбирая τ достаточно большим. Таким путем достигается то преимущество, что формулу (29.4) в каждом интервале длительностью τ можно будет применять и для функций $E(t)$, не интегрируемых абсолютно, например к плоским волнам постоянной интенсивности, не ограниченным во времени. При этом выражения (29.4) на разных интервалах времени τ (если τ выбрать достаточно большим), вообще говоря, не будут когерентны.

В случае интеграла Фурье формула (29.3) заменится на

$$\int_{-\tau/2}^{\tau/2} |E|^2\, dt = \tau \sum |a(\omega_n)\Omega|^2.$$

Если аппроксимировать сумму интегралом и учесть, что $\tau\Omega = 2\pi$, то получится

$$\int_{-\tau/2}^{\tau/2} |E|^2\, dt = 2\pi \int_0^\infty |a(\omega)|^2\, d\omega. \tag{29.6}$$

О физическом смысле этого соотношения говорится ниже в п. 5.

3. Приведем пример спектрального разложения, приводящий к важным обобщениям. Бесконечно длящееся синусоидальное колебание является идеализацией ограниченного ряда, или цуга синусоидальных волн, представленного на рис. 130 *а* «оборванной синусоидой». Пусть T_0 — «период», а ω_0 — «частота» этой «синусоиды». Тогда в комплексной форме

$$a(\omega) = \frac{1}{2\pi} \int_{-\tau/2}^{\tau/2} \frac{1}{i}\, e^{i(\omega_0-\omega)t}\, dt = \frac{2}{\pi i} \frac{\sin\left[(\omega-\omega_0)\tau/2\right]}{(\omega-\omega_0)\tau/2}.$$

В вещественной форме

$$E(t) = \frac{2}{\pi} \int_0^\infty \frac{\sin[(\omega - \omega_0)\tau/2]}{(\omega - \omega_0)\tau/2} \sin \omega t \, d\omega. \tag{29.7}$$

График функции $\frac{\sin \alpha}{\alpha}$, где $\alpha = \frac{(\omega - \omega_0)\tau}{2}$, приведен на рис. 130 *б*. Таким образом, оборванный цуг волн, изображенный на рис. 130 *а*, может быть представлен суперпозицией бесконечного множества синусоид, частоты которых непрерывно

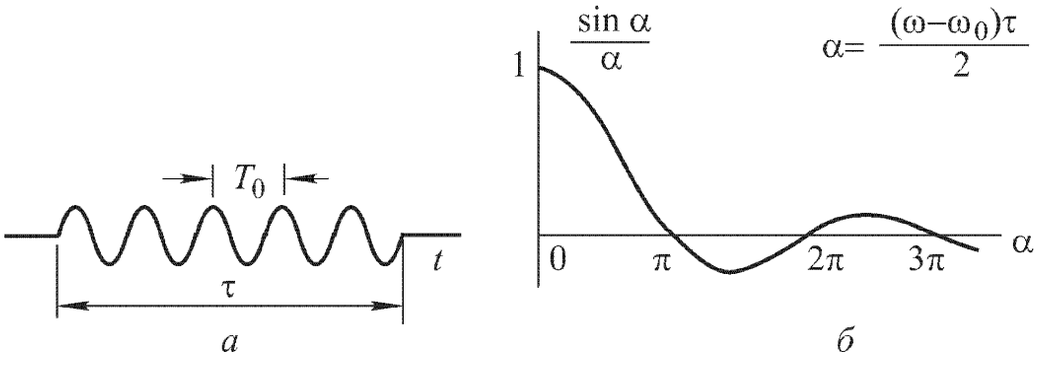


Рис. 130

заполняют бесконечный интервал $0 < \omega < +\infty$. Впрочем, основное значение имеет только интервал $-\frac{\pi}{2} < \alpha < +\frac{\pi}{2}$, или интервал $\omega_0 - \frac{\pi}{\tau} < \omega < \omega_0 + \frac{\pi}{\tau}$, где амплитуды колебаний велики. На всех остальных участках амплитуды малы. Если ими пренебречь, то можно сказать, что весь спектр частот практически сосредоточен в пределах интервала шириной $\Delta\omega$, который удовлетворяет условию

$$\Delta\omega \cdot \tau \gtrsim 2\pi. \tag{29.8}$$

Если ввести обычную частоту $\nu = \omega/(2\pi)$, то

$$\Delta\nu \cdot \tau \gtrsim 1. \tag{29.8а}$$

Это важное соотношение между шириной спектра $\Delta\omega$ (или $\Delta\nu$) и длительностью цуга τ имеет общий характер. Его можно также уяснить на следующем простом примере. Рассмотрим множество синусоид с одинаковыми амплитудами, но различными частотами, непрерывно и равномерно заполняющими интервал $\Delta\omega$. Пусть в точке $t = 0$ фазы всех синусоид совпадают, а следовательно, амплитуда колебаний максимальна. При $t = \tau/2$ разность фаз между крайними синусоидами будет $\Delta\omega \cdot \tau/2$. Если она сделается равной 2π, то в точке $t = \tau/2$ наложатся синусоиды со всевозможными фазами, непрерывно и равномерно заполняющими интервал шириной 2π. При таком наложении, как легко уяснить с помощью векторной диаграммы, синусоиды

погасят друг друга. При том же условии произойдет взаимное гашение синусоид и в точке $t = -\tau/2$. Таким образом, выделится интервал времени $(-\tau/2, +\tau/2)$, на обоих концах которого волновое поле обратится в нуль. На этом участке время τ связано с шириной спектра $\Delta\omega$ соотношением $\Delta\omega \cdot \tau = 4\pi$, которое по существу совпадает с (29.8).

4. Формула, аналогичная (29.8), имеет место и для пространственного распределения волнового поля, когда оно рассматривается в какой-либо фиксированный момент времени. Примером может служить «оборванная синусоида» на рис. 130 *а*, если по оси абсцисс откладывать координату x в направлении распространения волны. При таком рассмотрении снова можно пользоваться формулой (29.4), заменив в ней время t на координату x, а частоту ω на волновое число $k = \omega/v$. Поступив так, легко найти, что длина цуга волн Δx связана с соответствующим интервалом волновых чисел Δk соотношением

$$\Delta k \cdot \Delta x \gtrsim 2\pi. \tag{29.9}$$

Этот результат будет обобщен в § 44 на случай трех измерений. Соотношения (29.8) и (29.9) играют важную роль в квантовой механике, где они при несколько иной интерпретации выражают так называемый *принцип неопределенности Гайзенберга* (1901–1976).

5. Когда ширина спектра $\Delta\omega$ достаточно мала ($\Delta\omega \ll \omega$), то соответствующий свет называют *квазимонохроматическим.* Часто пользуются также «*спектроскопическим волновым числом*» $\widetilde{k} = k/(2\pi)$. Для него

$$\Delta\widetilde{k} \cdot \Delta x \gtrsim 1. \tag{29.9а}$$

Примером могут служить спектральные линии, излучаемые разреженными газами. Квазимонохроматический свет можно выделить из непрерывного спектра излучения (например, Солнца) призмой, дифракционной решеткой или другими приборами, осуществляющими спектральное разложение. За меру монохроматичности света можно принять отношение $|\omega/\Delta\omega|$ или равное ему отношение $|\lambda/\Delta\lambda|$. Наибольшей монохроматичностью обладает свет, излучаемый газовыми лазерами. Нельзя, однако, упускать из виду, что цуг волн или световой пучок, занимающий спектральную область $\Delta\omega$, должен обладать конечной длительностью, не меньшей $\tau \approx 2\pi/\Delta\omega$. В частности, не имеет смысла говорить о мгновенной объемной плотности энергии излучения в интервале $d\omega$, т.е. о величине $\sim |dE|^2 = |a(\omega)\, d\omega|^2$. Вместо мгновенной надо пользоваться средней плотностью энергии излучения за время порядка $\tau \approx 2\pi/d\omega$. Это значит, что интенсивность излучения в спектральном интервале $(\omega, \omega + d\omega)$ должна определяться величиной, пропорциональной $\tau|dE|^2$, т.е. $|a(\omega)|^2\, d\omega$. Ве-

личину $|a(\omega)|^2$ называют *спектральной плотностью излучения.* Именно такую величину имеют в виду, говоря о распределении интенсивности или энергии излучения в спектре. Вместо частоты можно, конечно, пользоваться длинами волн, представляя ту же величину в виде $|a_\lambda(\lambda)|^2\,d\lambda$. Очевидно, $a(\omega) = 2\pi c a_\lambda(\lambda)/\omega^2$.

ЗАДАЧА

Разложить в интеграл Фурье и найти спектральную плотность излучения для затухающего осциллятора, волновое поле которого определяется выражением

$$E(t) = e^{-t/\tau}\sin\omega_0 t \qquad (t \geqslant 0).$$

Ответ. $a(\omega) = \dfrac{1}{2\pi}\dfrac{1}{(\omega-\omega_0) - t/\tau}$,

$$|a(\omega)|^2 = \frac{1}{4\pi^2}\frac{1}{(\omega-\omega_0)^2 + (1/\tau)^2}. \tag{29.10}$$

Если $\omega - \omega_0 = 1/\tau$, т.е. $\Delta\omega\cdot\tau = 1$, то спектральная плотность излучения уменьшается в два раза.

Этот пример снова подтверждает общее соотношение (29.8). В примере, рассмотренном в тексте (см. рис. 130 *б*), спектральная плотность излучения $|a(\omega)|^2$ на краях интервала $-\dfrac{\pi}{2} < \alpha < \dfrac{\pi}{2}$ убывает в $(\pi/2)^2 \approx 2{,}5$ раза по сравнению с той же величиной при $\alpha = 0$.

§ 30. Влияние немонохроматичности света

1. Как и увеличение размеров источников, немонохроматичность света ведет сначала к *ухудшению контрастности* (видимости) интерференционных полос, а затем к полному исчезновению их. Чтобы не усложнять исследование учетом конечных размеров источника, будем предполагать, что источник света S точечный. Пусть S_1 и S_2 (см. рис. 113) — когерентные источники, являющиеся действительными или мнимыми изображениями источника S. Допустим сначала, что излучение источника S состоит из двух близких одинаково интенсивных спектральных линий с длинами волн λ и $\lambda' = \lambda + \delta\lambda$. Точка или линия экрана, где оптическая разность хода Δ интерферирующих лучей равна нулю, называется *центром интерференционной картины.*

Если начальные фазы источников S_1 и S_2 одинаковы, то в центр картины лучи с длинами волн λ и λ' придут в *одинаковых фазах.* Для обеих волн там получится светлая полоса. В другой точке экрана A, в которой $\Delta = N\lambda'$, где N — целое число (номер полосы или порядок интерференции), для длины волны λ' получится также светлая интерференционная полоса. Если $\Delta = (N + {}^1\!/_2)\lambda$, то в ту же точку A интерферирующие лучи

с другой длиной волны λ придут уже *в противоположных фазах*, и для такой длины волны интерференционная полоса будет темной. При этом условии в окрестности точки A светлые полосы с длиной волны λ' наложатся на темные полосы с длиной волны λ. Интерференционные полосы в указанной окрестности исчезнут. Условие первого исчезновения полос, таким образом, есть $N\lambda' = (N + {}^1\!/_2)\lambda$, или

$$N = \frac{\lambda}{2(\lambda' - \lambda)} = \frac{\lambda}{2\,\delta\lambda}. \tag{30.1}$$

Все изложенное остается верным и в том случае, когда фазы лучей, приходящих в центр интерференционной картины, противоположны. Только в этом случае центральная полоса будет темной.

Когда номер полосы мал по сравнению с величиной N, определяемой выражением (30.1), интерференционные полосы будут почти столь же отчетливы, что и в случае света с одной длиной волны. Когда номер полосы для длины волны λ' достигнет значения $2N$, номер соответствующей полосы для длины волны λ' сделается равным $(2N + 1)$. Тогда полосы интерференции сделаются столь же резкими, что и в центре интерференционной картины. При дальнейшем возрастании порядка интерференции будет наблюдаться периодическая смена резкости интерференционных полос от наибольшей отчетливости их до полного исчезновения.

Например, когда источником света является пламя натрия, то излучаются две узкие спектральные линии с длинами волн $\lambda = 589$ и $\lambda' = 589{,}6$ нм. При $N \approx \lambda/[2(\lambda' - \lambda)] = 490$ интерференционные полосы становятся очень неясными или совсем ненаблюдаемыми. При N, равном 980 или кратном этому числу, полосы опять становятся отчетливыми. При дальнейшем возрастании порядка интерференции полосы периодически меняют свой вид, становясь попеременно то отчетливыми, то размытыми. И так продолжается до тех пор, пока из-за конечной ширины обеих спектральных линий интерференционная картина не пропадет совсем.

2. Перейдем теперь к случаю, когда свет от источника S непрерывно и равномерно заполняет спектральный интервал $(\lambda, \lambda + \delta\lambda)$. В этом случае можно поступить так же, как мы поступали в § 28 (п. 2) при рассмотрении интерференционных полос от протяженного источника света. Весь спектральный интервал $(\lambda, \lambda + \delta\lambda)$ разобьем на множество пар бесконечно узких спектральных линий, находящихся на расстоянии $\delta\lambda/2$ друг от друга (на шкале длин волн). К каждой такой паре применима формула (30.1), если в ней сделать замену $\delta\lambda \to \delta\lambda/2$. Поэтому первое исчезновение интерференционных полос произойдет для

порядка интерференции

$$N = \lambda/\delta\lambda, \tag{30.2}$$

вдвое большего, чем в предыдущем случае. При дальнейшем возрастании порядка интерференции интерференционные полосы появятся вновь, но уже на светлом фоне, аналогично тому, как это было в случае протяженного монохроматического источника света (§ 28, п. 2, а также задача 3). Их контрастность будет незначительна, а при дальнейшем возрастании порядка интерференции и совсем исчезнет. Практически наблюдение интерференционных полос ограничится порядками интерференции, не превосходящими (30.2).

Формула (30.2) как оценочная остается верной и в случае произвольного распределения интенсивности света по длинам волн в интервале $\delta\lambda$. Таким образом, эта формула дает оценку максимально возможного порядка интерференции при заданной степени монохроматичности $\lambda/\delta\lambda$ используемого света. Два квазимонохроматических пучка света, когерентные при низких порядках интерференции, перестают быть таковыми при высоких порядках, превышающих примерно величину (30.2).

3. Максимально возможный порядок интерференции можно также найти из следующих соображений. Пусть используемый свет состоит из одинаковых цугов волн, следующих друг за другом через беспорядочно меняющиеся промежутки времени. Один из возможных цугов изображен на рис. 130 *а* (здесь показано колебание волнового поля во времени в фиксированной точке пространства; совершенно аналогично выглядит цуг в пространстве в фиксированный момент времени). Интерференционный прибор разделяет пучок света на два пучка, идущие к месту схождения по разным путям, так что между пучками возникает разность хода. Для когерентности пучков необходимо, чтобы разность хода между ними не превосходила длину цуга $L = c\tau$. В противоположном случае будет происходить наложение независимых цугов волн, испущенных в разные моменты времени, и интерференция не возникнет. Максимальный порядок интерференции не может превышать величину

$$N_{\text{макс}} = \frac{L}{\lambda} = \frac{\tau}{T}. \tag{30.3}$$

Разумеется, это условие должно согласовываться с условием (30.2), а потому должно быть

$$\frac{L}{\lambda} = \frac{\tau}{T} = \frac{\lambda}{\delta\lambda} = \frac{\omega}{\delta\omega}.$$

Если учесть, что $\lambda = 2\pi/k$ и $T = 2\pi/\omega$, то отсюда получится

$$\tau \cdot \delta\omega = 2\pi \quad \text{и} \quad L \cdot \delta k = 2\pi.$$

К таким соотношениям мы уже пришли в предыдущем параграфе из других соображений.

С изложенной точки зрения нарушения когерентности связаны с *запаздыванием* одного цуга волн по сравнению с другим. Поэтому здесь, в отличие от пространственной, говорят о *временной когерентности*, а длительность цуга τ называют *временем когерентности*. Временная когерентность — это то же самое, что и когерентность, связанная с *узостью спектрального интервала* $\Delta\omega$, занимаемого светом. В силу (29.8) время когерентности $\tau_{\text{ког}}$ связано с шириной спектрального интервала $\Delta\omega$ соотношением

$$\tau_{\text{ког}} \approx \frac{2\pi}{\Delta\omega} \approx \frac{1}{\Delta\nu}. \tag{30.4}$$

Чем меньше $\Delta\nu$, тем больше время когерентности и тем выше порядок интерференции, который может наблюдаться при заданном ν. Максимальная разность хода, при которой еще возможна интерференция, определяется соотношением

$$L \approx c\tau_{\text{ког}} = \lambda\frac{\nu}{\delta\nu} = \frac{\lambda^2}{\delta\lambda}.$$

Она называется *длиной когерентности*. Ей соответствует максимальный порядок интерференции

$$N_{\text{макс}} = \frac{\nu}{\delta\nu} = \frac{\lambda}{\delta\lambda}.$$

Значения параметров $\delta\nu$, τ и L очень резко различаются для света тепловых источников и света, генерируемого газовыми лазерами. Ширина спектральной линии $\delta\nu$ лучших «монохроматических» тепловых источников, которые могут быть созданы в лаборатории, порядка 10^8 Гц, тогда как в случае лазеров можно получить $\delta\nu \sim 10^2$ Гц или даже еще меньше. Соответствующие времена когерентности будут 10^{-8} и 10^{-2} с, а длины когерентности 1 и 10^6 м. Лазерные источники света позволяют наблюдать интерференцию при разности хода в несколько километров. Здесь максимальный порядок интерференции, который можно наблюдать, ограничивается не степенью монохроматичности лазерного излучения, а неоднородностью земной атмосферы и трудностями создания стабильной интерференционной схемы столь больших размеров.

4. Для белого света $\delta\lambda \sim \lambda$, т.е. $N \sim 1$. Казалось бы, что в белом свете интерференционные полосы наблюдаться не должны. Это действительно так, если пользоваться такими приемниками света, как фотоэлемент, болометр или термостолбик, которые обладают примерно одинаковой чувствительностью в различных участках спектра. Но глаз — *селективный приемник*, т.е. его чувствительность к различным длинам волн разная (см.

кривую видности человеческого глаза на рис. 82). Именно поэтому в белом свете глаз видит около десятка интерференционных полос. Полосы цветные, так как из-за различия в длинах волн полосы разного цвета имеют разную ширину и сдвинуты относительно друг друга. Только в центре картины, куда волны от обоих источников приходят в одинаковых фазах, соблюдается условие максимума для всех длин волн. Там получается *ахроматическая*, т.е. неокрашенная, светлая интерференционная полоса. В опыте с зеркалом Ллойда ахроматическая полоса темная. Это доказывает, что отражение света от зеркала сопровождается изменением фазы волны на π (см. § 33, 65). Конечно, при обычной постановке опыта (см. рис. 119) ахроматическую полосу увидеть нельзя, так как луч, отражающийся от зеркала, проходит больший путь, чем прямой луч. Чтобы получить эту полосу, надо на пути прямого луча поставить плоскопараллельную пластинку, смещающую вверх всю интерференционную картину.

§ 31. Корреляция и когерентность света

1. При точном количественном определении понятия когерентности надо учесть, что реальные световые колебания *не синусоидальны*. Напряженность поля в каждой точке пространства может быть представлена интегралом Фурье (29.4), т.е. в виде суперпозиции синусоидальных колебаний различных частот. Если область $\Delta\omega$, заполняемая этими частотами, мала по сравнению с самими частотами ω, входящими в суперпозицию, то результирующее колебание и представляемый им свет называются *квазимонохроматическими*. Выбрав внутри интервала $\Delta\omega$ произвольную частоту ω_0, запишем квазимонохроматическое колебание в виде

$$E(t) = a(t)e^{i\omega_0 t}, \tag{31.1}$$

где $a(t)$ — комплексная амплитуда, медленно меняющаяся по сравнению с быстро осциллирующей функцией $e^{i\omega_0 t}$. Она, конечно, определена не совсем однозначно, поскольку ее значение зависит от выбора частоты ω_0. Про колебание, представленное формулой (31.1), говорят, что оно *модулировано*.

При модуляции может медленно меняться (вещественная) амплитуда, колебания или его (начальная) фаза. В первом случае говорят об *амплитудной*, во втором — о *фазовой модуляции*. Могут также одновременно меняться и амплитуда, и фаза. В случае квазимонохроматического света, излучаемого реальными источниками, такие изменения происходят хаотически — амплитуда и фаза являются *случайными функциями времени*. Поэтому при изучении реального света, в том числе и квазимонохроматического, нельзя обойтись без использования *статистических методов*. Для простоты мы отвлечемся от векторного характера колебаний, считая их скалярными.

Важно отметить, что из-за очень высоких оптических частот все существующие приемники света не позволяют регистрировать быстрые изменения напряженности световых полей за времена порядка периода световых колебаний. Обычно они не позволяют следить и за быстрыми изменениями световых потоков, обусловленных случайными изменениями амплитуд и

фаз колебаний. Удается измерять только *квадраты напряженностей* световых полей, усредненные по промежуткам времени, весьма большим не только по сравнению с периодами световых колебаний, но и по сравнению с временами, в течение которых происходят случайные изменения амплитуд и фаз этих колебаний. Ниже предполагается, что световые потоки регистрируются именно такими «инерционными» приемниками.

Более того, мы будем предполагать, что световые потоки в среднем *стационарны*, т.е. значение среднего квадрата поля в каждой точке пространства одинаково для всех моментов времени и не зависит от положения на шкале времени временного интервала, по которому производится усреднение. Квадрат поля можно представить в виде

$$(\operatorname{Re} E)^2 = \left(\frac{E+E^*}{2}\right)^2 = \frac{1}{4}(E^2+E^{*2}) + \frac{1}{2}EE^*.$$

Если положить $a = a_0(t)e^{i\delta(t)}$, где $a_0(t)$ и $\delta(t)$ — *медленно меняющиеся* вещественная амплитуда и фаза, то

$$E^2 + E^{*2} = a_0^2[e^{2i(\omega_0 t+\delta)} + e^{-2i(\omega_0 t+\delta)}] = 2a_0^2\cos[2(\omega_0 t+\delta)].$$

Эта величина осциллирует во времени очень быстро и при усреднении пропадает. Результат усреднения величины $(\operatorname{Re} E)^2$ определяется только последним членом $^1/_2 EE^*$. Поэтому за меру интенсивности колебаний можно принять величину $\overline{EE^*}$.

2. Допустим теперь, что в точку наблюдения P в момент времени t приходят два колебания от источников света S_1 и S_2 (см. рис. 113). Чтобы прийти в P в момент t, эти колебания должны выйти из S_1 и S_2 в более ранние моменты времени $t-\theta_1$ и $t-\theta_2$, где θ_1 и θ_2 — времена, затрачиваемые светом на распространение от S_1 и S_2 до точки P. Чтобы отметить это, рассматриваемые колебания в точке P обозначим через $E_1(t-\theta_1)$ и $E_2(t-\theta_2)$ соответственно. При их сложении в точке P получится результирующее колебание

$$E \equiv E(P,t) = E_1(t-\theta_1) + E_2(t-\theta_2).$$

Для нахождения его интенсивности в точке P умножим это равенство на комплексно-сопряженное и произведем усреднение по времени. В результате получим

$$I = \overline{E_1(t-\theta_1)E_1^*(t-\theta_1)} + \overline{E_2(t-\theta_2)E_2^*(t-\theta_2)} + \\ + \overline{E_1(t-\theta_1)E_2^*(t-\theta_2) + E_1^*(t-\theta_1)E_2(t-\theta_2)}.$$

В силу предположения о стационарности (в среднем) световых потоков первое слагаемое справа не зависит от θ_1 и t. Оно представляет просто интенсивность I_1 первого колебания, пришедшего в точку P:

$$I_1 = \frac{1}{\tau}\int\limits_{-\tau/2}^{\tau/2} E_1(t-\theta_1)E_1^*(t-\theta_1)\,dt = \frac{1}{\tau}\int\limits_{-\tau/2}^{\tau/2} E_1(t)E_1^*(t)\,dt,$$

где τ — ширина временного интервала, по которому производится усреднение. Аналогично, второе слагаемое есть интенсивность I_2 второго колебания. По той же причине последнее слагаемое (интерференционный член) не зависит от t, а также от θ_1 и θ_2 в отдельности. Оно есть функция только

разности $\theta = \theta_2 - \theta_1$, т.е. времени запаздывания второго колебания относительно первого. Поэтому можно положить

$$\overline{E_1(t-\theta_1)E_2^*(t-\theta_2)} = \overline{E_1(t)E_2^*(t-\theta)} = F_{12}(\theta), \tag{31.2}$$

где $F_{12}(\theta)$ — комплексная функция, характеризующая *степень согласованности* рассматриваемых колебаний в точке P. Она называется *корреляционной функцией* колебаний $E_1(t-\theta_1)$ и $E_2(t-\theta_2)$ или *взаимной корреляционной функцией.*

В частном случае функции $E_1(t)$ и $E_2(t)$ могут оказаться *тождественными.* Это будет, например, когда оба колебания выходят из одного и того же источника, но приходят в точку P по различным путям. Тогда $F_{12}(\theta)$ называют *автокорреляционной функцией.* Ее можно обозначать через $F_{11}(\theta)$, но мы чаще будем применять обозначение $F(\theta)$. При $\theta = 0$ автокорреляционная функция переходит в $\overline{|E_1(t-\theta_1)|^2}$, т.е. в интенсивность колебания I_1 в точке P.

Функция $F_{12}(\theta)$ зависит от I_1 и I_2, т.е. от интенсивностей складываемых колебаний в точке P. Если положить

$$F_{12}(\theta) = \sqrt{I_1 I_2}\, f_{12}(\theta), \tag{31.3}$$

то получится *нормированная корреляционная функция*, которая зависит только от времени запаздывания θ, но уже не зависит от I_1 и I_2. Через эту функцию результирующая интенсивность в точке P представляется выражением

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\ \mathrm{Re}\,[f_{12}(\theta)]. \tag{31.4}$$

Для квазимонохроматического света $E_1(t) = a_1(t)e^{i\omega_0 t}$, $E_2(t) = a_2(t)e^{i\omega_0 t}$, так что

$$\overline{a_1(t)a_2^*(t-\theta)}e^{i\omega_0\theta} = \sqrt{I_1 I_2}\, f_{12}(\theta). \tag{31.5}$$

Как видно из этой формулы, величина $f_{12}(\theta)$ есть быстро меняющаяся функция времени запаздывания θ. Разделив ее на столь же быстро меняющуюся осциллирующую функцию $e^{i\omega_0\theta}$, получим уже медленно меняющуюся функцию

$$\gamma_{12}(\theta) = f_{12}(\theta)e^{-i\omega_0\theta}, \tag{31.6}$$

которая называется *комплексной степенью когерентности колебаний*, а ее модуль $|\gamma_{12}(\theta)| = |f_{12}(\theta)|$ — просто *степенью когерентности колебаний* в точке P. Таким образом,

$$\overline{a_1(t)a_2^*(t-\theta)} = \sqrt{I_1 I_2}\,\gamma_{12}(\theta), \tag{31.7}$$

т.е. $\gamma_{12}(\theta)$ есть *нормированная взаимная корреляционная функция для амплитуд* $a_1(t)$ *и* $a_2(t)$. Далее,

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\ \mathrm{Re}\,[\gamma(\theta)e^{i\omega_0\theta}]. \tag{31.8}$$

Полагая $\gamma_{12}(\theta) = |\gamma_{12}(\theta)|e^{i\delta}$, запишем последний результат в вещественной форме:

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\,|\gamma_{12}(\theta)|\cos(\omega_0\theta + \delta). \tag{31.9}$$

Эта формула отличается от аналогичной формулы (26.7) для строго синусоидальных колебаний добавочным множителем $|\gamma_{12}(\theta)|$ в интерференционном члене и добавочным, медленно меняющимся слагаемым $\delta(\theta)$ в разности

фаз. Значения вещественных амплитуд $|a_1|$ и $|a_2|$ и соответствующих интенсивностей I_1 и I_2 не зависят от выбора промежуточной частоты ω_0 в спектральном интервале $\Delta\omega$ квазимонохроматического света. Не может зависеть от выбора ω_0 и полная фаза $\omega_0\theta + \delta$, входящая в формулу (31.9). Но добавочная фаза δ, конечно, будет другой при другом выборе ω_0. Фаза $\omega_0\theta + \delta$ определяет наиболее быстрые изменения в пространстве интенсивности светового поля, т.е. изменения при переходе от одной интерференционной полосы к другой. Ввиду медленности изменения функции $|\gamma_{12}(\theta)|$, ее изменениями при таком переходе можно пренебречь. Тогда в максимумах $\cos(\omega_0\theta + \delta)$ будет равен $+1$, а в минимумах -1. Поэтому

$$I_{\text{макс}} = I_1 + I_2 + 2\sqrt{I_1 I_2}\,|\gamma_{12}(\theta)|, \quad I_{\text{мин}} = I_1 + I_2 - 2\sqrt{I_1 I_2}\,|\gamma_{12}(\theta)|.$$

Отсюда для видности интерференционных полос находим

$$V \equiv \frac{I_{\text{макс}} - I_{\text{мин}}}{I_{\text{макс}} + I_{\text{мин}}} = \frac{2\sqrt{I_1 I_2}}{I_1 + I_2}|\gamma_{12}(\theta)|. \tag{31.10}$$

Когда интенсивности складываемых колебаний одинаковы ($I_1 = I_2$), то $V = |\gamma_{12}(\theta)|$. По самому определению видность V не может быть больше единицы, а функция $\gamma_{12}(\theta)$ от интенсивностей пучков не зависит. Поэтому всегда $|\gamma_{12}(\theta)| \lesssim 1$. Когда $|\gamma_{12}(\theta)| = 0$, то $V = 0$, т.е. интерференционных полос не получается. В этом случае колебания называются *некогерентными*. Если при этом функция $\gamma_{12}(\theta)$ обращается в нуль при любых значениях θ, то некогерентность называется *полной*. Тогда всюду $I = I_1 + I_2$, т.е. имеет место *закон фотометрического сложения* интенсивностей. Такой случай осуществляется при наложении световых пучков от независимых источников света.

Если же $\gamma_{12}(\theta) \neq 0$, то наблюдается интерференция, и колебания называются *когерентными*. Когерентность называется *полной*, когда величина $|\gamma_{12}(\theta)|$ всюду достигает своего предельного значения 1. В этом случае интерференционные полосы наиболее контрастны, т.е. при заданных I_1 и I_2 видность V максимальна. Такой случай реализуется при наложении строго периодических, в частности монохроматических, пучков одинаковых периодов. Во всех остальных случаях (когда $0 < |\gamma_{12}(\theta)| < 1$) говорят о *частичной когерентности*. При перемещении точки наблюдения степень когерентности $|\gamma_{12}(\theta)|$ медленно изменяется. Вследствие этого медленно изменяется и видность интерференционных полос.

3. До сих пор речь шла о когерентности двух колебаний, происходящих в *одной и той же точке* пространства. Но можно говорить о когерентности *одного и того же волнового поля* в двух различных пространственно-временных точках $R_1(Q_1, t_1)$ и $R_2(Q_2, t_2)$. Этот вопрос сводится к предыдущему.

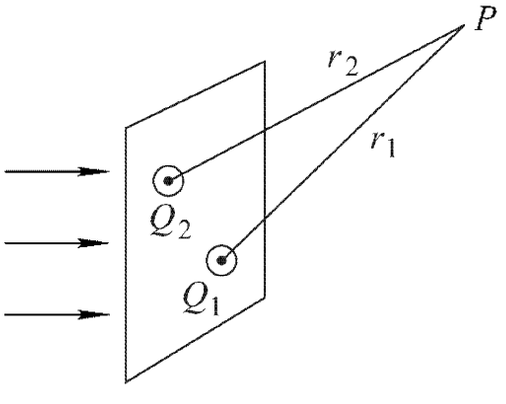


Рис. 131

Пусть Q_1 и Q_2 — какие-либо две точки пространства, находящиеся в рассматриваемом поле излучения (рис. 131). Пусть они являются центрами двух бесконечно малых отверстий в непрозрачном экране, поставленном на пути распространения света. Экран всюду загородит падающий свет, но пропустит свет через отверстия. Через отверстия пройдет не только прямой, но

и *дифрагированный свет*. Бесконечно малые отверстия в силу принципа Гюйгенса могут рассматриваться как точечные *вторичные источники*, посылающие свет за экран во всех направлениях. Возьмем за экраном удаленную точку наблюдения P. Пусть колебания, вышедшие из точек Q_1 и Q_2 в моменты времени t_1 и t_2, приходят в точку P *одновременно*. Тогда можно говорить о когерентности этих колебаний в том смысле, как это было разъяснено выше.

По определению мы называем колебания в точках Q_1 и Q_2 в моменты времени t_1 и t_2 (т.е. в пространственно-временных точках R_1 и R_2) *когерентными* или *некогерентными*, если когерентны или некогерентны соответствующие колебания в точке P. При этом степень когерентности $\gamma(\theta)$ мы определяем той же величиной, что и для колебаний в точке наблюдения.

В частности, если пространственные точки Q_1 и Q_2 совпадают, но свет попадает в P различными путями, то пространственно-временные точки $R_1(Q_1, t_1)$ и $R_2(Q_2, t_2)$ отличаются только моментами времени t_1 и t_2. В этом случае говорят о *временной когерентности*. При $t_1 = t_2$ степень временной когерентности равна единице. С увеличением разности этих времен степень когерентности убывает. Максимальное значение $|t_1 - t_2|$, при котором когерентность еще сохраняется, называется *временем когерентности*. Расстояние $v|t_1 - t_2|$, проходимое светом за это время, называется *длиной когерентности*.

В другом крайнем случае времена t_1 и t_2 одинаковы, но пространственные точки Q_1 и Q_2 не совпадают. Тогда говорят о *пространственной когерентности*. Сохраняя точку Q_1 неподвижной, будем поворачивать вокруг нее экран вместе с точкой Q_2. Тогда точка Q_2 будет перемещаться вокруг Q_1, а степень когерентности $|\gamma_{12}|$ будет меняться. Геометрическое место точек, где γ_{12} обращается в нуль, есть некоторая поверхность, окружающая точку Q_1. Объем, который она ограничивает, называется *объемом когерентности* вокруг точки P_1.

Вычисление степени временной когерентности может быть систематически использовано при определении допустимой ширины спектральной области, а пространственной когерентности — допустимых размеров источников света для возможности наблюдения интерференции.

4. Иллюстрируем понятие и свойства автокорреляционной функции и степени когерентности на простейшем примере, когда оба колебания представляются «оборванной синусоидой»: $E(t) = \sin \omega_0 t$ в интервале $0 < t < \tau$ и $E(t) = 0$ вне этого интервала. Перейдем к комплексной форме $E(t) = \frac{1}{i} e^{i\omega_0 t}$ и примем τ за промежуток времени, по которому производится усреднение. Тогда $I_1 = I_2 = 1$. Произведение $E(t)E^*(t-\theta)$ отлично от нуля только в интервале $\theta < t < \tau$. Поэтому только при $\theta < \tau$ функция $\gamma(\theta)$ может отличаться от нуля. При $\theta > \tau$ она обращается в нуль. В первом случае

$$\overline{E(t)E^*(t-\theta)} = \frac{1}{\tau} \int_0^\tau e^{i\omega_0 \theta}\, dt = \frac{\tau - \theta}{\tau} e^{i\omega_0 \theta} = F(\theta) = f(\theta).$$

Следовательно,

$$\gamma(\theta) = \begin{cases} 1 - \theta/\tau & \text{при} \quad \theta < \tau, \\ 0 & \text{при} \quad \theta > \tau. \end{cases} \tag{31.11}$$

К тому же результату мы пришли бы, если бы предположили, что источник света излучает цуги волн одинаковой длительности τ, беспорядочно

следующие друг за другом, причем каждый цуг разделяется на две части, идущие к точке наблюдения различными путями. Это непосредственно следует из того, что различные цуги, испускаемые источником, *статистически независимы* и поэтому не интерферируют между собой. Из формулы (31.11) следует физически очевидный результат, что колебания когерентны, если время запаздывания θ меньше длины цуга τ. В противоположном случае они некогерентны. Значит, τ есть время когерентности колебаний.

5. Модуль функции $|\gamma_{12}(\theta)|$ легко вычислить по формуле (31.10), измерив предварительно видность полос V и интенсивности I_1 и I_2 накладывающихся пучков в точке наблюдения. Значительно труднее измерить добавочную фазу δ, входящую в формулу (31.9). Особенно трудно это сделать, когда источниками света являются узкие спектральные линии. Для этого надо сравнить в *одном и том же месте* интерференционной картины номера интерференционных полос от рассматриваемого источника света с номерами полос от источника с частотой ω_0. Для номера максимума N-й интерференционной полосы от первого источника можно написать $\omega_0\theta+\delta = 2\pi N$. В том же месте второй источник, вообще говоря, не даст максимума. Этому месту будет соответствовать уже дробное число интерференционных полос, определяемое условием $\omega_0\theta = 2\pi N_0$. Отсюда $\delta = 2\pi(N - N_0)$. Таким путем в принципе можно экспериментально определить не только модуль, но и аргумент комплексной степени когерентности $\gamma_{12}(\theta)$. Вместе с тем можно определить и корреляционную функцию $F_{12}(\theta)$.

6. Автокорреляционная функция $F(\theta)$ связана важным соотношением со спектральной плотностью $I_\omega(\omega)$ излучения. Для установления этой связи пишем на основании определения автокорреляционной функции:

$$F(\theta) = \overline{E(t)E^*(t-\theta)} = \frac{1}{\tau}\int\limits_{-\tau/2}^{\tau/2} E(t)E^*(t-\theta)\,dt. \tag{31.12}$$

Подставим сюда

$$E^*(t-\theta) = \int\limits_0^\infty a^*(\omega)e^{-i\omega(t-\theta)}\,d\omega$$

и поменяем порядок интегрирования по t и ω. Использовав при этом формулу (29.5), получим

$$F(\theta) = \frac{2\pi}{\tau}\int\limits_0^\infty a^*(\omega)a(\omega)e^{i\omega\theta}\,d\omega.$$

Но $(2\pi/\tau)a^*(\omega)a(\omega)$ есть спектральная плотность излучения $I_\omega(\omega)$ (см. § 29, п. 5). Следовательно,

$$F(\theta) = \int\limits_0^\infty I_\omega(\omega)e^{i\omega\theta}\,d\omega. \tag{31.13}$$

Эта формула представляет фурье-разложение функции $F(\theta)$, а потому

$$I_\omega(\omega) = \frac{1}{2\pi}\int\limits_{-\infty}^{+\infty} F(\theta)e^{-i\omega\theta}\,d\theta. \tag{31.14}$$

Формулу (31.14) можно привести к другому виду. Для этого заметим, что, ввиду стационарности светового потока, в формуле (31.12) пределы интегрирования можно заменить любыми другими, сохраняя только неизменной ширину интервала интегрирования. Используя это, нетрудно доказать, что автокорреляционная функция удовлетворяет соотношению $F(-\theta) = F^*(\theta)$. После этого формула (31.14) приводится к виду

$$I_\omega(\omega) = \frac{1}{2\pi}\left[\int_0^\infty F(\theta)e^{-i\omega\theta}\,d\theta + \int_0^\infty F^*(\theta)e^{i\omega\theta}\,d\theta\right]. \tag{31.15}$$

Это соотношение можно записать в символическом виде

$$I_\omega(\omega) = \frac{1}{\pi}\int_0^\infty F(\theta)e^{-i\omega\theta}\,d\theta, \tag{31.16}$$

понимая его в том смысле, что левая часть равенства равна вещественной части правой. Соотношение (31.13) позволяет найти корреляционную функцию $F(\theta)$ по экспериментально измеренной спектральной плотности излучения $I_\omega(\omega)$. С помощью обратного соотношения (31.16) можно определить спектральную плотность $I_\omega(\omega)$, если экспериментально определить корреляционную функцию $F(\theta)$.

Именно так поступал Майкельсон, используя свой интерферометр для исследования структуры спектральных линий. Он измерял видность интерференционных полос в интерферометре и фазу δ, входящую в формулу (31.9), и по этим данным вычислял спектральную плотность излучения $I_\omega(\omega)$. В свое время из всех методов этот метод был наиболее точным. Позднее метод Майкельсона был вытеснен более простыми методами многолучевой интерферометрии.

§ 32. Теорема Ван-Циттера–Цернике

Определим комплексную степень пространственной когерентности γ_{12} для точек Q_1 и Q_2 экрана $\mathcal{Э}$, освещаемого протяженным квазимонохроматическим самосветящимся источником света (рис. 132). Сами точки Q_1 и Q_2 должны рассматриваться как вторичные источники волн в том смысле, как это было указано в п. 3 предыдущего параграфа. Будем предполагать, что точка наблюдения P равноудалена от Q_1 и Q_2. При этом условии вместо волн, накладывающихся в точке P, можно брать волновые поля, создаваемые первичным источником в точках Q_1 и Q_2, что и будет делаться в дальнейшем. Наши вычисления, следовательно, проводятся в предположении, что время запаздывания θ равно нулю. Однако, ввиду медленности изменения функции $\gamma_{12}(\theta)$, вычисленным значением ее можно пользоваться не только при $\theta = 0$, но и при малых значениях аргумента θ, пока видность интерференционных полос не претерпит существенных изменений.

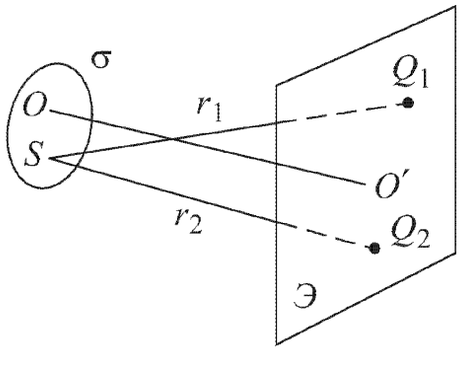


Рис. 132

Для простоты в качестве источника света возьмем малую площадку σ, плоскость которой параллельна плоскости экрана Э. Среда между источником σ и экраном Э предполагается однородной, а скорость света в ней обозначается через v. Линейные размеры площадки σ должны быть малы по сравнению с расстоянием ее до экрана. Предполагается также, что малы углы между «средней линией» OO' и прямыми, соединяющими произвольную точку S источника с точками Q_1 и Q_2.

Разобьем источник σ на малые площадки, линейные размеры которых малы по сравнению с длиной волны λ. Будем рассматривать их как некогерентные точечные источники, излучающие сферические волны. Волновые поля, создаваемые таким источником в точках Q_1 и Q_2, представятся выражениями

$$\frac{A_m(t-r_{1m}/v)}{r_{1m}}\,e^{\imath(\omega_0 t-kr_{1m})} \quad \text{и} \quad \frac{A_m(t-r_{2m}/v)}{r_{2m}}\,e^{\imath(\omega_0 t-kr_{2m})},$$

где r_{1m} и r_{2m} — расстояния m-го источника до точек Q_1 и Q_2 соответственно. Амплитуды результирующих колебаний в точках Q_1 и Q_2 будут

$$a_1(t)=\sum_m \frac{A_m(t-r_{1m}/v)}{r_{1m}}\,e^{-\imath kr_{1m}},$$

$$a_2(t)=\sum_m \frac{A_m(t-r_{2m}/v)}{r_{2m}}\,e^{-\imath kr_{2m}}.$$

Вычислим теперь взаимную корреляционную функцию амплитуд a_1 и a_2, т.е. среднее по времени от произведения $a_1(t)a_2^*(t-\theta)$, в предположении, что $\theta=0$. Перемножив почленно предыдущие суммы, заметим, что все слагаемые с различными m при усреднении обратятся в нуль, ввиду статистической независимости соответствующих элементарных источников света. В результате получим

$$a_1a_2^*=\sum_m \frac{\overline{A_m(t-r_{1m}/v)A_m^*(t-r_{2m}/v)}}{r_{1m}r_{2m}}\,e^{\imath k(r_{2m}-r_{1m})}.$$

(Усреднению подвергается только числитель, так как все прочие величины от времени не зависят.) Если предположить, что для всех m разности $r_{2m}-r_{1m}$ малы по сравнению с длиной когерентности, то различие аргументов $t-r_{1m}/v$ и $t-r_{2m}/v$ можно не учитывать. Следовательно, в силу предполагаемой однородности световых потоков можно опустить и сами аргументы, т.е.

$$\overline{A_m(t-r_{1m}/v)A_m^*(t-r_{2m}/v)}=\overline{A_mA_m^*}.$$

Во всех практически интересных случаях число элементарных излучателей света очень велико, так что их можно считать непрерывно распределенными по площадке σ с определенной поверхностной плотностью. Тогда от суммы можно перейти к интегралу. Если $I(S)$ — интенсивность света, создаваемая единицей площади источника на единичном расстоянии от него, то $\overline{A_mA_m^*}=I(S)\,dS$. В точках Q_1 и Q_2 соответственно

$$I_1\equiv I(Q_1)=\int\limits_\sigma \frac{I(S)\,dS}{r_1^2}, \quad I_2\equiv I(Q_2)=\int\limits_\sigma \frac{I(S)\,dS}{r_2^2}.$$

Введя еще нормирующий множитель $1/\sqrt{I_1I_2}$, получим окончательно

$$\gamma_{12}(0)=\frac{1}{\sqrt{I_1I_2}}\int \frac{I(S)}{r_1r_2}\,e^{\imath k(r_2-r_1)}\,dS. \tag{32.1}$$

Для наглядной интерпретации полученного результата воспользуемся следующей аналогией. Пусть точка Q_2 неподвижна, а точка Q_1 может занимать различные положения на экране $Э$. Заменим площадку σ отверстием σ' той же формы в непрозрачном экране. Допустим, что на него падает сферическая волна, сходящаяся в центре Q_2, волновое поле которой в точках отверстия представляется выражением

$$\psi = \frac{I(S)}{\sqrt{I_1 I_2}}\, e^{i(\omega_0 t + k r_2)} \equiv \psi_0(r_2) e^{i\omega_0 t}.$$

Пусть каждый элемент dS площади отверстия излучает по принципу Гюйгенса вторичную сферическую волну, поле которой в точке P_1 определяется выражением $\dfrac{1}{r_1}\,\psi_0 e^{i(\omega_0 t - k r_1)}$. Тогда в результате суперпозиции таких вторичных волн получится волна, комплексная амплитуда которой определяется формулой (32.1). Но именно так по принципу Гюйгенса решается задача о дифракции сферической волны на отверстии σ' (см. § 39). Следовательно, *комплексная степень взаимной когерентности в точках Q_1 и Q_2 ровна комплексный амплитуде в точке Q_1 соответствующей дифрагированной волны.*

В этом состоит *теорема Ван-Циттера–Цернике.* Она сводит вычисление степени взаимной когерентности γ_{12} к соответствующей задаче дифракции.

§ 33. Интерференция в пленках и пластинках

1. При освещении тонкой пленки происходит наложение волн от одного и того же источника, отразившихся от передней и задней поверхностей пленки. При этом может возникнуть интерференция. Если свет белый, то интерференционные полосы окрашены. Поэтому явление получило название *цветов тонких пленок.* Оно легко наблюдается на стенках мыльных пузырей, на тонких пленках масла или нефти, плавающих на поверхности воды, на пленках, возникающих на поверхности металлов при закалке (цвета побежалости), и т.д.

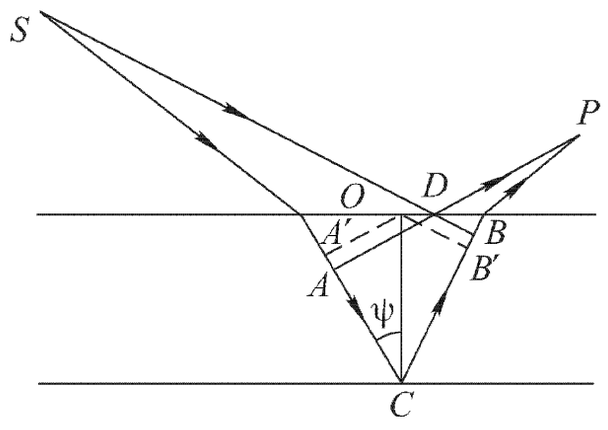


Рис. 133 *а*

2. Рассмотрим сначала плоскопараллельную пластинку толщины d с показателем преломления n, освещаемую точечным источником света S (рис. 133 *а*). При отражении от поверхностей пластинки возникает оптическая разность хода между соответствующими отраженными волнами. Вычислим ее в какой-либо произвольно выбранной точке P.

Пусть $SACBP$ и SDP — отраженные лучи, пересекающиеся в этой точке. Через точку D проведем плоскости DA и DB, перпендикулярные к соответствующим преломленным лучам в пластинке. Если пластинка тонкая, то для вычисляемой оптической разности хода можно написать: $\Delta = (SACBP) - (SDP)$. Через точку O, симметрично расположенную на верхней поверхности пластинки между точками A и B, проведем плоскости OA' и OB' (начерченные штриховыми линиями), параллельные плоскостям DA и DB. Тогда $\Delta = (A'CB') = 2(A'C) = 2nd\cos\psi$, где ψ — угол преломления.

Сюда необходимо, однако, ввести поправку. В пределе, когда толщина пластинки стремится к нулю, наша формула дает $\Delta = 0$. Поэтому в точке P должно было бы в этом случае происходить усиление колебаний. Но это невозможно, так как бесконечно тонкая пластинка вообще не может оказывать влияния на распространение света, так что никакого отражения возникнуть не может. Для этого волны, отраженные от передней и задней поверхностей пластинки, должны при интерференции гасить друг друга. Их фазы должны быть противоположны, т.е. оптическая разность хода Δ при $d \to 0$ должна стремиться к $\lambda/2$. Поэтому к прежнему выражению для Δ надо прибавить или отнять $\lambda/2$ (что совершенно безразлично). После этого получим [1])

$$\Delta = 2dn\cos\psi + \frac{\lambda}{2}. \tag{33.1}$$

Введение такой поправки, как будет показано в § 65, обусловлено изменением фазы на π, которое претерпевает волна при отражении на одной из поверхностей пластинки. Экспериментально это явление подтверждается наблюдением интерференции белого света в установке Ллойда (см. конец § 30).

Все рассуждения и результаты остаются в силе и для случая, когда точка P лежит по другую сторону пластинки (рис. 133 *б*). В этом случае отраженные лучи расходящиеся, в точке P пересекаются не сами лучи, а их продолжения за пластинку. Для наблюдения интерференции в точке P надо отраженные лучи сделать сходящимися с помощью собирающей линзы или вогнутого зеркала. Тогда интерференцию следует наблюдать в точке P', оптически сопряженной с точкой P.

Формула (33.1) справедлива и для тонких пластинок переменной толщины. Только в этом случае под d надо понимать толщину пластинки в том ее месте, где происходит отражение лучей, пересекающихся в точке P. Но в точку P от точечного

[1]) Неудовлетворительность приведенной аргументации была отмечена при выводе формулы (2.2).

источника света могут попадать, вообще говоря, только два луча. Отсюда следует, что в случае точечного монохроматического источника света каждая точка пространства характеризуется вполне определенной разностью хода приходящих в нее отра-

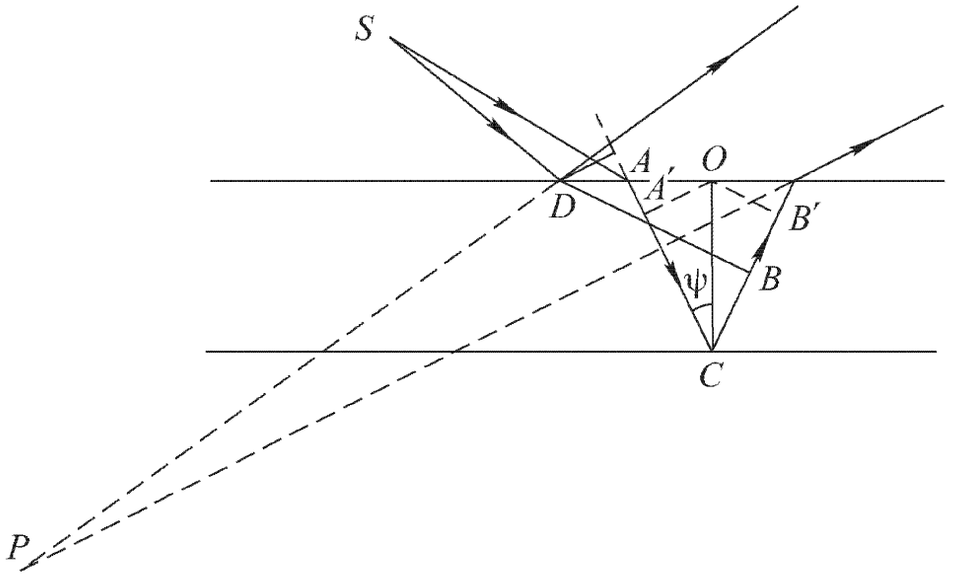


Рис. 133*б*

женных лучей. Поэтому устойчивая интерференция должна наблюдаться *в каждой точке пространства.* Про соответствующие полосы интерференции говорят, что они *не локализованы* (или локализованы всюду).

3. Положение существенно изменится, если источник света протяженный. Тогда в точку P будут попадать, вообще говоря, два отраженных луча от каждой точки источника. Место отражения на пластинке, углы падения и отражения, а следовательно, и разность хода для каждой пары лучей будут разными. Поэтому никакой устойчивой интерференции в точке P наблюдаться не будет.

Допустим, однако, что точка P находится близко от пластинки, например совпадает с точкой B. Тогда при малой толщине пластинки отражение будет происходить практически в одном и том же месте, т.е. при одной и той же толщине d, независимо от того, из какой точки источника исходят лучи. Если при этом лучи падают на пластинку почти нормально (так что $\cos\psi$ изменяется мало), то оптическая разность хода Δ будет зависеть только от толщины пластинки d в точке падения лучей. Линии на поверхности пластинки, где толщина d постоянна, будут также линиями постоянной разности фаз. На поверхности пластинки, если только степень монохроматичности света достаточна, появятся интерференционные полосы, каждая из которых характеризуется условием $d = \text{const}$. Они называются *полосами* или *линиями равной толщины.* Такие полосы как бы

нарисованы на самой пластинке. Про них говорят, что они *локализованы на пластинке*. Конечно, интерференционные полосы должны наблюдаться не только на пластинке, но и с обеих сторон вблизи от нее. Однако максимальной видностью обладают интерференционные полосы на самой пластинке.

Для наблюдения таких полос удобно воспользоваться собирательной линзой, с помощью которой можно получить изображение пластинки на экране. Так как линза не вносит дополнительной разности хода, то при этом на экране получается изображение и интерференционных полос. Линза как бы переносит место локализации интерференционной картины с поверхности пластинки на экран. При визуальном наблюдении полос равной толщины глаз надо аккомодировать на пластинку. Роль линзы выполняет хрусталик, а экрана — сетчатая оболочка глаза. Оптический прибор или глаз выполняет также и другую полезную функцию. Диафрагма прибора или зрачок глаза вырезают из отраженных лучей узкие пучки, в пределах которых угол ψ меняется незначительно. Тем самым создаются условия, благоприятные для получения полос равной толщины.

При наблюдении в белом свете полосы интерференции окрашены, так как разность фаз $\dfrac{4\pi\, dn\, \cos\psi}{\lambda} + \pi$ зависит от λ. В белом свете возможна интерференция только *низкого порядка*. Поэтому пленка должна быть тонкой — ее толщина не должна превышать примерно одной-двух длин волн. Когда толщина пленки составляет доли длины волны, то разность фаз приближается к π, и пленка становится темной.

Если обе поверхности пленки плоские, то интерференционные полосы прямолинейны и параллельны линии пересечения соответствующих плоскостей. Такие полосы наблюдаются, например, в клине, т.е. тонкой воздушной прослойке между плоскопараллельными стеклянными пластинками, когда с одного края между ними проложен, например, тонкий лист бумаги. Но если поверхности сложенных стеклянных пластинок неровные, то полосы равной толщины принимают неправильную, причудливую форму. На этом основан чувствительный интерференционный *метод контроля* поверхности на плоскопараллельность. В этом методе испытуемая поверхность прижимается к плоской и наблюдаются полосы интерференции в образовавшейся прослойке. Метод применяется и для контроля сферических или параболических поверхностей при шлифовке оптических зеркал и линз.

4. Интерференция в тонких пленках рассматривалась выше как *двухлучевая интерференция*. Мы ограничились интерференцией только двух волн, одна из которых получилась при од-

нократном отражении от верхней, а другая — от нижней поверхностей пленки. Многократными отражениями мы пренебрегли. Это можно делать, когда коэффициент отражения невелик. Если же коэффициент отражения близок к единице, то так поступать нельзя. Допустим, например, что при каждом отражении отражается 5 % падающего света, а остальные 95 % проходят. Примем интенсивность падающего света за единицу. Тогда, как нетрудно подсчитать, интенсивности отраженных и прошедших лучей (рис. 134) представятся следующей таблицей:

Лучи	1	2	3	4	1′	2′
Интенсивности	1	0,05	0,0451	0,00011	0,9025	0,00226

Из таблицы видно, что интенсивности отраженных лучей *2* и *3* почти одинаковы, а интенсивность луча *4* более чем в 100 раз меньше. Поэтому луч *4* и все отражения высших порядков можно не принимать во внимание. Из прошедших лучей интенсивность луча *2′* примерно в 400 раз меньше интенсивности луча *1′*. Поэтому интерференционные полосы в проходящем свете получаются *на светлом фоне* и по этой причине очень мало контрастны.

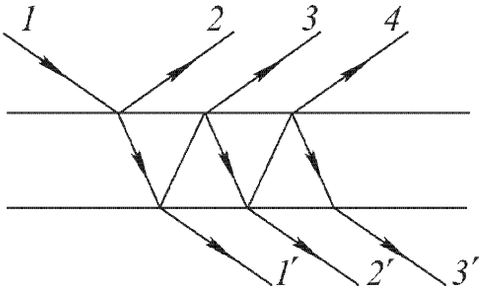


Рис. 134

5. Допустим теперь, что пластинка толстая и строго *плоскопараллельная*. Пусть она освещается параллельным пучком света. Формально это соответствует случаю точечного бесконечно удаленного источника S (см. рисунки 133 *а* и 133 *б*). Отраженные лучи будут также параллельными, т.е. точка наблюдения P удалится в бесконечность. При постоянной толщине пластинки d оптическая разность хода между отраженными лучами $2dn\cos\psi + \lambda/2$ зависит только от угла наклона падающих лучей. Угол ψ может принимать всевозможные значения, если источник света протяженный и имеет конечные угловые размеры. Практически это можно осуществить, поместив протяженный источник света в фокальной плоскости линзы, которая как бы удаляет источник света в бесконечность. Интерференционную картину следует наблюдать на бесконечно (т.е. достаточно) удаленном экране или в фокальной плоскости линзы, поставленной на пути отраженных лучей. Каждая интерференционная полоса на бесконечно удаленном экране характеризуется постоянством косинуса угла ψ. Поэтому интерференционные полосы при описанном способе наблюдения называют

полосами или линиями равного наклона [1]). *Они локализованы в бесконечности.*

Линза переносит область локализации из бесконечности в фокальную плоскость, оптически сопряженную с бесконечно удаленной плоскостью. Роль линзы может играть глаз, аккомодированный на бесконечность. В этом случае полосы равного наклона локализуются на *сетчатке глаза.* При наблюдении в зрительную трубу последняя должа быть установлена на бесконечность.

Для толстых пластинок оптическая разность хода Δ велика, т.е. содержит тысячи и десятки тысяч длин волн. Интерференционные полосы будут высокого порядка. Для их получения требуется высокая степень монохроматичности $\lambda/\delta\lambda$ падающего света.

6. Ньютон наблюдал интерференционные полосы равной толщины в воздушной прослойке между плоской поверхностью стеклянной пластинки и плоско-выпуклой линзой, прижатой к пластинке выпуклой стороной. Интерференционные полосы имели форму концентрических колец. Они получили название *колец Ньютона.* При постановке опыта радиус кривизны выпуклой поверхности линзы надо брать не менее приблизительно одного метра. Кольца удобно наблюдать в длиннофокусный микроскоп, сфокусированный на воздушную прослойку. Более контрастные кольца наблюдаются, конечно, в отраженном, а не в проходящем свете. Чтобы наблюдалось много колец, надо пользоваться светом сравнительно высокой монохроматичности. Подходящим может быть желтый свет натровой горелки или свет ртутной лампы. Вид колец показан на рис. 135.

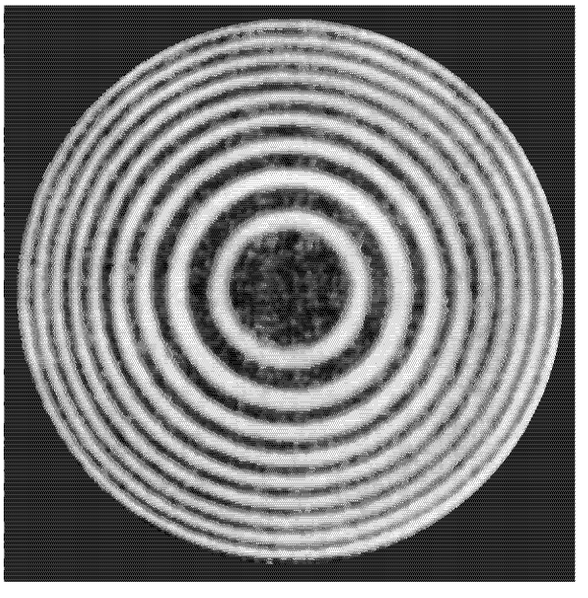


Рис. 135

Для вычисления радиусов колец дополним выпуклую поверхность линзы до полной сферы (рис. 136). Если BD — диаметр, то по известной геометрической теореме $AB \cdot AD = AO^2 = x^2$. Ввиду малости кривизны сферы и угла OCA, отрезок AB можно принять за толщину d воздушной

[1]) Этот общепринятый термин не совсем точно передает существо дела. Лучам одинакового наклона соответствует не линия, а точка — фокус линзы, в котором они собираются.

прослойки в точке A. Пренебрегая еще различием между AD и диаметром сферы $2R$, получим

$$d = \frac{x^2}{2R}, \quad \Delta = 2d + \frac{\lambda}{2} = \frac{x^2}{R} + \frac{\lambda}{2}.$$

Светлые кольца получатся при $\Delta = m\lambda$, где m — целое число. Из этого условия находим радиус x_m m-го светлого кольца:

$$x_m = \sqrt{(m - 1/2)\lambda R} = \sqrt{\lambda R/2}\sqrt{2m - 1}. \qquad (33.2)$$

Аналогично для радиуса m-го темного кольца

$$x_m = \sqrt{m\lambda R} = \sqrt{\lambda R/2}\sqrt{2m}. \qquad (33.3)$$

Таким образом, радиусы последовательных светлых колец пропорциональны квадратным корням из нечетных чисел 1, 3, 5, ..., а радиусы темных колец — квадратным корням из последовательных четных чисел 0, 2, 4, ... Эти закономерности экспериментально были установлены Ньютоном. Как указывалось в § 3 (п. 6), в интерференции Ньютон видел проявление *периодичности* световых процессов. Из наблюдения интерференционных колец Ньютон даже довольно точно вычислил величину, являющуюся количественной мерой указанной периодичности. В переводе на язык волновой теории вычисленная Ньютоном величина есть половина длины световой волны.

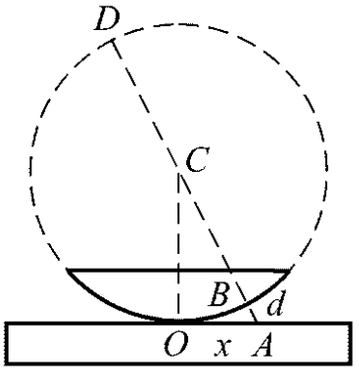


Рис. 136

Центр колец в отраженном свете темный, а в проходящем светлый. Это доказывает, что при отражении на одной из границ воздушной прослойки фаза отраженной волны меняется на π. Как будет показано в § 65, электрический вектор меняет фазу на π при отражении от среды с бóльшим показателем преломления. В противоположном случае изменения фазы не наблюдается. Это подтверждается следующим опытом Юнга. Он взял пластинку из флинта ($n = 1{,}7$), прижал к ней линзу из крона ($n = 1{,}5$), а прослойку между ними заполнил сассафрасовым маслом, показатель преломления которого имел промежуточное значение. Тогда фаза волны менялась на π при отражении как от верхней, так и от нижней поверхностей масляной прослойки. Благодаря этому центр ньютоновых колец получался светлым в отраженном и темным в проходящем свете.

7. Если в установке Ньютона линзу перемещать вверх параллельно самой себе, то из-за увеличения толщины воздушной прослойки каждая окружность постоянной (фиксированной) разности хода будет стягиваться к центру картины. Интерференционные кольца также станут стягиваться к центру, так как вдоль

каждого кольца оптическая разность хода остается постоянной. Достигнув центра, каждое кольцо превратится в кружок, исчезающий при дальнейшем перемещении линзы. Таким образом, центр картины будет становиться попеременно то светлым, то темным. Одновременно на периферии поля зрения будут зарождаться и перемещаться к центру новые интерференционные кольца, пока каждое из них не исчезнет в центре картины. Можно сказать, что при перемещении линзы непрерывно вверх пропадают кольца самых низких порядков интерференции и зарождаются кольца более высоких порядков. Таким путем удается наблюдать интерференцию все более и более высокого порядка. Понятно, что для этого требуется свет высокой степени монохроматичности.

Физо, впервые осуществивший подобный опыт, пользовался желтым светом натровой горелки (в опыте Физо интерференционные полосы получались между двумя плоскопараллельными пластинками). Физо заметил, что при увеличении числа полос N, прошедших в поле зрения, ухудшалась видимость полос. При $N = 490$ видимость достигала минимума, затем она опять улучшалась; при $N = 980$ полосы приобретали прежнюю отчетливость. Ближайший минимум наступал при $N = 1470$, следующий максимум при $N = 1960$ и т.д. Физо сумел заметить 52 таких максимума. На этом основании он пришел к заключению, что желтая линия натрия — двойная, т.е. состоит из двух близко расположенных спектральных линий с длинами волн λ и $\lambda' > \lambda$.

Легко найти расстояние между этими линиями $\lambda' - \lambda$. Как было показано в § 30, первое исчезновение видимости интерференционных полос происходит при условии $\lambda' - \lambda = \lambda/2N$. По наблюдениям Физо $N = 490$. Средняя длина волны желтой спектральной линии натрия $\approx 589{,}3$ нм. Поэтому $\lambda' - \lambda \approx 0{,}6$ нм.

Этот пример показывает, как интерференция высокого порядка может быть использована для разрешения сложных спектральных линий, состоящих из двух близко расположенных линий.

§ 34. Интерферометр Жамена

1. *Интерферометрами* называют оптические измерительные приборы, основанные на интерференции света. Они позволяют с высокой степенью точности измерять линейные и угловые расстояния, малые разности показателей преломления, исследовать структуру спектральных линий и т.п. Принцип действия интерферометров в основном один и тот же. Но в зависимости от предназначения им придают различные конструкции.

2. *Интерферометр Жамена* (1818–1886) состоит из двух толстых (толщина не менее 20 мм) плоскопараллельных пластин P_1

и P_2, изготовленных из весьма однородного стекла и смонтированных на массивной плите (рис. 137). Для исследований в далекой ультрафиолетовой области спектра пластины изготовляются из кварца или флюорита. Чтобы обе пластины имели строго одинаковую толщину, они вырезаются из одной и той же пластины. Пластины устанавливаются приблизительно под углом 45° к линии, соединяющей их центры. Угол между пластинами можно менять с помощью специальных установочных винтов, поворачивающих пластины вокруг вертикальной и горизонтальной осей. Чем меньше этот угол, тем шире интерференционные полосы. Задние отражающие поверхности пластин посеребрены. Луч света SA, отразившись от передней и задней поверхностей пластины P_1, разделяется на два параллельных луча AB и $C'D'$, между которыми возникает разность хода. Каждый из этих лучей в свою очередь раздваивается при отражении от пластины P_2. Между средними лучами DE и $D'E'$ возникает разность хода

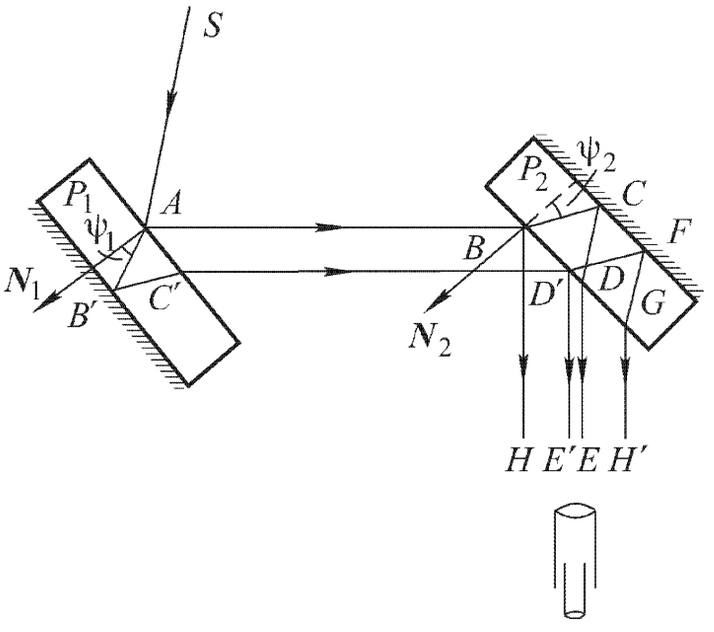


Рис. 137

$$\Delta = 2dn(\cos\psi_2 - \cos\psi_1), \qquad (34.1)$$

где d — толщина пластин, ψ_1 и ψ_2 — углы преломления в них, n — показатель преломления пластин относительно окружающего воздуха. Таким образом, на величину Δ пластины влияют противоположно. Потери фаз при отражении в формулу (34.1) не входят, так как они одинаковы для обоих интерферирующих лучей и компенсируют друг друга. Если $\psi_1 = \psi_2$, то $\Delta = 0$. Если же угол между пластинами мал, то и разность хода невелика. Это позволяет наблюдать интерференцию низкого порядка, используя белый свет. Лучи DE и $D'E'$ будут интерферировать так же, как и соответствующие лучи в случае только одной, но тонкой плоскопараллельной пластинки.

Назначение двух толстых пластин состоит в том, чтобы далеко развести лучи AB и $C'D'$, что позволяет ставить на пути этих лучей трубки с исследуемыми газами. Разность хода между средними и крайними лучами, а также между самими крайними лучами очень велика, и интерференцию в белом свете

от этих лучей получить невозможно. Источником света обычно служит освещаемая вертикальная щель шириной около одного или нескольких миллиметров, прикрываемая матовым стеклом. Невооруженным глазом видны три изображения щели. Самое яркое (справа) дают лучи, дважды отразившиеся от задних посеребренных поверхностей пластин. Среднее, промежуточной яркости, образуется лучами DE и $D'E'$. Наименее яркое (слева) соответствует лучам, дважды отразившимся от передних поверхностей пластин. Если фиксировать глаз на среднем изображении и вращением установочных винтов менять разность хода, то в белом свете можно получить цветные интерференционные полосы низкого порядка. Центральная ахроматическая полоса, т.е. полоса, которой соответствует разность хода $\Delta = 0$, получается белой. Наблюдая эту полосу, можно легко следить за смещением интерференционной картины.

Так как разность хода Δ зависит от направления падающих лучей, то в интерферометре Жамена получаются полосы равного наклона. Для их наблюдения удобно воспользоваться зрительной трубой небольшого увеличения, но большого поля зрения, установленной на бесконечность. Применение трубы не только увеличивает ширину полос, но и обладает еще тем преимуществом, что в трубу не попадают крайние лучи GH' и BH, создающие светлый фон и сильно уменьшающие контрастность интерференционных полос. В лекционных демонстрациях интерференционные полосы можно получать без дополнительных оптических приспособлений, направив лучи DE и $D'E'$ непосредственно на белый экран. Крайние лучи при этом должны быть экранированы.

3. Интерферометр Жамена предназначен для измерения небольших изменений показателей преломления. Поэтому его называют также *интерференционным рефрактометром.* Для уяснения принципа действия такого рефрактометра вообразим, что на пути одного из интерферирующих лучей AB или $C'D'$ (см. рис. 137) помещен плоскопараллельный слой какого-либо вещества толщины l с показателем преломления n_2. Тогда разность хода между интерферирующими лучами изменится на величину $(n_2 - n_1)l$, где n_1 — показатель преломления окружающего воздуха. В результате интерференционная картина сместится на m полос, причем $m = (n_2 - n_1)l/\lambda$. Число m (вообще говоря, дробное) можно найти, наблюдая интерференционные полосы в белом свете до и после внесения исследуемого вещества. Опыт показывает, что смещение на $1/10$ полосы ($m = 1/10$) наблюдается вполне уверенно и без труда.

Разность показателей преломления можно вычислить по формуле $n_2 - n_1 = m\lambda/l$. Если толщина слоя вещества $l = 10$ см,

а $\lambda = 500$ нм, то при $m = 1/10$ получаем $n_2 - n_1 = 5 \cdot 10^{-7} \sim 10^{-6}$. В специальных установках измерялись гораздо меньшие изменения показателя преломления. Интерферометр Жамена позволяет, таким образом, измерять ничтожные изменения показателя преломления, например, при изменении температуры газа или при добавлении к нему посторонних примесей. Измерение самих абсолютных значений показателя преломления этим прибором затруднительно. Обычно его используют для измерения разности показателей преломления исследуемого и хорошо изученного газа, например воздуха.

При практических измерениях на пути лучей AB и $C'D'$ ставятся по возможности одинаковые трубки. Сначала наблюдаются интерференционные полосы, когда обе трубки наполнены воздухом. Затем измеряется смещение полос интерференции, когда одна из трубок наполняется исследуемым газом. Для компенсации значительной разности хода между интерферирующими лучами служит компенсатор. Он состоит из двух совершенно одинаковых стеклянных пластинок, насаженных на общую ось, причем угол наклона между пластинками можно плавно менять. Одна из пластинок ставится на пути одного, вторая — на пути другого лучей. Вращая компенсатор при неизменном угле наклона между пластинками, можно привести в поле зрения центральную ахроматическую полосу, а затем по углу поворота определить число полос, прошедших через поле зрения.

4. Не так просто разобраться в вопросе о форме интерференционных полос, наблюдаемых в интерферометре Жамена. Для упрощения исследования этого вопроса введем несущественное предположение, что $n = 1$. Такому предположению соответствует замена каждой пластины парой бесконечно тонких параллельных отражающих плоскостей (из которых одна посеребрена). Тогда вместо (34.1) получится более простое выражение:

$$\Delta = 2d(\cos\psi_2 - \cos\psi_1).$$

Перейдем к векторной форме записи, обозначив через $\mathbf{N}_1$ и $\mathbf{N}_2$ единичные нормали к отражающим плоскостям, как это показано на рис. 137. Обозначим через $\mathbf{s}_1$, $\mathbf{s}_1'$, $\mathbf{s}$ единичные векторы лучей SA, AB и BH соответственно (те же векторы определяют направления и всех параллельных им лучей). Тогда

$$\cos\psi_1 = (\mathbf{N}_1\mathbf{s}_1') = -(\mathbf{N}_1\mathbf{s}_1'), \quad \cos\psi_2 = (\mathbf{N}_2\mathbf{s}).$$

По закону отражения $\mathbf{s}_1' = \mathbf{s} - 2(\mathbf{N}_2\mathbf{s})\mathbf{N}_2$, а потому

$$\cos\psi_1 = -(\mathbf{N}_1\mathbf{s}) + 2(\mathbf{N}_2\mathbf{s})(\mathbf{N}_1\mathbf{N}_2).$$

Следовательно,

$$\Delta = 2d(\mathbf{N}_1 + \mathbf{N}_2)\mathbf{s} - 4d(\mathbf{N}_2\mathbf{s})(\mathbf{N}_1\mathbf{N}_2).$$

Введем новые векторы

$$\mathbf{N} = \frac{1}{2}(\mathbf{N}_1 + \mathbf{N}_2), \quad \boldsymbol{\nu} = \frac{1}{2}(\mathbf{N}_2 - \mathbf{N}_1).$$

Из них первый направлен по биссектрисе угла между нормалями $\mathbf{N}_1$ и $\mathbf{N}_2$, а второй перпендикулярен к этой биссектрисе. Очевидно,

$$\mathbf{N}_1 = \mathbf{N} - \boldsymbol{\nu}, \quad \mathbf{N}_2 = \mathbf{N} + \boldsymbol{\nu}.$$

После подстановки этих выражений в формулу для Δ получим

$$\Delta = -4d(1 - \boldsymbol{\nu}^2)(\boldsymbol{\nu}\mathbf{s}) + 4d\boldsymbol{\nu}^2. \tag{34.2}$$

При работе с прибором угол между пластинами, а следовательно, и длина вектора $\boldsymbol{\nu}$ всегда малы. Поэтому квадратичными членами, содержащими $\boldsymbol{\nu}^2$, в формуле (34.2) обычно можно пренебречь.

5. Для светлых полос $\Delta = m\lambda$, для темных $\Delta = (m+{}^1\!/_2)\lambda$, где m — целое число (порядок интерференции). Если для наблюдения используется труба, то она сводит параллельные лучи в фокальной плоскости объектива, где и получается первичная интерференционная картина. Поле зрения трубы ограничено, так что используется лишь небольшой участок фокальной плоскости. При повороте трубы вокруг оптического центра объектива поворачивается и фокальная плоскость вместе с выделенным участком. Если последний мал, то без заметной ошибки можно принять все такие участки за элементы неподвижной сферы, центр которой совпадает с оптическим центром объектива. Назовем для краткости такую вспомогательную сферу *фокальной сферой объектива.* (Если объектив не является тонкой линзой, то все это можно повторить, заменив только оптический центр второй узловой точкой объектива.) Можно сказать, что первичная интерференционная картина как бы создается на неподвижной фокальной сфере, а наблюдатель рассматривает эту картину через окуляр трубы. Поворачивая трубу, он видит на фокальной сфере различные участки интерференционной картины. Разумеется, вид картины совершенно не зависит от того, вокруг какой точки вращалась труба, он определяется только *направлением* последней.

Таким образом, вопрос о форме видимых полос интерференции сводится к вопросу о форме их на вспомогательной фокальной сфере. Практически, однако, вращение трубы всегда производится вокруг вертикальной оси. Поэтому наблюдению доступен только узкий экваториальный поясок на фокальной сфере. Его приближенно можно считать поверхностью цилиндра с вертикальной осью, проходящей через оптический центр объектива. В дальнейшем ограничимся рассмотрением интерференционных полос только на таком «цилиндрическом фокальном пояске».

Фиксировав положение пластин интерферометра, отложим векторы $\mathbf{N}$ и $\boldsymbol{\nu}$, а также единичный вектор луча $\mathbf{s}$ из общего начала O, помещенного в оптическом центре объектива. Направления этих векторов будут играть роль координатных осей, перпендикулярную к ним прямую примем за ось Z. Различные единичные векторы лучей $\mathbf{s}$ будем откладывать из того же начала O. Как видно из (34.2), на поверхностях равных фаз $(\boldsymbol{\nu}\mathbf{s}) = \mathrm{const}$, т.е. эти поверхности представляют собой параллельные плоскости, перпендикулярные к вектору $\boldsymbol{\nu}$. Среди этих плоскостей находятся и равноотстоящие плоскости, на которых разность хода Δ содержит целое или полуцелое число длин волн. Они пересекают фокальную сферу вдоль некоторых кривых, которые и представляют собой интерференционные полосы, наблюдаемые в трубу. Ввиду малости поля зрения, при каждом положении трубы наблюдается лишь незначительный участок фокальной сферы, на котором интерференционные полосы практически неотличимы от прямолинейных.

В интерферометре Жамена первая пластина обычно неподвижна и установлена вертикально. Вторую пластину можно немного поворачивать вокруг вертикальной и горизонтальной осей, параллельных отражающим плоскостям. Если пластины строго параллельны друг другу, то $\boldsymbol{\nu} = 0$ и $\Delta = 0$. Следовательно, никаких интерференционных полос нет, все поле зрения освещено равномерно. В этом положении нормали $\mathbf{N}_1$ и $\mathbf{N}_2$, а следовательно, и вектор $\mathbf{N}$ лежат в горизонтальной плоскости.

Повернем теперь вторую пластину на малый угол вокруг вертикальной оси. Тогда вектор $\boldsymbol{\nu}$ расположится в той же го-

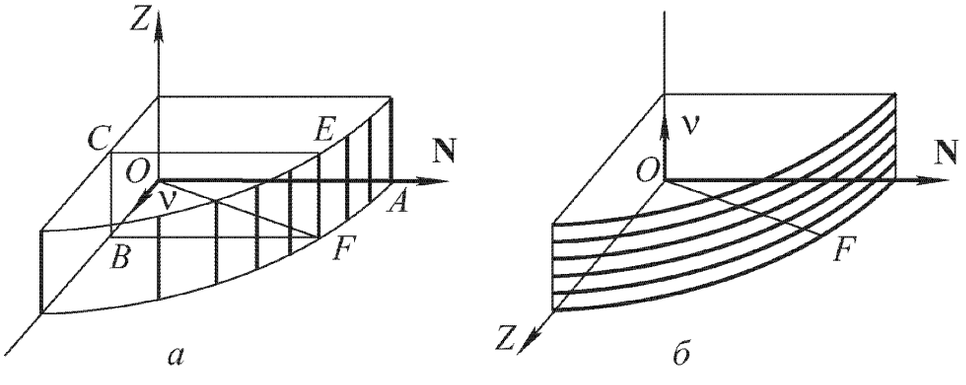


Рис. 138

ризонтальной плоскости (рис. 138 *а*). Перпендикулярные к нему плоскости равных фаз вертикальны. Одна из таких плоскостей $BCEF$ изображена на рис. 138 *а*. Она пересекает экваториальный цилиндрический поясок вдоль вертикального отрезка EF. Следовательно, интерференционные полосы, лежащие на этом пояске, будут вертикальны. На схематическом рис. 138 *а* они представлены жирными отрезками. Ширина интерференцион-

ных полос минимальна при наблюдении вдоль $\mathbf{N}$ и максимальна при наблюдении вдоль $\boldsymbol{\nu}$. Однако фактически ось трубы OF бывает направлена под углом $\approx 45°$ к нормали $\mathbf{N}$. Тогда вертикальные интерференционные полосы могут наблюдаться только в монохроматическом (квазимонохроматическом) свете, так как им соответствуют высокие порядки интерференции. Действительно, если наблюдение ведется вдоль $\mathbf{N}$, то $(\mathbf{s}\boldsymbol{\nu}) = 0$, и формула (34.2) дает $\Delta = 4d\boldsymbol{\nu}^2$. Ввиду малости $\boldsymbol{\nu}$ порядок интерференции получается низким. Нулевая полоса практически проходит через точку A.

Когда труба направлена вдоль $\mathbf{N}$ или близко к этому направлению, вертикальные интерференционные полосы можно увидеть и в белом свете. Но при удалении от этого направления скалярное произведение $(\boldsymbol{\nu}\mathbf{s})$, а с ним и порядок интерференции возрастают. При повороте на угол $\approx 45°$ порядок интерференции становится столь высоким, что вертикальные интерференционные полосы в белом свете наблюдаться уже не могут. Исключение было бы только в случае очень малых $\boldsymbol{\nu}$, когда порядок интерференции при любых положениях трубы низкий. Но тогда интерференционная полоса была бы столь широка, что не уместилась бы в поле зрения трубы. Поэтому этот случай не представляет никакого интереса.

Совсем иное положение возникнет, если вторую пластину из исходного положения повернуть вокруг горизонтальной оси. Тогда вектор $\boldsymbol{\nu}$ встанет вертикально (рис. 138 *б*), а перпендцкулярные к нему плоскости равных фаз сделаются горизонтальными. Интерференционные полосы определятся, если пересечь такими плоскостями поверхность экваториального цилиндрического пояска. На рис. 138 *б* они изображены жирными дугами окружностей, плоскости которых горизонтальны. При этом все они будут низкого порядка, а потому могут наблюдаться в белом свете, так как наблюдение ведется практически перпендикулярно к $\boldsymbol{\nu}$, когда $(\boldsymbol{\nu}\mathbf{s}) \approx 0$. В поле зрения трубы будут видны практически прямолинейные горизонтальные полосы. Легко сообразить, что при наклоне вектора $\boldsymbol{\nu}$ интерференционные полосы сделаются также наклонными к горизонту. Но в белом свете могут наблюдаться только интерференционные полосы малого наклона. Для получения полос с большим наклоном надо пользоваться монохроматическим (квазимонохроматическим) светом.

Все приведенные рассуждения можно повторить по той же схеме и не вводя предположения, что $n = 1$. Для избежания громоздкости вычислений надо только всюду ограничиться членами, линейными по $\boldsymbol{\nu}$, отбрасывая все члены более высоких степеней. При этом окончательные результаты не претерпят никаких существенных изменений.

6. Интерферометр Жамена обладает рядом недостатков, делающих его непригодным в некоторых особенно тонких оптических исследованиях. Изготовить пластины толще 5 см из вполне однородного стекла с точно параллельными плоскостями практически невозможно. Поэтому и интерферирующие лучи AB и $C'D'$ невозможно развести на расстояние больше $\sim$ 4 см. Но главный недостаток интерферометра Жамена состоит в том, что толстые пластины при освещении медленно прогреваются и медленно приходят в состояние теплового равновесия. Это приводит к тому, что интерференционные полосы часами медленно ползут в поле зрения трубы. Крайне затруднена работа в ультрафиолетовой области, требующая уникальных пластин из кварца или флуорита. Указанные недостатки устранены в интерферометре Д.С. Рождественского (1876–1940).

Интерферометр Рождественского принципиально не отличается от интерферометра Жамена. Каждая пластина в нем заменена комбинацией из двух параллельно установленных тонких пластинок M_1, P_1 и M_2, P_2 (рис. 139). Пластинки M_1 и M_2 посеребрены и выполняют роль зеркал. Пластинки P_1 и P_2, также посеребрены, полупрозрачны, они пропускают половину падающего света, а остальную половину отражают. Комбинация параллельных пластинок M_2, P_2 может поворачиваться на малый угол относительно другой комбинации также параллельных пластинок M_1, P_1. Ход лучей в интерферометре показан на схематическом рис. 139. Прибор действует так же, как и интерферометр Жамена. В нем наблюдаются интерференционные полосы равного наклона.

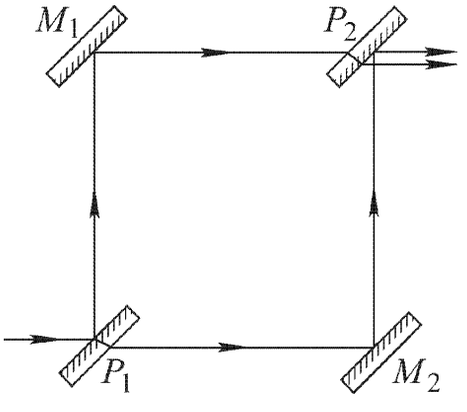


Рис. 139

Аналогично устроен построенный несколько раньше интерферометр Маха–Цендера. Отличие от интерферометра Рождественского состоит в том, что параллельно устанавливаются между собой зеркала M_1 и M_2, а также полупрозрачные пластинки P_1 и P_2. Одна пара пластин может поворачиваться на малый угол относительно другой пары.

§ 35. Интерферометр Майкельсона

С помощью *интерферометра Майкельсона* (1852–1931) впервые было проведено систематическое изучение тонкой структуры спектральных линий и выполнено первое прямое сравнение эталонного метра с длиной световой волны. Знаменитый опыт

Майкельсона–Морли, целью которого было обнаружение движения Земли относительно эфира (см. § 102), был выполнен также с помощью интерферометра Майкельсона. В настоящее время интерферометр Майкельсона устарел. Однако историческое значение указанных выше исследований, в особенности последнего, слишком велико. Поэтому необходимо хотя бы кратко остановиться на описании устройства и работы интерферометра Майкельсона.

Схема интерферометра приведена на рис. 140. Свет от протяженного источника S попадает на плоскопараллельную раз-

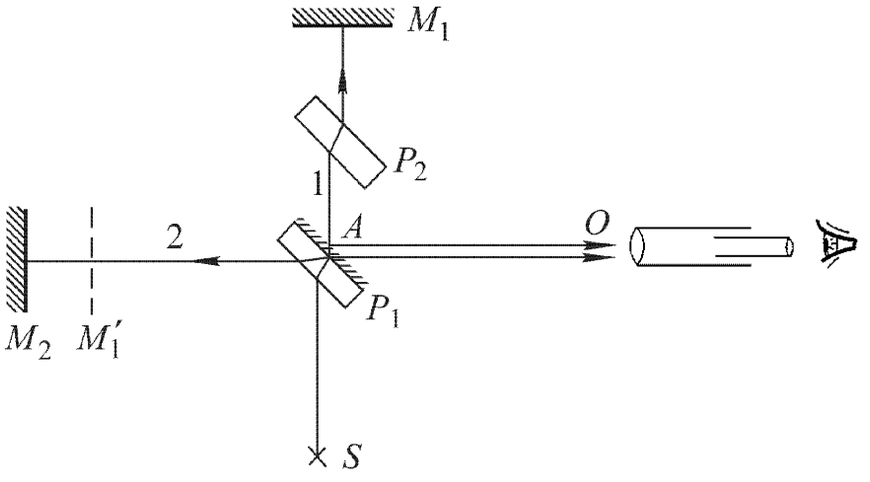


Рис. 140

делительную пластинку P_1, покрытую полупрозрачным тонким слоем серебра или алюминия. Эта пластинка частично пропускает, частично отражает свет, разделяя падающий пучок на два взаимно перпендикулярных пучка. Первый пучок, пройдя через пластинку P_1, отражается обратно зеркалом M_1, а затем частично отражается от пластинки P_1 в направлении AO. Второй пучок, отразившись от покрытой металлом поверхности пластинки P_1, направляется к зеркалу M_2, отражается от него, снова проходит через пластинку P_1 и далее идет в направлении AO, как и первый пучок.

Таким образом, от одного и того же источника получаются два пучка лучей одинаковой интенсивности, идущих в зрительную трубу, где и наблюдаются интерференционные полосы.

На пути первого пучка лучей ставится пластинка P_2, тождественная пластинке P_1. Она компенсирует разность хода между пучками, возникающую из-за того, что второй пучок пересекает разделительную пластинку P_1 три, а первый только один раз. Так как стекло обладает дисперсией, то без такой компенсации наблюдение интерференции в белом свете было бы невозможно.

Зеркало M_1 неподвижно, а M_2 с помощью микрометрического винта может перемещаться на специальных салазках строго

параллельно самому себе. В больших интерферометрах такое перемещение должно происходить на десятки сантиметров. Поэтому к механической части прибора предъявляются исключительно высокие требования. Зеркала интерферометра снабжены также установочными винтами, позволяющими придавать им правильное положение.

Пусть M_1' — изображение поверхности зеркала M_1 в отражающей плоскости разделительной пластинки P_1. Тогда интерференция будет происходить так же, как и в воздушном слое между двумя отражающими плоскостями M_2 и M_1'. Разность хода между отраженными лучами $\Delta = 2d\cos\varphi$, где d — толщина слоя, а φ — угол падения. Если слой плоскопараллелен, то будут получаться интерференционные полосы равного наклона, локализованные в бесконечности. Их можно наблюдать глазом, аккомодированным на бесконечность, или в трубу, установленную также на бесконечность. Получатся интерференционные кольца с центром в точке схождения лучей, нормально отраженных от поверхностей M_2 и M_1'. Этому направлению соответствует максимальная разность хода $\Delta = 2d$. Поэтому максимальный порядок интерференции будет наблюдаться в центре картины. Отсюда следует, что при увеличении толщины d воздушного зазора полосы интерференции будут перемещаться в направлении от центра (в противоположность тому, что было с кольцами Ньютона, см. § 33, п. 7).

При увеличении зазора d на $\lambda/2$ разность хода увеличится на λ, так что произойдет смещение на одну полосу (т.е. на место каждой светлой полосы станет такая же светлая соседняя полоса). При изменении угла падения на $\Delta\varphi$ разность хода изменится на $2d\sin\varphi\,\Delta\varphi$. Отсюда видно, что полосы интерференции получатся тем шире, чем меньше d. При $d = 0$ они стали бы бесконечно широкими, т.е. поле зрения было бы освещено равномерно.

При больших зазорах и высокой степени монохроматичности света с интерферометром Майкельсона наблюдалась интерференция очень высокого порядка (около 10^6). Если M_2 и M_1' близки друг к другу и образуют воздушный клин с небольшим углом, то полосы интерференции локализуются либо на поверхности клина, либо вблизи нее. Это — полосы равной толщины, имеющие вид равноотстоящих прямых, параллельных ребру клина.

§ 36. Многолучевая интерференция

1. Вернемся к исследованию интерференции в плоскопараллельных пластинках и учтем лучи, претерпевшие в них много-

кратные отражения (см. рис. 134). Обозначим через R коэффициент отражения света от границы раздела пластинки с воздухом, т.е. долю энергии падающего света, которая возвращается обратно при каждом отражении. При отсутствии поглощения оставшаяся доля $(1-R)$ приходит через эту границу. Если среды по обе стороны пластинки одинаковы (воздух), то, как будет показано в § 65, коэффициенты отражения на обеих поверхностях пластинки будут одинаковы. Допустим, что падающий свет монохроматический. Обозначим через I_0 его интенсивность. Тогда интенсивности прошедших пучков $1'$, $2'$, $3'$, ... будут

$$I_{1'} = (1-R)^2 I_0, \quad I_{2'} = R^2(1-R)^2 I_0, \quad I_{3'} = R^4(1-R^2)I_0, \ldots,$$

а соответствующие (вещественные) амплитуды

$$a_{1'} = (1-R)a_0, \quad a_{2'} = R(1-R)a_0, \quad a_{3'} = R^2(1-R)a_0, \ldots,$$

где a_0 — амплитуда падающего света. Разность хода между двумя соседними интерферирующими пучками составляет $\Delta = 2dn\cos\psi$, а разность фаз $\Phi = k\Delta = (4\pi/\lambda)dn\cos\psi$. Амплитуда прошедшей волны представится убывающей геометрической прогрессией

$$a_d = a_0(1-R)[1 + Re^{-i\Phi} + R^2 e^{-2i\Phi} + \ldots].$$

Если пластинка достаточно длинная, то прогрессию можно считать бесконечной. Тогда

$$a_d = \frac{1-R}{1-Re^{-i\Phi}} a_0.$$

Интенсивность прошедшей волны будет

$$I_d = \frac{(1-R)^2}{|1-Re^{-i\Phi}|^2} a_0^2 = \frac{(1-R)^2}{(1-R)^2 + 4R\sin^2(\Phi/2)} I_0. \tag{36.1}$$

Рассчитаем теперь интенсивность I_r отраженной волны. Интенсивности отраженных пучков 1, 2, 3, ... будут

$$I_1 = RI_0, \quad I_2 = R(1-R)^2 I_0, \quad I_3 = R^3(1-R)^2 I_0, \ldots,$$

а соответствующие (вещественные) амплитуды

$$a_1 = \sqrt{R}\, a_0, \quad a_2 = -\sqrt{R}\,(1-R)a_0, \quad a_3 = -\sqrt{R}\,R(1-R)a_0, \ldots$$

Знак минус в выражениях для амплитуд учитывает потерю полуволны при отражении на одной из поверхностей пластинки. Такую потерю полуволны не надо было учитывать при расчете амплитуды прошедшей волны, так как там все отражения происходят на границе стекло–воздух. В случае же отражения от пластинки луч 2 претерпевает отражение на границе воздух–стекло, а все остальные лучи — на границе стекло–воздух. Результирующая амплитуда отраженной волны представится геометрической прогрессией

$$a_r = \sqrt{R}\, a_0 - \sqrt{R}\,(1-R)a_0 e^{-i\Phi}[1 + Re^{-i\Phi} + R^2 e^{-2i\Phi} + \ldots].$$

Поступая, как раньше, найдем для интенсивности

$$I_r = \frac{4R\sin^2(\Phi/2)}{(1-R)^2 + 4R\sin^2(\Phi/2)} I_0. \qquad (36.2)$$

2. Из (36.1) и (36.2) видно, что $I_d + I_r = I_0$, как это и должно быть при отсутствии поглощения. Таким образом, картины распределения интенсивностей в прошедшем и отраженном свете — *взаимно дополнительные*: светлым местам и максимумам одной картины соответствуют темные места и минимумы другой. Для получения интерференционных полос, конечно, надо пользоваться протяженными источниками света. Полосы интерференции будут *полосами равного наклона*. Их можно спроецировать линзой на экран, помещенный в ее фокальной плоскости. Можно также пользоваться зрительной трубой, установленной на бесконечность. Интерференционные полосы будут иметь форму концентрических колец, центр которых находится в точке схождения лучей, нормальных к поверхности пластинки. Максимумы в проходящем свете получаются при $\Phi/2 = m\pi$, т.е. при $2dn\cos\psi = m\lambda$, где m — целое число. В отраженном свете этим местам соответствуют минимумы.

Таким образом, положения максимумов и минимумов определяются в точности теми же условиями, что и в случае простой двухлучевой интерференции. То же относится и к распределению интенсивностей, но при том существенном условии, что коэффициент отражения мал ($R \ll 1$). Тогда в (36.1) и (36.2) можно пренебречь квадратами R и произвести разложение по R. В первом порядке получится

$$I_d = 1 - 4R\sin^2(\Phi/2) = 1 - 2R(1-\cos\Phi),$$
$$I_r = 4R\sin^2(\Phi/2) = 2R(1-\cos\Phi).$$

Но к тому же результату мы пришли бы, если бы и не учитывали многократные отражения.

Однако распределение интенсивности существенно изменяется при увеличении коэффициента отражения R, в особенности когда этот коэффициент приближается к единице. На рис. 141 приведены кривые интенсивности прошедшего света для трех значений R. Интенсивность падающего света принята за единицу. По оси абсцисс отложена разность фаз $\Phi = (4\pi dn\cos\psi)/\lambda$. При $\Phi = 2m\pi$ (m — целое) получаются максимумы. Когда $R \ll 1$, максимумы пологие. Но уже при $R = 0{,}75$ они очень резкие. При приближении R к единице весь свет практически сосредоточивается в очень узких интерференционных полосах на темном фоне. В отраженном свете получаются столь же резкие, но темные интерференционные полосы на светлом фоне. Числитель в формуле (36.1) — величина постоянная. В максимуме ($\Phi = 2m\pi$) $I_{\text{макс}} = 1$.

Резкость интерференционных полос характеризуется их *полушириной*. Для полос в проходящем свете *полушириной называется расстояние между точками, лежащими по обе стороны максимума, в которых интенсивность составляет половину максимальной величины* $I_{\text{макс}}$. В окрестности максимума

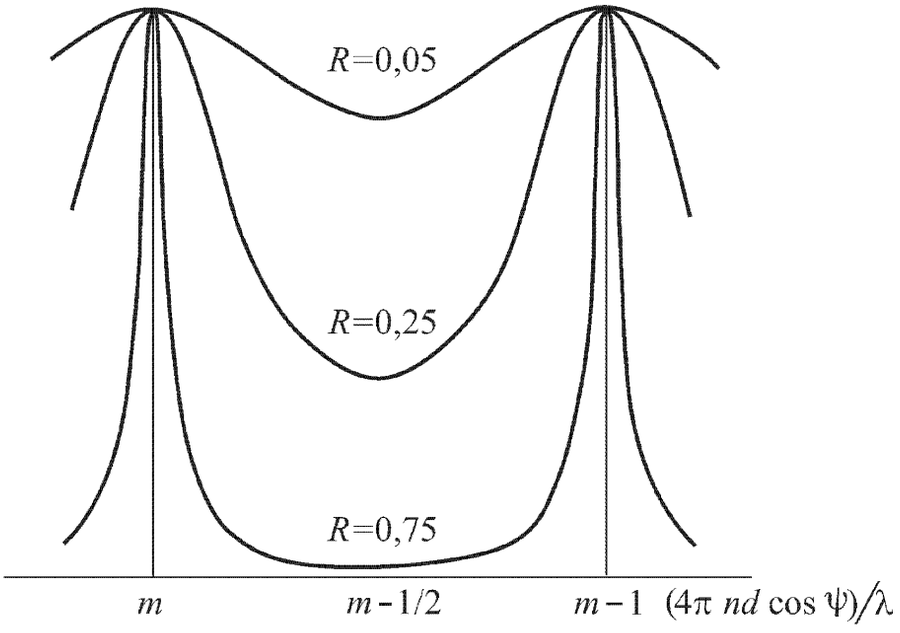


Рис. 141

m-го порядка Φ представим в виде $\Phi = m\pi + \varphi$. Тогда в формуле (36.1) Φ можно заменить на φ. Ввиду малости φ ту же формулу можно переписать так:

$$I_d = \frac{I_{\text{макс}}}{1 + R\varphi^2/(1-R)^2}. \tag{36.3}$$

Если $R\varphi^2/(1-R)^2 = 1$, то $I_d = {}^1\!/_2 I_{\text{макс}}$. Следовательно, полуширина определится выражением

$$\delta\Phi = 2\varphi = 2\frac{1-R}{\sqrt{R}}. \tag{36.4}$$

3. В применениях многолучевая интерференция всегда осуществляется в *толстых пластинках*, так что приходится иметь дело с интерференцией высоких порядков. Поэтому для получения интерференционных полос требуется *высокая монохроматичность света* ($\lambda/\delta\lambda \gtrsim m$). В интерференционной спектроскопии, где многолучевая интерференция применяется для изучения спектров, исследуемый свет, если он не обладает достаточной монохроматичностью, должен быть предварительно *монохроматизован*, например с помощью призменного спектрометра или другого монохроматора.

Обычно методами интерференционной спектроскопии исследуется структура тонких спектральных линий. Достоинством интерференционных спектроскопов является их высокая разре-

шающая способность при большой светосиле, простоте устройства, дешевизне и удобстве в обращении. *Разрешающая способность определяет наименьшее расстояние между близкими спектральными линиями, которые изображаются в виде раздельных спектральных линий.*

Для уяснения понятия разрешающей способности допустим, что применяемый свет состоит из двух близких, одинаково интенсивных спектральных линий с длинами волн λ и $\lambda' = \lambda + \delta\lambda$. Интерференционная картина будет состоять из двух систем интерференционных полос (максимумов) с теми же длинами волн. Одна система максимумов окажется сдвинутой относительно другой. Если сдвиг недостаточен, максимумы наложатся друг на друга, получится распределение с единственным максимумом, которое для глаза неотличимо от распределения, соответствующего одной спектральной линии. Тогда говорят, что рассматриваемая двойная линия *не разрешается спектральным прибором.*

Допустим теперь, что сдвиг равен полуширине интерференционной полосы. Соответствующее распределение интенсивности показано на рис. 142. Кривые λ и λ' изображают контуры спектральных линий, как они были бы видны в спектральном приборе в отдельности. В центре картины, где контуры пересекаются, интенсивность спектральных линий составляет $^1/_2 I_{\text{макс}}$. При сложении интенсивностей интенсивность в центре получится равной $I_{\text{макс}}$. В точке A интенсивность первой линии равна $I_{\text{макс}}$, а второй $^1/_5 I_{\text{макс}}$, как это видно из формулы (36.3). Таким образом, полная интенсивность в точке A будет $I_{\text{макс}} + {}^1/_5 I_{\text{макс}} = 1{,}2 I_{\text{макс}}$. Такая же интенсивность получится в точке A'. Кривая результирующей интенсивности будет состоять из двух максимумов с провалом в центре между ними. Интенсивность в месте провала будет составлять $1/1{,}2 = 83\,\%$ от интенсивности одного из максимумов на результирующей кривой.

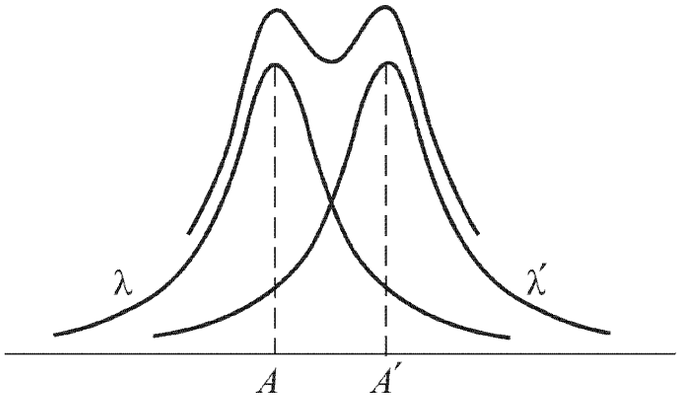


Рис. 142

Будет ли разрешаться спектральная линия при таком провале — это зависит от индивидуальных особенностей глаза. Опыт показывает, что нормальный глаз такие провалы обычно замечает без особого труда. Поэтому расстояние между максимумами спектральных линий, равное или большее полуширины интер-

ференционной полосы, мы примем (до некоторой степени условно) за *условие спектрального разрешения.*

Определим теперь минимальную разность длин волн $\delta\lambda = \lambda' - \lambda$, соответствующую такому критерию разрешения. Все величины, относящиеся к свету с длиной волны λ, будем обозначать нештрихованными, а к длине волны λ' — штрихованными буквами. Показатель преломления n будем считать одним и тем же для обеих длин волн. (Это строго выполняется только в интерферометре Фабри–Перо.) Тогда, независимо от длины волны, всем падающим лучам одного и того же направления будут соответствовать преломленные лучи также одного направления. В точке A' для длины волны λ' получается интерференционный максимум m-го порядка, а потому $\Phi' = 2m\pi$. В той же точке волна с длиной λ максимума уже не дает. Для такой волны разность фаз имеет значение $\Phi = 2m\pi + (1-R)/\sqrt{R}$, т.е. в рассматриваемой точке $\Phi' - \Phi \equiv \delta\Phi = (1-R)/\sqrt{R}$. Но ввиду одинаковости n и ψ для обеих волн из формулы $\Phi = (4\pi dn\cos\psi)/\lambda$ следует: $\delta\Phi/\Phi = |\delta\lambda/\lambda|$. Учтя, что в максимуме $\Phi = 2\pi m$, поэтому находим

$$\frac{\lambda}{\delta\lambda} = \frac{2\pi\sqrt{R}}{1-R}m. \tag{36.5}$$

Величина $\lambda/\delta\lambda$ называется *разрешающей способностью спектрального прибора.*

Интерференционные спектральные аппараты осуществлены в виде двух приборов, которые мы и рассмотрим.

4. И н т е р ф е р о м е т р и л и э т а л о н Ф а б р и–П е р о. Этот прибор, изобретенный Фабри (1867–1945) и Перо (1863–1925), является наиболее распространенным из интерференционных спектральных приборов. Он состоит из двух стеклянных или кварцевых пластинок P_1 и P_2, между которыми обычно

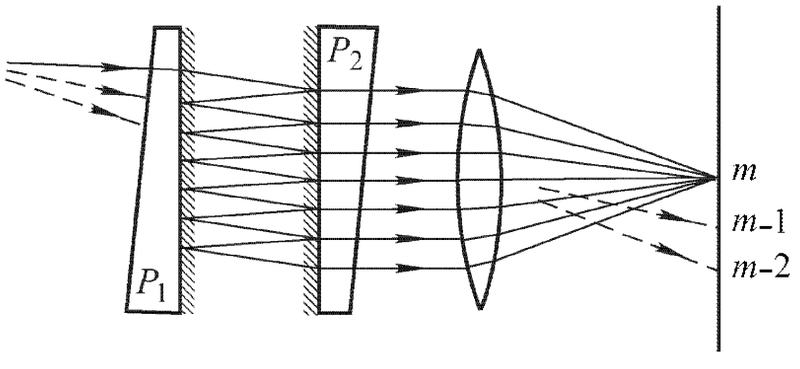


Рис. 143

находится воздух (рис. 143). Плоские поверхности пластинок, обращенные друг к другу, тщательно отшлифованы и покрыты высокоотражающими слоями (серебро, алюминий, многослой-

ные диэлектрические покрытия). В хороших приборах отступления внутренних поверхностей пластинок от идеальных плоскостей не превышают $0{,}01\lambda$. Отражательная способность зеркальных поверхностей пластинок при металлическом покрытии может быть доведена до 95 %, а при диэлектрических — до 98 %.

Параллельность зеркальных поверхностей достигается с помощью распорного кольца из инвара или плавленого кварца, помещаемого между пластинками. Кольцо снабжено тремя выступами с каждой стороны, к которым пластинки прижимаются при помощи трех пружин. Выступы подшлифованы так, что зеркальные поверхности устанавливаются параллельно друг другу. Небольшие отступления от параллельности устраняются нажимом соответствующей пружины.

Интерферометр, в котором используются распорные кольца, называется *эталоном Фабри–Перо*. Располагая набором эталонов с кольцами разной толщины, можно производить измерения при различных расстояниях между зеркальными поверхностями. В прежних моделях интерферометра Фабри–Перо можно было с помощью специальных микрометрических винтов менять расстояние между зеркалами. Но таким путем не удается достигнуть той высокой точности, которую можно получить с помощью эталона. Наружные поверхности пластинок обычно образуют небольшие углы с внутренними, чтобы отраженный от них светлый блик не мешал наблюдению основной интерференционной картины. Таким образом, интерферометр Фабри–Перо можно рассматривать как плоскопараллельную воздушную пластинку, на которой происходят многократные отражения световых лучей и последующая интерференция их.

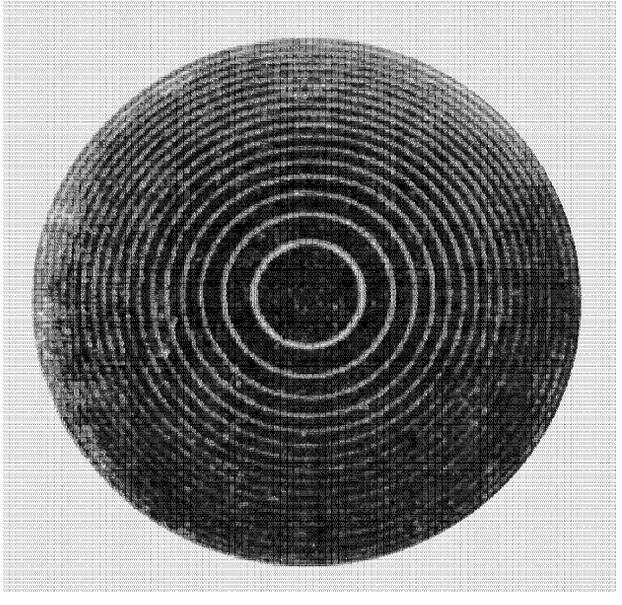


Рис. 144

Интерференционная картина состоит из концентрических колец равного наклона. Она приведена на рис. 144 в проходящем свете. Интерференционные максимумы тем у́же, чем больше отражательная способность зеркальных поверхностей пластинок интерферометра.

Расстояние между зеркалами h обычно составляет 1–100 мм, а в специальных эталонах доходит до 1 м. Поэтому порядки интерференции $m \approx 2h/\lambda$ очень высоки (при $h = 5$ мм $m \approx 20\,000$).

Ввиду малости угла ψ условие главного интерференционного максимума $2h\cos\psi = m\lambda$ можно записать в виде $h(2-\psi^2) = m\lambda$. Отсюда находим *угловую дисперсию* интерферометра Фабри–Перо:

$$\frac{d\psi}{d\lambda} = -\frac{m}{2h\psi} = -\frac{1}{\lambda\psi}. \tag{36.6}$$

При рабочих условиях ($\psi \approx 10^{-2}$ рад) угловая дисперсия интерферометра Фабри–Перо значительно превышает дисперсию других спектральных аппаратов. Это является основным преимуществом интерферометра Фабри–Перо.

Другим важным преимуществом интерферометра Фабри–Перо является его большая светосила. Благодаря этому и своей дешевизне такой интерферометр получил широкое распространение для спектральных исследований в оптической области спектра. Принцип интерферометра Фабри–Перо был использован также в объемных резонаторах оптических квантовых генераторов (см. § 120).

5. П л а с т и н к а Л у м м е р а–Г е р к е (1878–1960). Она представляет собой плоскопараллельную пластинку из очень однородного стекла или плавленого кварца толщиной от 3 до 10 мм и длиной до 30 см. Для направления световых лучей в пластинку на одном конце ее сбоку посажена на оптический

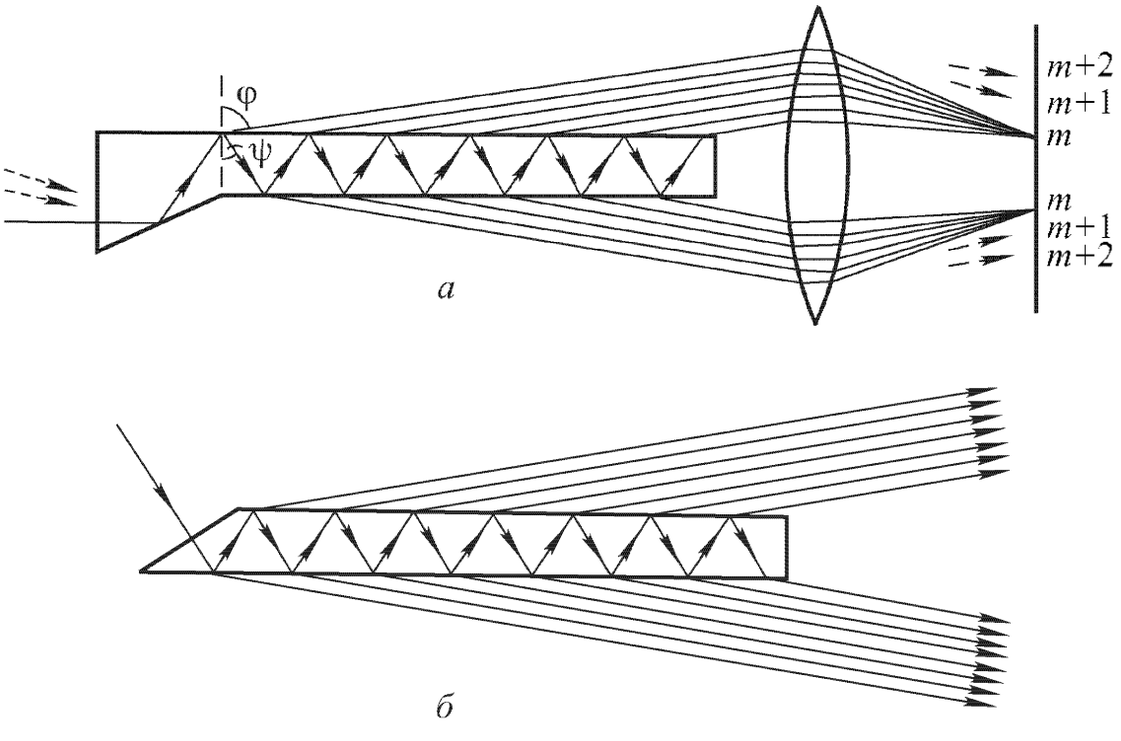


Рис. 145

контакт добавочная призмочка (рис. 145 *а*). При другом способе один конец пластинки скошен (рис. 145 *б*). В обоих случаях падающие лучи нормальны к поверхности стекла, чем достига-

ется уменьшение потерь света на отражение. Направление падающих лучей подбирается таким, чтобы угол падения на границе стекло–воздух был близок к предельному углу полного отражения. Тогда коэффициент отражения мало отличается от единицы. Пучки испытывают многократные отражения от плоскостей пластинки и выходят из нее с почти одинаковыми интенсивностями. Можно получить до 10–15 таких пучков с каждой стороны пластинки.

В пластинке Луммера–Герке наблюдаются интерференционные полосы равного наклона. Условие интерференционного максимума m-го порядка имеет вид

$$2hn\cos\psi = m\lambda, \tag{36.7}$$

где h — толщина пластинки, а ψ — угол преломления. Здесь не учтено изменение фазы при отражении, так как оно производит только несущественное смещение всей интерференционной картины. Число интерферирующих пучков по одну сторону пластинки равно $N = L/(2h\,\mathrm{tg}\,\psi)$, где L — длина пластинки. А так как свет падает почти под предельным углом, то

$$N = \frac{L}{2h}\sqrt{n^2-1}. \tag{36.8}$$

Вопрос о разрешающей способности пластинки Луммера–Герке будет разобран в § 48.

6. Затронем еще два вопроса, имеющих отношение к многолучевой интерференции. Рассмотрим сначала много равноотстоящих параллельных отражающих плоскостей с малыми коэффициентами отражения (рис. 146). Почти весь свет, падающий на каждую плоскость, проходит через нее без потерь. В результате получится много параллельных отраженных лучей *1*, *2*, *3*, ... почти одинаковой интенсивности. Разность хода между соседними лучами составляет $2d\cos\varphi$, где d — расстояние между соседними плоскостями. Если выполнено условие

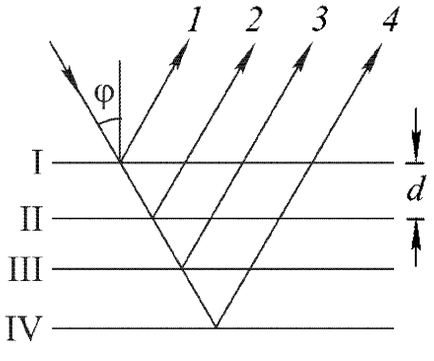


Рис. 146

$$2d\cos\varphi = m\lambda \tag{36.9}$$

(m — целое число), то все лучи будут находиться в фазе. Возникнет сильное отражение, называемое *интерференционным отражением*. При заданном m условие (36.9) может выполняться только для определенных длин волн. Поэтому при освещении рассматриваемой системы плоскостей белым светом интерференционное отражение будет наблюдаться лишь для некоторых (а при $d < \lambda$ — для одной) сравнительно узких спектральных линий. Система плоскостей, таким образом, при-

обретает свойство *избирательного отражения*. При этом, так как разность хода зависит от угла падения, цвет избирательно отраженного света будет меняться с изменением угла падения.

Отметим также, не входя ни в какие подробности, что надлежащей комбинацией диэлектрических слоев с различными показателями преломления удается осуществить на поверхности стекла *многослойные покрытия* (пленки) с высокими коэффициентами отражения (до 98–99 % для определенных длин волн). Такие покрытия применяются, например, в зеркалах интерферометров Фабри–Перо. Следует, однако, иметь в виду, что отражающие многослойные пленки в белом свете окрашены, так как интерференционное условие сильного отражения точно выполняется только для какой-либо одной длины волны.

§ 37. Стоячие световые волны

1. При интерференции монохроматических волн, распространяющихся навстречу друг другу, образуются *стоячие волны* (см. т. III, § 140). В этом случае, как видно из формулы (26.12), ширина полосы Δx равна $\lambda/2$, т.е. расстоянию между соседними пучностями или узлами. В случае электромагнитных волн пучности (узлы) электрического вектора $\mathbf{E}$ совпадают с узлами (пучностями) магнитного вектора $\mathbf{B}$. Таким образом, в стоячей волне электрическое и магнитное поля можно *пространственно разделить* и исследовать свойства и действия этих полей в отдельности. Моменты прохождения электрического и магнитного полей через максимум в стоячей волне не совпадают, а сдвинуты относительно друг друга на половину периода световых колебаний. Получение стоячих световых волн наталкивается на трудности, связанные с малостью длины волны. Эти трудности впервые были преодолены О. Винером (1862–1927) в 1890 г.

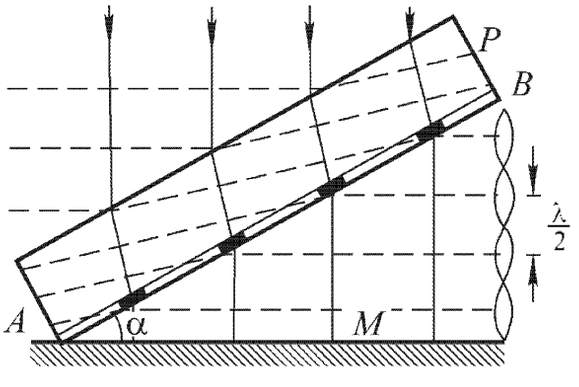


Рис. 147

Стеклянная пластинка P (рис. 147), покрытая тонким фоточувствительным слоем толщиной $\sim \dfrac{\lambda}{20} - \dfrac{\lambda}{30}$, приводилась в соприкосновение с хорошим металлическим зеркалом M, образуя с ним очень малый угол α (порядка $1'$ или меньше). Пластинка освещалась монохроматическими лучами, нормальными к поверхности зеркала.

В результате отражения от зеркала в воздушном клине между зеркалом и фотослоем образовывались стоячие световые волны. Их пучности обнаруживались в фотослое после его проявления в виде равноотстоящих почерневших полос, параллельных ребру воздушного клина. Расстояние между этими полосами равно $\lambda/(2\sin\alpha)$. Полосы можно видеть глазом, если угол α взять достаточно малым. В опытах Винера расстояние между полосами составляло 1–2 мм. По этим данным можно было оценить длину световой волны. Удадось также показать, что для красного света длина волны примерно вдвое больше, чем для синего.

Выдвигалось возражение, что при интерпретации опыта не было учтено обратное отражение света от фотослоя. Чтобы снять это возражение, Винер заполнял пространство между фотослоем и металлическим зеркалом бензолом, показатель преломления которого близок к показателю преломления желатины фотослоя. Тем самым устранялось влияние отражения от фотослоя. Оказалось, что при этом вид интерференционных полос не менялся.

Главное значение опытов Винера состоит в том, что они позволили установить, какой вектор производит фотохимическое действие: электрический или магнитный. Опыты показали, что первый слой почернения получается на расстоянии $\lambda/4$ от поверхности металлического зеркала. Здесь поэтому находится пучность вектора, вызывающего фотографическое действие. Но при отражении электромагнитной волны от металла на его поверхности получается узел электрического вектора (см. т. III, § 45). Первая пучность электрического вектора находится на расстоянии $\lambda/4$ от поверхности металлического зеркала — там же, где наблюдается почернение фотографического слоя. Отсюда следует, что фотографическое (и вообще фотохимическое) действие производится электрическим, а не магнитным полем световой волны.

Позднее, в 1892 г., Друде (1863–1906) и Нернст (1864–1941) повторили опыт Винера, заменив фотографический слой тонкой пленкой флуоресцирующего вещества. Айвс в 1933 г. сделал то же самое с фотоэлектрическим слоем. Оказалось, что флуоресценция и фотоэффект вызываются также электрическим полем. Этого и следовало ожидать с точки зрения электронных представлений о строении вещества.

Действительно, действие света на вещество сводится к силам, с которыми на электроны действуют электрические и магнитные поля. Первая сила равна $e\mathbf{E}$, вторая $\frac{e}{c}[\mathbf{vB}]$, где $\mathbf{v}$ — скорость электрона. Магнитная сила много меньше электрической, так как $v \ll c$. Таким образом, во всех действиях электрический

вектор световой волны является более важным, чем магнитный. Поэтому его нередко называют *световым вектором*.

2. Винер исследовал также стоячие световые волны в поляризованном свете. При отражении от прозрачной среды под определенным углом отраженный свет оказывается линейно поляризованным (см. § 65). Про такой свет теперь говорят, что он *поляризован перпендикулярно к плоскости падения*. Со времен Френеля (1821 г.) до конца XIX века во всех эфирных теориях света обсуждался вопрос, как направлен световой вектор: параллельно плоскости поляризации или перпендикулярно к ней. Опыт Винера и должен был решить этот вопрос.

Винер изменил постановку своего предыдущего опыта, заставив линейно поляризованный свет отражаться от металлического зеркала под углом 45°. Тогда угол между направлениями распространения падающей и отраженной волн будет 90°. Поэтому при интерференции этих волн расстояние между соседними пучностями или узлами стоячей волны окажется равным $\lambda/\sqrt{2}$, как это видно из формулы (26.12). Если световой вектор перпендикулярен к плоскости падения, то колебания в падающей и отраженной волнах будут параллельными. Тогда возникнет интерференция этих волн и на фотографической пластинке, наклоненной к поверхности металлического зеркала, получатся такие же темные и светлые полосы, как и в первом опыте Винера. Если же световой вектор параллелен плоскости падения, то интерференция невозможна, так как в этом случае колебания в падающем и отраженном свете будут взаимно перпендикулярны (см. § 26, п. 5).

Опыт показал, что полосы получаются, когда свет поляризован перпендикулярно к плоскости падения. Если же он поляризован в плоскости падения, то получается равномерное почернение пленки. Отсюда следует, что *световой вектор лежит в плоскости поляризации*. А так как предыдущими опытами доказано, что он является электрическим вектором, то последний также лежит в плоскости поляризации.

3. На явлении стоячих волн основан метод *цветной фотографии*, разработанный Липпманом (1845–1922) в 1891 г. Стеклянная пластинка покрыта с одной стороны толстым слоем мелкозернистой сухой фотографической эмульсии высокой чувствительности. Пластинка составляет стенку сосуда, в который наливается ртуть. Эмульсией пластинка должна быть обращена внутрь сосуда, чтобы фотографический слой находился в контакте с ртутью. Пластинка освещается снаружи нормально падающими лучами определенной длины волны. При отражении света от поверхности ртути в толще эмульсии образуются стоячие волны, которые и экспонируют фотопластинку. После это-

го пластинка вынимается, проявляется и фиксируется. В местах пучностей электрического вектора выделяются тонкие слои металлического серебра, находящиеся на расстоянии $\lambda/2$ друг от друга. Получается плоскослоистая среда из отражающих параллельных плоскостей, подобная той, которая изображена на рис. 146.

Осветим теперь со стороны эмульсии так обработанную пластинку нормально падающим светом. Малая часть света отразится от слоя I, остальная часть пройдет прямо. От второго слоя отразится столь же малая часть, и т.д. Получится столько же отраженных пучков, сколько металлических слоев отложилось в толще эмульсии. Интенсивность каждого пучка мала, но их число велико. Если длины волн освещающего света и света, применявшегося при экспонировании, в точности одинаковы, то каждый последующий отраженный пучок будет запаздывать по фазе относительно предыдущего на 2π. В этом случае произойдет интерференционное усиление пучков, так что результирующая интенсивность отраженного света окажется большой. Если же каждый пучок будет сдвинут по фазе относительно предыдущего на $\pm 2\pi/N$, где N — общее число пучков, то произойдет их взаимное гашение. Вообще, интенсивно будут отражаться только такие длины волн, которые равны длине волны света, применявшегося при экспонировании, или очень близки к ней. Поэтому при освещении пластинки белым светом она в отраженном свете будет казаться окрашенной в соответствующий цвет.

Спектральная область $\delta\lambda$ отраженного света тем у́же, чем толще слой эмульсии, чем больше число отложившихся слоев металлического серебра. Если на эмульсию подышать, то она немного разбухнет, расстояния между металлическими слоями увеличатся, что вызовет смещение цвета в красную сторону. При увеличении угла падения φ условие интерференционного усиления $2d\sin\varphi = \lambda$ будет выполняться для более коротких волн. Поэтому при наклоне пластинки ее цвет меняется в сторону фиолетового конца спектра.

Для получения цветных фотографий предметов при экспонировании пластинки надо получать цветное изображение фотографируемого предмета на светочувствительном слое, как во всяком фотоаппарате. Липпмановские цветные фотографии получаются высокого качества в чистых спектральных цветах, например при фотографировании спектров. Смешанные цвета, как правило, дают плохие результаты. Метод Липпмана не получил распространения. Современная техническая цветная фотография основана на принципе светофильтров — использовании соответствующих красящих пигментов, вводимых в эмульсию фотопластинок.

Нейхаусс в конце 90-х годов прошлого века на опыте убедился в слоистой структуре пленок Липпмана. Была получена по методу Липпмана фотография в красном свете. После этого пленка была отделена от стекла и произведен поперечный разрез ее. Затем была получена микрофотография одного из разрезов пленки с увеличением в 1000 раз. На микрофотографии было видно около десяти темных полос, чередовавшихся со светлыми. Темные полосы получились в местах отложения металлического серебра. Расстояние между серединами темных полос получилось ≈ 350 нм. Необходимо заметить, что никакой микроскоп не позволяет различить детали предмета, много меньшие длины световой волны. Успех опыта отчасти определялся тем, что пленка заметно разбухла при обработке до ее разрезания. Кроме того, разрез производился в косом направлении.

§ 38. Излучение Вавилова–Черенкова

1. Интерференция света — отнюдь не редкое явление, как это может показаться с первого взгляда. По существу все, что связано с распространением света в веществе, относится к классу явлений интерференции. Действительно, световая волна, падая на среду, возбуждает электрические колебания в атомах и молекулах. Вследствие этого эти частицы начинают излучать вторичные электромагнитные волны, также воздействующие на атомы и молекулы среды. Падающая волна и возбужденные ею вторичные волны интерферируют, и этой интерференцией определяются все особенности распространения света в среде.

Действие линзы и вогнутого зеркала можно рассматривать также как интерференционный эффект. Действительно, возмущения от точечного источника, если они попадают на различные участки линзы, в дальнейшем распространяются по различным путям вдоль лучей к фокусу. Так как оптические длины всех лучей от источника до фокуса одинаковы, то возмущения приходят в фокус в одинаковых фазах и при наложении усиливают друг друга. Этим интерференционным эффектом и объясняется фокусирующее действие линзы. Если бы при своем распространении световые возмущения строго следовали законам лучевой (геометрической) оптики, то в фокусе получилась бы бесконечная интенсивность светового поля. То обстоятельство, что этого не происходит, связано с отступлениями от геометрической оптики и прежде всего с дифракцией.

2. Рассмотрим более интересный пример. Допустим, что в прозрачной однородной среде движется электрон или какая-либо другая заряженная частица с постоянной скоростью V. Своим полем движущийся электрон возбуждает атомы и моле-

кулы среды, и они становятся центрами излучения электромагнитных волн. При равномерном движении электрона эти волны *когерентны* и могут интерферировать между собой. Если скорость электрона V больше фазовой скорости света в среде v, то волны, исходящие от электрона в различные моменты времени, при определенных условиях могут приходить в точку наблюдения одновременно.

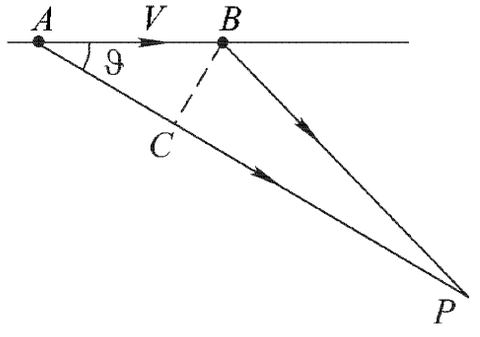


Рис. 148

Действительно, пусть A и B (рис. 148) — точки, через которые электрон проходил в моменты времени t_1 и t_2 соответственно. На прохождение расстояния AB электрон затратил время $t_2 - t_1 = AB/V$. В точку наблюдения P волны из A и B придут в моменты времени $t_1 + AP/v$ и $t_2 + BP/v$. Разность этих времен равна

$$\Delta t = (t_2 - t_1) - \frac{AP - BP}{v} = \frac{AB}{V} - \frac{AP - BP}{v}.$$

Если точка P удалена достаточно далеко, то $AP - BP \approx AC = AB\cos\vartheta$, так что $\Delta t = AB\left(\frac{1}{V} - \frac{\cos\vartheta}{v}\right)$. Так как по предположению $v < V$, то существует угол ϑ, удовлетворяющий условию

$$\cos\vartheta = \frac{v}{V} = \frac{1}{n\beta}, \tag{38.1}$$

где $\beta \equiv V/c$, а $n \equiv c/v$ — показатель преломления среды. Если это условие выполнено, то все волны придут в точку P *одновременно*, какова бы ни была длина отрезка AB. В этом случае при интерференции произойдет взаимное усиление их. Во всех других случаях путь электрона можно разбить на такие отрезки, чтобы от крайних точек каждого из них волны приходили в точку P с разностью хода λ. Волны, приходящие от всех точек каждого из таких отрезков, полностью погасят друг друга из-за интерференции. Следовательно, то же произойдет с волнами, приходящими от всех точек среды, лежащих на пути движения электрона. Таким образом, в направлениях, определяемых условием (38.1), электрон (точнее — среда, в которой он движется) будет излучать электромагнитные волны, а в остальных направлениях излучения не будет.

Такое излучение экспериментально было обнаружено в 1934 г., а затем подробно исследовано Ф.А. Черенковым (р. 1904), в то время аспирантом С.И. Вавилова (1891–1951). Черенков показал, что все без исключения жидкие и твердые тела при прохождении через них быстрых электронов, помимо флуоресценции, имеющей место в некоторых случаях, всегда ис-

пускают слабый видимый свет с непрерывным спектром. Свечение частично поляризовано, причем электрический вектор лежит преимущественно в плоскости, образованной световым лучом излучения и направлением движения электрона. Излучение направлено главным образом вперед и максимально вдоль образующих конуса, ось которого задается направлением движения электрона, а угол при вершине 2ϑ определяется формулой (38.1). Свечение не удается потушить ни температурным воздействием, ни прибавлением к светящейся среде веществ, тушащих флуоресценцию. Попытки определения времени нахождения излучающих атомов и молекул в возбужденном состоянии показали, что это время равно нулю: свечение прекращается одновременно с прекращением прохождения электронов через среду. Эти особенности свечения привели С.И. Вавилова к заключению, что оно не является люминесценцией, для которой характерна конечная длительность свечения.

Ввиду очень слабой яркости свечения, Черенков при количественных измерениях в большинстве случаев применял метод Вавилова фотометрирования по порогу зрения.

В дальнейшем было показано, что свечение вызывается также протонами, мезонами и другими быстрыми заряженными частицами. Свечение, вызываемое радиоактивными излучениями, было известно и ранее, но оно во всех случаях неправильно считалось одним из видов люминесценции.

3. Ведущая роль в руководстве экспериментальными исследованиями, направленными на выяснение природы свечения, принадлежит С.И. Вавилову. Качественное объяснение явления излучения Вавилова–Черенкова, приведенное выше, было дано И.Е. Таммом (1895–1971) и И.М. Франком (р. 1908) в 1934 г. Тогда же ими была создана и количественная теория, согласующаяся с наблюдаемыми фактами. Несколько позднее, в 1940 г., В.Л. Гинзбург (р. 1916) построил квантовую теорию, основанную на законах сохранения энергии и импульса.

Из рассуждений, приведенных выше, ясно, что в формулу (38.1) должна входить *фазовая*, а не какая-либо другая скорость света, так как именно она определяет фазы колебаний, а с ними и условие интерференционного усиления волн. Из формулы (38.1) видно, что излучение на частотах ω, для которых $n(\omega) < 1/\beta$, невозможно. Поэтому спектр излучения Вавилова–Черенкова должен обрываться на коротких волнах, где условие (38.1) перестает выполняться из-за дисперсии света. В частности, невозможно излучение Вавилова–Черенкова рентгеновских волн, так как для них $n < 1$.

Если бы электрон двигался в среде строго равномерно, то излучение было бы точно сосредоточено на поверхности конуса, определяемого условием (38.1). Из-за неравномерности движе-

ния появляется размытие этой поверхности. Но даже и при наличии такого размытия направленность излучения указывает на то, что молекулы и атомы среды, возбуждаемые электроном, излучают *когерентно* если и не на всем пути движения электрона, то во всяком случае на его части порядка длины световой волны. Следовательно, рассматриваемое явление *практически не зависит от атомного строения среды* и может быть рассмотрено в рамках *макроскопической электромагнитной теории*. Это и было сделано Таммом и Франком. Ограничимся приведением основного результата. Полная энергия, излучаемая электроном в единицу времени на единице его пути, определяется формулой

$$\frac{dW}{dt} = \frac{e^2}{c^2} \int\limits_{\beta n > 1} \left(1 - \frac{1}{\beta^2 n^2}\right) \omega \, d\omega. \tag{38.2}$$

Интегрирование производится по частотам, для которых $\beta n(\omega) > 1$. Той же формулой очевидным образом определяется и распределение излучаемой энергии по спектру частот. Формула справедлива при выполнении условия

$$T(\omega) \left|\frac{dV}{dt}\right| \ll v(\omega), \tag{38.3}$$

т.е. когда скорость электрона V за период световых колебаний $T = 2\pi/\omega$ меняется мало по сравнению с фазовой скоростью $v(\omega)$. По оценке Тамма и Франка полная потеря энергии электрона на излучение Вавилова–Черенкова в жидкостях и твердых телах по порядку величины составляет несколько тысяч электрон-вольт на сантиметр пути, т.е. пренебрежимо мала по сравнению с потерями энергии по другим причинам.

Излучение электрона, конечно, приводит к его торможению. Само по себе ускоренное движение электрона вызывает излучение. Однако из изложенного выше следует, что это излучение не имеет ничего общего с интерференцией, определяющей излучение Вавилова–Черенкова. Если бы к электрону приложить силу, уравновешивающую все тормозящие силы, то ускорение исчезло бы, а излучение Вавилова–Черенкова осталось бы. Именно так надо понимать утверждение, что электрон, равномерно движущийся в среде, излучает, если его скорость больше фазовой скорости света в этой среде.

4. Явления, аналогичные излучению Вавилова–Черенкова, давно известны в гидро- и аэродинамике. Например, если скорость судна превышает минимальную фазовую скорость на поверхности воды, то оно начинает непрерывно генерировать волновые цуги, даже когда его скорость остается постоянной. В этом причина так называемого *волнового сопротивления*, встречаемого судами при движении по поверхности воды. То же самое происходит при полете снарядов и самолетов в воздухе. Они

начинают излучать ударные волны, называемые *волнами Маха* (1838–1916), если движутся со сверхзвуковыми скоростями. Потери энергии на это излучение столь значительны, что они обусловливают основное сопротивление, испытываемое снарядами и самолетами при сверхзвуковых движениях. Указанные явления сложнее излучения Вавилова–Черенкова из-за нелинейности гидродинамических уравнений.

5. Излучение Вавилова–Черенкова может вызываться не только движущимися частицами, но и каким-либо возбуждением, распространяющимся со скоростью, превышающей фазовую скорость света в среде. Допустим, например, что на границу раздела двух сред падает волна с плоским фронтом AB (рис. 149). Вдоль границы раздела побежит возмущение со скоростью $V = v_1/\sin\varphi$, где v_1 — фазовая скорость света в первой среде. Оно возбудит во второй среде излучение Вавилова–Черенкова под углом ϑ к границе раздела. Угол ϑ определяется формулой (38.1), т.е. $\cos\vartheta = v_2/V$, где v_2 — фазовая скорость во второй среде. Замечая, что $\cos\vartheta = \sin\psi$, отсюда находим $\sin\psi/\sin\varphi = v_2/v_1$. Это — закон преломления Снеллиуса. Таким образом, преломление света можно трактовать как эффект Вавилова–Черенкова, возбуждаемый во второй среде падающей волной. Так же можно рассматривать и отражение света. Распространение света в однородной среде тоже подчиняется закону (38.1). В этом случае скорость волнового фронта V совпадает с фазовой скоростью v, так что формула (38.1) дает $\vartheta = 0$, т.е. волновой фронт распространяется без изменения направления.

Рис. 149

6. Излучение Вавилова–Черенкова нашло широкое применение в ядерной физике и физике высоких энергий. На нем основано действие так называемых *черенковских счетчиков*, т.е. детекторов релятивистских заряженных частиц, излучение которых регистрируется с помощью фотоумножителей. Основное назначение черенковских счетчиков — разделение релятивистских частиц с одинаковыми импульсами, но различными скоростями. Пусть, например, пучок, состоящий из релятивистских протонов и π-мезонов, проходит через однородное поперечное магнитное поле. Направления траекторий прошедших частиц будут определяться только их импульсами, но не будут зависеть от их скоростей. С помощью диафрагм можно выделить протоны и π-мезоны с одинаковыми импульсами. Из-за различия масс скорости π-мезонов v_π окажутся несколько больше скоростей про-

тонов v_p. Если полученный пучок направить в газ и подобрать показатель преломления n газа так, чтобы было $v_\pi > \frac{c}{n} > v_p$, то π-мезоны будут давать излучение Вавилова–Черенкова, а протоны — нет. Таким образом, счетчик будет регистрировать только π-мезоны, но не будет регистрировать протоны. Показатель преломления газа в камере счетчика n можно менять, изменяя его давление.

7. В заключение остановимся на вопросе, может ли возникнуть электромагнитное излучение, когда заряженная частица движется равномерно в среде с досветовой скоростью, т.е. со скоростью, меньшей фазовой скорости света в рассматриваемой среде. Если среда однородна, то ответ будет отрицательным. Действительно, движущаяся заряженная частица на своем пути, конечно, будет возбуждать атомы и молекулы среды. Последние начнут излучать. Однако эти излучения в результате интерференции погасят друг друга, так как при равномерном движении частицы и однородности среды их амплитуды одинаковы, а фазы возрастают линейно с расстоянием, пройденным частицей. Но если среда неоднородна, то гашения не будет и появится излучение. Такое излучение, на существование которого было указано В.Л. Гинзбургом и И.М. Франком в 1944 г., называется *переходным излучением.*

Особенно наглядно происхождение переходного излучения можно понять на следующем примере. Если электрон (или другая заряженная частица) находится перед плоской границей идеального металла, то электрическое поле вне металла можно рассматривать как поле диполя, состоящего из электрона и его «электрического изображения» в поверхности металла (см. т. III, § 23). Если электрон приближается к металлу, то электрическое изображение движется к нему навстречу. При этом электрический момент диполя уменьшается. Вследствие этого и возникает излучение. В момент, когда электрон пересекает границу металла, происходит как бы аннигиляция электрона и его электрического изображения. Так же возникает переходное излучение, когда электрон выходит из металла в вакуум.

Аналогично объясняется возникновение переходного излучения, когда заряженная частица переходит через границу двух диэлектриков. В этом случае также применим метод электрических изображений, хотя и в несколько измененной форме (см. т. III, § 24).

Переходное излучение наблюдалось уже давно в виде свечения анодов рентгеновских трубок. Конечно, природа этого свечения была выяснена много позднее.

Переходное излучение также используется в счетчиках релятивистских частиц для определения их скоростей.

ГЛАВА IV

ДИФРАКЦИЯ СВЕТА

§ 39. Принцип Гюйгенса–Френеля. Зоны Френеля

1. Под *дифракцией света* понимают всякое уклонение от прямолинейного распространения света, если оно не может быть истолковано как результат отражения, преломления или изгибания световых лучей в средах с непрерывно меняющимся показателем преломления. Если в среде имеются мельчайшие частицы постороннего вещества (туман) или показатель преломления заметно меняется на расстояниях порядка длины волны, то в этих случаях говорят о *рассеянии света* и термин «дифракция» не употребляется. Явления дифракции для своего истолкования и количественного рассмотрения не требуют никаких новых принципов. Всякая дифракционная задача, если ее рассматривать строго, сводится к нахождению решения уравнений Максвелла, удовлетворяющего соответствующим граничным условиям. Однако в такой строгой постановке дифракционные задачи, ввиду их сложности, допускают аналитические решения лишь в простейших идеализированных случаях. В оптике значительно большее значение имеют нестрогие методы решения дифракционных задач, основанные на принципе Гюйгенса в обобщенной формулировке Френеля или Кирхгофа.

2. В § 3 мы указывали, что принцип Гюйгенса в том виде, в каком он был сформулирован самим Гюйгенсом, есть не более чем *геометрический рецепт* для построения волновых фронтов. Во всех применениях вторичные волны Гюйгенса выступают не как реальные волны, а как *вспомогательные сферы*, используемые для такого построения. Эти сферы, построенные из точек волнового фронта как из центров, проявляют свое действие *только на огибающей*, которая и дает новое положение волнового фронта. При этом оставалось необъясненным, почему при распространении волны не возникает *обратная волна*.

Искусственную гипотезу об огибающей вторичных волн Френель заменил физически ясным положением, согласно которому вторичные волны при наложении *интерферируют* друг с другом. Свет должен наблюдаться во всех местах пространства, где при интерференции вторичные волны усиливаются; в тех же местах, где они взаимно гасят друг друга, должна наблюдаться темнота. Тем самым выясняется и физический смысл огибающей. К огибающей все вторичные волны приходят в *одинаковых фазах*, и их интерференция приводит к большой интенсивности света. Становится ясным, по крайней мере качественно, и отсутствие обратной волны. Вторичные волны, идущие от волнового фронта вперед, вступают в свободное от возмущений пространство. Они интерферируют только друг с другом. Напротив, вторичные волны, идущие назад, вступают в пространство, где уже есть волновое возмущение — прямая волна. При интерференции вторичные волны *гасят прямую волну*, так что после прохождения волны пространство за ней оказывается невозмущенным.

Френель дал следующую формулировку принципа Гюйгенса, несколько обобщенную Рэлеем (1842–1919). *Окружим все источники света S_1, S_2, S_3, ... произвольной замкнутой поверхностью F* (рис. 150). *Каждую точку такой поверхности можно рассматривать как источник вторичных волн, распространяющихся во всех направлениях. Эти волны когерентны, поскольку все они возбуждаются одними и теми же первичными источниками. Световое поле, возникающее в результате их интерференции, в пространстве вне поверхности F совпадает с полем реальных источников света.*

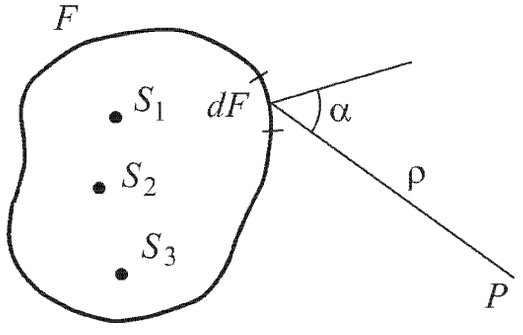


Рис. 150

Таким образом, действительные источники света можно как бы заменить окружающей их светящейся поверхностью F с непрерывно распределенными по ней когерентными вторичными источниками. Отличие этой поверхности от реальной поверхности излучающего тела состоит в том, что она абсолютно прозрачна для всякого излучения. В такой формулировке принцип Гюйгенса–Френеля выражает весьма общее положение. Он означает, что волна, отделившаяся от своих источников, в дальнейшем ведет *автономное существование*, совершенно не зависящее от наличия источников.

3. При математической формулировке принципа Гюйгенса–Френеля будем предполагать, что источники света — монохроматические с одной и той же частотой ω. По предположению Френеля каждый элемент площади dF поверхности F (рис. 150)

испускает вторичную сферическую волну, а волновое поле в точке наблюдения P представляется суперпозицией таких волн в виде интеграла

$$E = \int \frac{a\,dF}{\rho} e^{i(\omega t - k\rho)} \tag{39.1}$$

по всей (замкнутой) поверхности F. Френель предположил, что амплитуда вторичной волны пропорциональна амплитуде первичной волны, приходящей к элементу dF, а также площади самого элемента dF. Кроме того, она убывает с возрастанием угла между нормалью к поверхности F и направлением излучения вторичной волны. Конкретного выражения для амплитуды вторичной волны Френель не дал. Это было сделано только в 1883 г. Кирхгофом (1824–1887) при строгой формулировке принципа Гюйгенса–Френеля (см. § 43). Несмотря на это, Френелю удалось получить правильные решения многих дифракционных задач, пользуясь простыми интуитивными соображениями и нестрогими методами расчета. Эти простые и наглядные методы, приводящие быстро к цели, до сих пор используются при элементарном изложении теории дифракции. Они взяты за основу и в настоящей книге. Мы изменим только те (не влияющие на окончательный результат) предположения Френеля, которые оказались неверными.

4. Рассмотрим сначала свободное распространение сферической волны

$$E_0 = \frac{1}{r_0} e^{i(\omega t - k r_0)} \tag{39.2}$$

в однородной среде (рис. 151). Перед этим выражением можно было бы поставить какую-то постоянную амплитуду A_0. Мы не делаем этого, выбирая единицы измерения так, чтобы было $A_0 = 1$. В качестве вспомогательной поверхности F в формуле (39.1) выберем сферический волновой фронт радиуса r_0. Согласно гипотезе Френеля, $a = K(\alpha)e^{-ikr_0}/r_0$, где функция $K(\alpha)$, помимо длины волны, зависит только от угла между нормалью к волновому фронту и направлением излучения вторичной волны Гюйгенса. Полное поле в точке наблюдения P представится интегралом

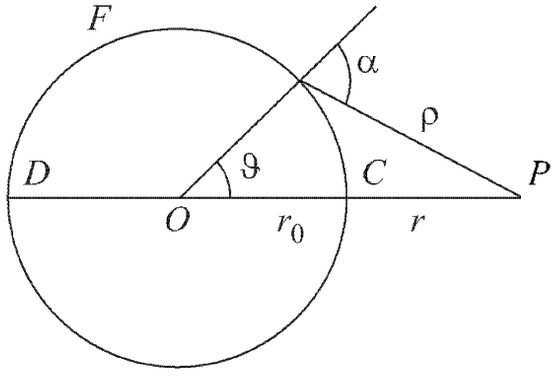


Рис. 151

$$E = \int \frac{K(\alpha)}{r_0 \rho} e^{i(\omega t - k r_0 - k\rho)}\, dF.$$

Если за элемент площади dF принять площадь кольца, выре-

заемого из волнового фронта двумя бесконечно близкими концентрическими сферами с центрами в точке наблюдения P, то $dF = 2\pi r_0^2 \sin\vartheta\, d\vartheta$. Примем за переменную интегрирования расстояние ρ. Дифференцируя соотношение $\rho^2 = r_0^2 + (r_0 + r)^2 - 2r_0(r_0 + r)\cos\vartheta$ при постоянных r_0 и r, находим $\sin\vartheta\, d\vartheta$, а затем и элемент площади dF. В результате получим

$$E = \frac{2\pi}{r_0 + r}\, e^{i(\omega t - kr_0)} \int\limits_{r}^{r_{\text{макс}}} K(\rho) e^{-ik\rho}\, d\rho, \tag{39.3}$$

где прежняя функция $K(\alpha)$ теперь рассматривается как функция ρ. Верхний предел интеграла равен $r_{\text{макс}} = r + 2r_0$. Однако, имея в виду дальнейшие применения, мы не будем его конкретизировать. Точное вычисление интеграла (39.3), конечно, невозможно без знания вида функции $K(\rho)$. Однако Френель, используя малость длины световой волны, дал метод приближенного вычисления подобных интегралов при весьма общих предположениях относительно функции $K(\rho)$. Опишем из точки P, как из центра, концентрические сферы с радиусами r, $r + \lambda/2$, $r + 2(\lambda/2)$, $r + 3(\lambda/2)$, ... (рис. 152). Они разобьют волновой фронт F на кольцевые области, получившие название *зон Френеля*. Ввиду малости длины волны, аргумент ρ, а с ним и функция $K(\rho)$ в пределах одной зоны могут считаться постоянными. В этом приближении интеграл (39.3) по n-й зоне будет равен

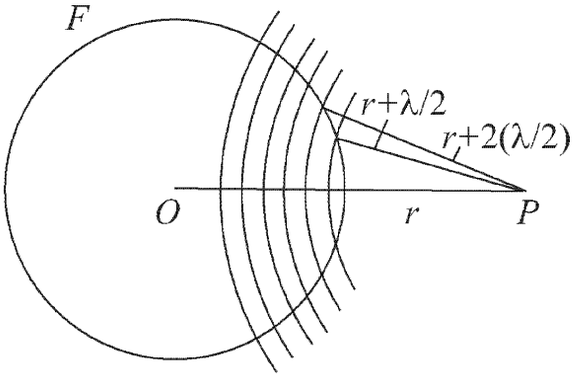


Рис. 152

$$K_n \int\limits_{r + (n-1)\lambda/2}^{r + n\lambda/2} e^{-ik\rho}\, d\rho = (-1)^{n+1} \frac{2K_n}{ik}\, e^{-ikr},$$

где через K_n обозначено среднее значение функции $K(\rho)$ для n-й зоны. Если N — общее число зон, посылающих вторичные волны в точку наблюдения, то поле E представится знакопеременной суммой

$$E = E_1 + E_2 + E_3 + \ldots + E_N. \tag{39.4}$$

Общий член ее имеет вид

$$E_n = (-1)^{n+1} \frac{4\pi K_n}{ik(r_0 + r)}\, e^{i[\omega t - k(r_0 + r)]}. \tag{39.5}$$

То обстоятельство, что знаки рядом стоящих слагаемых в сумме (39.4) противоположны, означает, что колебания, вносимые

соседними зонами Френеля, противоположны по фазе. Этого и следовало ожидать, так как из самого построения зон Френеля видно, что одно из этих двух колебаний запаздывает относительно другого на полволны. По предположению Френеля абсолютные значения множителей K_n, а с ними и членов суммы (39.4) медленно убывают с возрастанием номера n. Уберем в сумме (39.4) слагаемое E_1 и добавим слагаемое E_{N+1}. Ввиду медленности убывания членов суммы (39.4), интуитивно можно ожидать, что от этого абсолютное значение суммы (39.4) почти не изменится, а ее знак заменится на противоположный. Приняв это, наряду с (39.4) можно также написать:

$$E = -E_2 - E_3 - \ldots - E_N - E_{N+1}.$$

Тогда $2E = E_1 - E_{N+1}$. Но в том же приближении $E_N = -E_{N-1}$, и следовательно,

$$E = \frac{1}{2}(E_1 + E_n). \tag{39.6}$$

Таким образом, волновое возмущение, создаваемое N первыми зонами Френеля, равно полусумме возмущений, вносимых крайними зонами.

Возьмем теперь все зоны на сферическом волновом фронте F (см. рис. 151). Введем предположение (обосновываемое при строгой формулировке принципа Гюйгенса, данной Кирхгофом), что при $\alpha = \pi$, т.е. в точке D волнового фронта, функция $K(\alpha)$ обращается в нуль. Тогда $E_N = 0$, и из формулы (39.6) получаем

$$E = \frac{1}{2}E_1 = \frac{2\pi K_1}{ik(r_0 + r)}\, e^{i[\omega t - k(r_0 + r)]}, \tag{39.7}$$

т.е. *при свободном распространении волны волновое возмущение от всего волнового фронта составляет половину возмущения, даваемого одной только первой зоной Френеля.* Дело происходит так, как если бы из всего волнового фронта действующей осталась только часть первой зоны Френеля [1]).

5. Формула (39.7) интересна также в том отношении, что она позволяет точно определить множитель K_1. Действительно, при свободном распространении сферической волны (39.2)

[1]) В полученном результате авторы многих книг видят доказательство прямолинейного распространения света, утверждая, что свет распространяется как бы в узком прямолинейном канале, поперечное сечение которого порядка размеров первой френелевой зоны. Однако не надо забывать, что этот вывод относится к *свободному* распространению волны, когда проблема прямолинейного распространения света вообще не возникает. Она возникает лишь *при наличии препятствий* на пути распространения света. Это есть *проблема теней*, являющаяся частным случаем *проблемы дифракции света.*

напряженность поля заранее известна в любой точке пространства. В частности, в точке наблюдения

$$E = \frac{1}{r_0 + r}\, e^{i[\omega t - k(r_0 + r)]}.$$

Сравнивая это выражение с (39.7), находим

$$K_1 = \frac{ik}{2\pi} = \frac{k}{2\pi}\, e^{i\pi/2}. \tag{39.8}$$

Таким образом, множитель K_1, а также все остальные множители K_2, K_3, ..., — чисто мнимые. Значит, *вторичные волны Гюйгенса опережают по фазе колебания поля* в точках волнового фронта на $\pi/2$. Если бы такого опережения не было, то вторичные волны из центра первой френелевой зоны приходили бы в точку наблюдения с фазой $\omega t - k(r_0 + r)$, а от ее краев — с меньшей фазой $\omega t - k(r_0 + r) - \pi$. Фаза результирующего колебания, возбуждаемого всей первой зоной, была бы равна полусумме этих значений, т.е. оказалась бы меньше правильного значения на $\pi/2$.

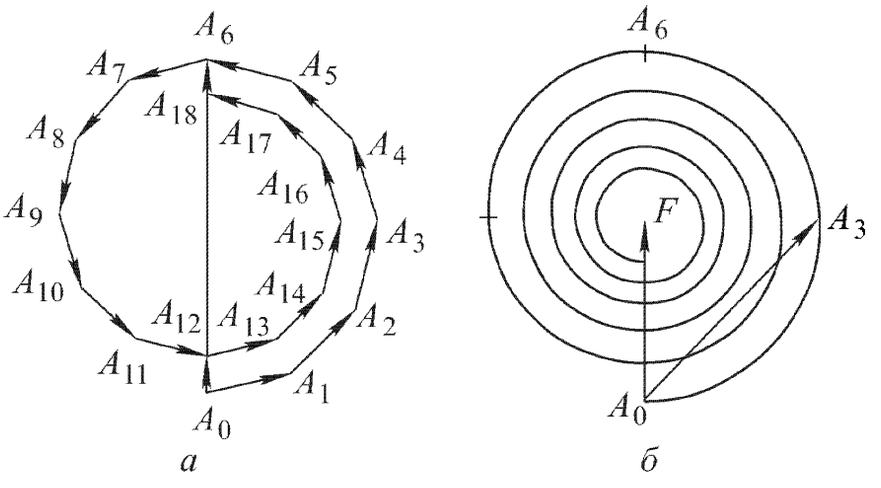


Рис. 153

6. Результаты (39.6) и (39.7) имеют основное значение в методе зон Френеля. Для лучшего уяснения интерпретируем их на векторной диаграмме (см. т. III, § 126). Разобьем каждую зону Френеля на m кольцевых подзон (на рис. 153 *а* построение выполнено для $m = 6$). Колебания, возбуждаемые в точке наблюдения такими подзонами, на векторной диаграмме изобразятся векторами $\overrightarrow{A_0A_1}$, $\overrightarrow{A_1A_2}$, $\overrightarrow{A_2A_3}$ и т.д., образующими ломаную линию. Колебание, возбуждаемое несколькими соседними подзонами, представится геометрической суммой таких векторов. Например, вектор $\overrightarrow{A_0A_6}$ представляет «действие» первой френелевой зоны, вектор $\overrightarrow{A_0A_{12}}$ — совместное действие двух первых френелевых зон, и т.п. Если число подзон N устремить к бесконечности, то в пределе ломаная перейдет в непрерывную спи-

раль, вьющуюся вокруг фокуса F (рис. 153 *б*). На этой спирали действие первой френелевой зоны представляется вектором $\overrightarrow{A_0A_6}$, а всего волнового фронта — вектором $\overrightarrow{A_0F}$, который почти точно вдвое короче предыдущего. Действие центральной половины первой зоны представляется наклонным вектором $\overrightarrow{A_0A_3}$, и т.п.

7. В § 27 (п. 8) указывалось, что прохождение сходящейся сферической волны через фокус (т.е. ее центр) сопровождается изменением фазы на π. Приведем новое истолкование этого явления, используя метод зон Френеля.

Как всегда, зоны Френеля надо строить с той стороны волнового фронта, *куда он распространяется*. В рассматриваемом случае это будет вогнутая сторона. Конечно, и здесь вторичные волны будут опережать по фазе колебания поля на волновом фронте, так как это есть свойство *самих элементарных источников вторичных волн*, не зависящее от формы волнового фронта, на котором они находятся. Действие же всего волнового фронта в точке наблюдения при свободном распространении волны по-прежнему будет равно половине действия первой френелевой зоны. В частности, вторичные волны, исходящие от этой зоны, при интерференции всегда дают правильную фазу, независимо от положения точки наблюдения. Если точка наблюдения P лежит перед центром (фокусом) O сходящейся сферической волны (рис. 154), то в наших предыдущих рассуждениях ничего не изменится, так как в этом случае колебания от краев A и B первой френелевой зоны приходят в P позже, чем от ее середины C. Но положение изменится, когда точка наблюдения P' будет находиться по другую сторону от центра O (тогда, конечно, и первая зона $A'B'$ будет другой). Теперь края первой зоны A' и B' окажутся ближе на полволны к точке P', чем ее центр C. Вследствие этого, помимо обычного опережения по фазе на $\pi/2$, вторичные волны, приходящие в P' от краев A' и B' первой зоны, получат дополнительное приращение фазы на π, так что полное опережение по фазе вследствие указанных двух причин составит $\dfrac{1}{2}\left(\dfrac{\pi}{2}+\dfrac{3\pi}{2}\right)=\pi$. Его надо добавить к фазе волны $\omega t - kr$, прошедшей через точку O. В результате мы и приходим к явлению изменения фазы сходящейся сферической волны при ее прохождении через фокус.

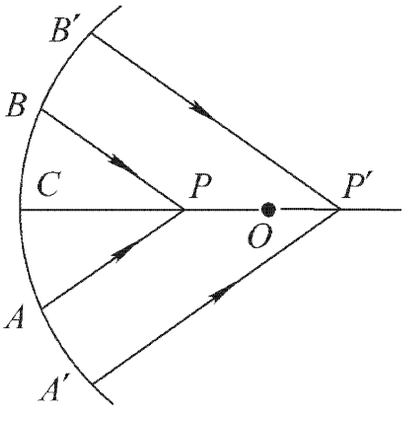


Рис. 154

Когда точка наблюдения P приближается к фокусу O, френелевы зоны расширяются. В некотором положении P_0 этой точки первая френелева зона охватывает всю сферу. При дальнейшем приближении к O построение зон Френеля становится невозможным. Можно сказать, что вся сфера становится как бы частью первой зоны Френеля. Применяя к этому случаю наши предыдущие рассуждения, нетрудно убедиться, что появляется опережение по фазе, заключенное между нулем и $\pi/2$, когда точка наблюдения находится между P_0 и O. В точке O оно составляет $\pi/2$, а при переходе через нее возрастает до π. Таким образом, изменение фазы волны на π при переходе через фокус совершается не скачком, а непрерывно.

§ 40. Дифракция на оси от круглого отверстия и экрана. Зонная пластинка

1. Поставим между точечным источником S и точкой наблюдения P непрозрачный экран с круглым отверстием, плоскость которого перпендикулярна к оси SP, а центр O расположен на той же оси (рис. 155). Согласно Френелю, действие такого препятствия сводится к тому, что экран как бы устраняет ту часть волнового фронта, которую он прикрывает. На открытой же части волнового фронта световое поле не изменяется. Такое предположение соответствует приближению геометрической оптики, а потому оно может быть приближенно верным только тогда, когда радиус отверстия очень велик по сравнению с длиной световой волны. Это и предполагается в дальнейшем. Будем предполагать, кроме того, что размеры отверстия можно менять, что дает возможность открывать любое число зон Френеля.

Рис. 155

Пусть a_0 и I_0 означают амплитуду и интенсивность света в точке P при свободном распространении волны, т.е. в отсутствие экрана. Общий световой поток, поступающий через отверстие, строго пропорционален его площади. Но он будет распределяться по-разному по освещаемой поверхности, в зависимости от того, сколько, зон Френеля укладывается в отверстии. В одних местах может получиться интенсивность меньше, в других больше I_0. Никакого противоречия с законом сохранения энергии в этом нет.

Если открыть первую зону Френеля, то амплитуда и интенсивность света в той же точке будут $a_1 = 2a_0$, $I_1 = (2a_0)^2 = 4I_0$. Таким образом, интенсивность в центре P дифракционной картины получится в четыре раза большей, чем было бы при свободном распространении волны. При удалении от центра P интенсивность будет монотонно убывать. При расширении отверстия в точку P начнут приходить вторичные волны. Их интерференция с ранее пришедшими волнами вызовет уменьшение интенсивности в той же точке. При определенных размерах отверстия центр P перестанет быть точкой максимальной интенсивности. Вокруг точки P начнет образовываться светлое кольцо, к которому и переместится максимум интенсивности. Когда отвер-

стие откроет две первые зоны Френеля, то их действия в точке P практически полностью уничтожат друг друга из-за интерференции. В точке P получится темный кружок, окруженный светлым кольцом.

При дальнейшем увеличении размеров отверстия действия двух первых зон Френеля в точке P остаются компенсированными. Все поле в P создается только частью третьей зоны Френеля. В центре картины появляется светлое пятнышко, а центральный темный кружок расширяется и переходит в темное кольцо, окружающее это пятнышко. Когда число N открытых зон равно трем, интенсивность света в точке P будет такой же, какая получилась бы, если бы была открыта только одна третья зона. Центр картины будет практически столь же светлым, что и при одной открытой первой зоне Френеля. При $N = 4$ светлый центр сменится темным. Вообще, при нечетном N центр дифракционных колец светлый, а при четном — темный. На схематическом рис. 156 для различных N показано распределение интенсив-

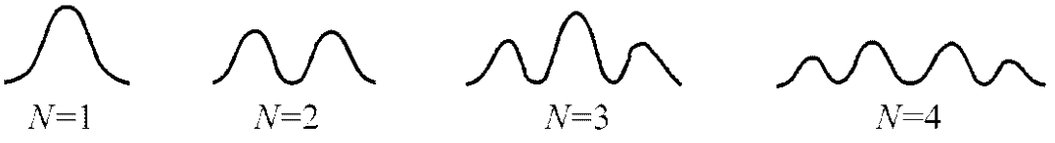


Рис. 156

ности света в зависимости от расстояния до центра картины, а рис. 157 воспроизводит фотографии наблюдаемой картины колец при нечетном ($а$) и четном ($б$) N.

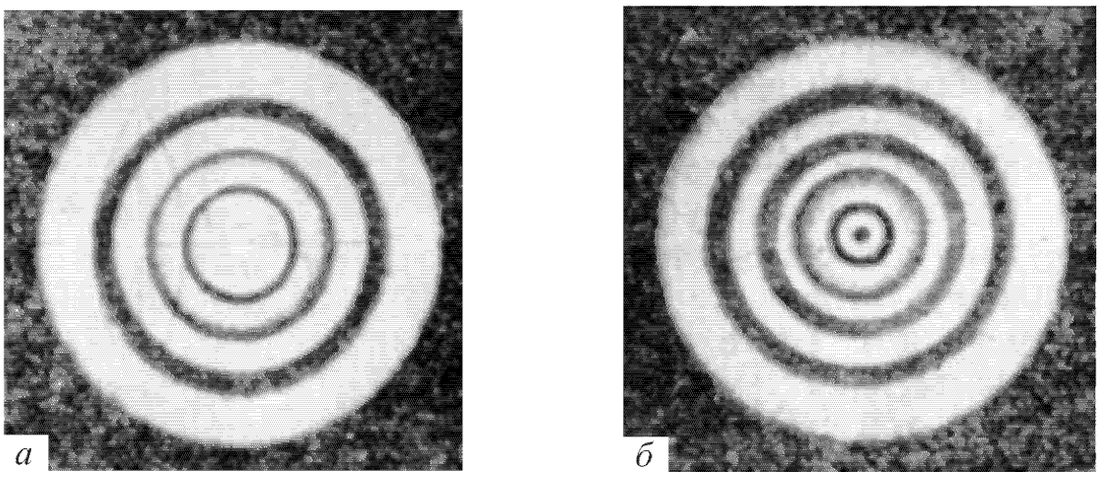


Рис. 157

2. Определим теперь размеры и число m зон Френеля, укладывающихся в отверстии AB (см. рис. 155). Пусть D — диаметр отверстия, а a и b — расстояния от его центра до точек S и P. Из точек S и P как из центров опишем сферы, проходящие че-

рез край отверстия AB. Пренебрегая квадратами отрезков OE и OF, по известной геометрической теореме можем написать:

$$(D/2)^2 \approx OF \cdot 2a, \quad (D/2)^2 = OE \cdot 2b.$$

Отсюда

$$EF = EO + OF = \frac{D^2}{8}\left(\frac{1}{a} + \frac{1}{b}\right).$$

Число m найдется делением этого отрезка на $\lambda/2$. Оно равно

$$m = \frac{D^2}{4\lambda}\left(\frac{1}{a} + \frac{1}{b}\right). \tag{40.1}$$

Если m целое, то D будет диаметром, а $R_m = D/2$ — радиусом m-й зоны, точнее — ее внешнего края. Следовательно,

$$R_m = \sqrt{\frac{ab}{a+b}\, m\lambda}. \tag{40.2}$$

Например, если $a = b = 1$ м, $\lambda = 600$ нм, то $R_1 \approx 0{,}55$, $R_2 \approx 0{,}77$, $R_3 \approx 0{,}95$ мм и т.д.

3. Интенсивность света в точке наблюдения P можно во много раз усилить, прикрыв все четные или все нечетные зоны Френеля. Оставшиеся неприкрытыми зоны будут усиливать действие друг друга. Прикрытие можно осуществить, поместив в плоскости отверстия так называемую *зонную пластинку* (рис. 158). Ее можно изготовить, начертив на листе бумаги темные кольца, а затем сфотографировав их в уменьшенном масштабе. Внутренние радиусы колец должны быть пропорциональны квадратным корням из последовательных нечетных чисел, а внешние — из четных. Тогда получится пластинка, центр которой светлый. Можно изготовить аналогичную пластинку с темным центром. Ширина всех колец должна быть велика по сравнению с длиной волны. Тогда при надлежащих размерах колец пластинка со светлым центром будет удалять из волнового фронта все четные, а пластинка с темным центром — все нечетные зоны Френеля.

Рис. 158

Усиление интенсивности света зонной пластинкой аналогично фокусирующему действию линзы. Более того, расстояния от

пластинки до источника S и «изображения» P связаны тем же соотношением, что и соответствующие величины для линзы. Это видно, если формулу (40.1) переписать в виде

$$\frac{1}{a}+\frac{1}{b}=\frac{1}{f}, \tag{40.3}$$

где «фокусное расстояние» определяется формулой

$$\frac{1}{f}=\frac{4m\lambda}{D^2}=\frac{m\lambda}{R_m^2}. \tag{40.4}$$

Если центр зонной пластинки светлый, то число m — нечетное, в этом случае в формулу (40.4) входит (внешний) радиус светлого кольца пластинки. Если же центр пластинки темный, то число m — четное и под R_m следует понимать (внешний) радиус темного кольца. Какой номер брать при вычислении f — это, конечно, не имеет значения.

С помощью зонной пластинки можно даже получать оптические изображения, хотя и весьма низкого качества.

В отличие от линзы, зонная пластинка имеет *несколько фокусов*. Действительно, найдем положение точки наблюдения P, при котором в центральном круге пластинки уместятся первые три зоны Френеля. Тогда в следующее кольцо пластинки попадут четвертая, пятая и шестая зоны и т.д. Если центр зонной пластинки светлый, то поле в точке P представится суммой

$$E=(E_1+E_2+E_3)+(E_7+E_8+E_9)+(E_{13}+E_{14}+E_{15})+\ldots,$$

или $E=E_1+E_7+E_{13}+\ldots$, так как действия соседних зон практически уничтожают друг друга. Таким образом в P получится максимум (фокус). Фокусное расстояние найдется по формуле $f_1=R_1/(3\lambda)=f/3$. Аналогично находятся фокусы высших порядков:

$$f_n=\frac{f}{2n+1}, \tag{40.5}$$

где n — целые числа, которым можно придавать не только положительные, но и отрицательные значения. Отрицательным значениям соответствуют расходящиеся волны и *мнимые фокусы*.

Рэлей указал, что интенсивность света в точке наблюдения P увеличится в четыре раза, если изменить на π фазы вторичных волн, исходящих от всех зон Френеля с четными (или нечетными) номерами. Такая *«зонная пластинка с обращением фазы»* была изготовлена Вудом (1868–1955) путем травления поверхности стеклянной пластинки. Ее действие вполне эквивалентно действию линзы, поскольку в обоих случаях вторичные волны от всех точек волнового фронта приходят в P в одинаковых фазах.

4. Поставим теперь между источником S и точкой наблюдения P непрозрачный круглый экран AB (рис. 159), плоскость

которого перпендикулярна к оси SP. Пусть DA и BE — неприкрытые части волнового фронта сферической волны, исходящей из источника S. Разобьем ее на кольцевые зоны Френеля, начав их построение от края экрана. Рассуждая как и раньше, представим напряженность поля излучения в точке P в виде половины напряженности, создаваемой в этой точке вторичными волнами первой кольцевой зоны Френеля. Следовательно, каков бы ни был диаметр диска, в центре P его геометрической тени должно наблюдаться *светлое пятнышко.* Такой вывод был сделан Пуассоном (1781–1840) и показался последнему столь абсурдным, что он выдвинул его в качестве возражения против волновой теории света Френеля. Араго (1786–1853) немедленно поставил опыт и обнаружил пятнышко в соответствии с выводом Пуассона [1]). Явление получило название *пятна Араго–Пуассона.* Наблюдаемая картина дифракции от круглого экрана приведена на рис. 160.

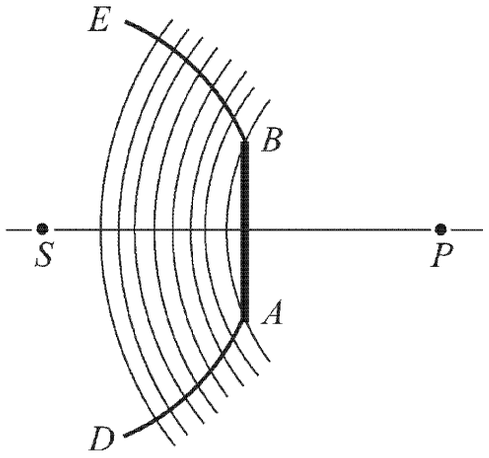


Рис. 159

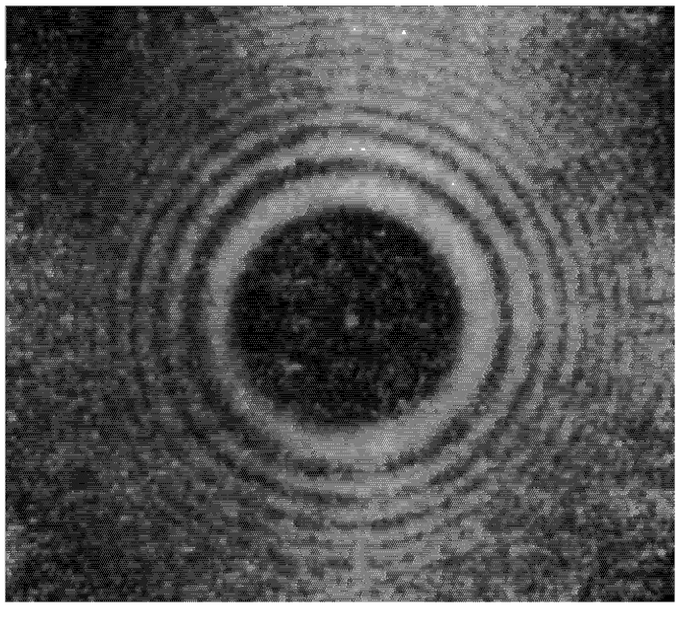


Рис. 160

Если точка наблюдения P не находится в центре картины, то кольцевые зоны Френеля, конечно, можно построить и для нее. Однако зоны, расположенные ближе к центру, окажутся неприкрытыми лишь частично. Это сильно усложняет вычисление интенсивности света. Можно только сказать, что дифракционная картина должна обладать осевой симметрией. Вне геометрической тени получается система концентрических светлых и темных колец. Внутри самой

[1]) Маральди наблюдал такое пятнышко еще в 1723 г. Вероятно, еще раньше (1715 г.) оно наблюдалось Делилем, хотя указания последнего недостаточно ясны. Однако эти наблюдения остались незамеченными и были забыты, поскольку природа явления не была понята.

геометрической тени также могут получиться дифракционные кольца, в особенности когда экран прикрывает небольшое число зон Френеля. Но эти кольца мало контрастны, а распределение света в них сложное.

Рис. 161

Демонстрацией явления пятна Араго–Пуассона может служить опыт Поля. Поль сфотографировал ярко освещенный шаблон, заменив объектив гладким металлическим шаром. Эта фотография воспроизведена на рис. 161, а параметры установки приведены в задаче в конце этого параграфа. Ангерер сфотографировал изображение человеческого лица, заменив шар металлическим диском.

5. Применим метод зон Френеля к объяснению *теней*, т.е. *прямолинейного распространения света*. Поскольку речь идет о законе, от которого принципиально должны наблюдаться отступления, наши рассуждения не могут претендовать на строгость.

Пусть на пути распространяющейся волны поставлен экран или отверстие произвольной формы. Их размеры должны быть велики по сравнению с длиной волны. Разобьем волновой фронт на кольцевые зоны Френеля. Некоторые зоны могут оказаться открытыми полностью, другие частично, третьи совсем закрытыми.

Допустим сначала, что точка наблюдения лежит вне геометрической тени, далеко от ее границы. Первые члены ряда (39.4) получатся такими же, как и при свободном распространении волны. Последующие члены начнут изменяться из-за частичного экранирования соответствующих им зон. В зависимости от формы края экрана эти изменения будут носить более или менее нерегулярный характер. Если обнаружится тенденция убывания их по абсолютной величине, то правдоподобно допустить, что напряженность поля в точке наблюдения окажется равной половине напряженности, создаваемой центральной зоной. В случае точечного источника и ровного края экрана это может оказаться и не совсем так. Однако если источник не совсем точечный, а края экрана не совсем ровные, то произойдет *статистическое сглаживание* при наложении дифракционных картин от точечных источников, на которые можно разложить протяженный источник. Тогда вдали от края экрана получится такая же освещенность, какая получилась бы при свободном распространении волны.

Допустим теперь, что точка наблюдения лежит внутри геометрической тени, опять далеко от ее границы. Первые зоны Френеля будут полностью закрыты. Нумерацию зон начнем с

первой (частично) открытой зоны. Представим поле рядом

$$E = (E_1 + E_2 + \ldots + E_N) + (E_{N+1} + E_{N+2} + \ldots),$$

в котором $N+1$ означает номер первой целиком открытой зоны. Ко второй скобке применимы рассуждения, применявшиеся выше в случае свободного распространения волны. Поэтому вторая скобка приближенно равна $^1/_2 E_{N+1}$. В первой скобке слагаемые меняются более или менее нерегулярно, обнаруживая в среднем тенденцию возрастания по абсолютной величине. При статистическом усреднении (с учетом неполной точечности источника и неровностей краев экрана) эти нерегулярности сглаживаются, так что первую скобку можно принять равной $^1/_2(E_1 + E_N)$, или $^1/_2 E_N$, поскольку величина E_1 должна считаться близкой к нулю из-за малости открытой части соответствующей зоны. Таким образом, $E = {}^1/_2(E_N + E_{N+1}) \approx 0$, так как для соседних зон $E_N \approx \approx -E_{N+1}$. Итак, при погружении в область геометрической тени интенсивность света обращается в нуль.

Если источник точечный, а края экрана резкие, то граница геометрической тени расщепляется в дифракционные полосы, как это мы видели при рассмотрении дифракции на круглых отверстии и экране. Однако если края экрана неровные, то полосы начинают размываться, а при увеличении размеров источника переходят в полутень.

ЗАДАЧИ

1. В опыте Поля, описанном в тексте, диаметр шара $D = 40$ мм, расстояние от фотографируемого шаблона до шара $a = 12$ м, расстояние от шара до изображения $b = 18$ м, размер шаблона $y = 7$ мм. Определить размер его изображения y'. При каких условиях опыт удастся с шаром, поверхность которого испещрена множеством неправильных царапин?

О т в е т. $y' = \dfrac{b}{a} y = 10{,}5$ мм. Для удачи опыта необходимо, чтобы глубина царапин h не превосходила ширины крайней френелевой зоны:

$$h = \frac{\lambda}{D} \frac{ab}{a+b} = 180\lambda = 0{,}1 \text{ мм}.$$

2. Оценить максимальные угловые размеры α предмета, который можно сфотографировать с помощью непрозрачного диска с идеально гладкими краями.

Р е ш е н и е. При смещении точечного объекта в сторону с главной оптической оси диск представится эллипсом. Малая полуось эллипса будет отличаться от радиуса диска на $r\left(1 - \cos\dfrac{\alpha}{2}\right)$. Это отличие не должно превышать ширины крайней френелевой зоны. Из этого условия для максимально допустимых угловых размеров предмета находим

$$\alpha < \frac{4}{D}\sqrt{\frac{\lambda ab}{a+b}} \approx 0{,}2 \text{ рад} \approx 10^\circ,$$

где использованы обозначения и численные данные из предыдущей задачи.

3. Зонная пластинка применяется для фотографирования предмета, который виден из места нахождения пластинки под углом $\alpha = 0{,}1$ рад. Оценить оптимальное число зон пластинки для получения наибольшей яркости и отчетливости изображения.

Р е ш е н и е. Допустим сначала, что источник света S точечный, а зонная пластинка CD наклонена к оптической оси под углом $(\pi - \alpha)/2$ (рис. 162). Из рисунка видно: $x^2 = a^2 + R^2 + 2aR\sin\dfrac{\alpha}{2}$, и аналогично для y. Извлекая квадратные корни и пренебрегая всеми степенями радиуса R, начиная с третьей, получим для разности хода между лучами SCP и SOP:

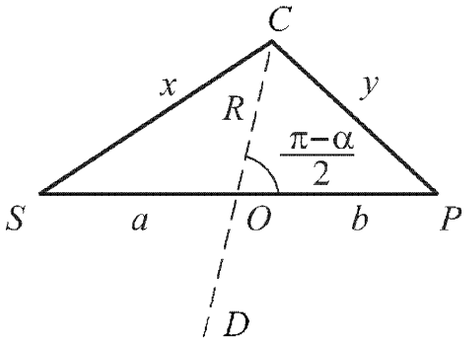


Рис. 162

$$\Delta = (x + y) - (a + b) = \frac{R\cos^2(\alpha/2)}{2}\left(\frac{1}{a} + \frac{1}{b}\right),$$

или $\Delta = \Delta_0 - \delta\Delta$. Здесь Δ_0 — значение Δ при отсутствии наклона зонной пластинки, а $\delta\Delta$ — приращение величины Δ, обусловленное наклоном:

$$\delta\Delta = \Delta_0\left(1 - \cos^2\frac{\alpha}{2}\right) = \Delta_0\sin^2\frac{\alpha}{2} \approx \frac{1}{4}\Delta_0\alpha^2.$$

Если $\delta\Delta \ll \lambda$, то наклон пластинки не скажется существенно на работе зон, расположенных в пределах круга радиуса R. Если же $\delta\Delta \approx \lambda/2$, то все зоны, расположенные выше этого круга, становятся бесполезными и даже вредными. Из этого условия находится предельное значение разности хода: $\Delta_0 \approx 2\lambda/\alpha^2$. Соответствующее число зон Френеля будет

$$N \approx \frac{\Delta_0}{\lambda/2} \approx \frac{4}{\alpha^2}.$$

Допустим теперь, что фотографируемый предмет не точечный, причем его центр расположен на оси зонной пластинки. Для периферийных точек предмета, не лежащих на оси пластинки, последняя действует как наклоненная под углом $\alpha/2$. Поэтому предельное число зон Френеля, при котором должно получиться наиболее отчетливое изображение, будет

$$N \sim 4/\alpha^2 \sim 400.$$

§ 41. Метод Френеля решения дифракционных задач. Дифракция Фраунгофера и Френеля

1. Принципа Гюйгенса еще недостаточно для решения дифракционных задач, если даже пользоваться точной формулировкой его, данной Кирхгофом (см. § 43). Этот принцип сводит только любую дифракционную задачу к определению волнового поля на произвольной замкнутой поверхности F, окружающей все источники света (см. рис. 150). Но точное определение поля на поверхности F возможно лишь после нахождения его во всем

пространстве. Эту трудность Френель преодолел введением специальной гипотезы, которой мы, в сущности, уже пользовались в предыдущем параграфе. Сформулируем ее на примере непрозрачного плоского экрана с отверстиями, поставленного на пути распространяющейся волны.

Выберем в качестве вспомогательной поверхности F заднюю (т.е. неосвещаемую) сторону экрана. Примем, что на всех участках этой поверхности, которые прикрыты экраном, волновое поле равно нулю, а на отверстиях оно определяется приближенными законами геометрической оптики, т.е. поле такое, какое получилось бы в отсутствие экрана. Тем самым интегрирование в (39.1) распространяется только на отверстия, где волновое поле считается известным. Отметим недостатки введенной гипотезы.

1) Распространение гипотезы на неплоские экраны неоднозначно, так как в этом случае неясно, как провести вспомогательную поверхность через отверстия экрана.

2) Гипотеза противоречива. Если с помощью принципа Гюйгенса вычислить волновое поле во всем пространстве, то на поверхности F оно не совпадает с исходным полем, принятым при вычислении. На задних сторонах экранов вычисленное поле не обратится в нуль, а на отверстиях не будет совпадать с полем свободно распространяющейся волны.

3) Гипотеза допускает разрыв волнового поля на краях отверстий, что противоречит граничным условиям электродинамики Максвелла.

4) Гипотеза противоречит поперечности световых волн. Это противоречие не устраняется заменой скалярных колебаний векторными. Колебания, как и при изучении явлений интерференции, мы считали скалярными лишь ради простоты. При переходе к векторным колебаниям математическая формулировка принципа Гюйгенса и полученные из него результаты по существу не изменятся. Как строго показал Кирхгоф, для этого достаточно только скаляр E всюду заменить вектором $\mathbf{E}$.

Допустим теперь, что на непрозрачный экран с отверстием нормально падает плоская линейно поляризованная электромагнитная волна. На вспомогательной поверхности F вектор $\mathbf{E}$ будет иметь одно и то же направление, параллельное плоскости экрана. Принцип Гюйгенса сводит задачу о дифракции к суперпозиции коллинеарных векторных колебаний того же направления. Поэтому следует ожидать, что в дифрагированной волне вектор $\mathbf{E}$ всюду будет параллелен плоскости экрана. Это будет так и вдали от экрана, где дифрагированные волны разных направлений расходятся и перестают накладываться друг на друга. Так будет и в волне, дифрагировавшей косо к плоскости экрана. Но в действительности вектор $\mathbf{E}$ перпендикулярен

к дифрагирующим лучам и образует с вычисленным направлением угол, равный «углу дифракции» ϑ (рис. 163).

Метод Френеля решения дифракционных задач может быть оправдан, когда размеры отверстий и непрозрачных промежутков между ними велики по сравнению с длиной световой волны, а потому заметная интенсивность света наблюдается лишь при малых углах дифракции. Действительно, в отверстиях законы геометрической оптики нарушаются лишь в непосредственной близи от их краев. Применяя гипотезу Френеля, мы при вычислении интеграла (39.1) пользуемся неправильными значениями подынтегральной функции только внутри узких полосок вблизи краев отверстий. Ширина этих полосок порядка длины световой волны. В основной области интегрирования используются правильные значения подынтегрального выражения. Автоматически снимается и возражение (4), связанное с поперечностью световых колебаний, так как различием направлений вычисленного и действительного полей при малых углах дифракции можно пренебречь. Все это в основном подтверждается совпадением результатов теоретических расчетов с опытом. Но для малых препятствий, сравнимых с длиной волны, и, следовательно, для больших углов дифракции гипотезой Френеля уже нельзя пользоваться.

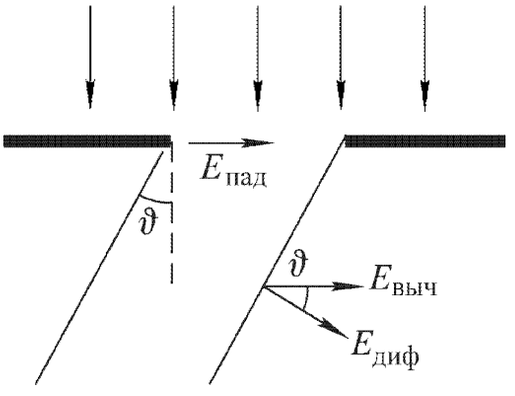


Рис. 163

Отметим еще, что гипотеза Френеля приводит к выводу о независимости дифрагированной волны от материала экрана. Этот вывод в основном также подтверждается опытом. Только более точные опыты обнаруживают и влияние материала экрана на дифракцию света.

2. Явления дифракции принято классифицировать в зависимости от расстояний источника и точки наблюдения (экрана) от препятствия, поставленного на пути распространения света. Если эти расстояния очень велики (бесконечно велики), то дифракция называется *дифракцией в параллельных лучах* или *дифракцией Фраунгофера* (1787–1826). В противоположном случае говорят о *дифракции в непараллельных лучах* или *дифракции Френеля*. Практически для осуществления дифракции Фраунгофера точечный источник света помещают в фокусе собирательной линзы. Получающийся параллельный пучок света дифрагирует на каком-то препятствии. Дифракционная картина наблюдается в фокальной плоскости линзы, поставленной на пути дифрагированного света, или в зрительную трубу, установлен-

ную на бесконечность. Ясно, что между фраунгоферовой и френелевой дифракциями нет принципиального различия и резкой границы. Одна непрерывно переходит в другую.

Для лучшего уяснения приведенной классификации начнем с примера. Рассмотрим круглое отверстие и точечный источник на его оси. Пусть сначала точка наблюдения также находится на оси. Если в отверстии укладывается небольшая часть первой зоны Френеля, то дифракция будет фраунгоферовой. В этом случае все колебания в плоскости отверстия совершаются и приходят в точку наблюдения практически в одинаковых фазах. При смещении точки наблюдения вбок появляются разности фаз между вторичными волнами, приходящими в точку наблюдения от различных точек отверстия. Этим и обусловлено появление дифракционных колец. Если отверстие заменить непрозрачным экраном, то этот случай, по соображениям, которые выяснятся в п. 4, также относят к дифракции Фраунгофера. Если же в отверстии или экране (для точки наблюдения, лежащей на оси системы) укладывается заметная часть первой зоны или несколько зон Френеля, то дифракция считается френелевой.

3. Рассмотрим теперь общий случай. Пусть препятствием является непрозрачный экран с отверстием (рис. 164), S — точечный источник монохроматического света, P — точка наблюдения. За начало координат примем произвольную точку O в плоскости отверстия. По принципу Гюйгенса–Френеля волновое поле в точке P представляется интегралом

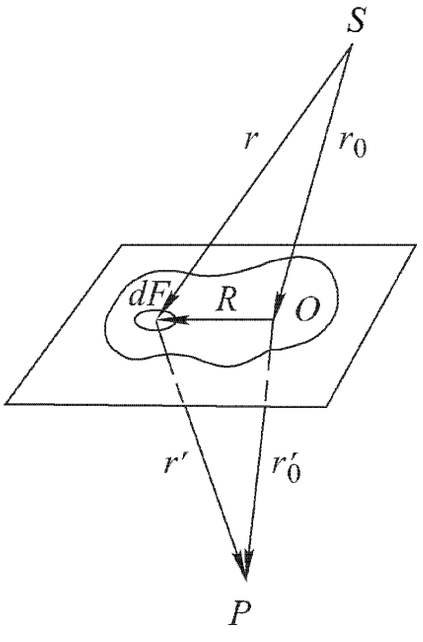


Рис. 164

$$E_P = \int \frac{1}{rr'} e^{i\Phi(\mathbf{R})}\, dF, \qquad (41.1)$$

где $\Phi = \omega t - k(r + r')$. Ввиду малости углов дифракции в подынтегральном выражении (41.1) опущен множитель $K(\alpha)$, определяющий зависимость поля вторичной волны от направления ее излучения. Знаменатель rr' можно считать постоянным, так как размеры отверстия предполагаются очень малыми как по сравнению с r, так и по сравнению с r'. Но зависимость фазы Φ от радиуса-вектора $\mathbf{R}$, определяющего положение элемента площади dF, конечно, должна быть учтена, так как фаза быстро меняется уже на малых расстояниях (порядка ширины френелевой зоны). Интегрирование в (41.1) во всех случаях выполняется по практически небольшому числу (не бо-

лее нескольких десятков) полных или неполных зон Френеля (см., например, следующий параграф). Если расстояния r_0 и r'_0 настолько велики, что разложение фазы $\Phi(\mathbf{R})$ по степеням $\mathbf{R}$ можно оборвать на членах первой степени, то такие расстояния можно считать бесконечно большими. В этом случае имеет место дифракция Фраунгофера. Если же такая точность недостаточна, то дифракция будет френелевой. Обычно в случае френелевой дифракции разложение фазы $\Phi(\mathbf{R})$ по степеням $\mathbf{R}$ достаточно оборвать на членах второй степени.

Найдем количественный критерий, при выполнении которого в фазе Φ можно пренебречь квадратичными членами и, следовательно, считать дифракцию фраунгоферовой. Как видно из рис. 164, $\mathbf{r} = \mathbf{r}_0 + \mathbf{R}$, откуда

$$r^2 = r_0^2 + 2(\mathbf{r}_0\mathbf{R}) + R^2.$$

Извлекая квадратный корень с точностью до квадратичных членов включительно, получим

$$r = r_0 + (\mathbf{sR}) + \frac{1}{2r_0}[R^2 + (\mathbf{sR})^2],$$

где через $\mathbf{s}$ обозначен единичный вектор $\mathbf{s} = \mathbf{r}_0/r_0$. Аналогично вычисляется расстояние r'. В результате получаем

$$\Phi = \omega t - k(r_0 + r'_0) + k(\mathbf{s} - \mathbf{s}')\mathbf{R} + \\ + \frac{k}{2}\left[\left(\frac{1}{r_0} + \frac{1}{r'_0}\right)R^2 + \frac{1}{r_0}(\mathbf{sR})^2 + \frac{1}{r'_0}(\mathbf{s}'\mathbf{R})^2\right]. \quad (41.2)$$

В отличие от постоянного вектора $\mathbf{s}$, единичный вектор $\mathbf{s}' = \mathbf{r}'_0/r'_0$ меняется при изменении положения точки наблюдения. Этим и обусловлено появление дифракционных картин, т.е. изменение интенсивности светового поля от точки к точке. Слагаемое $-k(r_0 + r'_0)$ не играет роли. Оно вносит в выражение для E_P только постоянный фазовый множитель $e^{-ik(r_0+r'_0)}$, не влияющий на интенсивность светового поля в точке P. Квадратичные члены имеют порядки

$$kR^2/r_0 = 2\pi R^2/(\lambda r_0) \quad \text{и} \quad kR^2/r'_0 = 2\pi R^2/(\lambda r'_0).$$

Максимальные значения эти величины принимают при $R = D$, где D — диаметр отверстия. Практически этими членами можно пренебречь, если они меньше $\pi/2$. Действительно, характерной величиной для фазы является $\pi/2$, так как при изменении углов на такую величину синусы и косинусы претерпевают существенные изменения. Все слагаемые в фазе, много меньшие $\pi/2$, не играют роли и могут быть отброшены без особого ущерба для точности. Для максимальных изменений, обусловленных квадратичными членами, мы (несколько условно) заменим требование «много меньше» более слабым требованием «меньше».

Таким путем получим

$$r_0 > \frac{4D^2}{\lambda}, \quad r_0' > \frac{4D^2}{\lambda}. \tag{41.3}$$

Этими условиями и определяется область фраунгоферовой дифракции. Например, если $D = 1$ мм, $\lambda = 500$ нм, то фраунгоферова дифракция наступает при r_0, $r_0' > 4D^2/\lambda = 8$ м.

Легко также написать условия, при которых разложение фазы $\Phi(\mathbf{R})$ можно оборвать на членах второй степени:

$$r_0 > \sqrt{4D^3/\lambda}, \quad r_0' > \sqrt{4D^3/\lambda}. \tag{41.4}$$

В приведенном примере они удовлетворяются при r_0, $r_0' > 9$ см. Впрочем, квадратичным приближением (41.2) пользуются даже тогда, когда условия (41.4) не выполняются. Это делается при вычислении интеграла (41.1) в несущественных частях области интегрирования, где точное знание фазы Φ не имеет значения. Пример такого рода будет приведен в следующем параграфе при рассмотрении дифракции на прямолинейном крае экрана.

Теперь становится качественно ясным, как ведет себя световой пучок, прошедший через диафрагму, на различных расстояниях r от нее. Для определенности будем считать, что диафрагма освещается параллельным пучком света. На малых расстояниях $r \ll D^2/\lambda$ приближенно применима геометрическая оптика (прямолинейное распространение света). Затем с увеличением расстояния r наблюдается сложная картина френелевой дифракции. На больших расстояниях $r \gg D^2/\lambda$ дифракционная картина упрощается и переходит в дифракцию Фраунгофера.

4. Следствием гипотезы Френеля является полезная *теорема Бабине* (1794–1872) о свойствах так называемых *дополнительных экранов*. Возьмем плоский экран, освещаемый каким-либо источником света. Пусть $E_{\text{вх}}$ — поле падающей волны, какое получилось бы в точке (x, y) на передней поверхности экрана при отсутствии последнего, а $E_{\text{вых}}$ — поле в той же точке на задней поверхности экрана (экран предполагается тонким, и его толщина не играет роли). Пусть эти поля связаны соотношением $E_{\text{вых}} = \alpha_1 E_{\text{вх}}$. Коэффициент α_1 называется *пропускаемостью экрана*. Он может зависеть от координат x, y, структуры экрана и длины волны, но не зависит от напряженности волнового поля. Для другого экрана той же геометрической формы $E_{\text{вых}} = \alpha_2 E_{\text{вх}}$. Эти два экрана называются *дополнительными*, если $\alpha_1 + \alpha_2 = 1$. Примером могут служить два непрозрачных экрана с отверстиями, причем отверстия одного экрана совпадают с непрозрачными частями другого. Согласно (41.1), волновые поля в точке наблюдения при использовании дополнительных

экранов могут быть записаны в виде

$$E_P^{(1)} = \int \frac{\alpha_1}{rr'} E_{\text{вх}} e^{i\Phi}\, dF, \quad E_P^{(2)} = \int \frac{\alpha_2}{rr'} E_{\text{вх}} e^{i\Phi}\, dF,$$

причем в обоих случаях интегрирование распространяется по всей поверхности экрана. (Для определенности источник света предполагается точечным, что не имеет никакого значения.) Сложив эти два выражения и учтя, что $\alpha_1 + \alpha_2 = 1$, получим

$$E_P^{(1)} + E_P^{(2)} = \int \frac{1}{rr'} E_{\text{вх}} e^{i\Phi}\, dF.$$

Но последний интеграл есть поле E_P в точке P, которое получилось бы при свободном распространении волны. Таким образом,

$$E_P^{(1)} + E_P^{(2)} = E_P. \tag{41.5}$$

Эта формула и выражает теорему Бабине в ее наиболее общей форме.

Обычно теорема Бабине понимается в более узком смысле — применительно к случаю фраунгоферовой дифракции. Пусть фраунгоферова дифракционная картина наблюдается в фокальной плоскости линзы. Если бы на пути параллельных лучей не было препятствий, то световое поле в этой плоскости было бы всюду равно нулю, за исключением фокуса линзы. Таким образом, согласно (41.5), во всякой точке фокальной плоскости, за исключением фокуса, должно быть $E_P^{(1)} + E_P^{(2)} = 0$. Так как интенсивность света I пропорциональна $|E_P|^2$, то отсюда получаем

$$I_1 = I_2. \tag{41.6}$$

Следовательно, поскольку наблюдению доступна только интенсивность светового поля, а не его фаза, *фраунгоферовы дифракционные картины от дополнительных экранов, получаемые в фокальной плоскости линзы, всюду одинаковы, за исключением самого фокуса.*

ЗАДАЧА

На черный экран падает плоская световая волна. Из-за дифракции за экраном, наряду с неотклоненной волной, появятся волны всевозможных направлений (рассеянный свет). Показать, что количество рассеянной энергии равно количеству энергии, поглощенному экраном.

Р е ш е н и е. Заменим экран дополнительным, т.е. отверстием той же величины и формы. От этого по теореме Бабине интенсивность светового поля в бесконечности сохранится неизменной во всех направлениях, за исключением направления первичной волны. Но на любое строго фиксированное направление за отверстием приходится нулевая интенсивность света, так как отверстие рассеивает весь падающий на него свет. С другой стороны, экран по предположению полностью поглощает весь падающий

свет. Отсюда непосредственно получается требуемый результат. Разумеется, он применим только в тех случаях, когда размеры экрана очень велики по сравнению с длиной волны.

§ 42. Зоны Шустера и спираль Корню

1. В одномерных задачах, например при рассмотрении дифракции на прямоугольной щели, разбиение волнового фронта на кольцевые зоны нецелесообразно. Лучше разбивать волновой фронт на *полосатые зоны*, называемые *зонами Шустера* (1851–1934). Ограничимся случаем, когда волновой фронт плоский, хотя обобщение на случай сферического фронта и не встречает никаких затруднений. Пусть плоскость волнового фронта AB перпендикулярна к плоскости рис. 165. Обозначим через b длину перпендикуляра PO, опущенного из точки наблюдения на волновой фронт. Проведем цилиндрические коаксиальные поверхности, ось которых проходит через точку P перпендикулярно к плоскости рисунка, а радиусы равны b, $b+\lambda/2$, $b+2(\lambda/2)$, ... Тогда волновой фронт разобьется на прямоугольные полосы, которые и называются зонами Шустера. Центральную зону условимся считать за две зоны: одна расположена справа, а другая слева от точки O. Тогда $r_n^2 = b^2 + x_n^2$, $r_{n-1}^2 = b^2 + x_{n-1}^2$, а потому $r_n^2 - r_{n-1}^2 = x_n^2 - x_{n-1}^2$. Приближенно

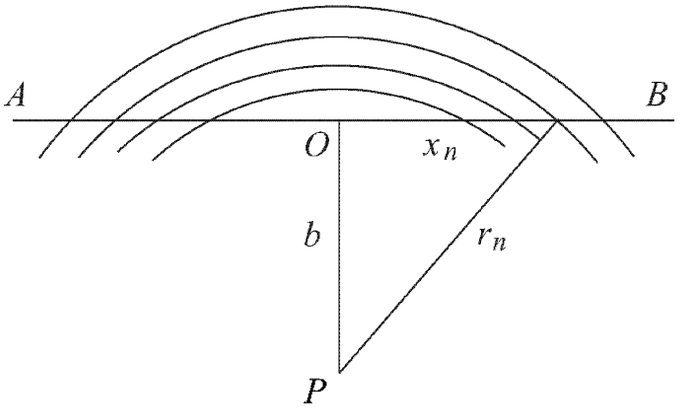


Рис. 165

$$r_n^2 - r_{n-1}^2 = (r_n + r_{n-1})(r_n - r_{n-1}) = 2b(\lambda/2) = b\lambda.$$

Таким образом, получаем рекуррентное соотношение

$$x_n^2 - x_{n-1}^2 = b\lambda, \tag{42.1}$$

из которого могут быть найдены все x_n. Так как $x_0 = 0$, то

$$x_1 = \sqrt{b\lambda}, \quad x_2 = \sqrt{2b\lambda}, \quad \ldots, \quad x_n = \sqrt{nb\lambda}. \tag{42.2}$$

Ширины последовательных зон Шустера будут

$$\sqrt{b\lambda}, \quad (\sqrt{2}-1)\sqrt{b\lambda}, \quad (\sqrt{3}-\sqrt{2})\sqrt{b\lambda}, \ldots \tag{42.3}$$

Они монотонно убывают и в пределе, когда $r \to \infty$, стремятся к $\lambda/2$, как это ясно из их построения. (Впрочем, высшие зоны не

играют роли. Имеют значение только несколько десятков первых зон Шустера.)

Как и в случае зон Френеля, применим теперь графический метод (см. рис. 153). Каждую зону Шустера разобьем на узкие полоски и будем изображать колебание в точке P, вносимое отдельной полоской, вектором на векторной диаграмме. Затем перейдем к пределу, устремляя к нулю ширину каждой полоски. В результате получится плавная кривая, называемая *спиралью Корню* (1841–1902) (рис. 166). Она состоит из двух симметричных ветвей, бесконечное число раз обвивающихся во-

X

круг «фокусов» F и F' и неограниченно приближающихся к ним. Верхняя ветвь представляет действие правой половины волнового фронта, нижняя — левой. Отличие каждой из ветвей от соответствующей спирали на рис. 153 обусловлено более быстрым убыванием начальных зон Шустера, чем зон Френеля. Колебание, возбуждаемое первой правой зоной Шустера, изображается вектором $\overrightarrow{OA}$, второй правой — вектором $\overrightarrow{A2}$, двумя первыми правыми зонами вместе — вектором $\overrightarrow{O2}$ и т.д. (все эти векторы на рис. 166 не проведены). Колебание, возбуждаемое всем волновым фронтом, представляется вектором $\overrightarrow{F'F}$, соединяющим фокусы спирали Корню. По мере приближения к фокусам амплитуды колебаний становятся все меньше и меньше и в пределе обращаются в нуль.

2. При нахождении уравнения спирали Корню надо учесть, что реально всегда приходится иметь дело не с бесконечными, а

с ограниченными волновыми фронтами, причем заметная интенсивность наблюдается лишь при малых углах дифракции. Поэтому в формуле (41.1) изменения знаменателей r и r' (а также уже отброшенных ранее ослабляющих множителей $K(\alpha)$) можно не принимать во внимание. Если нас интересует только относительное распределение интенсивности, то можно положить $rr' = 1$. В плоскости волнового фронта фазу можно представить в виде $\Phi = \omega t - kr$ (здесь произведено переобозначение: в формуле (41.1) расстояние r обозначалось через r').

Примем волновой фронт за координатную плоскость XY, а начало координат поместим в точке O. Тогда $r^2 = b^2 + (x^2 + y^2)$, а следовательно, $r - b = (x^2 + y^2)/(2b) + \ldots$ Члены высших степеней можно отбросить, если даже они добавляют в фазу слагаемые порядка π и больше. Дело в том, что такие члены, как это видно из формы спирали Корню, не меняя общего характера дифракционной картины, производят в ней только практически незаметные смещения высших дифракционных максимумов и минимумов. Кроме того, высшие дифракционные максимумы и минимумы следуют друг за другом столь часто, что для их реального осуществления требуются точечные источники света высокой степени монохроматичности. В противном случае все дифракционные полосы высших порядков размываются и переходят в равномерно освещенный фон. Отбросим все фазовые множители, не влияющие на относительное распределение интенсивности светового поля. Тогда поле в точке наблюдения P представится интегралом

$$E_P = \iint e^{-ik(x^2+y^2)/(2b)}\, dx\, dy.$$

Интегрирование должно быть выполнено по всей открытой поверхности волнового фронта. Допустим, что в направлении оси Y она простирается достаточно далеко в обе стороны. Тогда интегрирование по y можно выполнить в пределах от $-\infty$ до $+\infty$, в результате чего появится постоянный множитель, не представляющий интереса. Интегрирование по x произведем от нуля, считая верхний предел x переменным (он может быть и положительным, и отрицательным). Вместо x, как это принято, введем новую переменную s по формуле $kx^2/b = \pi s^2$. Тогда

$$E_P = \int_0^s e^{-i\pi s^2/2}\, ds, \tag{42.4}$$

$$E_P^* = \int_0^s e^{i\pi s^2/2}\, ds. \tag{42.5}$$

При изображении колебаний можно пользоваться как выражением (42.4), так и комплексно сопряженным с ним (42.5). При построении спирали Корню обычно применяют выражение (42.5). Оно и представляет уравнение спирали Корню в комплексной форме. Если координатные оси выбраны так, как указано на рис. 166, то в прямоугольных координатах уравнение спирали Корню запишется в виде

$$X(s) = \int_0^s \cos\frac{\pi s^2}{2}\,ds, \quad Y(s) = \int_0^s \sin\frac{\pi s^2}{2}\,ds. \tag{42.6}$$

Входящие сюда интегралы называются *интегралами Френеля*. Очевидно,

$$X(s) = -X(-s), \quad Y(s) = -Y(-s),$$

т.е. кривая (42.6) симметрична относительно начала координат.

Полагая $s = \infty$, находим координаты фокусов спирали Корню:

$$X_F = Y_F = 1/2, \quad X_{F'} = Y_{F'} = -1/2.$$

Впрочем, для многих целей проще пользоваться непосредственно комплексной формой (42.5). В частности, для дифференциала дуги спирали Корню из (42.5) находим: $|e^{i\pi s^2/2}\,ds| = |ds|$. Отсюда следует, что параметр s есть длина дуги спирали, отсчитываемая от начала координат O.

Если τ — угол между касательной к спирали Корню и осью X, то $\operatorname{tg}\tau = dY/dX = \operatorname{tg}(\pi s^2/2)$, а потому

$$\tau = \pi s^2/2. \tag{42.7}$$

При $s = 0$ угол $\tau = 0$, т.е. в начале координат кривая касается оси X. При $s = 1$ касательная вертикальна и идет вверх. При $s = \sqrt{2}$, $\tau = \pi$ касательная снова горизонтальна, но идет в отрицательном направлении оси X. При $s = \sqrt{3}$, $\tau = {}^3\!/_2\pi$ она вертикальна и идет вниз. При $s = 2$, $\tau = 2\pi$ касательная принимает исходное — горизонтальное — направление. Формула (42.7) позволяет наглядно проследить, как кривая обвивается вокруг фокусов F и F', делая при этом бесконечное число оборотов. Эта формула особенно полезна в том отношении, что она позволяет по заданному параметру s легко находить соответствующую точку на спирали Корню.

Из формулы (42.7) получаем формулу для кривизны спирали Корню:

$$\frac{1}{R} = \frac{d\tau}{ds} = \pi s. \tag{42.8}$$

Длина всей спирали Корню бесконечна, а потому при приближении к фокусам ее кривизна стремится к бесконечности.

3. При работе со спиралью Корню надо знать значение параметра s. Его легко найти, зная на экране расстояние x точки наблюдения от центра картины O (см. рис. 165). Вычислив ширину первой зоны Шустера $\sqrt{\lambda b}$, находим далее $s = x\sqrt{2/(\lambda b)}$.

Рассмотрим в качестве примера дифракционную картину от прямолинейного края экрана (рис. 167). Где бы ни находилась точка наблюдения P, для нее всегда будет открыт правый край волнового фронта. На векторной диаграмме (см. рис. 166) колебание в точке наблюдения представится вектором $\overrightarrow{M_nF}$, конечная точка которого всегда находится в верхнем фокусе F, а начальная M_n лежит где-то на спирали Корню. Если, сохраняя неизменным положение конечной точки F, перемещать начальную точку M_n вдоль спирали Корню (положения M_1, M_2, M_3, ...), то таким путем можно получить распределение амплитуд и интенсивности колебаний света по всему экрану.

Обозначим через $a_0 = |FF'|$ и $I_0 = a_0^2$ амплитуду и интенсивность волны, когда открыт весь волновой фронт. Когда точка наблюдения P находится на границе геометрической тени, то колебание представится вектором $\overrightarrow{OF} = {}^1\!/_2\overrightarrow{F'F}$. Ему соответствует амплитуда $a_0/2$ и интенсивность $I_0/4$. При перемещении точки P в освещенную область экрана изображающая точка M_n начнет перемещаться по нижней ветви спирали Корню, а амплитуда и интенсивность колебаний будут последовательно проходить через максимумы и минимумы. Максимальная амплитуда, как видно из рис. 166, составляет $1{,}12a_0$, а интенсивность $1{,}25I_0$; минимальные значения их соответственно $0{,}89a_0$ и $0{,}78I_0$. При дальнейшем продвижении в освещенную область интенсивность асимптотически приближается к I_0. При погружении точки P в область геометрической тени изображающая точка M_n перемещается по верхней ветви спирали Корню. При этом по мере погружения в указанную область интенсивность света монотонно убывает и асимптотически стремится к нулю.

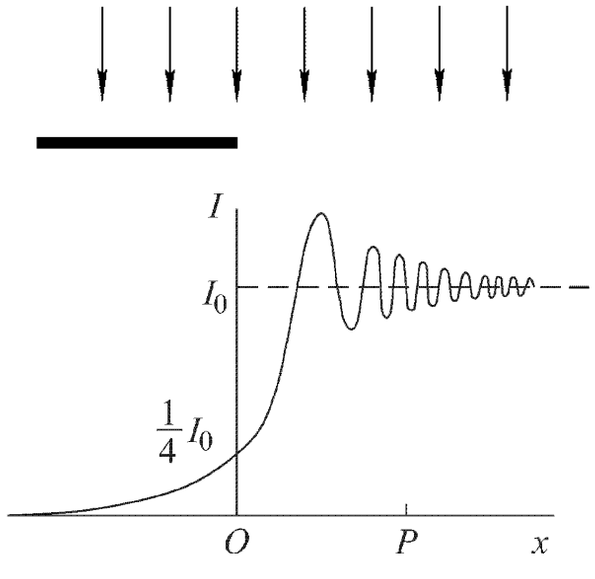


Рис. 167

Распределение интенсивности графически представлено на рис. 167. Таким образом, нет резкой границы между светом и тенью: в области геометрической тени интенсивность света убывает непрерывно

и монотонно, а освещенная область расщепляется в дифракционные полосы. На рис. 168 показана дифракционная картина, наблюдаемая при дифракции света на крае экрана. Таким же путем можно рассчитать дифракционную картину на щели или длинном прямоугольном экране. На рис. 169 показана тень проволоки от точечного (или линейного) источника.

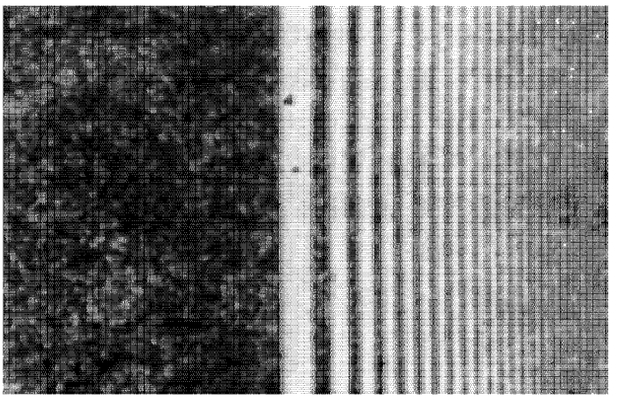


Рис. 168

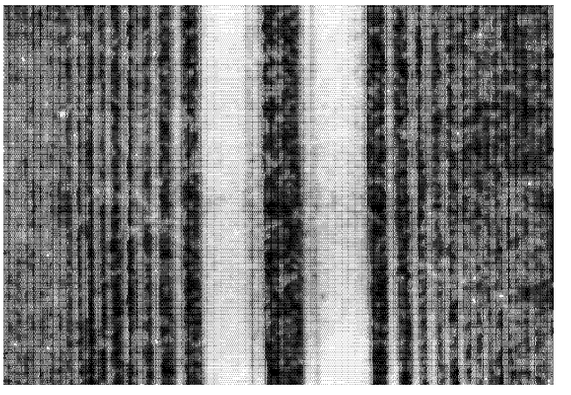


Рис. 169

ЗАДАЧИ

1. Положение светлых и темных полос при дифракции плоской волны на крае экрана можно приближенно (но с достаточной точностью) определить по точкам пересечения нижней ветви спирали Корню с прямой $F'F$, соединяющей ее фокусы, приняв во внимание, что в этих точках прямая FF' практически перпендикулярна к спирали Корню. Найти таким путем координаты указанных полос.

О т в е т.

$$x_n = {}^1\!/_2 \sqrt{\lambda r(8n-5)} \quad \text{(светлые полосы)},$$
$$x_n = {}^1\!/_2 \sqrt{\lambda r(8n-1)} \quad \text{(темные полосы)},$$

где $n = 1, 2, 3, \ldots$

2. Когда звезда проходит мимо края Луны, получаются дифракционные полосы. Определить скорость их движения v по земной поверхности и оценить порядок их ширины Δx. Для наблюдения полос можно воспользоваться телескопом, в фокусе которого помещен фотоэлемент. Полосы, проходя перед объективом телескопа, возбуждают переменные электрические токи, которые могут быть усилены и с помощью осциллографа записаны на движущейся ленте. Оценить длительность t прохождения полос перед объективом.

Принимая звезду за равномерно светящийся диск, можно теоретически рассчитать распределение освещенности в дифракционной картине. Сравнивая его с распределением освещенности, найденным экспериментально, можно определить угловой диаметр звезды, что и было фактически выполнено в 1946 г. Уайтфордом для четырех звезд на стодюймовом рефлекторе Маунт-Вильсоновской обсерватории. Оценить угловые размеры звезд, для которых может быть пригоден этот метод.

О т в е т. $v \approx 500$ м/с; $\Delta x \sim \sqrt{b\lambda} \approx 10$ м, где $b = 3{,}8 \cdot 10^5$ км — среднее расстояние до Луны. Метод пригоден, когда угловые размеры звезд лежат примерно в интервале от 10^{-4} до 10^{-3} угловой секунды.

3. Для получения фотографий дифракционных картин в тех случаях, когда источник света и экран расположены очень далеко, В.К. Аркадьевым (1884–1953) был применен *метод подобия*, в котором вместо действительных препятствий, стоящих на пути лучей, используются их уменьшенные и подобные модели, но длина волны сохраняется неизменной.

Требуется получить фотографию дифракционной картины от диска диаметром $D = 50$ см, когда на его оси расположен точечный источник света на расстоянии $A = 25$ км, а экран удален от него на $B = 50$ км (плоскость экрана перпендикулярна к оси диска). С этой целью диск заменили уменьшенной моделью с диаметром $d = 1$ см. Пользуясь методом зон Френеля, определить, на каких расстояниях a и b следует поместить источник света и экран, чтобы получилась подобная и уменьшенная в $n = 50$ раз дифракционная картина.

О т в е т. $a = A/n^2 = 10$ м, $b = B/n^2 = 20$ м.

§ 43. Принцип Гюйгенса в формулировке Кирхгофа

1. До Кирхгофа принцип Гюйгенса–Френеля оставался гипотезой. Кирхгоф в 1883 г. вывел формулу, которую можно рассматривать как уточненную формулировку указанного принципа. Приведем вывод формулы Кирхгофа, хотя в дальнейшем и не будем ею пользоваться. Читатель может опустить его без ущерба для понимания дальнейшего.

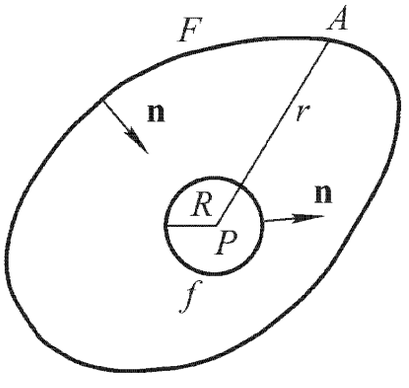


Рис. 170

Допустим, что среда, в которой распространяется свет, однородна. Будем характеризовать световое поле какой-то величиной E. Под E можно понимать либо вектор $\mathbf{E}$, либо вектор $\mathbf{B}$, либо одну из их проекций на декартовы оси координат. Эта величина удовлетворяет волновому уравнению

$$\Delta E - \frac{1}{v^2}\frac{\partial^2 E}{\partial t^2} = 0,$$

которое в случае монохроматического поля переходит в

$$\Delta E + k^2 E = 0. \tag{43.1}$$

Найдем значение E в произвольной точке пространства P (рис. 170). Обозначим через r переменное расстояние какой-либо точки A от P. Величина

$$\chi = \frac{1}{r} e^{-\imath k r}, \tag{43.2}$$

рассматриваемая как функция точки A, также удовлетворяет уравнению

$$\Delta\chi + k^2\chi = 0. \tag{43.3}$$

Окружим точку P произвольной замкнутой поверхностью F, и притом такой, что в окружаемом ею пространстве нет источников света. Функция χ обращается в бесконечность в точке P. Исключим эту точку, окружив ее сферой f достаточно малого радиуса R с центром в P. Тогда во всем

пространстве между сферой f и поверхностью F функции E и χ, а также их производные будут конечны и непрерывны. К ним можно применить формулу Грина (1793–1841)

$$\int (E\Delta\chi - \chi\Delta E)\,dV = -\int\limits_{F+f} \left(E\frac{\partial\chi}{\partial n} - \chi\frac{\partial E}{\partial n}\right) dF, \tag{43.4}$$

где V — объем пространства между поверхностями f и F, а $\mathbf{n}$ — внутренняя нормаль по отношению к этому пространству. Так как в указанном пространстве источников света нет, то в нем справедливы уравнения (43.1) и (43.3), Следовательно,

$$E\,\Delta\chi - \chi\,\Delta E = -k^2(E\chi - \chi E) = 0,$$

а потому

$$\oint\limits_f \left(E\frac{\partial\chi}{\partial n} - \chi\frac{\partial E}{\partial n}\right) df = -\oint\limits_F \left(E\frac{\partial\chi}{\partial n} - \chi\frac{\partial E}{\partial n}\right) dF.$$

Перейдем в этом равенстве к пределу, устремляя радиус сферы f к нулю. Правая часть равенства при этом не будет меняться. Что касается левой, то, взяв радиус R настолько малым, чтобы $kR \ll 1$, можно заменить экспоненциальный множитель e^{-ikR} единицей. Тогда левая часть примет вид

$$\oint\limits_f \left[E\frac{\partial}{\partial R}\left(\frac{1}{R}\right) - \frac{1}{R}\frac{\partial E}{\partial R}\right] df.$$

Так как величины E и $\dfrac{\partial E}{\partial R}$ в окрестности точки P конечны, то интеграл от второго слагаемого будет порядка $-4\pi R\dfrac{\partial E}{\partial R}$, т.е. при $R \to 0$ он обратится в нуль. Интеграл же от первого слагаемого в пределе перейдет в $-4\pi E_P$. Окончательно имеем

$$E_P = \frac{1}{4\pi}\oint\limits_F \left[E\frac{\partial}{\partial n}\left(\frac{e^{-ikr}}{r}\right) - \frac{e^{-ikr}}{r}\frac{\partial E}{\partial n}\right] dF, \tag{43.5}$$

где E_P — значение функции E в точке P.

2. Формула (43.5) по виду напоминает принцип Гюйгенса–Френеля. И тут и там поле в точке P выражается интегралом по замкнутой поверхности F. Однако у Френеля источники света лежат *внутри* замкнутой поверхности F, а точка P *вне* этой поверхности. Формула же (43.5), наоборот, предполагает, что точка P лежит *внутри* поверхности F, а источники *вне* ее. Легко, однако, преобразовать формулу (43.5), чтобы указанное различие исчезло. Для этого предположим, что все источники света S_1, S_2, S_3, ... лежат в конечной области пространства. Окружим эту область замкнутой поверхностью F (рис. 171). Пусть точка P находится в пространстве вне поверхности F. Опишем из P как из центра сферу f настолько большого радиуса, чтобы она целиком окружала поверхность F. Тогда в пространстве

между f и F не будет источников света, а потому можно для вычисления E в точке P применить формулу (43.5):

$$E_P = \frac{1}{4\pi} \int\limits_{F+f} \left(E \frac{\partial \chi}{\partial n} - \chi \frac{\partial E}{\partial n} \right) dF.$$

Докажем, что интеграл по сфере f стремится к нулю, когда ее радиус стремится к бесконечности. Для этого необходимо выяснить поведение функции E на бесконечности. Предположим, что в пространстве, ограниченном F, находится один или несколько точечных источников света: S_1, S_2, S_3, ... Тогда поле этих источников представится в виде

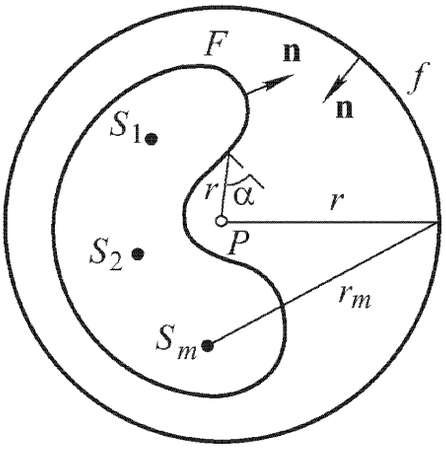


Рис. 171

$$E = \sum C_m \frac{e^{-\imath kr} m}{r_m},$$

где C_m — постоянные коэффициенты. Если r стремится к ∞, то r_m также стремится к ∞, однако разность $r_m - r$ будет оставаться конечной. Представим E в виде

$$E = \sum C_m \frac{e^{-\imath k(r+a_m)}}{r+a_m} = \frac{e^{-\imath kr}}{r} \sum \frac{A_m}{1+a_m/r},$$

где A_m — новые постоянные. Разлагая выражение под знаком суммы в ряд по степеням $1/r$, получим

$$E = \frac{e^{-\imath kr}}{r} \left\{ \sum A_m - \frac{1}{r} \sum A_m a_m + \frac{1}{r^2}(\ldots) + \ldots \right\},$$

или

$$E = \frac{e^{-\imath kr}}{r}(C + \Phi) = (C + \Phi)\chi,$$

где C — постоянная, а Φ стремится к нулю по крайней мере как $1/r$. Подставляя это значение E в интеграл по сфере f, получим

$$\oint\limits_f \left(E \frac{\partial \chi}{\partial n} - \chi \frac{\partial E}{\partial n} \right) df = - \oint\limits_f \chi^2 \frac{\partial \Phi}{\partial n} df.$$

Подынтегральное выражение в последнем интеграле стремится к нулю по крайней мере как $1/r^3$, тогда как поверхность сферы обращается в бесконечность как r^2. Поэтому при $r \to \infty$ весь интеграл стремится к нулю. Таким образом, если сферу f удалить в бесконечность, то получится

$$E_P = \frac{1}{4\pi} \oint\limits_F \left[E \frac{\partial}{\partial n} \left(\frac{e^{-\imath kr}}{r} \right) - \frac{e^{-\imath kr}}{r} \frac{\partial E}{\partial n} \right] dF. \tag{43.6}$$

Эти рассуждения приводят также к следующему важному результату. Если некоторый участок поверхности F удаляется в бесконечность, то часть

интеграла (43.6) по этому участку стремится к нулю. При этом предполагается, что все источники света находятся в *конечной области пространства.*

3. Формулы (43.5) и (43.6) и выражают *принцип Гюйгенса в формулировке Кирхгофа.* В обеих формулах $\mathbf{n}$ означает внутреннюю нормаль по отношению к тому пространству, в котором находится точка наблюдения P.

Выполнив дифференцирование по $\mathbf{n}$ и приняв во внимание, что $\partial r/\partial n = -\cos\alpha$, где α — угол между нормалью $\mathbf{n}$ и направлением из площадки dF на точку P, получим

$$E_P = \oint K(\alpha, r)\frac{e^{-ikr}}{r}\,dF. \tag{43.7}$$

Здесь введено обозначение

$$K(\alpha, r) = \frac{1}{4\pi}\left[\left(ik + \frac{1}{r}\right)E\cos\alpha - \frac{\partial E}{\partial n}\right]. \tag{43.8}$$

Тем самым установлена связь формулы Кирхгофа с принципом Гюйгенса: подынтегральное выражение в формуле (43.8) может рассматриваться как *вторичная волна*, распространяющаяся от площадки dF к точке P. Множитель K, однако, зависит *не только от угла* α, как предполагал Френель, но также и *от расстояния* r. В противном случае вторичная волна не могла бы удовлетворять волновому уравнению. Таким образом, вторичные волны не обладают шаровой симметрией. Они сферические только в том смысле, что их волновые фронты имеют форму сфер. Амплитуды же зависят от направления распространения и меняются с расстоянием иначе, чем $1/r$. Только в «волновой зоне», когда расстояние точки P от излучающего центра dF очень велико по сравнению с длиной волны, можно в выражении (43.8) пренебречь $1/r$ по сравнению с ik. Тогда

$$E_P = \frac{1}{4\pi}\oint \left(ikE\cos\alpha - \frac{\partial E}{\partial n}\right)\frac{e^{-ikr}}{r}\,dF. \tag{43.9}$$

Благодаря малости длин световых волн такой упрощенной формой принципа Гюйгенса в оптике можно пользоваться при решении всех конкретных задач.

4. Чтобы составить на примере более конкретное представление о вторичных волнах, рассмотрим свободное распространение сферической волны от точечного источника. В качестве поверхности F возьмем сферу радиуса r_0 с центром в источнике O (рис. 172). Поле на поверхности F представим выражением

$$E_0 = \frac{e^{i(\omega t - kr_0)}}{r_0}.$$

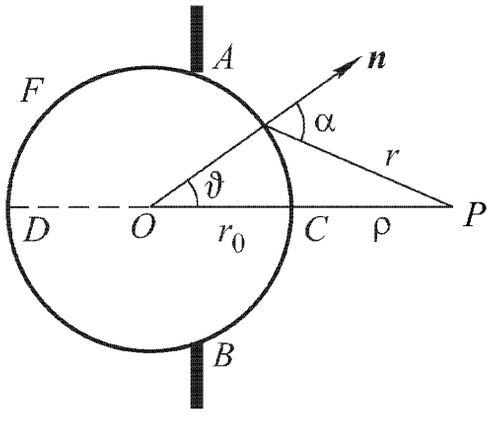


Рис. 172

Предполагая, что радиус r_0 очень велик по сравнению с длиной волны, отсюда найдем $\partial E/\partial n = \partial E_0/\partial r_0 = -ikE_0$. Подставляя это значение в формулу (43.9), получим

$$E_P = \frac{ik}{4\pi}\oint (1 + \cos\alpha)E_0\frac{e^{-ikr}}{r}\,dF. \tag{43.10}$$

Сравнение этой формулы с формулой (43.7) дает

$$K(\alpha) = \frac{ik}{4\pi}(1 + \cos\alpha). \tag{43.11}$$

Это и есть «ослабляющий множитель» $K(\alpha)$, введенный в § 39 ad hoc. Из теории автоматически получается, что он чисто мнимый и с возрастанием α монотонно убывает по абсолютной величине. Он обращается в нуль при $\alpha = \pi$, т.е. в точке D сферы, диаметрально противоположной точке наблюдения P.

Интеграл (43.10) теперь можно вычислить, поскольку он не содержит никаких неизвестных функций. Произведем это вычисление, так как таким путем можно получить строгое обоснование метода зон Френеля и результатов, полученных этим методом. Взяв за переменную интегрирования r и использовав значение E_0, преобразуем интеграл (43.10) к виду

$$E_P = \int_{r_1}^{r_2} A(r)e^{-\imath kr}\,dr, \tag{43.12}$$

где введено обозначение

$$A(r) = \frac{ik(1 + \cos\alpha)}{2(\rho + r_0)}e^{\imath(\omega t - kr_0)}, \tag{43.13}$$

а через r_1 и r_2 обозначены наибольшее и наименьшее значения, принимаемые r. Интегрируя (43.12) по частям, получим

$$E_P = -\frac{A(r)e^{-\imath kr}}{ik}\Bigg|_{r_1}^{r_2} + \frac{1}{ik}\int_{r_1}^{r_2}\frac{dA}{dr}e^{-\imath kr}\,dr. \tag{43.14}$$

Здесь единственной величиной, зависящей от r, является $\cos\alpha$. Найдем ее производную. Из рис. 172 следует, что $(\rho + r_0)^2 = r^2 + r_0^2 + 2rr_0\cos\alpha$, откуда $\dfrac{d\cos\alpha}{dr} = -\left(\dfrac{1}{r_0} + \dfrac{\cos\alpha}{r}\right)$. Взяв от (43.13) логарифмическую производную по r и воспользовавшись предыдущей формулой, найдем

$$\frac{1}{ik}\frac{dA}{dr} = \frac{i}{1 + \cos\alpha}\left(\frac{1}{kr_0} + \frac{\cos\alpha}{kr}\right)A(r).$$

Такой величиной, а следовательно, и интегралом в формуле (43.14) следует пренебречь, так как это уже было сделано при выводе исходной формулы (43.9). Таким образом, в формуле (43.14) остается только первое слагаемое, т.е.

$$E_P = \frac{A(r_1)e^{-\imath kr_1} - A(r_2)e^{-\imath kr_2}}{ik}. \tag{43.15}$$

Существенно, что величина E_P представляется разностью одной и той же функции, но при различных значениях аргумента r, причем при изменении r на $\lambda/2$ знак этой функции меняется на противоположный. Этого достаточно для обоснования основного результата (39.6), на котором основан метод зон Френеля. Действительно, применив формулу (43.15) к первой зоне Френеля, найдем, что действие этой зоны может быть представлено в виде $E_1 = (u_1 + u_2)$, действие второй зоны — в виде $E_2 = -(u_2 + u_3)$, и т.д. Явный

вид выражений u_i для доказательства не имеет значения. Действие первых N зон выразится суммой

$$E = (u_1 + u_2) - (u_2 + u_3) + \ldots + (-1)^{N+1}(u_N + u_{N+1}),$$

т.е.

$$E = u_1 + (-1)^{N+1} u_{N+1}.$$

Так как по величине действия двух соседних зон почти одинаковы, то $u_{N+1} = u_N$. С той же степенью точности $^1/_2 E_1 = u_1$, $^1/_2 E_N = (-1)^{N+1} u_N = (-1)^{N+1} u_{N+1}$. Следовательно,

$$E = \frac{1}{2}(E_1 + E_N).$$

§ 44. Дифракция Фраунгофера на щели

1. Дифракционные явления Фраунгофера имеют в оптике значительно большее практическое значение, чем дифракционные явления Френеля. При практическом осуществлении дифракции Фраунгофера источник света S помещается в фокусе линзы (рис. 173). (Линза L_1 не нужна, если источником света служит лазер, поскольку от него исходит уже параллельный пучок света.) Дифракция возникает на каком-либо препятствии AB, поставленном на пути световых лучей, прошедших через линзу L_1. Дифракционная картина наблюдается в фокальной плоскости другой линзы L_2. Ее можно также наблюдать в зрительную трубу, установленную на бесконечность. Но в теоретических рассуждениях удобнее от всех этих вспомогательных приспособлений отвлечься, предполагая, что на препятствие AB падает параллельный пучок лучей, а дифракция наблюдается «в бесконечности».

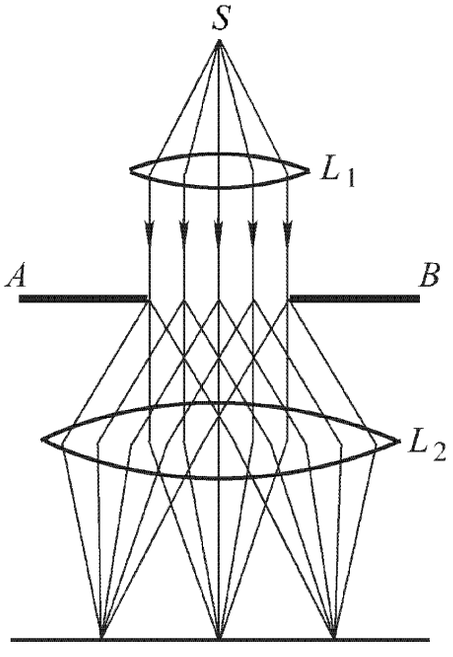


Рис. 173

Простейшим для расчета и практически очень важным случаем является фраунгоферова дифракция на длинной прямоугольной щели. Ширину щели обозначим через b, ее длину будем считать бесконечной. Пусть на щель нормально падает плоская монохроматическая волна (рис. 174). Световое поле за щелью найдется по принципу Гюйгенса как результат интерференции когерентных вторичных волн, исходящих из различных точек волнового

фронта на щели. Вторичные волны, излучаемые полоской волнового фронта ширины dx, параллельной щели, складываясь, дают цилиндрическую волну, осью которой является эта полоска. Зависимость этой волны от направления ее распространения, определяемого углом ϑ, можно не учитывать, так как задача решается методом Френеля, а потому угол дифракции ϑ должен предполагаться малым. Однако необходимо учесть разности фаз между волнами, исходящими из различных полосок. Разумеется, речь идет о фазах колебаний на бесконечном расстоянии от щели. Волна, исходящая из dx под углом ϑ, опережает по фазе волну того же направления, исходящую из середины щели O, на $kx\sin\vartheta$. Поэтому результирующее поле в бесконечности, создаваемое всей щелью, представится интегралом

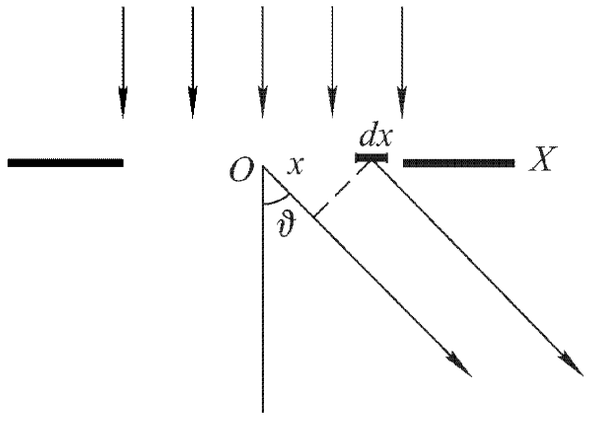


Рис. 174

$$E = \int\limits_{-b/2}^{+b/2} e^{ikx\sin\vartheta}\,dx.$$

Здесь опущены все множители, не влияющие на относительное распределение волнового поля по направлениям. Вычислив интеграл, получим

$$E = b\frac{\sin\alpha}{\alpha}, \tag{44.1}$$

где введено обозначение

$$\alpha = \frac{kb\sin\vartheta}{2} = \frac{\pi b\sin\vartheta}{\lambda}. \tag{44.2}$$

Отсюда для распределения интенсивности света по направлениям найдем

$$I = I_0\left(\frac{\sin\alpha}{\alpha}\right)^2, \tag{44.3}$$

где I_0 — интенсивность в направлении падающей волны. На рис. 175 представлены графики функций $\frac{\sin\alpha}{\alpha}$ (штриховая кривая) и $\left(\frac{\sin\alpha}{\alpha}\right)^2$ (сплошная кривая). Обе функции обращаются в максимум, равный единице, при $\alpha = 0$. При $\alpha = m\pi$, где $m = \pm1, \pm2, \ldots$, они равны нулю, т.е. в этих точках наблюдаются минимумы интенсивности. Между двумя соседними минимумами располагаются максимумы различных порядков. Их положе-

ния определяются трансцендентным уравнением $\alpha\cos\alpha-\sin\alpha=0$. Практически можно считать, что максимумы располагаются посередине между соседними минимумами.

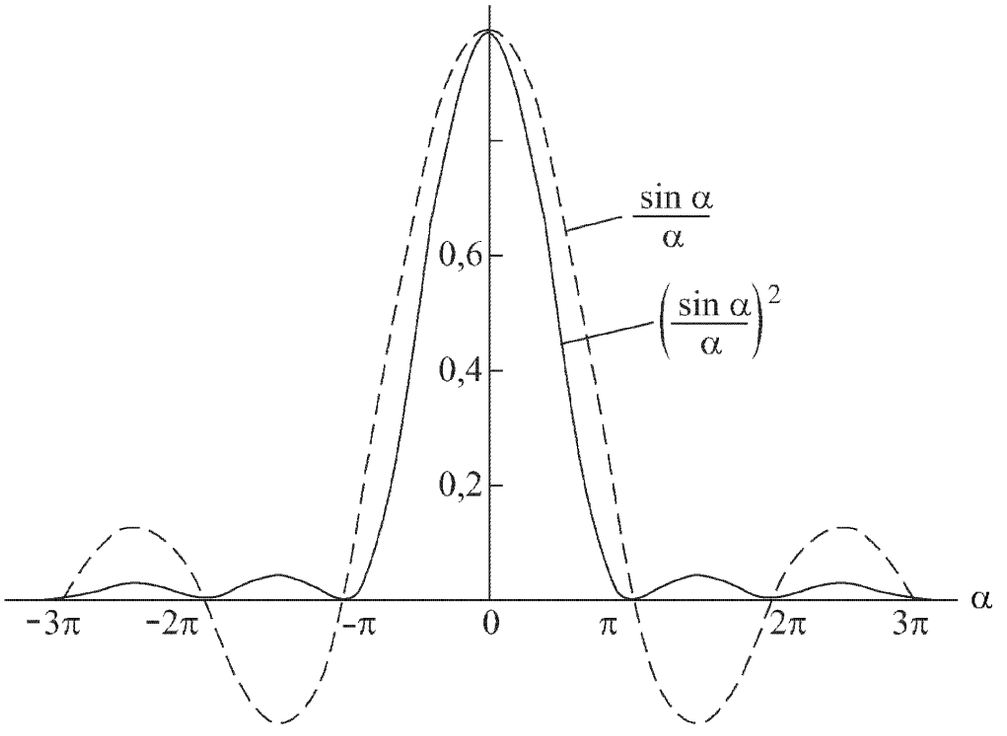


Рис. 175

2. Условие минимума $\alpha=m\pi$ можно также записать в виде

$$b\sin\vartheta=m\lambda. \tag{44.4}$$

Оно означает, что разность хода между волнами, исходящими от крайних точек щели, должна содержать целое число волн. Этот результат легко уяснить без всяких вычислений. Допустим сначала, что $b\sin\vartheta=\lambda$. Разобьем щель на две части одинаковой ширины. Тогда волны от этих частей придут в (удаленную) точку наблюдения в противоположных фазах и погасят друг друга при интерференции, — получится минимум интенсивности. Если $b\sin\vartheta=m\lambda$, то разбиение щели надо произвести на $2m$ частей одинаковой ширины. Волны от частей с нечетными номерами придут в фазах, противоположных фазам волн, пришедших от частей с четными номерами. А так как, ввиду симметрии, части щели, равноудаленные от противоположных краев ее, совершенно равноправны, то ясно, что эффект, вызванный нечетными частями, погасится при интерференции эффектом, вызванным четными частями.

Приведенное обоснование формулы (44.4) не только проще предыдущего, но, что особенно важно, и обладает большей общностью. Действительно, в предыдущем выводе щель должна предполагаться широкой ($b\gg\lambda$), чтобы можно было применять приближенный метод Френеля. Последний же вывод использует только свойства симметрии щели и применим не только при

малых, но и при любых углах дифракции ϑ. Требуется только, чтобы было $b > \lambda$. В противном случае из условия (44.4) мы получили бы $\sin\vartheta > 1$, и все дифракционные минимумы с нулевой интенсивностью были бы невозможны.

3. Когда падающая волна плоская, то все дифрагированные лучи перпендикулярны к оси щели. Если на их пути поставить линзу, то каждый дифрагированный пучок параллельных лучей соберется в маленькое пятнышко. Такие пятнышки — максимумы и минимумы интенсивности — расположатся вдоль прямой, перпендикулярной к оси щели и лежащей в фокальной плоскости линзы. В этой плоскости и надо поместить экран для наблюдения.

Если в качестве источника света взять светящуюся линию или узкую освещаемую коллиматорную щель, параллельную щели, на которой происходит дифракция, то каждая точка источника даст на экране дифракционную картину, описанную выше. В результате наложения таких картин каждое дифракционное пятнышко вытянется в полоску. Образуется система дифракционных полос, показанная на рис. 176. Центральная полоса светлая и примерно вдвое шире остальных светлых и тем-

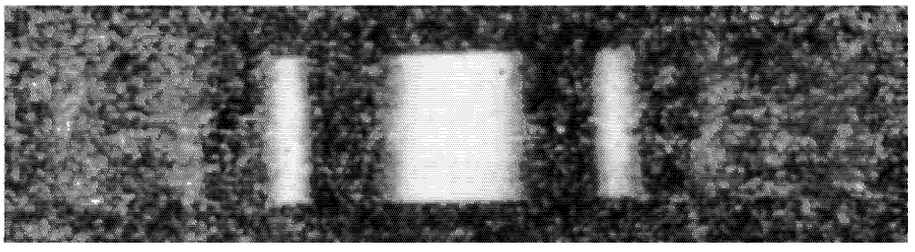


Рис. 176

ных полос. Максимальный порядок минимума, который может наблюдаться, определяется условием $\sin\vartheta < 1$, т.е. $m < b/\lambda$, как это видно из формулы (44.4). Чем шире щель, тем ярче картина, тем у́же дифракционные полосы, а число самих полос больше. При сужении щели картина расширяется, а ее яркость уменьшается. Когда $b = \lambda$, минимумы первого порядка получаются при $\vartheta = \pi/2$ и исчезают при дальнейшем сужении щели.

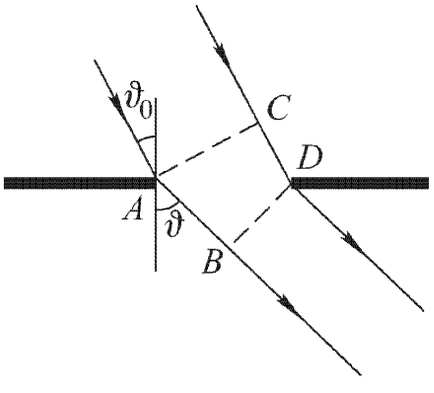


Рис. 177

4. При наклонном падении света под углом ϑ_0 (рис. 177) разность хода между крайними интерферирующими лучами $AB - CD$ составляет $b(\sin\vartheta - \sin\vartheta_0)$. Поэтому условие дифракционного минимума (44.4) переходит в

$$b\,(\sin\vartheta - \sin\vartheta_0) = m\lambda. \qquad (44.5)$$

Если углы ϑ и ϑ_0 малы, то $b(\vartheta - \vartheta_0) \approx m\lambda$. Основная доля света сосредоточена в центральной дифракционной полосе, т.е. между минимумами первого и минус первого порядков. Этим можно воспользоваться для оценки дифракционной расходимости световых пучков, выделяемых тем или иным способом, например в результате прохождения плоской волны через диафрагму. Световые лучи, прошедшие через диафрагму, отклоняются от своего исходного направления на угол

$$\delta\vartheta \sim \lambda/D, \tag{44.6}$$

где D — поперечное сечение пучка (в направлении, где оно минимально). Такое дифракционное уширение пучка обусловлено волновой природой света и принципиально не может быть устранено при заданной ширине пучка. Поэтому не существует строго параллельных пучков лучей. Это — идеализация, предполагающая, что поперечное сечение пучка бесконечно велико. Если пучок проходит путь l, то на этом пути он претерпевает дифракционное уширение $\sim l\,\delta\vartheta = l\lambda/D$. Таким уширением можно пренебрегать только тогда, когда оно мало по сравнению с шириной самого пучка, т.е. когда $l\lambda/D \ll D$, или $l \ll D^2/\lambda$. Только на таких расстояниях пучок может рассматриваться как луч геометрической оптики (см. § 41, п. 3).

5. Допустим, что источником света является бесконечно удаленная светящаяся точка или линия, параллельная длине щели. Угол ϑ на дифракционный минимум m-го порядка определяется условием (44.5). Ближайший к нему максимум сдвинут на угол $\delta\vartheta$, причем

$$b\,[\sin(\vartheta + \delta\vartheta) - \sin\vartheta_0] = m\lambda + \frac{\lambda}{2}.$$

Вычитая отсюда (44.5) и заменяя разности дифференциалами, получим $b\cos\vartheta \cdot \delta\vartheta = \lambda/2$. Пусть другой такой же (некогерентный) источник света сдвинут относительно первого как раз на угол $\delta\vartheta$. Тогда дифракционные максимумы от одного из этих источников наложатся на минимумы от другого. В результате дифракционные полосы пропадут. Угловое расстояние $\delta\vartheta$ между точечными (или линейными) источниками, когда это произойдет, определяется формулой

$$\delta\vartheta = \lambda/2b \tag{44.7}$$

(при условии, что угол ϑ мал).

Допустим теперь, что источник света протяженный и имеет форму равномерно светящейся полосы с угловой шириной $\delta\Theta$. Разобьем ее на две одинаковые части. Каждую часть в свою очередь разобьем на бесконечно узкие полоски одинаковой ширины. Скомбинируем их в пары полосок, сдвинутых относительно друг друга на угол $\delta\Theta/2$. Если $\delta\Theta/2 = \delta\vartheta$, то, очевидно, дифракционных полос не получится. Таким образом, условие исчезновения

дифракционных полос от протяженного источника рассматриваемой формы будет

$$\delta\Theta = \lambda/b. \tag{44.8}$$

От угловых размеров $\delta\Theta$ источника легко перейти к его линейным поперечным размерам δl. Если источник помещен в главном фокусе коллиматорной линзы с фокусным расстоянием f, то $\delta l = f\,\delta\Theta$, т.е.

$$\delta l = f\frac{\lambda}{b}. \tag{44.9}$$

6. Когда ширина щели становится меньше или порядка длины волны, приближенный метод Френеля, которым мы пользовались выше, становится неприменимым. Тогда волновое поле в плоскости щели уже нельзя отождествлять с неискаженным полем падающей волны, как это делается в методе Френеля. Задачу надо решать математически строго с использованием уравнений Максвелла и соответствующих им граничных условий.

Впервые такой метод был осуществлен в 1896 г. Зоммерфельдом (1868–1951) в задаче о дифракции плоской волны на прямолинейном крае экрана. Зоммерфельд рассмотрел идеально проводящий (а потому непрозрачный) экран, толщина которого пренебрежимо мала по сравнению с длиной волны. Хотя в оптике такой случай и невозможно осуществить, решение Зоммерфельда имеет большое значение, так как оно позволяет судить о точности и границах применимости приближенных методов. В 1897 г. Рэлей решил задачу о дифракции на узкой щели ($b \ll \lambda$) в бесконечно тонком идеально проводящем экране. В курсе общей физики нет возможности приводить эти решения [1]). Сравним только их результаты с тем, что дает простой метод Френеля, чтобы составить более конкретное представление о границах применимости этого метода.

Предположим, что электрический вектор падающей волны параллелен щели. Как показывает строгий расчет, ход амплитуды дифрагированной волны качественно сохраняется, но выражается через функции Бесселя. Амплитудная кривая с ростом ϑ спадает круче прежней функции $(\sin\alpha)/\alpha$. В максимуме значение амплитуды в $4\lambda/(b\pi^2)$ раз меньше значения, определяемого формулой (44.1). Так, при $b = {}^1\!/_{10}\lambda$, амплитуда уменьшается в четыре раза [2]). Расхождение с приближенной теорией уменьшается при дальнейшем увеличении b.

---

[1]) Интересующихся отсылаем к книге: *А. Зоммерфельд.* Оптика. — М.: ИЛ, 1953.

[2]) Если электрический вектор падающей волны перпендикулярен к длине щели, то амплитуда уменьшается слабее. Это легко понять, обратившись к соответствующему опыту Герца с проволочной решеткой (см. т. III, § 142, п. 7).

Численные расчеты, выполненные Морзе и Рубинштейном (1938), показали, что при ширине щели около λ или больше результаты, полученные по приближенному методу Френеля, могут считаться достаточно удовлетворительными. Таким образом, даже в случае наиболее тонких современных дифракционных решеток (при щелях порядка 1000–2000 нм) приближенный метод Френеля не ведет еще к заметным ошибкам.

§ 45. Дифракция Фраунгофера на отверстиях

1. Допустим, что свет падает перпендикулярно к плоскости непрозрачного экрана с отверстиями. (Обобщение на случай наклонного падения не встречает никаких затруднений.) Координатную плоскость XY совместим с плоскостью экрана, через dF обозначим элемент площади в этой плоскости. Направление дифрагированного света будем характеризовать единичным вектором $\mathbf{s}$ (рис. 178). Разность хода между лучами, вышедшими в этом направлении из элемента площади dF и из начала координат O, т.е. длина отрезка OA, равна $(\mathbf{rs})$, где $\mathbf{r}(x, y)$ — радиус-вектор элемента dF. Соответствующая разность фаз будет $k(\mathbf{rs})$. Результирующее поле в фраунгоферовой дифракционной картине представится интегралом

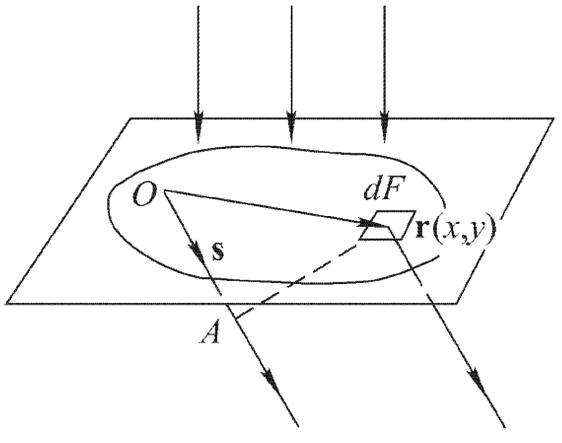


Рис. 178

$$E = \int e^{ik(\mathbf{sr})}\, dF, \tag{45.1}$$

распространенным по всем отверстиям.

В случае прямоугольного отверстия удобно перейти к прямоугольным координатам, предполагая, что координатные оси параллельны сторонам отверстия. Если a и b — длины этих сторон, то

$$E = \int\limits_{-a/2}^{+a/2} \int\limits_{-b/2}^{+b/2} e^{ik(s_x x + s_y y)}\, dx\, dy. \tag{45.2}$$

Это — в точности такие же интегралы, которые встречались нам при рассмотрении дифракции на щели. Выполнив интегрирование, получим

$$E = ab\, \frac{\sin\alpha}{\alpha}\, \frac{\sin\beta}{\beta}, \tag{45.3}$$

где

$$\alpha = \frac{1}{2}\,kas_x = \frac{\pi a s_x}{\lambda}, \quad \beta = \frac{1}{2}\,kbs_y = \frac{\pi b s_y}{\lambda}. \tag{45.4}$$

Заменив в одном из интегралов пределы интегрирования бесконечными, получим предельный случай бесконечно длинной щели. Интенсивность определяется формулой

$$I = I_0 \left(\frac{\sin\alpha}{\alpha}\right)^2 \left(\frac{\sin\beta}{\beta}\right)^2. \tag{45.5}$$

Дифракционную картину можно получить, если наложить друг на друга две взаимно перпендикулярные дифракционные картины, одна из которых получена при дифракции на щели ширины a, а другая — на щели ширины b (рис. 179). Картина вытянута в направлении более короткой стороны прямоугольного отверстия.

Случай круглого отверстия на практике представляет большой интерес, так как все оправы линз и объективов имеют обычно круглую форму. В этом случае при вычислении интеграла (45.1) естественно перейти к полярным координатам. При малых углах дифракции ин-

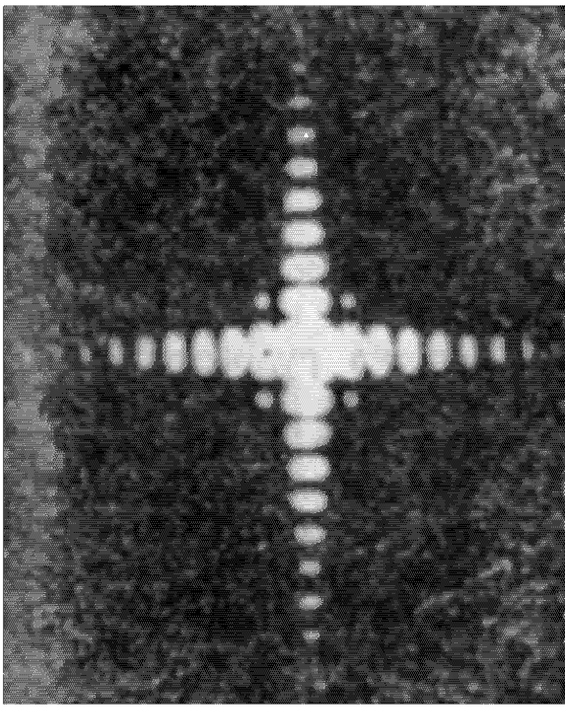


Рис. 179

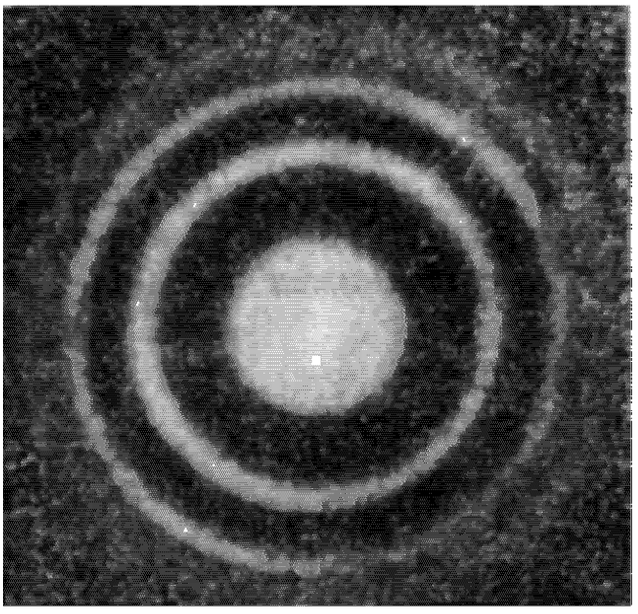


Рис. 180

теграл выражается через бесселеву функцию первого порядка $J_1(\alpha)$, где $\alpha = kR\vartheta = 2\pi R\vartheta/\lambda$ (R — радиус отверстия, ϑ — угол дифракции). Опуская вычисления, приведем окончательные результаты. Дифракционная картина, естественно, имеет вид концентрических светлых и темных колец (рис. 180). Центр картины светлый, так как в него все вторичные волны приходят в оди-

наковых фазах. Распределение амплитуд (штриховая кривая) и интенсивностей (сплошная кривая) в зависимости от угла дифракции ϑ (или, что то же, расстояния от центра картины) приведено на рис. 181. Соответствующие кривые мало отличаются от кривых рис. 175 не только качественно, но и количественно. Приближенно угловые радиусы темных колец определяются формулой

$$\vartheta_m = \left[0{,}61 + \frac{m-1}{2}\right]\frac{\lambda}{R}. \qquad (45.6)$$

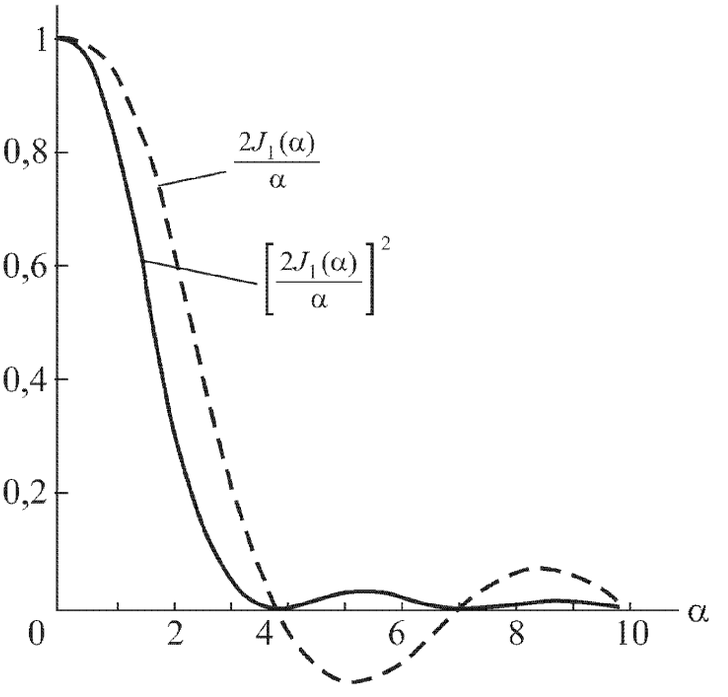


Рис. 181

Более точные данные приведены в табл. 5. Из нее видно, что около 98 % света приходится на центральный максимум. Если исключить центральный максимум, то остальные темные и светлые кольца практически равноотстоящие.

Таблица 5

Минимумы	Максимумы	Интенсивность в максимумах
$\vartheta_1 = 0{,}61$	$\vartheta'_1 = 0$	1
$\vartheta_2 = 1{,}12$	$\vartheta'_2 = 0{,}81$	0,0175
$\vartheta_3 = 1{,}62$	$\vartheta'_3 = 1{,}33$	0,0042
$\vartheta_4 = 2{,}12$	$\vartheta'_4 = 1{,}85$	0,0016

2. Таким же путем, по крайней мере численно, можно рассчитать дифракционную картину Фраунгофера на отверстии любой формы. При решении подобных задач полезно руководствоваться *соображениями подобия*. Представим волновое поле интегралом вида (45.2), но распространенным по области F, занимаемой рассматриваемым отверстием. Введем новые координаты $x' = \mu x$, $y' = y$, где μ — постоянная. Область F плоскости XY преобразуется в область F' плоскости $X'Y'$. Она получается из

F равномерным растяжением в μ раз в направлении оси X. Интеграл (45.2) преобразуется в

$$E(\mathbf{s}) = \frac{1}{\mu} \iint\limits_{F'} e^{i(s'_x x' + s'_y y')}\, dx'\, dy',$$

где единичный вектор $\mathbf{s}'$ определяется своими проекциями $s'_x = s_x/\mu$, $s'_y = s_y$. (Постоянная μ должна быть такой, чтобы $s'^2_x + s'^2_y < 1$.) Но последний интеграл представляет волновое поле $E'(\mathbf{s}')$ (в штрихованной системе координат) в направлении единичного вектора $\mathbf{s}'$ при фраунгоферовой дифракции на отверстии F', а потому

$$E'(\mathbf{s}') = \mu E(\mathbf{s}). \tag{45.7}$$

Таким образом, по известной дифракционной картине на каком-либо отверстии можно без новых вычислений получить новые дифракционные картины. Для этого надо отверстие равномерно вытянуть (сжать) в каком-либо направлении. Тогда, как видно из формул $x' = \mu x$, $s'_x = s_x/\mu$, дифракционная картина сожмется (вытянется) в том же направлении. Так, при растяжении круглого отверстия оно переходит в эллиптическое, а дифракционные кольца сжимаются, также принимая эллиптическую форму. Конечно, отверстие можно вытянуть или сжать и вдоль каких-либо двух направлений.

3. Рассмотрим теперь случай, когда в экране имеется большое число N одинаковых и одинаково ориентированных отверстий. Волновое поле в бесконечности представится суммой $E = \sum E_i$, где E_i — поле, которое возникло бы при дифракции при наличии одного только i-го отверстия. Для интенсивности получим

$$I = \overline{EE^*} = \sum I_i + \sum_{i \neq j} \sum \overline{E_i E_j^*}.$$

При фраунгоферовой дифракции распределение интенсивности в дифракционной картине определяется только *направлением лучей*, а не положением световых пучков. При боковом смещении последних интенсивность не меняется. *Распределение интенсивности не изменится, если отверстие в плоскости экрана сместить в сторону без изменения его ориентации.* Поэтому в последней сумме все интенсивности одинаковы: $I_i = I_1$.

Рассмотрим особо два случая: 1) отверстия расположены хаотически; 2) отверстия расположены «правильно», в определенном порядке. В первом случае среди членов двойной суммы с $i \neq j$ в среднем найдется столько же положительных членов, сколько и отрицательных. При сложении таких членов в сред-

нем получится нуль. Поэтому $I = NI_1$. Получается такая же дифракционная картина, что и от одного отверстия, но усиленная по интенсивности в N раз. Интенсивности отдельных картин арифметически складываются, но сами картины не интерферируют между собой. Во втором случае, напротив, члены с $i \neq j$ не компенсируются. Они могут интерферировать и существенно влиять на дифракционную картину. Этот случай будет рассмотрен в следующем параграфе на примере дифракционной решетки.

Первый случай легко демонстрировать на стеклянной пластинке, запыленной спорами ликоподия. Последние имеют форму шариков практически одинаковых размеров. При освещении пластинки параллельным пучком лучей на удаленном экране появляется дифракционная картина, состоящая из концентрических колец. При освещении белым светом внешние края колец окрашены в красный, а внутренние — в фиолетовый цвет. Это указывает на дифракционную природу явления, так как длины волн красных лучей больше, чем фиолетовых.

4. Описанное явление наблюдается в природе в виде *венцов*. Так называются светлые туманные кольца на небесном своде вокруг Солнца или Луны. Иногда венцы наблюдаются вокруг ярких звезд или планет, а также вокруг земных источников света. Венцы возникают в результате дифракции света на водяных капельках (или кристалликах льда), когда перед светилом проходит полупрозрачное облако (чаще всего высококучевое) или туман. Угловые радиусы венцов обычно не превосходят 5°. Дифракционная природа колец в явлении венцов подтверждается тем, что наружные края колец имеют красноватый цвет, а внутренние — синеватый. При наличии в атмосфере капель всевозможных размеров кольца венцов налагаются друг на друга и образуют общее белое сияние вокруг диска светила, называемое в случае Солнца *околосолнечным ореолом*.

От венцов следует отличать *гало*. Так называется группа оптических явлений в атмосфере, возникающая при преломлении или отражении лучей Солнца или Луны на плавающих в воздухе кристалликах льда, образующих перистые облака. Угловые радиусы гало значительно больше угловых радиусов венцов и составляют 22° или 46°. Наружный край колец гало имеет синеватую, а. внутренний — красноватую окраску. Это указывает на рефракционную природу явления (дисперсия). В отличие от венцов, угловые размеры которых могут меняться с изменением радиусов водяных капелек, угловые размеры колец гало строго постоянны, так как они зависят только от углов между гранями кристалликов, которые при изменении размеров последних остаются неизменными.

§ 46. Дифракционная решетка

1. *Дифракционная решетка* — важнейший спектральный прибор, предназначенный для *разложения света в спектр и измерения длин волн.* Она представляет собой плоскую стеклянную или металлическую поверхность, на которой делительной машиной нарезано очень много (до сотен тысяч) прямых равноотстоящих штрихов. На стеклянных решетках наблюдения можно производить как в *проходящем*, так и в *отраженном свете*, на металлических — только в *отраженном.* Применяются *вогнутые металлические решетки*, в которых штрихи наносятся на вогнутой сферической поверхности.

Сначала рассмотрим простейшую идеализированную решетку, состоящую из одинаковых равноотстоящих параллельных щелей, сделанных в непрозрачном экране. Ширину щели обозначим через b, ширину непрозрачной части экрана между двумя соседними щелями — через a. Величина $d = a + b$ называется *периодом решетки.* В решетке осуществляется *многолучевая интерференция* когерентных дифрагированных пучков света, исходящих от щелей решетки при ее освещении. Дифракционная картина наблюдается по методу Фраунгофера, т.е. либо на бесконечно удаленном экране, либо в фокальной плоскости линзы, поставленной на пути дифрагированного света.

Пусть на решетку перпендикулярно к ее поверхности падает плоская монохроматическая волна (рис. 182). Разность хода

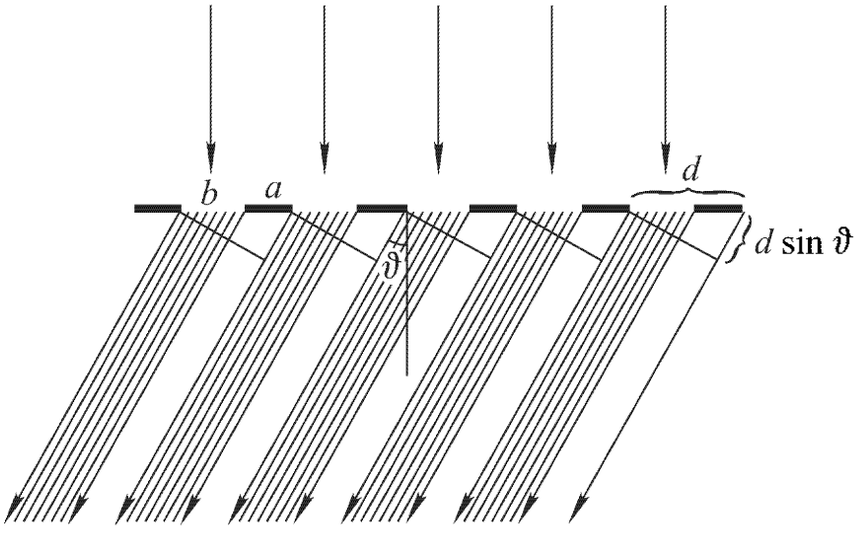


Рис. 182

между вторичными волнами, исходящими из соседних щелей решетки, будет $d\sin\vartheta$, а разность фаз $\delta = kd\sin\vartheta = \dfrac{2\pi}{\lambda}d\sin\vartheta$, где ϑ — угол дифракции. Обозначим через E_1 поле в точке наблюдения, излучаемое первой щелью. Оно определяется формулой

$E_1 = b\dfrac{\sin\alpha}{\alpha}$. Поля, излучаемые остальными щелями, представляются выражениями:

$$E_2 = E_1 e^{-i\delta}, \quad E_3 = E_1 e^{-2i\delta}, \ldots, E_N = E_1 e^{-i(N-1)\delta},$$

где N — общее число щелей. Полное поле, излучаемое всеми щелями, представляется суммой

$$E = E_1[1 + e^{-i\delta} + e^{-2i\delta} + \ldots + e^{-i(N-1)\delta}] = E_1\frac{1 - e^{-iN\delta}}{1 - e^{-i\delta}},$$

откуда

$$E = E_1\frac{\sin(N\delta/2)}{\sin(\delta/2)}e^{-i(N-1)\delta/2},$$

$$A = A_1\frac{\sin(N\delta/2)}{\sin(\delta/2)}, \tag{46.1}$$

где A_1 — вещественная амплитуда волны от одной щели, A — от всей решетки. Для интенсивностей получаем

$$I = I_1\left[\frac{\sin(N\delta/2)}{\sin(\delta/2)}\right]^2. \tag{46.2}$$

Формулы (46.1) и (46.2) — основные в теории дифракционной решетки.

2. Когда $\vartheta = 0$, то $\delta = 0$. В этом случае выражения (46.1) и (46.2) принимают неопределенный вид $0/0$. Раскрыв неопределенность, получим

$$A_{\text{гл}} = NA_1, \quad I_{\text{гл}} = N^2 I_1. \tag{46.3}$$

Тот же результат получается в случае $\delta/2 = m\pi$, т.е. при

$$d\sin\vartheta = m\lambda \qquad (m = 0, \pm1, \pm2, \ldots). \tag{46.4}$$

В направлениях, определяемых этим условием, получаются максимумы, интенсивность которых в N^2 раз превосходит интенсивность волны от одной щели в том же направлении. Они называются *главными максимумами*. Целое число m называют *порядком главного максимума* или *порядком спектра*. Условие (46.4) определяет направления, в которых излучения от всех щелей решетки приходят в точку наблюдения в одинаковых фазах, а потому усиливают друг друга. В таких направлениях при отдельных значениях m могут и не возникнуть максимумы. Это будет, когда $I_1 = 0$, т.е. в направлениях на дифракционные минимумы от одной щели. Например, если $a = b$, то все главные максимумы четных порядков не появятся. Действительно, условие появления главного максимума порядка $2n$ имеет вид $d\sin\vartheta = 2n\lambda$. При $d = 2b$ оно переходит в $b\sin\vartheta = n\lambda$, т.е. в условие *дифракционного минимума* на щели. Таким образом, в рассматриваемом направлении ни одна щель, а потому и решетка в целом не излучают.

3. Выражения (46.1) и (46.2) обращаются в нуль, если $\sin(N\delta/2) = 0$, но $\sin(\delta/2) \neq 0$, т.е. при $N\delta/2 = (Nm + p)\pi$, или

$$d\sin\vartheta = \left(m + \frac{p}{N}\right)\lambda \qquad (p = 1, 2, \ldots, N-1). \tag{46.5}$$

В соответствующих направлениях получаются дифракционные минимумы, в которых интенсивность света равна нулю. Между двумя соседними минимумами получается максимум. Такие максимумы называются *второстепенными* или *добавочными*. Между двумя соседними главными максимумами располагается $(N-1)$ минимумов и $(N-2)$ добавочных максимумов. На эти максимумы и минимумы накладываются минимумы, возникающие при дифракции от отдельной щели, в которых функция E_1 обращается в нуль.

Второстепенные максимумы находятся примерно посередине между соответствующими дифракционными минимумами. Величину δ, определяющую направление на какой-либо из второстепенных максимумов, можно поэтому вычислить по приближенной формуле

$$N\frac{\delta}{2} = \frac{(Nm + p) + (Nm + p + 1)}{2}\,\pi,$$

или

$$\frac{\delta}{2} = \left(m + \frac{2p+1}{2N}\right)\pi. \tag{46.6}$$

Пользуясь этой формулой, найдем приближенное выражение для интенсивности второстепенных максимумов в окрестности соответствующего главного максимума (т.е. при малых p). В положениях второстепенных максимумов числители в формуле (46.2) равны единице. Если число щелей решетки N очень велико, а номер второстепенного максимума p невелик, то угол $\delta/2$ будет мал, и можно положить

$$\sin\frac{\delta}{2} = \pm\sin\frac{2p+1}{2N}\,\pi \approx \pm\frac{2p+1}{2N}\,\pi.$$

Это дает

$$I = \frac{I_1}{\pi^2}\left(\frac{2N}{2p+1}\right)^2 = \frac{4}{(2p+1)^2\pi^2}\,I_{\text{гл}}. \tag{46.7}$$

Таким образом, интенсивности главного максимума и ближайших к нему второстепенных максимумов находятся в отношениях

$$1 : \frac{4}{9\pi^2} : \frac{4}{25\pi^2} : \frac{4}{49\pi^2} : \ldots = 1 : 0{,}045 : 0{,}016 : 0{,}0083 : \ldots$$

Второстепенные максимумы слабы по сравнению с главными максимумами. При большом числе щелей они обычно не играют

роли. Второстепенные максимумы создают более или менее равномерный слабый фон, на нем выступают узкие и резкие главные максимумы, в которых концентрируется практически весь дифрагированный свет. Распределение интенсивности в дифрагированном свете представлено на схематическом рис. 183 для

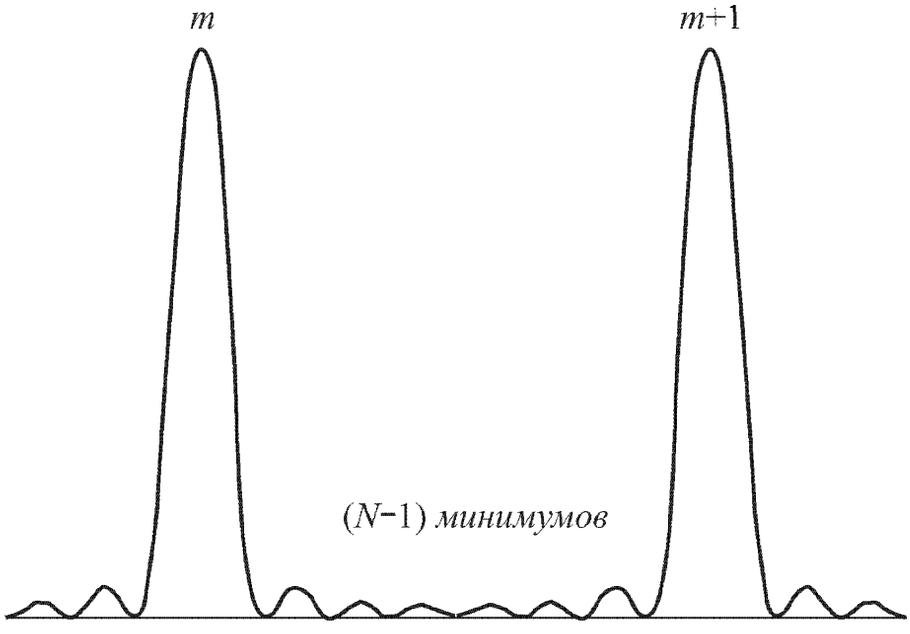


Рис. 183

$N = 8$. Величина множителя I_1, входящего в формулу (46.2), не указана. Между соседними главными максимумами находится $N - 2 = 6$ второстепенных максимумов. Относительные интенсивности главных и второстепенных максимумов представляются числами 100; 5,0; 2,25; 1,6; 1,6; 2,25; 5,0; 100.

Красивую демонстрацию можно получить, направив на решетку яркий узкий пучок света от лазера. При дифракции пучок расщепляется на много хорошо видимых в воздухе ярких пучков, веерообразно расходящихся от решетки. Падая на потолок и стены аудитории, эти пучки оставляют на них яркие светлые пятна (световые зайчики).

4. Происхождение второстепенных или добавочных максимумов и минимумов легко уяснить с помощью векторной диа-

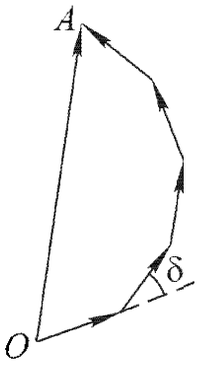


Рис. 184

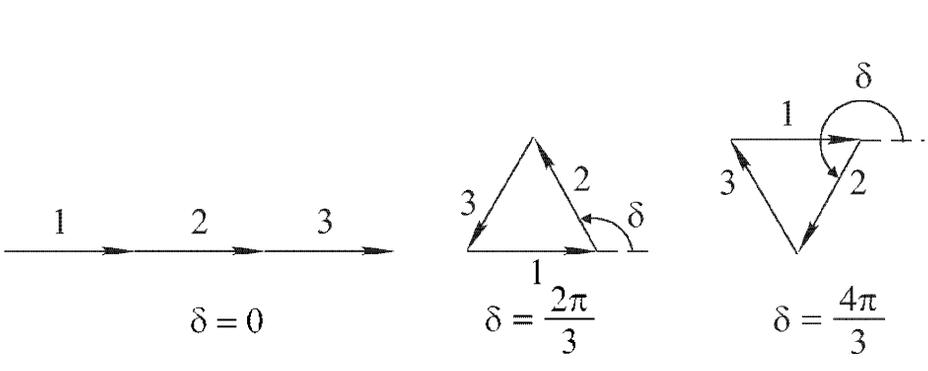


Рис. 185

граммы. На векторной диаграмме колебания от отдельных щелей решетки представляются векторами равной длины. Из-за сдвига фаз такие векторы повернуты один относительно другого на один и тот же угол $\delta = \frac{2\pi}{\lambda} d \sin\vartheta$ (рис. 184), образуя звенья правильной ломаной линии. Замыкающая этой ломаной OA изображает колебание, возбуждаемое всей решеткой. Если ломаная замкнутая, то $E = 0$, т.е. получается дифракционный минимум. Первый раз это произойдет при $\delta = 2\pi/N$, второй — при $\delta = 4\pi/N$ и т.д. Легко видеть, что таким путем получается условие дифракционного минимума (46.5). Все это проиллюстрировано на рис. 185 для $N = 3$ и на рис. 186 для $N = 4$.

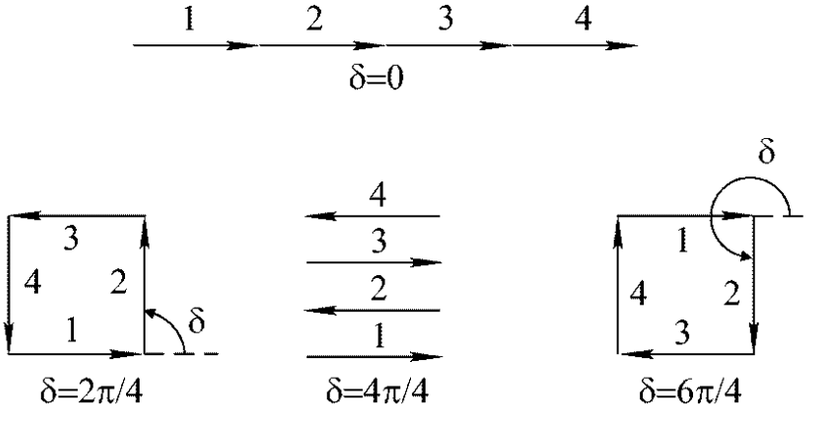


Рис. 186

5. Если волна падает на решетку наклонно под углом ϑ_0 (рис. 187), то разность хода между соседними пучками становится равной $AD - CB = d(\sin\vartheta - \sin\vartheta_0)$. Характер дифракционной картины в основном сохраняется. Положение главных максимумов определяется условием

$$d(\sin\vartheta - \sin\vartheta_0) = m\lambda, \quad (46.8)$$

а дифракционных минимумов — условием

$$d(\sin\vartheta - \sin\vartheta_0) = \left(m + \frac{p}{N}\right)\lambda \quad (p = 1, 2, \ldots, N-1). \quad (46.9)$$

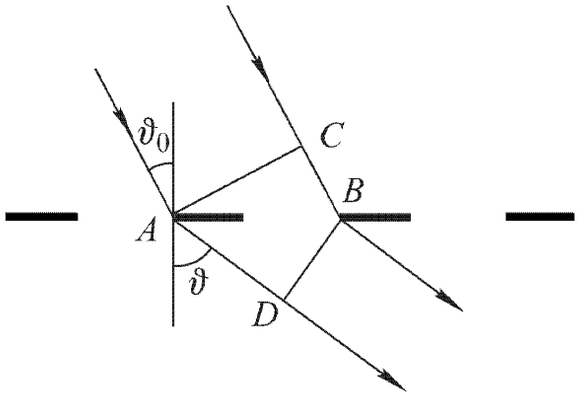


Рис. 187

Зная положение главных максимумов, можно вычислить длину волны по формуле (46.4) или (46.8). Решетка есть прибор для измерения именно *длины волны*, а не частоты колебаний, как это иногда неверно утверждают.

6. Форма штрихов, нанесенных на решетку, материал, из которого она изготовлена, и т.п. сказываются лишь на виде волны

E_1 *от отдельного штриха.* Положение же главных максимумов определяется исключительно периодом решетки, а дифракционных минимумов — кроме того еще числом штрихов N. При выводе формул (46.4), (46.8) и (46.9) из принципа Гюйгенса–Френеля нет необходимости применять приближенные методы. Из рассуждений п. 4 ясно, что существенна только *периодичность светового поля на выходе* (т.е. задней стороне) *решетки*, обусловленная в свою очередь периодичностью структуры последней. Ничего другого о поле на выходе решетки знать не надо. Поэтому перечисленные формулы *совершенно точны.*

Под дифракционной решеткой в широком смысле слова понимается всякая структура, обладающая *пространственной периодичностью.* Если свойства структуры периодически меняются только в одном направлении, то решетка называется *одномерной* или *линейной.* Если же периодичность решетки имеет место в двух или трех направлениях, то решетка называется соответственно *двух-* или *трехмерной.* В последнем случае ее называют также *пространственной.* В этом и следующих параграфах рассматриваются только линейные решетки.

Не обязательно, чтобы при прохождении через решетку менялась амплитуда волны. Существенно только, чтобы на выходе решетки периодически менялось *волновое поле в целом.* Можно различать два крайних идеализованных случая: 1) решетка вносит периодические изменения в амплитуду волны, не влияя на ее фазу; такая решетка называется *амплитудной*; 2) решетка вносит периодические изменения в фазу волны, но не влияет на ее амплитуду; такую решетку называют *фазовой.*

Всякая реальная решетка, строго говоря, не является чисто амплитудной или чисто фазовой. Она периодически меняет на выходе как амплитуду, так и фазу волнового поля. Приблизительно амплитудной решеткой является рассмотренная выше совокупность равноотстоящих щелей в непрозрачном экране (рис. 187). Приближением фазовой решетки может служить стеклянная пластинка, представленная на рис. 188. В обоих случаях период решетки должен быть велик по сравнению с длиной волны. Примером фазовой отражательной решетки может служить решетка, изготовленная С.М. Рытовым (р. 1908) и И.Л. Фабелинским (р. 1911). Она представляет собой равнобочную стеклянную призму с преломляющим углом 90° (рис. 189). На гипотенузной поверхности AB напылены узкие равноотстоящие полоски серебра, параллельные преломляющему ребру призмы. Свет падает нормально на одну из боковых граней призмы. Попадая на гипотенузную грань AB, он испытает либо полное отражение от непосеребренных полосок, либо отразится от посеребренных полосок. В последнем случае отражение также практически полное благодаря высокому коэффициенту

отражения серебра. В обоих случаях отражение сопровождается изменением фазы, но эти изменения разные (см. § 66 и 73). Таким образом, амплитуда волны при отражении не меняется, а в фазу вносятся периодические изменения.

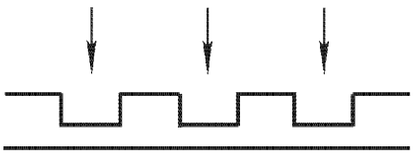


Рис. 188

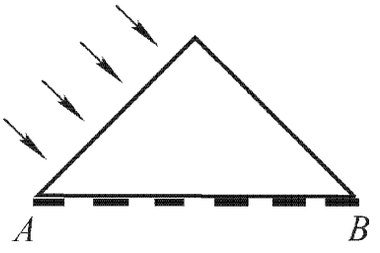


Рис. 189

7. Не все целые числа m, входящие в формулы (46.4) и (46.8), допустимы. Так как $|\sin\vartheta| < 1$, то на m должно быть наложено ограничение

$$\left|m\frac{\lambda}{d}+\sin\vartheta_0\right| \leqslant 1. \tag{46.10}$$

В частности, при нормальном падении

$$|m| \leqslant d/\lambda. \tag{46.11}$$

При $d < \lambda$ может возникнуть только главный максимум нулевого порядка. Плоская волна после прохождения через решетку продолжает оставаться плоской, и никаких боковых волн не возникает. Решетка ведет себя подобно плоскопараллельной пластинке. Но это справедливо только вдали от решетки. В тонком поверхностном слое вблизи решетки существуют неоднородные волны, быстро затухающие при удалении от нее (см. § 52).

8. Допустим, что решетка грубая ($d \gg \lambda$), а свет падает на нее под небольшим углом ϑ_0. Тогда углы дифракции будут малы, так что вместо (46.8) можно написать приближенно

$$d\cdot(\vartheta-\vartheta_0) = m\lambda. \tag{46.12}$$

На очень грубых решетках обнаружить дифракцию трудно из-за малости углов дифракции $(\vartheta-\vartheta_0)$. Но этого можно добиться, применяя скользящее падение света, когда угол падения близок к 90°. При малой разности $(\vartheta-\vartheta_0)$ формулу (46.8) можно переписать в виде

$$d\cos\vartheta_0\cdot(\vartheta-\vartheta_0) = m\lambda. \tag{46.13}$$

По виду формула (46.13) совпадает с формулой (46.12). Роль периода d играет величина $d\cos\vartheta_0$, которая может быть сделана очень малой. Следовательно, скользящее падение как бы уменьшает период решетки и увеличивает углы дифракции. Таким путем удается получать отчетливые дифракционные спектры даже от очень грубых решеток, например от граммофонных

пластинок. Последние позволяют в демонстрационной аудитории получать в белом свете даже довольно красивые дифракционные спектры разных порядков.

Большое значение метод скользящего падения имеет в рентгеновской спектроскопии. Изготовление дифракционной решетки для рентгеновской области спектра встречает большие трудности из-за исключительной малости длин рентгеновских волн (порядка 0,1 нм и меньше). Но метод скользящего падения позволяет получать великолепные дифракционные картины в рентгеновском свете от обычных отражательных оптических решеток. Задача облегчается тем, что показатель преломления рентгеновских лучей меньше единицы. Это позволяет применять такие углы падения, при которых рентгеновские лучи испытывают полное отражение. Таким образом, с помощью оптической дифракционной решетки можно точно измерить длину волны монохроматического рентгеновского излучения. Изучая далее дифракцию этого излучения на естественном кристалле, можно измерить в абсолютных единицах постоянную решетки этого кристалла. После этого такой кристалл может быть использован в рентгеновском спектрографе для измерения длин волн рентгеновских лучей в абсолютных единицах (см. § 61).

ЗАДАЧИ

1. Показать, что для обычной щелевой дифракционной решетки $I_{\text{диф}} \leqslant \leqslant {}^1\!/_4 I_{\text{пад}}$, где $I_{\text{пад}}$ — интенсивность падающего, а $I_{\text{диф}}$ — дифрагированного света, приходящегося на все дифракционные максимумы, кроме максимума нулевого порядка. Максимальное значение интенсивности дифрагированного света достигается, когда ширина щелей решетки равна половине ее периода.

Р е ш е н и е. Если I_0 — интенсивность света, сосредоточенного в максимуме нулевого порядка, а $I_{\text{диф}}$ — дифрагированного света, то $I_{\text{прош}} = I_0 + I_{\text{диф}}$, где $I_{\text{прош}}$ — интенсивность прошедшего света. Аналогично, для дополнительной решетки $I'_{\text{прош}} = I'_0 + I'_{\text{диф}}$. По теореме Бабине $I_{\text{диф}} = I'_{\text{диф}}$. Кроме того,

$$\frac{I_{\text{прош}}}{I'_{\text{прош}}} = \frac{b}{d-b}, \quad \frac{I_0}{I'_0} = \left(\frac{b}{d-b}\right)^2.$$

С учетом этого получаем

$$I_{\text{диф}} = \frac{b}{d}\,\frac{d-b}{d} I_{\text{пад}}.$$

Максимум этого выражения достигается при $b = {}^1\!/_2 d$, а потому $I_{\text{диф}} \leqslant \leqslant {}^1\!/_4 I_{\text{пад}}$.

2. При каком условии можно наблюдать зеркальное отражение от шероховатой поверхности при малых и больших углах падения?

Р е ш е н и е. Обозначим через ϑ угол, который падающие лучи образуют с нормалью к плоскости, изображенной на рис. 190 штриховой линией.

Разность хода вторичных волн, исходящих от поверхности тела под углом ϑ' к нормали, определяется выражением

$$\Delta = AD - BC = a(\sin\vartheta' - \sin\vartheta) + h(\cos\vartheta' + \cos\vartheta).$$

Здесь a может принимать любые значения. Поэтому, если $\vartheta' \neq \vartheta$, то Δ может принимать также любые значения, и притом для неправильно отражающей поверхности они с одинаковой вероятностью могут быть и положительными, и отрицательными. Отсюда следует, что правильное отражение, если таковое возможно, может происходить лишь под углом $\vartheta' = \vartheta$. В таком случае $\Delta = 2h\cos\vartheta$. Следовательно, каково бы ни было h, можно подобрать достаточно большой угол ϑ, для которого $\Delta \ll \lambda$. При этом условии отражение будет правильное. При нормальном падении $\Delta = 2h$, и правильное отражение возможно лишь при соблюдении условия $h \ll \lambda$. Таким образом, если угол падения близок к $\pi/2$, зеркальное отражение наблюдается всегда. При малых углах падения зеркальное отражение может наблюдаться, если шероховатость поверхности $\ll\lambda$.

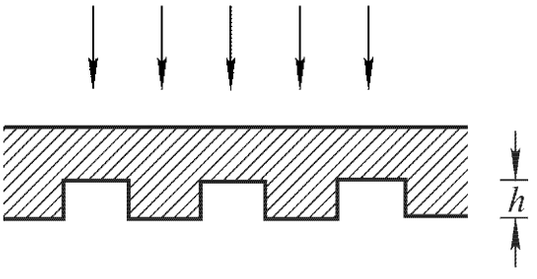


Рис. 191

3. Прозрачная одномерная периодическая структура, профиль которой изображен на рис. 191, освещается сверху плоской монохроматической волной, падающей нормально на верхнюю границу. Ширины уступов и впадин структуры одинаковы. При заданном показателе преломления n подобрать глубину h таким образом, чтобы главные фраунгоферовы дифракционные максимумы первого порядка имели наибольшую интенсивность. Какова при этом интенсивность нулевого главного максимума?

О т в е т. $h = \dfrac{2m-1}{2(n-1)}\lambda$, где $m = 1, 2, 3, \ldots$

Интенсивность нулевого главного максимума равна нулю.

4. Сложная излучающая система состоит из конечного ряда густо расставленных параллельных вибраторов с равномерным распределением фаз колебаний вдоль ряда. Как должен изменяться со временем сдвиг фаз $\Delta\varphi$ между двумя соседними вибраторами, чтобы главный лепесток диаграммы направленности всей системы совершал круговой обзор местности с постоянной угловой скоростью Ω (при отсутствии вращения самой системы)?

О т в е т. Если угол между нормалью к ряду вибраторов и направлением на главный лепесток $\vartheta = \Omega t + \delta$, то

$$\Delta\varphi = -\frac{2\pi}{\lambda}\sin(\Omega t + \delta) + 2m\pi,$$

где

$$m = 0, \pm 1, \pm 2, \ldots$$

5. Над центром граммофонной пластинки O_1 помещен точечный источник монохроматического света на высоте $h_1 = 1$ см. Глаз наблюдателя расположен на высоте $h_2 = 10$ см на расстоянии $a = 110$ см от оси пластинки. Помимо геометрического изображения источника глаз видит систему дифракционных полос на поверхности пластинки. Определить расстояние Δx между ними, если расстояние между бороздками $d = 0{,}5$ мм. Длина световой волны $\lambda = 550$ нм.

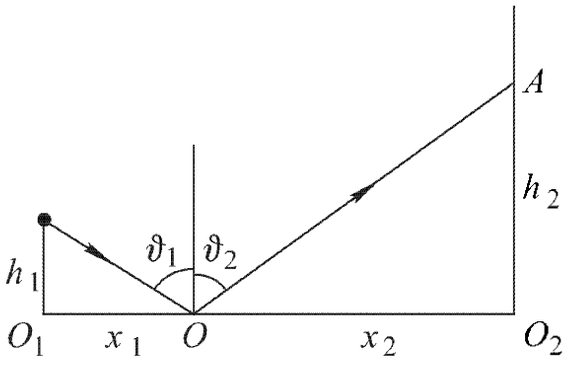


Рис. 192

Р е ш е н и е. Условие максимума m-го порядка (рис. 192) $d(\sin\vartheta_1 - \sin\vartheta_2) = m\lambda$, а максимума $(m+1)$-го порядка $d(\sin\vartheta_1' - \sin\vartheta_2') = (m+1)\lambda$. При переходе от одного максимума к другому углы ϑ_1 и ϑ_2 получают приращения $\Delta\vartheta_1$ и $\Delta\vartheta_2$, связанные соотношением $d(\cos\vartheta_1\,\Delta\vartheta_1 - \cos\vartheta_2\,\Delta\vartheta_2) = \lambda$. Кроме того, $x_1 = h_1\operatorname{tg}\vartheta_1$, $x_2 = h_2\operatorname{tg}\vartheta_2$, причем $x_1 + x_2 = a = \text{const}$. Из этого условия получается второе соотношение

$$\frac{h_1}{\cos^2\vartheta_1}\Delta\vartheta_1 + \frac{h_2}{\cos^2\vartheta_2}\Delta\vartheta_2 = 0.$$

Из этих двух соотношений находим $\Delta\vartheta_1$ и $\Delta\vartheta_2$, а после этого — расстояние между максимумами:

$$\Delta x = \Delta x_1 = \Delta x_2 = \frac{h_1}{\cos^2\vartheta_1}\Delta\vartheta_1 = -\frac{h_2}{\cos^2\vartheta_2}\Delta\vartheta_2.$$

Результат можно упростить, заметив, что углы ϑ_1 и ϑ_2 мало отличаются от угла падения ϑ_0, соответствующего зеркальному отражению. Заменив эти углы на ϑ_0, находим

$$\Delta x = \frac{h_1 h_2}{h_1 + h_2}\,\frac{\lambda}{d\cos^3\vartheta_0}.$$

При этом

$$\cos\vartheta_0 \approx \frac{h_1}{x_1} \approx \frac{h_2}{x_2}, \quad \text{или} \quad \cos\vartheta_0 = \frac{h_1 + h_2}{x_1 + x_2} = \frac{h_1 + h_2}{a} = \frac{1}{10}.$$

Подстановка числовых значений дает $\Delta x = 1$ см.

§ 47. Дифракционная решетка как спектральный прибор

1. Положение главных максимумов в дифракционной решетке зависит от длины волны. Исключение составляют только

главные максимумы нулевого порядка ($m = 0$), положения которых от длины волны не зависят. Белый и всякий сложный свет можно рассматривать как суперпозицию монохроматических волн с различными длинами. Эти волны при дифракции на решетке ведут себя *независимо*. Поэтому решетка в каждом порядке $m \neq 0$ разложит падающий свет в спектр, в котором отдельные монохроматические компоненты окажутся пространственно разделенными. Главные дифракционные максимумы, соответствующие $m = 1$, образуют *спектр первого порядка*. За ним идет *спектр второго* ($m = 2$), *третьего* ($m = 3$) и *высших порядков*. Если падающий свет белый, то спектр каждого порядка имеет вид цветной полосы, в которой встречаются все цвета радуги. В такой полосе наиболее отклоненными будут красные лучи, наименее отклоненными — фиолетовые.

Положение спектральных линий в спектрах дифракционной решетки определяется простыми соотношениями (46.4) или (46.8). В этом отношении дифракционные спектры выгодно отличаются, например, от спектров призматических, получаемых разложением света дисперсионными призмами. В призматических спектрах положение спектральной линии определяется сложной зависимостью показателя преломления материала призмы от длины волны. Спектр называется *нормальным*, если координата x, характеризующая положение спектральной линии в спектре, линейно меняется с длиной волны. При малых углах дифракции, когда изменением косинуса угла ϑ можно пренебречь, дифракционная решетка дает нормальный спектр.

Важными характеристиками дифракционной решетки и других спектральных аппаратов являются *угловая дисперсия*, *дисперсионная область* и *разрешающая способность*.

2. Угловая дисперсия. *Угловой дисперсией называется производная* $d\vartheta/d\lambda$. Чем больше угловая дисперсия, тем больше расстояние в спектре между двумя спектральными линиями с фиксированными длинами волн. Дифференцируя формулу (46.8) при постоянном ϑ_0, находим для решетки

$$\frac{d\vartheta}{d\lambda} = \frac{m}{d\cos\vartheta} = \frac{\sin\vartheta - \sin\vartheta_0}{\lambda\cos\vartheta}. \tag{47.1}$$

Следовательно, *угловая дисперсия не зависит от параметров решетки, а определяется, помимо длины волны, только углами* ϑ *и* ϑ_0. При заданных d и m она возрастает с увеличением угла ϑ. Наличие косинуса в знаменателе объясняет выгоду скользящего падения (см. конец предыдущего параграфа).

3. Дисперсионная область. Если спектры соседних порядков перекрываются, то спектральный аппарат становится непригодным для исследования соответствующего участка спектра. *Максимальная ширина спектрального интервала* $\Delta\lambda$, *при*

которой еще нет перекрытия, называется дисперсионной областью спектрального аппарата. Найдем дисперсионную область для дифракционной решетки. Пусть длины волн падающего излучения лежат в спектральном интервале от λ до $\lambda' = \lambda + \Delta\lambda$. Пусть правый конец спектра $(m+1)$-го порядка для длины волны λ совпадает по своему положению с левым концом спектра m-го порядка для длины волны λ' (см. рис. 182). Тогда

$$d(\sin\vartheta - \sin\vartheta_0) = m\lambda',$$
$$d(\sin\vartheta - \sin\vartheta_0) = (m+1)\lambda.$$

Отсюда $m\lambda' = (m+1)\lambda$, а следовательно,

$$\lambda' - \lambda \equiv \Delta\lambda = \lambda/m. \tag{47.2}$$

Это и есть дисперсионная область дифракционной решетки в рассматриваемом участке спектра. При заданной длине волны она определяется только порядком спектра m. Чем больше m, тем у́же дисперсионная область. В дифракционных решетках используются спектры *низких порядков* (обычно второго или третьего). Поэтому дифракционные решетки характеризуются широкими областями дисперсии и пригодны для исследования *широких участков спектра.* В этом основное преимущество дифракционных решеток перед интерференционными спектральными аппаратами, у которых из-за высоких порядков m дисперсионные области очень узкие.

4. Разрешающая способность. Большая дисперсия еще не означает, что две спектральные линии с близкими длинами волн λ и $\lambda' = \lambda + \delta\lambda$ разрешаются спектральным аппаратом, т.е. при их наблюдении воспринимаются как раздельные спектральные линии. Каждая спектральная линия, как бы узка она ни была, изображается спектральным аппаратом не в виде линии, а в виде более или менее размытой дифракционной картины с максимумами и минимумами интенсивности. Дисперсия определяет расстояние, на которое спектральный аппарат разводит центры дифракционных картин, возникающих от двух спектральных линий с различными длинами волн. Если сами картины размыты и имеют значительную ширину, то даже при сравнительно большом разведении их результирующая картина, возникающая от их наложения, неотличима от дифракционной картины, возникающей от одиночной спектральной линии. Чем у́же дифракционные картины от двух близких спектральных линий, тем на меньшее расстояние требуется развести их центры, чтобы разрешить эти спектральные линии. *Наименьшая разность длин волн двух спектральных линий $\delta\lambda$, при которой спектральный аппарат разрешает эти линии, называется спектральным разрешаемым расстоянием, а величина $R = \lambda/\delta\lambda$ — разрешающей способностью аппарата.*

Разрешаемое спектральное расстояние $\delta\lambda$ относится к числу не вполне точно определенных понятий и может быть указано лишь ориентировочно. Для дифракционной решетки Рэлей предложил следующий критерий спектрального разрешения. *Спектральные линии с близкими длинами волн λ и λ' считаются разрешенными, если главный максимум дифракционной картины для одной длины волны совпадает по своему положению с первым дифракционным минимумом в том же порядке для другой длины волны.* Если такой критерий выполняется, то на основании формулы (46.9) можно написать

$$d(\sin\vartheta - \sin\vartheta_0) = \left(m + \frac{1}{N}\right)\lambda,$$

$$d(\sin\vartheta - \sin\vartheta_0) = m\lambda'.$$

Отсюда $(m + 1/N)\lambda = m\lambda'$, и следовательно, $\delta\lambda \equiv \lambda' - \lambda = \lambda/(Nm)$,

$$R = \frac{\lambda}{\delta\lambda} = Nm. \tag{47.3}$$

С целью обоснования критерия Рэлея начертим распределение интенсивности для спектральных линий λ и λ', сдвинутых одна относительно другой на расстояние $\delta\lambda$ (рис. 193). Так как спектральные линии с различными длинами волн некогерентны, то результирующая интенсивность найдется простым сложением интенсивностей обеих линий. Она представлена на рис. 193 сплошной кривой, когда интенсивности обеих спектральных линий одинаковы, а расстояние между ними $\delta\lambda$ соответствует критерию Рэлея. В центре кривой получился провал или минимум, интенсивность в котором, как легко подсчитать, составляет около 80 % максимальной. При наличии такого провала нормальный глаз обычно воспринимает получившуюся картину как «двойную спектральную линию».

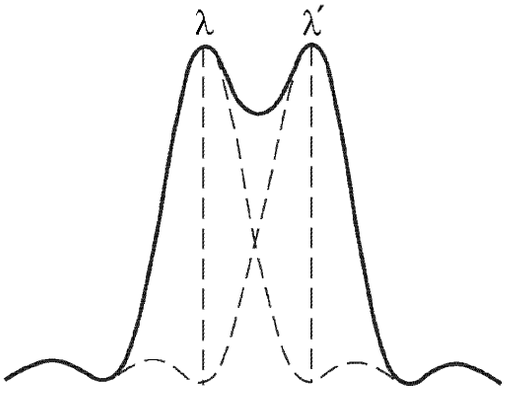


Рис. 193

5. Для повышения разрешающей способности можно либо увеличивать число штрихов N, либо повышать порядок интерференции m. Первый путь используется в дифракционных решетках, второй — в интерференционных спектральных приборах. В Советском Союзе изготовляются плоские и вогнутые дифракционные решетки различных размеров и с различным числом штрихов на миллиметр. Для ультрафиолетовой и видимой областей изготовляются дифракционные решетки, имеющие 1200 и 600 штрихов на миллиметр при размерах

100×100 мм^2 и 150×150 мм^2, а для инфракрасной области — от 300 до 1 штриха на миллиметр при размерах от 150×150 мм^2 до 300×300 мм^2 Таким образом, общее число штрихов доходит приблизительно до 200 000, а разрешающая способность во втором порядке — до 400 000. Важным достоинством дифракционной решетки является малый порядок спектра m. Благодаря этому дифракционные решетки обладают широкими дисперсионными областями $\Delta\lambda = \lambda/m$ и пригодны для исследования *широких интервалов спектра*. Недостатками дифракционных решеток являются малая светосила и сложность в обращении.

В интерференционных спектральных приборах число интерференционных пучков N относительно невелико (несколько десятков, в интерферометре Майкельсона $n = 2$), тогда как порядки спектров m очень высоки (около 10 000 и больше). Поэтому интерференционные спектральные приборы имеют малые дисперсионные области. Они могут применяться для исследования только *очень узких участков спектра*, например для изучения структуры отдельных спектральных линий, выделенных каким-либо другим спектральным аппаратом с большей дисперсионной областью, но с недостаточной разрешающей способностью. Однако эти приборы более просты в обращении и имеют бóльшую светосилу, чем дифракционные решетки.

6. К решетке, как точному спектральному прибору, предъявляются очень высокие требования. Надо нанести десятки или сотни тысяч совершенно идентичных штрихов с идеальной периодичностью. Поэтому техника изготовления дифракционных решеток совершенствовалась довольно медленно. Первая дифракционная решетка, по-видимому, была изготовлена в 1785 г. американским астрономом Риттенгаузом. Но ни самим Риттенгаузом и никем другим она не использовалась для получения и изучения спектров. Решетка была вновь открыта в 1821 г. Фраунгофером, который заложил основы дифракции в параллельных лучах и выполнил первые исследования с помощью дифракционного спектроскопа (в частности, открыл темные линии в спектре солнечного излучения). Фраунгофер построил решетки с числом штрихов от 15 до 130 на сантиметр, наматывая тонкую проволоку на два параллельных винта. До недавнего времени подобные проволочные решетки применялись в области длинных (инфракрасных) волн. Затем Фраунгофер стал изготовлять более совершенные решетки штриховкой слоя золота на поверхности стекла и, наконец, штриховкой самого стекла алмазным острием. Наилучшие решетки были получены последним методом. Лучшая решетка имела ширину около 12 мм и период $3 \cdot 10^{-3}$ мм (3300 штрихов на сантиметр).

После Фраунгофера многие искусные механики уделяли много внимания штриховке решеток. Особо следует отметить астронома-любителя Л.М. Резерфорда, большая часть решеток которого была изготовлена в 1880 г. Его решетки значительно превосходили все предшествующие. Резерфорд ввел в практику отражательные решетки, нанося делительной машиной штрихи на зеркальной поверхности металла. Металл более мягок, чем стекло, и поэтому значительно меньше изнашивает алмазное острие, от постоянства которого так сильно зависит качество решетки.

Но наиболее значительные усовершенствования были сделаны Роулэндом (1848–1901). Он усовершенствовал способы изготовления винтов для делительной машины и первый стал изготовлять вогнутые отражательные решетки, выполняющие одновременно функции решетки и собирающей линзы. Решетки Роулэнда имели до 8000 штрихов на сантиметр при ширине до 10 см и превосходном качестве. Они сделали возможным выполнение важнейших спектроскопических исследований в конце XIX и начале XX веков.

Дальнейшие улучшения в машины Роулэнда ввели Андерсон и Вуд, которые после смерти Роулэнда заменили его в его лаборатории. Из этой лаборатории и поныне выходят наиболее совершенные решетки, имеющие до 12 000 штрихов на сантиметр. С таких гравированных решеток получают дешевые копии (реплики) путем изготовления отпечатков на желатине или специальных пластмассах. В СССР налажено производство дифракционных решеток, а также реплик высокого качества. Их параметры были указаны выше.

§ 48. Эшелон Майкельсона и интерференционные спектральные приборы

1. Дифракционная решетка, как уже указывалось, есть спектральный прибор, в котором осуществляется многолучевая интерференция идентичных световых пучков, из которых каждый сдвинут по фазе относительно предыдущего на одну и ту же величину. Ко всем приборам, работающим по этому принципу, применимы общие результаты, изложенные в двух предыдущих параграфах. В частности, их спектральная разрешающая способность выражается формулой $\lambda/\delta\lambda = Nm$, т.е. равна *разности хода между крайними интерферирующими лучами, выраженной в длинах волн.* Высокая разрешающая способность достигается как увеличением числа интерферирующих пучков N, так и повышением порядка интерференции m. В дифракционных решетках число интерферирующих пучков очень велико (до 200 000 и больше), а порядки интерференции m низкие (не выше 3). В так называемых *интерференционных спектральных приборах*, наоборот, порядки интерференции m очень высокие (до 10^5 и выше), а число интерферирующих пучков сравнительно невелико (20–40). Одним из таких приборов является ступенчатая решетка, или *эшелон Майкельсона.*

2. Эшелон Майкельсона состоит из нескольких (30–40) пластин из очень однородного стекла, толщина h которых порядка 1–3 см. Толщина пластин должна быть совершенно одинакова с точностью до сотой длины волны. Поэтому их вырезают из одной и той же плоскопараллельной пластины, постоянство толщины которой контролируется интерференционными методами. Пластины сложены в виде ступенчатой лестницы (рис. 194).

Ширина ступеньки a должна быть одной и той же на протяжении всей «лестницы». Получается как бы сплошной кусок однородного стекла, ступенчатая поверхность которого действует как дифракционная решетка. Пусть на эшелон нормально падает пучок параллельных лучей. Разность хода между вторичными волнами Гюйгенса, исходящими под углом ϑ от соседних ступенек эшелона и приходящими в соответствующую точ-

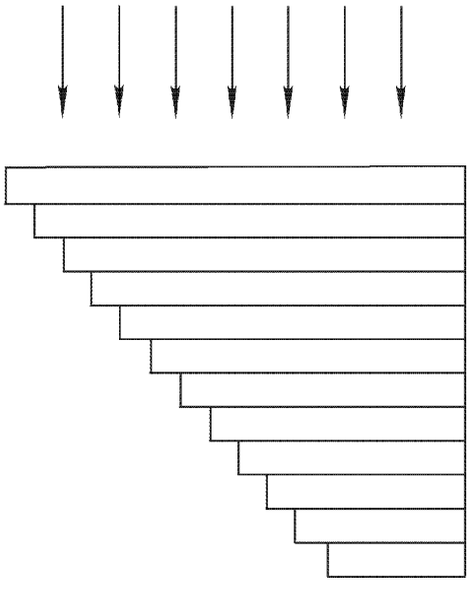


Рис. 194

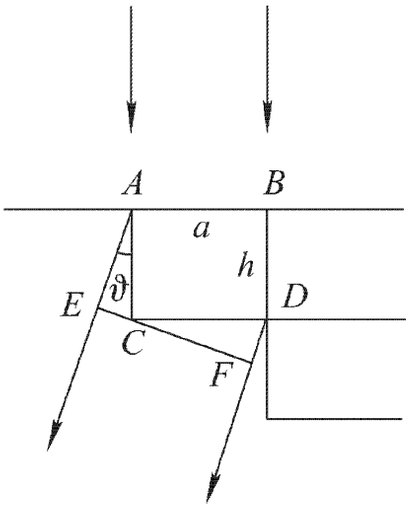


Рис. 195

ку фраунгоферовой дифракционной картины, будет $(BDF) - AE = nh + a\sin\vartheta - h\cos\vartheta$ (рис. 195). Положения главных максимумов определяются условием

$$nh + a\sin\vartheta - h\cos\vartheta = m\lambda, \tag{48.1}$$

где m — целое число.

Из (48.1) дифференцированием находим угловую дисперсию

$$\frac{d\vartheta}{d\lambda} = \frac{m}{a\cos\vartheta + h\sin\vartheta}, \tag{48.2}$$

или, ввиду малости угла ϑ, $d\vartheta/d\lambda = m/a$, или

$$\frac{d\vartheta}{d\lambda} = \frac{h(n-1)}{a\lambda}. \tag{48.3}$$

Так как $h \sim a$, то угловая дисперсия эшелона относительно велика. Поэтому при незначительных изменениях λ получаются заметные изменения угла дифракции ϑ.

Дисперсионная область эшелона определяется выражением

$$\Delta\lambda = \frac{\lambda}{m} = \frac{\lambda^2}{h(n-1)}. \tag{48.4}$$

Она очень мала, и это является недостатком эшелона. Например, при $h = 1$ см, $n = 1{,}5$, $\lambda = 600$ нм получаем $\Delta\lambda \approx 0{,}07$ нм,

что примерно в 10 раз меньше расстояния между компонентами желтой D-линии натрия.

Разрешающая способность дается формулой

$$\frac{\lambda}{\delta\lambda} = Nm = \frac{Nh(n-1)}{\lambda}, \tag{48.5}$$

где N — число пластин. При $N = 40$ в предыдущем примере получим $\lambda/\delta\lambda \approx 3{,}3 \cdot 10^5$. Со штриховой решеткой в спектре первого порядка такой разрешающей способности можно было бы достигнуть при числе штрихов порядка 300 000.

Уточним формулу (48.5), приняв во внимание дисперсию показателя преломления стекла. Пусть при угле дифракции ϑ получается дифракционный максимум m-го порядка для длины волны $\lambda' = \lambda + \delta\lambda$. Тогда

$$hn' + a\sin\vartheta - h\cos\vartheta = m\lambda' = m(\lambda + \delta\lambda),$$

где n' — показатель преломления стекла для той же длины волны. Согласно Рэлею, $\delta\lambda$ будет равна разрешаемому спектральному расстоянию, если для длины волны λ в том же направлении получится первый дифракционный минимум, т.е.

$$hn + a\sin\vartheta - h\cos\vartheta = m\lambda + \lambda/N.$$

Вычитая это соотношение из предыдущего, получим

$$\frac{\lambda}{\delta\lambda} = N\left(m - h\frac{dn}{d\lambda}\right), \tag{48.6}$$

или после подстановки $m = h(n-1)/\lambda$

$$\frac{\lambda}{\delta\lambda} = \frac{Nh}{\lambda}\left[(n-1) - \lambda\frac{dn}{d\lambda}\right]. \tag{48.7}$$

Так как в прозрачной области спектра $dn/d\lambda < 0$ (нормальная дисперсия), то дисперсия показателя преломления повышает разрешающую способность эшелона.

Рассуждая аналогично, легко получить следующие выражения для дисперсионной области и угловой дисперсии эшелона:

$$\Delta\lambda = \frac{\lambda}{m - h(dn/d\lambda)}, \tag{48.8}$$

$$\frac{d\vartheta}{d\lambda} = \frac{m - h(dn/d\lambda)}{a\cos\vartheta + h\sin\vartheta}. \tag{48.9}$$

Дисперсия показателя преломления увеличивает угловую дисперсию эшелона, но уменьшает его дисперсионную область.

Интенсивность света в дифракционной картине, разумеется, определяется общей формулой (46.2). В ней $I_1(\vartheta)$ означает интенсивность, которая получилась бы при дифракции на отдельной ступеньке эшелона ширины a. Функция $I_1(\vartheta)$ такая же, что и при дифракции на щели. Поэтому весь дифрагированный свет

практически концентрируется в интервале углов ϑ между $\vartheta_1 = -\lambda/a$ и $\vartheta_2 = +\lambda/a$. Заметную интенсивность будут иметь только такие главные максимумы, которые попадают в указанный интервал. Ввиду малости угла ϑ можно пренебречь его квадратами, представив формулу (48.1) в виде $(n-1)h + a\vartheta = m\lambda$, или

$$\vartheta = \frac{m\lambda - (n-1)h}{a}. \tag{48.10}$$

Допустим, что показатель преломления таков, что для определенного m числитель в этой формуле обращается в нуль. Тогда в указанный интервал попадут только три главных максимума, порядки которых равны $(m-1)$, m, $(m+1)$. Крайние из них приходятся на концы интервала, их интенсивность равна нулю. Поэтому дифрагированный свет будет состоять только из спектра одного порядка m. Если же числитель в формуле (48.10) не равен нулю, то будут видны спектры двух соседних порядков, вообще говоря, разной интенсивности. Их интенсивности будут одинаковы, когда числитель в формуле (48.10) по абсолютной величине равен $\lambda/2$. Таким образом, возможны такие установки эшелона, при одной из которых получается только один дифракционный максимум, а при другой — два одинаковых симметричных максимума. Чтобы перейти от одной установки к другой, эшелон обычно помещают в герметическую камеру и изменяют в ней давление, а с ним и относительный показатель преломления n, пока не будут достигнуты требуемые для работы условия.

Майкельсону принадлежит также идея использования эшелона как отражательной решетки. Такой эшелон был построен Вильямсом в 1933 г. из пластин плавленого кварца. Тщательно очищенные пластины кварца сажались на оптический контакт и нагревались до температуры, значительно меньшей температуры плавления кварца. В результате пластины прилипали друг к другу настолько прочно, что после охлаждения их удавалось отделить одна от другой только при значительном усилии. Отражающие ступеньки эшелона покрывались алюминием, напыленным в вакууме.

3. В эшелоне Майкельсона интерферирующие пучки возникают в результате дифракции на ступеньках. В чисто интерференционных спектроскопах — интерферометре Фабри–Перо и пластинке Луммера–Герке — они образуются в результате отражений света. Эти приборы уже были рассмотрены в § 36. Получаемый в них спектр определенного порядка есть не что иное, как интерференционная полоса равного наклона. Дисперсионная область, конечно, определяется общей формулой (47.2). Для интерферометра Фабри–Перо порядок спектра m можно вычислить по формуле $m = 2h/\lambda$, где h — расстояние между отражающими поверхностями интерферометра. Для пластинки

Луммера–Герке то же самое можно сделать по формуле (36.7), заменяя в ней ψ предельным углом полного отражения. Это дает

$$m = \frac{2h\sqrt{n^2-1}}{\lambda}. \tag{48.11}$$

Остановимся еще на разрешающей способности интерферометра Фабри–Перо и пластинки Луммера–Герке. В них интерферируют пучки, интенсивность которых медленно убывает с возрастанием номера пучка. Если бы число пучков было бесконечно, как предполагалось при вычислениях в § 36, то спектр содержал бы только одни главные максимумы (см. рис. 141) и никаких добавочных максимумов и минимумов [1]). В этом случае критерий спектрального разрешения Рэлея теряет смысл. Поэтому в § 36 был дан другой критерий. Для разрешающей способности интерферометра Фабри–Перо он приводит к формуле (36.5), имеющей тот же вид, что и формула (47.3). Роль числа интерферирующих пучков N играет величина $N = 2\pi\sqrt{R}/(1-R)$, практически равная $2\pi/(1-R)$. В этом нет ничего неожиданного, так как интуитивно следует ожидать, что убывание интенсивности эквивалентно ограничению числа эффективно действующих пучков без учета их ослабления. Число таких эффективных пучков, очевидно, пропорционально $1/(1-R)$.

Что касается пластинки Луммера–Герке, то для нее в формулу (36.5) необходимо ввести поправку на зависимость показателя преломления стекла от длины волны λ, а также на конечное число интерферирующих пучков. Поэтому мы повторим вывод формулы (36.5) с этим уточнением, считая сначала число интерферирующих пучков бесконечным. Сохраняя обозначения прежнего вывода, пишем, как и раньше, $\delta\Phi = (1-R)/\sqrt{R}$, причем в максимуме m-го порядка $\Phi = 2\pi m$, а во всех остальных точках $\Phi = (4\pi dn\cos\psi)/\lambda$. Однако теперь величины n и ψ имеют разные значения для λ и λ'. Общим для них является только угол падения (так как свет исследуется вне пластинки), а потому $n\sin\psi = n'\sin\psi' = \sin\varphi$. Следовательно, должно быть

$$\frac{\delta\Phi}{\Phi} = \frac{\delta n}{n} - \frac{\sin\psi\,\delta\psi}{\cos\psi} - \frac{\delta\lambda}{\lambda}$$

при условии

$$\frac{\delta n}{n} + \frac{\cos\psi}{\sin\psi}\,\delta\psi = 0.$$

Исключив $\delta\psi$, получим

$$\frac{\delta\Phi}{\Phi} = \frac{1}{\cos^2\psi}\frac{\delta n}{n} - \frac{\delta\lambda}{\lambda}.$$

[1]) Реально они, конечно, есть из-за ограниченности числа интерферирующих пучков.

Подставим сюда значения $\delta\Phi$, $\Phi = 2m\pi$, а также $\delta n = (dn/d\lambda)\delta\lambda$. Тогда

$$\left|\frac{\lambda}{\delta\lambda}\right| = \frac{2\pi\sqrt{R}\,m}{1-R}\left(1 - \frac{1}{\cos^2\psi}\frac{\lambda}{n}\frac{dn}{d\lambda}\right). \tag{48.12}$$

Эта формула выведена в предположении бесконечного числа интерферирующих пучков. Она была бы применима к интерферометру Фабри–Перо, если бы в нем пространство между отражающими зеркалами было заполнено однородной средой, показатель преломления n которой отличен от единицы (для такого интерферометра можно было бы положить $\cos\varphi = 1$). Для пластинки Луммера–Герке формула неприменима из-за ограниченного числа интерферирующих пучков. В этом случае, как ясно из изложенного выше, формулу надо заменить на

$$\left|\frac{\lambda}{\delta\lambda}\right| = Nm\left(1 - \frac{1}{\cos^2\psi}\frac{\lambda}{n}\frac{dn}{d\lambda}\right). \tag{48.13}$$

После подстановки значений N, m и $\cos\psi$ получаем

$$\frac{\lambda}{\delta\lambda} = L\left(\frac{n^2-1}{\lambda} - n\frac{dn}{d\lambda}\right). \tag{48.14}$$

Аналогично, для дисперсионной области

$$\Delta\lambda = \frac{\lambda^2}{2h\sqrt{n^2-1}\left(1 - \dfrac{\lambda n}{n^2-1}\dfrac{dn}{d\lambda}\right)}. \tag{48.15}$$

Из-за дисперсии показателя преломления стекла разрешающая способность пластинки Луммера–Герке возрастает, а дисперсионная область уменьшается.

§ 49. Разрешающая способность призмы

1. Действие призмы как спектрального прибора основано на зависимости показателя преломления вещества от длины волны. Для определения разрешающей способности призмы необходимо учесть дифракцию света на краях диафрагмы или самой призмы, ограничивающих ширину падающего светового пучка. Допустим сначала, что на призму падает монохроматический параллельный пучок лучей, ограниченный диафрагмой AA' (рис. 196). Пусть волновой фронт падающей волны совпадает с плоскостью диафрагмы AA'. Возьмем за призмой произвольный волновой фронт BB'. По определению волнового фронта оптические длины $(ACDB)$ и $(A'C'D'B')$ одинаковы:

$$AC + na + DB = A'C' + nb + D'B', \tag{49.1}$$

где a и b — геометрические длины CD и $C'D'$, проходимые светом в веществе призмы.

Если крайние точки B и B' сместить бесконечно мало вдоль плоскости BB' и соединить их с A и A' бесконечно близкими виртуальными лучами, то в силу принципа Ферма равенство (49.1) сохранится с точностью до членов высших порядков относительно этих боковых смещений. Это приводит к следующему правилу построения волнового фронта. Надо от точек исходного волнового фронта провести действительные или бесконечно близкие к ним виртуальные лучи и отложить на них отрезки одинаковой оптической длины. Геометрическое место концов таких отрезков и даст новое положение волнового фронта. В случае плоских волновых фронтов достаточно ограничиться построением крайних лучей $ACDB$ и $A'C'D'B'$. При построении волновых фронтов, отличающихся бесконечно мало своими направлениями, можно пользоваться одними и теми же световыми путями, хотя истинные пути света и отличаются друг от друга. То же справедливо для световых пучков, бесконечно мало отличающихся длинами волн. Это используется ниже для упрощения вычислений.

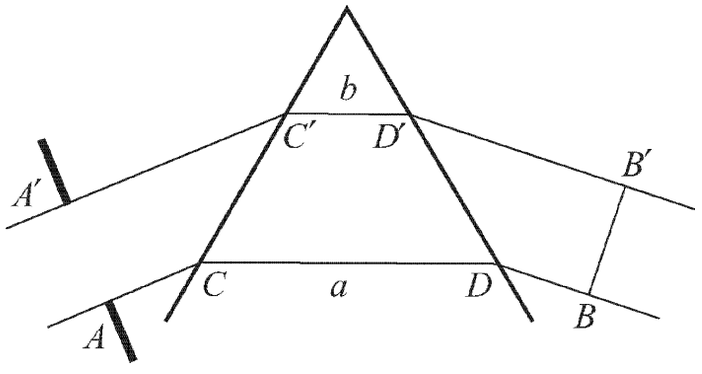


Рис. 196

Учтем теперь дифракцию света на краях диафрагмы AA' и определим за призмой направления на нулевой дифракционный максимум и дифракционные минимумы. Проведем крайние лучи $ACDB$ и $A'C'D'B'$, которые могут быть либо действительными, либо бесконечно близкими к ним виртуальными, и пересечем их произвольной плоскостью BB'. (Мы не предполагаем теперь, что плоскость BB' перпендикулярна к световым лучам.) Если разность оптических длин $(ACDB)$ и $(A'C'D'B')$ равна нулю, то плоскость BB' будет одним из волновых фронтов. Нормаль к ней укажет направление на главный максимум, т.е. максимум нулевого порядка. Если же эти оптические длины отличаются на целое число длин волн, то, как и при дифракции на щели, нормаль к плоскости BB' укажет направление на дифракционный минимум соответствующего порядка. В частности, если оптические длины отличаются на λ, то получится дифракционный минимум первого порядка.

2. Теперь нетрудно вывести формулу для разрешающей способности призмы. Предположим, что на диафрагму падает параллельный пучок света с двумя длинами волн λ и λ'. Пусть плоскость BB' есть волновой фронт для длины волны λ'. Нор-

маль к ней определяет направление на нулевой максимум для длины волны λ'. Пусть то же направление есть направление на дифракционный минимум первого порядка для длины волны λ. Согласно критерию Рэлея разность $\delta\lambda = \lambda' - \lambda$ есть минимальное расстояние, разрешаемое призмой. На основании сказанного выше

$$AC + n(\lambda')a + DB = A'C' + n(\lambda')b + D'B',$$
$$AC + n(\lambda)a + DB = A'C' + n(\lambda)b + D'B' + \lambda.$$

Вычитая почленно, получим

$$(a - b)[n(\lambda') - n(\lambda)] = -\lambda,$$

или

$$(a - b)\frac{dn}{d\lambda}\delta\lambda = -\lambda.$$

Отсюда для разрешающей способности призмы получаем

$$\frac{\lambda}{\delta\lambda} = -(a - b)\frac{dn}{d\lambda}. \tag{49.2}$$

Из вывода ясно, что под λ следует понимать длину волны в вакууме. Предполагается, что дисперсия стекла нормальная, т.е. $dn/d\lambda < 0$.

Разрешающая способность призмы зависит только от дисперсии показателя преломления $dn/d\lambda$ и разности путей a и b, проходимых в призме крайними лучами пучка. Поэтому для полного использования разрешающей способности необходимо, чтобы световой пучок покрывал всю боковую поверхность призмы, на которую он падает. В этом случае $b = 0$, а a означает длину основания призмы. Например, для стеклянной призмы из тяжелого флинта в желтой области спектра $dn/d\lambda \approx -956$ см$^{-1}$. Разрешающая способность призмы при $a = 1$ см в указанной области спектра будет $\lambda/\delta\lambda \approx 956$. Это минимальная разрешающая способность, при которой может быть разрешена двойная D-линия натрия ($\lambda = 589{,}0$ нм, $\lambda' = 589{,}6$ нм).

Для сложной призмы, состоящей из нескольких простых призм, поставленных одна за другой, формула (49.2) должна быть заменена на

$$\frac{\lambda}{\delta\lambda} = -\sum_i (a_i - b_i)\frac{dn_i}{d\lambda}. \tag{49.3}$$

Здесь суммирование производится по всем призмам, составляющим систему.

3. Призму можно рассматривать как предельный случай эшелона Майкельсона (см. рисунки 194 и 195). Допустим, что ширина и высота ступеньки эшелона стремятся к нулю, а общее число ступенек — к бесконечности. При этом последовательно будут исчезать спектры высших порядков, пока не оста-

нется только спектр нулевого порядка. Начиная с этого момента, эшелон будет действовать как призма. При $m = 0$ формула (48.6) перейдет в формулу (49.2), так как Nh есть длина основания призмы. Для дисперсионной области в этом случае формула (48.8) дает $\Delta\lambda = \infty$. Этого и следовало ожидать, так как весь свет, выходящий из призмы, концентрируется в спектре нулевого порядка, а потому вопрос о наложении спектров разных порядков не возникает.

ЗАДАЧИ

1. Какую разрешающую силу должен иметь спектральный аппарат для разрешения дублета D-линии натрия ($\lambda_1 = 589{,}0$ нм, $\lambda_2 = 589{,}6$ нм)? Подсчитать минимальное число штрихов решетки, которая может разрешить его в спектре первого порядка. При каком наименьшем основании a можно сделать то же самое с помощью призмы, изготовленной из стекла с дисперсией $dn/d\lambda = -956$ см$^{-1}$?

О т в е т. 1) $\lambda/\delta\lambda > 1000$; 2) $N \sim 1000$; 3) $a \approx 1$ см.

2. Изменяется ли разрешающая сила решетки при изменении наклона первичного пучка, падающего на нее?

О т в е т. Нет.

3. Изменится ли разрешающая способность и дисперсионная область дифракционной решетки, если, закрепив неподвижно трубу, в которую наблюдаются дифракционные спектры, закрыть через одну щели решетки?

О т в е т. Разрешающая способность не изменится. Дисперсионная область уменьшится вдвое.

4. На плоскую отражательную решетку, содержащую $N = 50\,000$ штрихов, нормально падает свет от двойной линии натрия ($\lambda_1 = 589{,}0$ нм, $\lambda_2 = 589{,}6$ нм). Число штрихов, приходящееся на 1 см длины решетки, равно $n = 5000$. Какой максимальный порядок спектра m можно получить от такой решетки и каково минимальное расстояние $\delta\lambda$ между спектральными линиями, которое способна разрешить решетка в указанной области спектра? Спектр максимального порядка фотографируется на фотопластинке с помощью объектива с фокусным расстоянием $f = 50$ см. Какое расстояние Δx между спектральными линиями λ_1 и λ_2 получится на фотопластинке?

О т в е т. $m = 3$; $\delta\lambda = \dfrac{\lambda}{Nm} = 0{,}004$ нм; $\Delta x = \dfrac{fm\,\delta\lambda}{\sqrt{d^2 - m^2\lambda^2}} = 1$ мм.

5. Как изменится разрешающая способность и дисперсионная область пластинки Луммера–Герке, если 1) ее толщину увеличить вдвое, а длину оставить без изменения; 2) длину увеличить вдвое, а толщину оставить без изменения?

О т в е т. 1) Разрешающая способность не изменится, дисперсионная область уменьшится вдвое. 2) Разрешающая способность увеличится вдвое, дисперсионная область не изменится.

6. Какую минимальную длину должна иметь пластинка Луммера–Герке ($n = 1{,}5$), чтобы разрешить дублетную структуру водородной линии H_α ($\lambda = 656{,}3$ нм)? Разность длин волн линий дублета равна 0,014 нм. Пренебречь величиной $dn/d\lambda$.

О т в е т. $L = \dfrac{\lambda^2}{(n^2 - 1)\,\delta\lambda} = 2{,}5$ см.

7. Какая интерференционная картина будет наблюдаться при освещении монохроматическим светом двух скрещенных пластинок Луммера–Герке, т.е. двух пластинок, расположенных одна за другой вдоль прямой линии, но повернутых друг относительно друга на 90°?

О т в е т. Будет наблюдаться система пятен, расположенных в узлах прямоугольной сетки.

8. Какой должна быть длина a основания стеклянной призмы, чтобы она имела такую же разрешающую способность, как и пластинка Луммера–Герке длиной $L = 20$ см? Показатель преломления пластинки $n = 1{,}5$; дисперсия показателя преломления призмы $dn_{\text{пр}}/d\lambda = -956$ см$^{-1}$; $\lambda = 600$ нм.

О т в е т. $a = -\dfrac{L(n-1)}{\lambda\, dn_{\text{пр}}/d\lambda} = 44$ м.

9. Эшелон Майкельсона состоит из $N = 30$ стеклянных пластинок с показателем преломления $n = 1{,}5$; толщина каждой из них $h = 1$ см. Какова должна быть длина a основания стеклянной призмы, чтобы она имела такую же разрешающую способность, что и рассматриваемый эшелон? Дисперсия показателя преломления призмы $dn_{\text{пр}}/d\lambda = 956$ см$^{-1}$; $\lambda = 600$ нм.

О т в е т. $a = \dfrac{Nh(n-1)}{\lambda\, dn_{\text{пр}}/d\lambda} = 2{,}6$ м.

10. Стеклянная призма с основанием $a = 10$ см изготовлена из тяжелого флинта, дисперсия которого в окрестности $\lambda = 600$ нм равна $dn/d\lambda = 1000$ см$^{-1}$. Какую максимальную разрешающую способность может иметь дифракционная решетка, ширина заштрихованной части которой равна длине основания этой призмы? Сравнить разрешающую способность такой решетки с разрешающей способностью призмы.

Р е ш е н и е. Из формулы решетки $d(\sin\vartheta - \sin\vartheta_0) = m\lambda$ следует $m\lambda \leqslant 2d$. Отсюда умножением на $N = a/d$ получаем для максимальной разрешающей способности решетки

$$R_{\text{реш}} = Nm = \frac{2a}{\lambda} = 3\cdot 10^5 \quad \text{и} \quad \frac{R_{\text{реш}}}{R_{\text{пр}}} = \frac{2/\lambda}{dn/d\lambda} = 33.$$

Из этого примера видно, что при одних и тех же геометрических размерах разрешающая способность решетки, вообще говоря, много больше, чем призмы.

11. Оценить ширину коллиматорной щели, при которой практически полностью используется теоретическая разрешающая способность призмы.

Р е ш е н и е. Предельная разрешающая способность призмы найдется из формулы (49.2), если положить $b = 0$. Пусть на призму падает плоская волна, содержащая длины волн λ и λ', расстояние между которыми $\delta\lambda$ равно минимальному расстоянию, разрешаемому призмой. Разность показателей преломления для этих длин волн $\delta n = (dn/d\lambda)\delta\lambda$. Подставляя сюда $\delta\lambda = \lambda \Big/ \left(a\dfrac{dn}{d\lambda}\right)$, получим $\delta n = \dfrac{\lambda}{a}$. По выходе из призмы первоначально параллельный пучок света из-за дисперсии сделается расходящимся. Рассчитаем угловое расхождение вышедшего пучка. Как видно из рис. 197, $\sin\varphi_1 = n\sin\psi_1$. Дифференцируя при постоянном φ_1, находим: $\delta n\sin\psi_1 + n\cos\psi_1\cdot\delta\psi_1 = 0$. Так как $\psi_1 + \psi_2 = A = \text{const}$ и, следовательно,

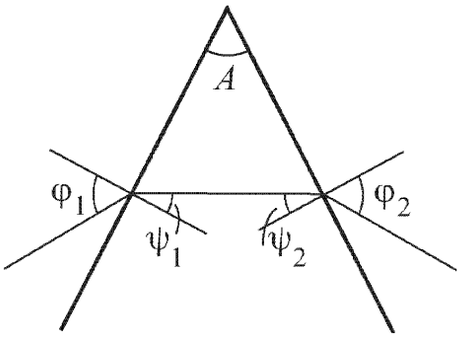


Рис. 197

$\delta\psi_1 + \delta\psi_2 = 0$, то $\delta n \sin\psi_1 = n\cos\psi_1 \cdot \delta\psi_2$. Из закона преломления $\sin\varphi_2 = n\sin\psi_2$ находим искомое угловое расхождение вышедшего пучка:

$$\delta\varphi_2 = \frac{\sin\psi_2}{\cos\varphi_2}\,\delta n + \frac{n\cos\psi_2}{\cos\varphi_2}\,\delta\psi_2 = \left(\frac{\sin\psi_2}{\cos\varphi_2} + \frac{n\cos\psi_2}{\cos\varphi_2}\,\frac{\sin\psi_1}{n\cos\psi_1}\right)\delta n.$$

При работе призма должна быть установлена на угол наименьшего отклонения, а потому $\varphi_1 = \varphi_2 = \varphi$, $\psi_1 = \psi_2 = \psi$,

$$\delta\varphi_2 = 2\,\frac{\sin\psi}{\cos\varphi}\,\delta n = 2\,\frac{\sin\psi}{\cos\varphi}\,\frac{\lambda}{a}.$$

Если h — ширина щели, f — фокусное расстояние коллиматора, то угловая ширина щели будет $\alpha = h/f$. Для полного использования разрешающей способности призмы угол α должен быть мал по сравнению с углом $\delta\varphi_2$. Практически достаточно, чтобы угол α был меньше приблизительно половины угла $\delta\varphi_2$. Это дает

$$h < \frac{\sin\psi}{\cos\varphi}\,\frac{f\lambda}{a}. \tag{49.4}$$

Так как $\psi = A/2$, то

$$h < \frac{\sin(A/2)}{\sqrt{1 - n^2\sin^2(A/2)}}\,\frac{f\lambda}{a}. \tag{49.5}$$

При $a = 10$ см, $A = 60°$, $n = 1{,}73$, $f = 25$ см, $\lambda = 500$ нм последняя формула дает $h < 1{,}2 \cdot 10^{-3}$ мм.

§ 50. Действие спектрального аппарата на световые импульсы

1. В предыдущих параграфах спектральный аппарат рассматривался как *анализатор*, пространственно разделяющий монохроматические волны, уже имеющиеся (согласно теореме Фурье) в световом возмущении, попадающем в этот прибор. Но на него можно также смотреть как на *генератор*, который под действием непериодических возмущений (импульсов) вырабатывает периодические возмущения и пространственно разделяет их. Проще всего такой подход разъяснить на примере дифракционной решетки.

Предположим, что решетка состоит из бесконечного числа равноотстоящих щелей в непрозрачном экране. Пусть на нее нормально падает какое-либо плоское непериодическое возмущение. Его форма для последующих рассуждений не имеет значения. Для наглядности будем изображать его в виде бесконечно короткого прямоугольного импульса (рис. 198). Достигнув в момент t_0 всех точек решетки, импульс возбудит вторичные цилиндрические волны, исходящие от щелей решетки. Вместо одиночного импульса за решеткой получится бесконечное множество следующих друг за другом пространственно разделенных

импульсов с цилиндрическими волновыми фронтами (рис. 199). Сечения волновых фронтов плоскостью чертежа в любой фиксированный момент времени t будут окружностями одного и то-

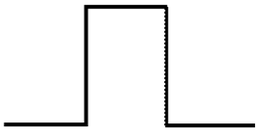


Рис. 198

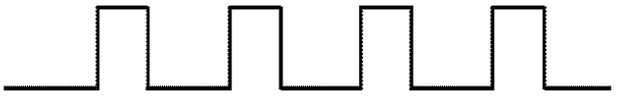


Рис. 199

го же радиуса $c(t-t_0)$, описанными из щелей решетки как из центров (рис. 200). На больших расстояниях от решетки малые участки волновых фронтов могут рассматриваться как плоские и следующие друг за другом через равные промежутки времени

$$\tau = \frac{d\sin\vartheta}{c}. \tag{50.1}$$

На таких расстояниях возмущение представляется периодической функцией $E(t)$, период которой τ зависит от угла ϑ, т.е. от направления излучения. Вместо непериодического импульса решетка создает пространственно разделенные периодические возмущения различных периодов. Исключение составляют импульсы, которые за решеткой распространяются под углом $\vartheta = 0$. В этом случае разложения не будет. Эти периодические возмущения, вообще говоря, не синусоидальны. Но, как и всякие периодические функции, они могут быть разложены в ряд Фурье с основным периодом τ.

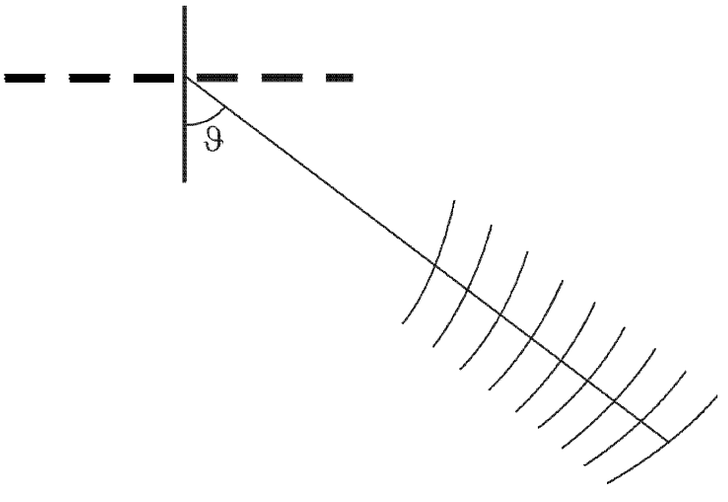


Рис. 200

Для понимания действия решетки такое разложение не обязательно. Но оно целесообразно, когда приемниками излучения в спектроскопах и спектрографах являются глаз, фотографическая пластинка и другие селективные приемники. Они действуют подобно набору гармонических осцилляторов с различными собственными частотами.

Ряд Фурье, в который может быть разложено периодическое возмущение $E(t)$, вообще говоря, будет содержать монохроматическую волну с основным периодом τ. Наряду с ней могут войти гармоники с периодами $\tau/2$, $\tau/3$, ... Не все эти гармоники воспринимаются приемником. В простейшем случае только одна из

них, например гармоника с периодом τ/m, действует на приемник. Обозначим ее период через T. Тогда $T = \frac{\tau}{m} = \frac{d\sin\vartheta}{cm}$, или

$$d\sin\vartheta = m\lambda. \tag{50.2}$$

Это — основная формула дифракционной решетки. Но могут быть и такие случаи, когда разложение Фурье содержит две или больше гармоник, действующих на приемник. Тогда имеет место перекрытие спектров разных порядков. Это, однако, возможно только тогда, когда падающее излучение заполняет спектральный интервал, ширина которого превосходит дисперсионную область решетки.

Рассмотрим теперь воздействие на решетку единичной монохроматической волны с периодом T. Вторичные волны, выходящие в этом случае из щелей решетки, а потому и их суперпозиция будут монохроматическими с тем же периодом T. Значит, монохроматическая волна после прохождения через спектральный аппарат остается монохроматической без изменения периода. Отсюда и из принципа суперпозиции следует, что *в спектре решетки не могут появиться монохроматические составляющие с частотами, отсутствующими в спектре падающего излучения*. Если спектральная область, занимаемая падающим излучением, не превосходит дисперсионную область, перекрытия спектров различных порядков совсем не будет. При этом в случае неограниченной решетки в спектре каждого порядка каждому направлению луча будет соответствовать строго определенная длина волны λ. Наложения волн различных частот не будет. Это значит, что разрешающая способность неограниченной решетки бесконечно велика.

2. Допустим теперь, что длина решетки L и число штрихов N конечны. Тогда периодические возмущения, формирующиеся решеткой, будут состоять из N импульсов, следующих один за другим через время $\tau = (d\sin\vartheta)/c$. Время

$$\Theta = N\tau = \frac{Nd\sin\vartheta}{c} = \frac{L\sin\vartheta}{c} \tag{50.3}$$

приблизительно равно времени от момента появления первого до момента появления N-го импульсов. Оно называется *временем затягивания импульса*, так как за решеткой вместо одиночного короткого импульса получается волновой процесс длительностью Θ. Время Θ называют также *временем установления спектрального аппарата*, так как по прошествии такого времени начинают действовать все N штрихов решетки. Волновое возмущение в точке наблюдения может быть представлено в виде

$$E(t) = f_1(t) + f_2(t) + \ldots + f_N(t),$$

где слагаемые в правой части представляют импульсы, приходящие от отдельных штрихов. Так как эти импульсы запаздывают

один относительно другого на время τ, а в остальном совершенно тождественны, то $f_2(t) = f_1(t-\tau)$, $f_3(t) = f_1(t-2\tau)$ и т.д. Результирующее возмущение представится в виде

$$E(t) = f_1(t) + f_1(t-\tau) + \ldots + f_1[t-(N-1)\tau].$$

Разложим его в интеграл Фурье

$$E(t) = \int_0^\infty A(\omega) e^{i\omega t}\, d\omega, \quad \text{где} \quad A(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} E(t) e^{-i\omega t}\, dt.$$

Подставив сюда выражение для $E(t)$ и учтя, что $f_1(t-\tau) = f_1(t)$, получим

$$A(\omega) = \frac{1}{2\pi}[1 + e^{-i\omega\tau} + \ldots + e^{-i\omega(N-1)\tau}] \int_{-\infty}^{+\infty} f_1(t) e^{-i\omega t}\, dt = \\ = F(\omega)\, \frac{\sin(N\omega\tau/2)}{\omega\tau/2}, \quad (50.4)$$

где введено обозначение

$$F(\omega) = \frac{1}{2\pi}\, e^{-i(N-1)\omega\tau/2} \int_{-\infty}^{+\infty} f_1(t) e^{-i\omega t}\, dt.$$

Величина $|F(\omega)|^2\, d\omega$ с точностью до постоянного множителя представляет интенсивность падающего импульса, приходящуюся на интервал частот $(\omega, \omega + d\omega)$. Такой же смысл имеет величина $|A(\omega)|^2\, d\omega$ для излучения, выходящего из решетки под углом ϑ. Теперь дифрагированное излучение, распространяющееся от решетки в определенном направлении, состоит не из одной монохроматической волны, как было в случае неограниченной решетки, а непрерывно распределено по спектру частот. Главные максимумы в этом распределении приходятся на частоты, при которых знаменатель в выражении (50.4) обращается в нуль, т.е. $\omega\tau/2 = m\pi$. Это приводит к соотношению (50.2). Главные максимумы не появятся, если $F(\omega) = 0$, т.е. когда в падающем импульсе не представлено излучение соответствующей частоты. Ближайший минимум в распределении интенсивности по частотам приходится на частоту, определяемую условием $N\omega\tau/2 = Nm\tau + \pi$, так как при такой частоте числитель в формуле (50.4) обращается в нуль. Величина

$$\delta\omega = \frac{2\pi}{N\tau} = \frac{2\pi}{\Theta} \quad (50.5)$$

определяет порядок ширины спектральной области, которую заполняют волны, выходящие из решетки в одном и том же направлении. По своему содержанию соотношение (50.5), очевид-

но, совпадает с ранее выведенной и подробно проанализированной формулой (29.8).

Допустим теперь, что возмущение, падающее на решетку, состоит из одной только неограниченной синусоиды. Тогда частота ω в формуле (50.4) будет фиксирована. Но угол ϑ может меняться, а формула (50.4) определит зависимость амплитуды дифрагированного света от этого угла. Так как $\omega\tau = (2\pi/\lambda)d\sin\vartheta$, то эта зависимость представится формулой

$$A(\vartheta) = F(\vartheta)\frac{\sin N\dfrac{\pi d\sin\vartheta}{\lambda}}{\sin\dfrac{\pi d\sin\vartheta}{\lambda}}, \tag{50.6}$$

тождественной с формулой (46.1). Поэтому отпадает надобность в дальнейших рассуждениях. Заметим только, что разрешающая способность R связана с временем установления Θ соотношением

$$R = \Theta/T. \tag{50.7}$$

3. По описанной схеме действует всякий спектральный прибор. Это совершенно очевидно для интерференционных спектральных приборов, но не столь очевидно для призмы. В случае призмы время распространения света от волнового фронта AB (рис. 201) до другого волнового фронта $A'B'$ одно и то же для

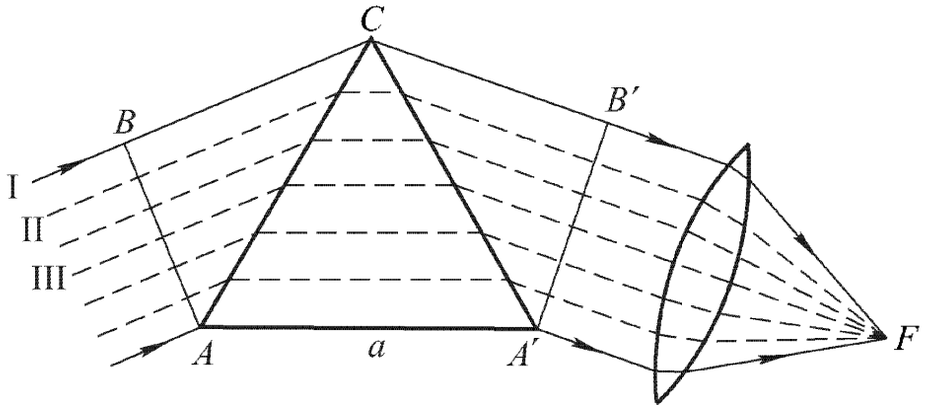


Рис. 201

всех лучей. Однако это заключение относится к *фазе волны* и из него не следует, что световой импульс придет в точку наблюдения целиком (что было бы, если бы вещество призмы не обладало дисперсией), а не разобьется на ряд следующих друг за другом тождественных импульсов. В диспергирующей среде импульсы, вышедшие одновременно от различных точек волнового фронта AB, придут в соответствующие точки волнового фронта $A'B'$ не одновременно. Импульс растянется во времени, и можно говорить о времени затягивания Θ.

Наиболее просто работу призмы можно проанализировать, когда падающий на нее импульс представляет группу волн, т.е. заполняет узкую область частот. Этого предположения достаточно для анализа наиболее важного вопроса — о разрешающей способности призмы. Как известно (см. § 8), при распространении в диспергирующей среде группа непрерывно деформируется, периодически восстанавливая свою форму через время восстановления $\tau = d\lambda/dv$. Разобьем падающий пучок на ряд пучков I, II, III, ... равноотстоящими плоскостями, параллельными лучам самого пучка. Разбиение произведем так, чтобы каждый пучок приходил в фокус F с запаздыванием τ по отношению к предыдущему пучку. Тогда волновые импульсы, связанные с рассматриваемыми пучками, по выходе из призмы будут иметь одну и ту же форму. Через фокус F последовательно будут проходить тождественные световые импульсы, следующие друг за другом через одно и то же время τ. Призма преобразовала падающий на нее импульс в последовательность равноотстоящих импульсов, т.е. в периодическое возмущение, а это и требуется от спектрального аппарата. Время затягивания, очевидно, равно $\Theta = AA'/u - BCB'/c$. По определению волнового фронта $BCB'/c = AA'/v = a/v$. Здесь v означает фазовую, а u — групповую скорости света в веществе призмы. Таким образом,

$$\Theta = a\left(\frac{1}{u} - \frac{1}{v}\right) = a\,\frac{v-u}{uv},$$

или на основании формулы (8.8)

$$\Theta = -a\,\frac{\dfrac{\lambda}{n}\dfrac{dn}{d\lambda}}{v\left(1+\dfrac{\lambda}{n}\dfrac{dn}{d\lambda}\right)} = -\frac{a}{v}\,\frac{\lambda\,dn}{d(n\lambda)}. \tag{50.8}$$

Поделив это выражение на T, найдем разрешающую способность призмы:

$$R = -a\,\frac{dn}{d(n\lambda)}. \tag{50.9}$$

Здесь λ означает длину волны в призме, тогда как в формуле (49.2) той же буквой обозначена длина волны в вакууме. Обозначив последнюю через λ_0, получим

$$R = -a\,\frac{dn}{d\lambda_0}, \tag{50.10}$$

что по содержанию совпадает с формулой (49.2). С изложенной точки зрения особенно ясно действие системы призм, разрешающая способность которой выражается формулой (49.3).

§ 51. Вогнутая отражательная решетка

1. Свет, дифрагировавший на обычной плоской решетке с равноотстоящими штрихами, должен направляться на линзу или другое фокусирующее устройство, чтобы в фокальной плоскости получилась картина резких спектральных линий. Однако можно обойтись и без линзы, если штрихи на плоской поверхности нанести неравномерно, надлежащим образом монотонно уменьшая или увеличивая расстояние между ними. Можно также придать поверхности решетки вогнутую сферическую форму и нанести на нее равноотстоящие штрихи.

Фокусирующее действие многих решеток было замечено давно, но до работ Роулэнда оно рассматривалось как недостаток решетки. Роулэнд превратил его в достоинство решетки, построив *вогнутые отражательные решетки*. Наряду с интерферометром Фабри–Перо такие решетки до настоящего времени являются основными спектральными аппаратами, широко используемыми в точных спектроскопических исследованиях. Спектральные линии, получаемые с помощью вогнутых отражательных решеток, по своей резкости значительно превосходят те же линии в решетках, где фокусировка осуществляется с помощью линз.

2. Пусть G_0G (рис. 202) — отражающая сферическая поверхность радиуса ρ с центром в точке C, на которой нанесены штрихи равноотстоящими плоскостями, перпендикулярными к плоскости рисунка и параллельными диаметру OC, а d — расстояние между этими плоскостями (период решетки). Поместив источник света в точке A_0, выясним, при каких условиях спектр m-го порядка можно сфокусировать в заданной точке A, лежащей в плоскости A_0OC. При исследовании ограничимся рассмотрением лучей, лежащих в той же плоскости, приняв ее за координатную плоскость XY. Начало координат поместим в центре решетки O, ось X направим вдоль диаметра OC, ось Y — по касательной к сфере, лежащей в рассматриваемой плоскости. Пусть $P(x, y)$ — точка на поверхности решетки, находящаяся у какого-либо штриха. Введем обозначения: $A_0O = r_0$, $AO = r$, $A_0P = u_0$, $AP = u$. Для фокусировки в точке A спектра m-го порядка необходимо и достаточно, чтобы разность хода между лучами, приходящими от соседних штрихов, составляла $m\lambda$. Предполагая (что не влияет на результат), что начало O находится также у штриха, запишем это условие в виде $u_0 + u = r_0 + r \pm nm\lambda$, где n — номер штриха P. При описанном выше способе нанесения штрихов $n = \pm y/d$, а потому при любом y должно быть

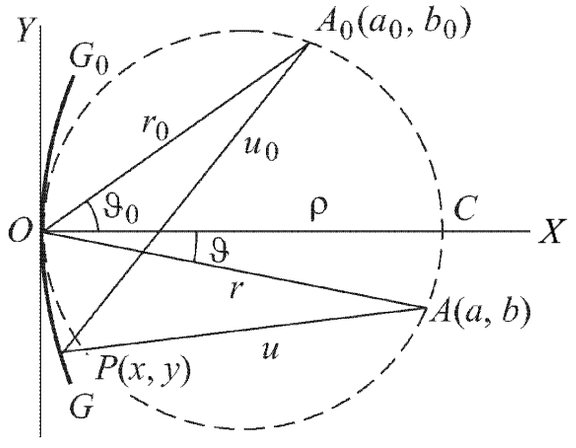


Рис. 202

$$u_0 + u = r_0 + r \pm \frac{m\lambda}{d} y. \tag{51.1}$$

Обозначим через a_0, b_0 координаты точки A_0; a, b — координаты точки A. Тогда

$$u_0^2 = (y - b_0)^2 + (x - a_0)^2 = r_0^2 + x^2 + y^2 - 2b_0y - 2a_0x. \tag{51.2}$$

Уравнение окружности, по которой поверхность решетки пересекается с плоскостью рисунка, имеет вид

$$x^2 + y^2 = 2\rho x. \tag{51.3}$$

Подставив отсюда $2x$ в предыдущее уравнение и использовав соотношение $r_0^2 - b_0^2 = a_0^2$, найдем

$$u_0^2 = \left(r_0 - \frac{b_0 y}{r_0}\right)^2 + \left(\frac{a_0}{r_0^2} - \frac{1}{\rho}\right) a_0 y^2 + \left(1 - \frac{a_0}{\rho}\right) x^2. \tag{51.4}$$

Второй член — величина второго, а последний — четвертого порядка малости по y. Сохраняя только члены второго порядка, получим

$$u_0 = r_0 - \frac{b_0 y}{r_0} + \frac{1}{2r_0}\left(\frac{a_0}{r_0^2} - \frac{1}{\rho}\right) a_0 y^2. \tag{51.5}$$

Аналогично,

$$u = r - \frac{by}{r} + \frac{1}{2r}\left(\frac{a}{r^2} - \frac{1}{\rho}\right) a y^2. \tag{51.6}$$

Подставим эти выражения в формулу (51.1) и приравняем нулю члены с первыми и вторыми степенями y. Тогда получим два условия:

$$\frac{b_0}{r_0} + \frac{b}{r} = \mp\frac{m\lambda}{d}, \tag{51.7}$$

$$\frac{a_0}{r_0}\left(\frac{a_0}{r_0^2} - \frac{1}{\rho}\right) + \frac{a}{r}\left(\frac{a}{r^2} - \frac{1}{\rho}\right) = 0, \tag{51.8}$$

при выполнении которых, по крайней мере с рассматриваемой степенью точности, решетка будет обладать требуемыми фокусирующими свойствами.

3. Условие (51.7) определяет *направления на дифракционные максимумы* m-го порядка. Действительно, если ϑ_0 — угол падения центрального луча на решетку, а ϑ — угол между соответствующим дифрагированным лучом и нормалью к ней (рис. 202), то $\sin\vartheta_0 = b_0/r_0$, $\sin\vartheta = b/r$, и уравнение (51.7) принимает вид

$$d(\sin\vartheta + \sin\vartheta_0) = \pm m\lambda, \tag{51.9}$$

такой же, что и при дифракции на плоской отражательной решетке.

Второе условие, (51.8), определяет расстояние r точки A от центра решетки O. Вместе с уравнением (51.9) оно, таким образом, однозначно определяет положение этой точки, если известно положение источника света A_0. Используя соотношения $a_0 = r_0\cos\vartheta_0$ и $a = r\cos\vartheta$, уравнение (51.8) можно представить в виде

$$\frac{\cos^2\vartheta_0}{r_0} + \frac{\cos^2\vartheta}{r} = \frac{\cos\vartheta_0 + \cos\vartheta}{\rho}. \tag{51.10}$$

При малых углах ϑ_0 и ϑ получается общеизвестное уравнение вогнутого зеркала радиуса ρ. Таким образом, каково бы ни было положение источника света A_0, всегда существует такая точка A, в которой приближенно собираются дифрагированные лучи. Однако можно достигнуть большей точности и простоты в установке решетки, а также лучшей резкости спектральных линий, если распорядиться выбором точки A_0 так, чтобы в выражении

(51.4) исчезли члены второго порядка. Для этого должно быть $\frac{a_0}{r_0^2} - \frac{1}{\rho} = 0$. Тогда, ввиду (51.8), будет $\frac{a}{r^2} - \frac{1}{\rho} = 0$. Эти соотношения можно записать в виде

$$r_0^2 = a_0\rho, \quad \text{или} \quad r_0 = \rho\cos\vartheta_0, \tag{51.11}$$

$$r^2 = a\rho, \quad \text{или} \quad r = \rho\cos\vartheta. \tag{51.12}$$

Соотношение (51.11) означает, что источник света должен лежать на окружности диаметра ρ, касающейся поверхности решетки в точке O. Из (51.12) тогда следует, что дифрагированные лучи соберутся в точке A, лежащей на той же окружности. Эта окружность называется *кругом Роулэнда*. На рис. 202 она изображена штриховой линией.

Угловая дисперсия решетки $d\vartheta/d\lambda = m/(d\cos\vartheta)$ не будет зависеть от угла ϑ, если спектры получаются при малых ϑ вблизи $\vartheta = 0$. Тогда $\cos\vartheta \approx 1$, угол ϑ будет линейной функцией длины волны, т.е. получится нормальный спектр. Поэтому в обычных условиях решетка устанавливается так, чтобы изучаемый спектр получался вблизи ее главной оптической оси.

Большие решетки в первом порядке имеют разрешающую способность около 200 000. Угловое расстояние между спектральными линиями, еще разрешаемыми решеткой, составляет $m\delta\lambda/(d\cos\vartheta) = m\lambda/(Rd\cos\vartheta)$, а линейное

$$\Delta x = \frac{m}{d\cos\vartheta}\frac{\rho\lambda}{R} \approx \frac{m}{d}\frac{\rho\lambda}{R}. \tag{51.13}$$

Для эффективного использования разрешающей способности решетки необходимо, чтобы величина Δx была не меньше определенного предела. Положим $\Delta x = 0{,}01$ мм, $m = 1$, $R = 200\,000$, $\lambda = 500$ нм, $d = 1/600$ мм. Тогда по предыдущей формуле найдем, что радиус кривизны решетки ρ должен быть не меньше ≈ 7 м, т.е. очень велик.

4. Вычисления, приведенные выше, применимы только к лучам, лежащим в плоскости рисунка. Более подробное исследование показывает, что лучи, выходящие из этой плоскости, дают *астигматические изображения*: точечный источник, помещенный в A_0, изображается в виде отрезка, перпендикулярного к плоскости рисунка и проходящего через точку A. Астигматизм увеличивается по мере возрастания угла падения ϑ_0, т.е. с увеличением порядка спектра. Когда источником света является щель, то наличие астигматизма ведет к уменьшению резкости изображения и понижению разрешающей способности, если щель установлена не совсем параллельно штрихам решетки. Поэтому необходимо добиваться тщательной установки щели на параллельность штрихам решетки. По той же причине нельзя делать штрихи решетки очень длинными. Обычно наносят короткие штрихи вдоль широких участков, а не стремятся покрыть ими большую площадь, как в случае плоской решетки, не обладающей астигматизмом.

5. Роулэнду принадлежит и теоретически наиболее совершенный способ установки дифракционной решетки, показанный на схематическом рис. 203. Решетка G_0G и фотографическая камера устанавливаются на противоположных концах твердого стержня OA, длина которого равна радиусу кривизны решетки ρ. Они могут скользить вдоль взаимно перпендикулярных рельсов A_0O и A_0A на салазках, шарнирно соединенных со стержнем OA. Для достижения лучшей отчетливости изображения фотопластинку P_0P изгибают по окружности, совпадающей с кругом Роулэнда. Щель помещается в точке пересечения A_0 направляющих рельсов. При фотографировании спектров разных порядков, а также различных участков одного и того

же спектра щель A_0 и источник света L остаются неподвижными. Так как точки O, A_0, A лежат на круге Роулэнда, то при надлежащем положении стержня OA изображение щели A_0 получится в точке A. При перемещении концов стержня вдоль направляющих рельсов через точку A будут последовательно проходить фокусы различных длин волн в спектре рассматриваемого порядка. Так как в установке Роулэнда $\vartheta = 0$, то $d \sin \vartheta_0 = m\lambda$. С другой стороны, $OA_0 = \rho \sin \vartheta_0$. Комбинируя эти две формулы, находим

$$OA = \frac{\rho m}{d} \lambda. \qquad (51.14)$$

На стержне, связанном с рельсом OA_0, можно нанести равномерную шкалу длин волн. Тогда длина волны, попадающей в центр поля зрения A, может быть отсчитана непосредственно по этой шкале.

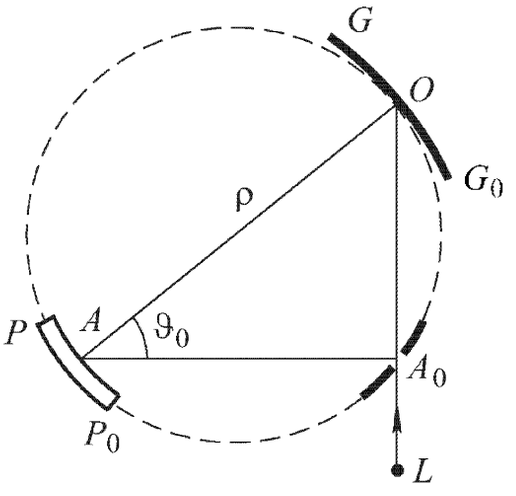


Рис. 203

При установке решетки по способу Роулэнда объем, занимаемый приборами, настолько велик, что в нем трудно поддерживать постоянство температуры, необходимое при длительных экспозициях. Кроме того, установка Роулэнда довольно дорога. По этим и по ряду других причин она применяется редко. Пашеном, Иглом и др. были разработаны другие способы; с ними можно ознакомиться по специальным руководствам по технике спектроскопии.

§ 52. Дифракция на решетке как краевая задача

1. В предыдущих параграфах волновое поле за решеткой вычислялось путем *суммирования волн*, исходящих от штрихов решетки. Для некоторых целей более предпочтителен другой способ. Допустим сначала, что решетка бесконечна. Переднюю поверхность ее будем называть *входом*, заднюю — *выходом*. Эта терминология применима и для отражательной решетки. Для нее входом и выходом служит одна и та же (передняя) поверхность. Без потери общности можно рассуждать так, как если бы решетка была бесконечно тонкой. Примем плоскость решетки за координатную плоскость XY. Ось X направим перпендикулярно к штрихам, а ось Z — в сторону распространения дифрагированного света (рис. 204). Как и раньше, отвлечемся от векторного характера волнового поля, считая его скалярным. Это всегда можно сделать, если электрический или магнитный вектор падающей волны параллелен штрихам решетки. Надо только в первом случае световое поле характеризовать электрическим, а во втором — магнитным вектором. В общем случае падающую волну можно разложить на две составляющие, из которых одна поляризована в плоскости падения, а другая пер-

пендикулярно к ней. Тогда задача о дифракции волны сведется к двум независимым задачам, отличающимся состоянием поляризации падающего света.

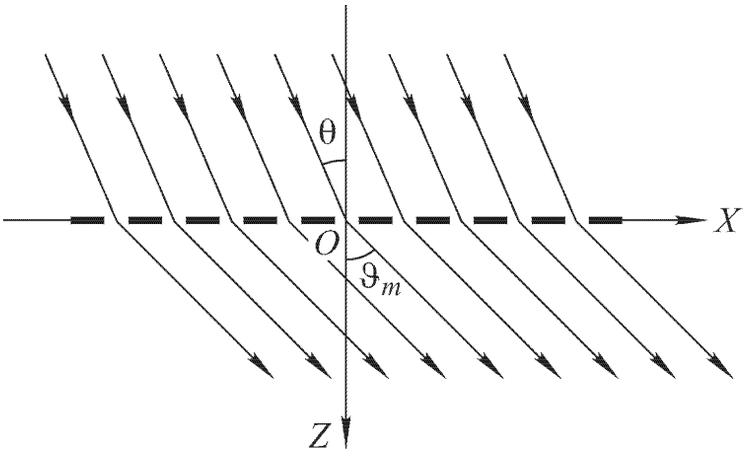


Рис. 204

Падающую волну представим в виде

$$E_0 = Ae^{i(\omega t - \mathbf{kr})}. \tag{52.1}$$

Полагая здесь $z = 0$, найдем поле $E_{\text{вх}}$ на входе решетки. В силу линейности уравнений, определяющих распространение волн, суммарное поле прошедшей (отраженной) и дифрагированных волн на выходе можно представить в виде $E_{\text{вых}} = DE_{\text{вх}}$. Коэффициент D, вообще говоря, зависит от длины волны и называется *пропускаемостью* или *амплитудной прозрачностью решетки.* Это есть характеристика *только самой решетки.* Для двумерной решетки величина $D = D(x, y)$ есть функция двух координат, для одномерной — одной координаты: $D = D(x)$.

Задача о дифракции на всякой решетке сводится к нахождению амплитудной прозрачности последней. Действительно, знание этой величины позволяет определить поле на выходе решетки, а этого достаточно, чтобы по принципу Гюйгенса–Френеля вычислить поле дифрагированной волны. Для нахождения самого решения, конечно, не обязательно пользоваться принципом Гюйгенса–Френеля. Достаточно найти решение волнового уравнения

$$\Delta E + k^2 E = 0, \tag{52.2}$$

которое в плоскости $z = 0$ переходило бы в функцию $E_{\text{вых}}(x, y) = D(x, y)E_{\text{вх}}(x, y)$ и удовлетворяло некоторым дополнительным требованиям, определяющимся физическими условиями задачи. Такой метод в литературе известен под названием *метода Рэлея*, хотя он и является только частным случаем общего *метода краевых задач* математической физики. Здесь этот метод применяется к одномерной дифракционной решетке.

2. Поскольку для решетки D есть периодическая функция x с периодом d, ее можно представить рядом Фурье

$$D = \sum_{m=-\infty}^{+\infty} D_m e^{-impx} \qquad \left(p = \frac{2\pi}{d}\right). \tag{52.3}$$

Поэтому

$$E_{\text{вых}} = DE_{\text{вх}} = \sum_{m=-\infty}^{m=+\infty} AD_m e^{i[\omega t - (k_x + mp)x]}. \tag{52.4}$$

На это соотношение надо смотреть как на *граничное условие*, которому должно удовлетворять за решеткой решение уравнения (52.2). Частным решением этого уравнения является плоская волна $E = ae^{i(\omega t - \mathbf{qr})}$, волновой вектор $\mathbf{q}$ которой удовлетворяет условию $\mathbf{q}^2 = k^2$. Он может быть и вещественным (однородная волна), и комплексным (неоднородная волна). Суперпозицией таких решений можно получить общее решение:

$$E = \sum_{m=-\infty}^{+\infty} a_m e^{i(\omega t - q_m \mathbf{r})}. \tag{52.5}$$

Чтобы при этом удовлетворить граничному условию (52.4), необходимо положить

$$q_{mx} = k_x + mp, \quad q_{my} = 0. \tag{52.6}$$

Тогда $q_{mz} = \sqrt{k^2 - q_{mx}^2}$. Двузначность квадратного корня устраняется физическими соображениями. Вещественным значениям q_{mz} соответствуют однородные волны, которые могут только *уходить от решетки*, а не приходить к ней. Этому условию удовлетворяют только положительные значения q_{mz}. Мнимым соответствуют неоднородные волны. При удалении от решетки они могут *только затухать*. Поэтому коэффициент при мнимой части должен быть отрицательным. Таким образом, соотношение (52.6) надо пополнить условиями

$$q_{mz} = +\sqrt{k^2 - q_{mx}^2} \quad \text{и} \quad q_{mz} = -i\sqrt{k^2 - q_{mx}^2}, \tag{52.6а}$$

в которых квадратные корни понимаются в *арифметическом смысле*. После этого волновые векторы $\mathbf{q}_m$ определяются однозначно, а для поля на выходе можно написать

$$E_{\text{вых}} = \sum_{-\infty}^{+\infty} a_m e^{i(\omega t - q_{mx} x)}.$$

Сравнение этого выражения с (52.4) дает

$$a_m = AD_m, \tag{52.7}$$

чем и завершается решение задачи.

3. Полагая в (52.6) $k_x = (2\pi/\lambda)\sin\theta$, $q_{mx} = (2\pi/\lambda)\sin\vartheta_m$, $p = 2\pi/d$, получим

$$d(\sin\vartheta_m - \sin\theta) = m\lambda, \tag{52.8}$$

т.е. основную формулу дифракционной решетки. Однако в полученном решении содержатся и новые результаты. Во-первых, весь спектр за решеткой состоит из *одних только главных максимумов*, никаких второстепенных максимумов нет. Это, конечно, объясняется тем, что полученные результаты относятся к неограниченной решетке. Во-вторых, в формулу (52.5), наряду с однородными, входят и *неоднородные волны*, а потому она определяет волновое поле *на любых расстояниях от решетки*, а не только в волновой зоне, как было в случае формулы (46.2). При нормальном падении света наивысший порядок однородных волн m определяется условием $m \leqslant d/\lambda$. При $m > d/\lambda$ составляющая q_{mz} мнимая. Соответствующая неоднородная волна экспоненциально затухает по закону $\exp(-\varkappa_m z)$, где

$$\varkappa_m = \sqrt{k^2 - q_{mz}^2} = \frac{2\pi}{d}\sqrt{m^2 - (d/\lambda)^2}. \tag{52.9}$$

На расстоянии $z = 1/\varkappa_m$ амплитуда волны убывает в e раз, ее квадрат — в e^2 раз. Наибольшее из этих расстояний порядка d. Поэтому при $z \gg d$ поле за решеткой состоит только из однородных волн.

В частности, если $d < \lambda$, то при нормальном падении поле вдали от решетки состоит только из одной плоской волны нулевого порядка ($m = 0$). Никаких боковых дифракционных максимумов не появляется. Исследование прошедшей или отраженной волны при $z \gg \dfrac{\lambda d}{2\pi\sqrt{\lambda^2 - d^2}}$, или практически на расстояниях $z \gg \lambda$, не даст никакой информации о структуре решетки. Рассматривая решетку в микроскоп, мы не обнаружим на ней штрихов. По той же причине свет, распространяющийся в прозрачной среде, например в кристалле, не отклоняется в стороны, хотя среда и построена из дискретных атомов или молекул — кристалл является периодической структурой и в этом смысле может рассматриваться как дифракционная решетка (фактически наблюдаемое рассеяние света вызывается неоднородностями среды и тепловыми флуктуациями). В оптическом отношении кристалл ведет себя как однородная сплошная среда. Причина этого в том, что период молекулярной структуры кристалла меньше длины волны. Но как только нарушится это условие, появятся дифракционные максимумы. Возникнет интерференционное рассеяние волн. Такой случай реализуется в рентгеновской области спектра.

4. В непосредственной близи от решетки волновое поле имеет очень сложный вид. На расстояниях $z \gg 1/\varkappa_m$ поле становится более простым, поскольку из него выпадают неоднородные волны. Однако картина поля все еще довольно сложна, так как однородные волны, из которых оно состоит, накладываются друг на друга. При удалении от решетки на некоторое минимальное расстояние плоские волны различных порядков пространственно разделяются. На сравнительно малых расстояниях от решетки к пространственно разделенным волнам различных порядков можно применять геометрическую оптику. Направления на главные максимумы излучения определяются требованием, чтобы все волны, исходящие из различных щелей решетки, при интерференции усиливали друг друга. При этом явно или неявно всегда имеется в виду, что интерференция происходит либо в фокусе собирающей линзы, либо на бесконечном расстоянии от решетки. Однако поскольку дифрагированное поле всюду представляется суперпозицией плоских волн, условие интерференционного усиления определяет также направление распространения таких плоских волн на любых расстояниях от решетки.

5. Если в разложении пропускаемости D в ряд Фурье отсутствует какой-либо член, то будет отсутствовать и главный максимум соответствующего порядка в волновом поле за решеткой. Рассмотрим в качестве примера синусоидальную решетку, называемую также решеткой Рэлея. Ее пропускаемость определяется выражением

$$D = \alpha + \beta \cos px = \alpha + \frac{\beta}{2}\,(e^{ipx} + e^{-ipx})$$

с постоянными a, β и p. В этом случае световое поле за решеткой состоит из трех плоских однородных волн: неотклоненной волны (максимум нулевого порядка) с амплитудой $a_0 = \alpha A$ и двух боковых максимумов первого порядка с амплитудами $a_1 = a_{-1} = {}^1\!/_2\beta A$. Других главных максимумов не будет. Если поверхности тела придать форму синусоиды, амплитуда которой мала по сравнению с длиной волны, то такая поверхность приближенно будет вести себя как решетка Рэлея. Это удалось сделать Вуду путем травления (примерно в течение 10 с) поверхности стеклянной дифракционной решетки слабым раствором фтористоводородной кислоты. Обычные стеклянные решетки дают очень слабые спектры, так как штрихи, нанесенные алмазом, узки по сравнению со светлыми промежутками между ними. Если штрихи расширить, смочив поверхность решетки слабой фтористоводородной кислотой, то удается достигнуть увеличения интенсивности спектров в несколько раз (до десяти).

Возможны и такие структуры, которые формально могут рассматриваться как дифракционные решетки, дающие только

один максимум. Примером может служить призма. Если на одну из боковых граней призмы падает плоская монохроматическая волна, то фаза волнового поля на выходе (на второй грани призмы) будет меняться с координатой по линейному закону. Соответствующее граничное условие удовлетворяется только одной уходящей плоской волной. Это и есть волна, прошедшая через призму и отклоненная ею. Проблема отражения и преломления света на плоской границе раздела двух сред может решаться как краевая задача по методу Рэлея. При таком подходе плоская граница, раздела ведет себя как дифракционная решетка, дающая только один максимум.

6. Ограниченная решетка может быть рассмотрена также методом Рэлея. В пределах решетки ее пропускаемость периодична, а вне обращается в нуль. Разложение пропускаемости в интеграл Фурье имеет вид

$$D(x) = \int\limits_{-\infty}^{+\infty} C(f) e^{-\imath f x}\, df. \tag{52.10}$$

Поле на выходе решетки представится выражением

$$E_{\text{вых}} = A \int\limits_{-\infty}^{+\infty} C(f) e^{-f(k_x+f)x}\, df.$$

Решение, удовлетворяющее этому и всем прочим условиям задачи, имеет вид

$$E = A \int\limits_{-\infty}^{+\infty} C(f) e^{-\imath \mathbf{qr}}\, df, \tag{52.11}$$

где $q_x = k_x + f$, $q_y = 0$, а q_z определяется прежними формулами (52.6а). Зададим теперь пропускаемость решетки выражением

$$D(x) = \begin{cases} 0, & \text{если } -\infty < x < -L, \\ \sum\limits_{m=-\infty}^{+\infty} D_m e^{-impx}, & \text{если } -L < x < +L, \\ 0, & \text{если } +L < x < +\infty. \end{cases} \tag{52.12}$$

Вычисляя коэффициент Фурье $C(f)$, получим

$$C(f) = \frac{1}{2\pi} \int\limits_{-\infty}^{+\infty} D(x) e^{ifx}\, dx = \frac{1}{\pi} \sum_{m=-\infty}^{+\infty} D_m \frac{\sin[L(f-mp)]}{f-mp}.$$

В результате поле за решеткой представится в вида

$$E = \sum E_m, \tag{52.13}$$

где

$$E_m = \frac{A}{\pi}\int\limits_{-\infty}^{+\infty} D_m \frac{\sin[L(f-mp)]}{f-mp}\, e^{-i\mathbf{qr}}\, df =$$

$$= \frac{Ad}{\pi}\int\limits_{-\infty}^{+\infty} D_m \frac{\sin\left[(N/a)(fd-2m\pi)\right]}{fd-2m\pi}\, e^{-i\mathbf{qr}}\, df. \quad (52.14)$$

По-прежнему волновое поле за решеткой складывается из волн различных порядков m. Однако волна E_m уже не сводится к единственной плоской волне определенного направления, а состоит из суперпозиции *непрерывного множества* плоских волн различных направлений. Распределение интенсивности по направлениям в отдельной волне E_m определяется той же кривой, что и в волне, дифрагированной на щели (см. рис. 175). Интенсивность достигает максимума, когда знаменатель обращается в нуль, т.е. когда $q_x = k_x + mp = 0$. Это условие определяет направление на главный максимум m-го порядка. В рассматриваемом случае

$$\frac{\sin\left[(N/2)(fd-2m\pi)\right]}{fd-2m\pi} = \frac{N}{2},$$

а потому интенсивность волны E_m в главном максимуме пропорциональна N^2. Когда $\sin\left[(N/2)(fd-2m\pi)\right] = 0$, но $fd-2m\pi \neq 0$, интенсивность волны обращается в нуль. Этим определяются направления на дифракционные минимумы, число которых между двумя соседними главными максимумами равно $N-1$. Таким образом, получаются все результаты, найденные выше методом суммирования волн.

§ 53. Примеры на применение метода Рэлея

Ниже предполагается, что период решетки значительно больше длины световой волны, так что пропускасмость $D(x)$ можно вычислить в приближении геометрической оптики.

1. Начнем с плоской амплитудной решетки, состоящей из прозрачных щелей ширины b и непрозрачных промежутков между ними ширины a. В приближении геометрической оптики пропускаемость $D(x)$ равна единице на щели и нулю в промежутках между ними. Начало координат поместим в середине щели. Тогда коэффициент Фурье D_m представится выражением

$$D_m = \frac{1}{d}\int\limits_{-d/2}^{+d/2} D(x)e^{impx}\, dx = \frac{1}{d}\int\limits_{-b/2}^{+b/2} e^{impx}\, dx = \frac{b}{d}\frac{\sin(\pi mb/d)}{\pi mb/d}. \quad (53.1)$$

При вычислении интенсивности дифрагированного пучка в принципе следовало бы учесть изменение его поперечного сечения из-за наклона к плоскости решетки. Однако это было бы превышением точности, так как наши вычисления применимы лишь при малых углах дифракции, когда $\cos\vartheta \approx 1$. В этом приближении относительная интенсивность m-го дифракционного пучка $I_m = D_m^2$. (Интенсивность падающей волны принята за

единицу.) Для спектра нулевого порядка $D_0 = b/d$, $I_0 = (b/d)^2$. Полная интенсивность, пропускаемая решеткой, $I_{\text{прош}} = b/d$. Разность этих величин дает суммарную интенсивность света, приходящуюся на спектры прочих порядков:

$$\frac{b}{d} = \frac{b^2}{d^2} + \frac{2}{\pi^2} \sum_{m=1}^{\infty} \frac{1}{m^2} \sin^2 \frac{\pi m b}{d}. \tag{53.2}$$

(Это — чисто математическое соотношение, доказываемое в теории рядов Фурье.)

Относительная доля дифрагированного света будет

$$\frac{I_{\text{прош}} - I_0}{I_{\text{прош}}} = 1 - \frac{b}{d}. \tag{53.3}$$

Она максимальна и стремится к единице, когда $b/d \to 0$. Однако в этом случае сама интенсивность $I_{\text{прош}}$ также стремится к нулю. Интенсивность m-го дифракционного пучка равна

$$I_m = D_m^2 = I_0 \left[\frac{\sin(\pi m b/d)}{\pi m b/d} \right]^2. \tag{53.4}$$

Когда $b/d \to 0$, интенсивности всех дифракционных пучков становятся одинаковыми и равными I_0. Однако, как уже отмечено выше, в этом случае каждая из этих интенслвностей сама стремится к нулю.

Если $\pi m b/d = n\pi$, т.е. $b/d = n/m$, где n — целое число, меньшее m и взаимно простое с ним, то обращаются в нуль интенсивности спектров с порядками m, $2m$, $3m$, ... Так, при $b/d = 1/2$ пропадают все спектры четных порядков. Смысл этого результата, как уже отмечалось в § 46, станет очевидным, если заметить, что условие m-го главного максимума $d \sin\vartheta = m\lambda$ умножением на $b/d = n/m$ преобразуется в $b \sin\vartheta = n\lambda$, т.е. в условие n-го дифракционного минимума при дифракции на отдельной щели. Таким образом, под углом ϑ каждая щель, а потому и решетка в целом света не посылают.

2. Рассмотрим более общий случай. Допустим, что на участках длины b пропускаемость решетки равна β, а на участках длины a она равна α. Величины α и β постоянны, но могут быть комплексными. Таким образом, решетка является *амплитудно-фазовой*. Когда α и β — числа вещественные, то решетка будет *амплитудной*. Если же они — числа вида $e^{i\rho}$ (ρ вещественно), то решетка становится чисто *фазовой*. Рассматриваемая амплитудно-фазовая решетка эквивалентна плоскопараллельной пластинке с пропускаемостью α и наложенной на нее дифракционной решетке. Пропускаемость последней на участках b равна $(\beta - \alpha)$, а на участках a — нулю. Разумеется, величины α и β, a и b можно поменять местами и получить вторую эквивалентную систему. Математически обе системы отличаются одна от другой только обозначениями, а потому достаточно рассмотреть лишь одну из них, например, первую.

Вычисление коэффициентов Фурье D_m сводится к предыдущей задаче. Для плоскопараллельной пластинки все коэффициенты Фурье обращаются в нуль, за исключением нулевого, который равен α. Поэтому, поместив начало координат в центре одного из отрезков b и воспользовавшись формулой (53.1), получим

$$D_m = (\beta - \alpha) \frac{b}{d} \frac{\sin(\pi m b/d)}{\pi m b/d} + \alpha \delta_m, \tag{53.5}$$

где $\delta_m = 1$ при $m = 0$ и $\delta_m = 0$ при $m \neq 0$. При $\alpha = 0$, $\beta = 1$ получаются результаты предыдущей задачи.

3. Рассмотрим теперь частные случаи чисто амплитудной и чисто фазовой решеток. Для амплитудной решетки величины α и β вещественны и положительны. Все коэффициенты D_m также вещественны. Знаки этих коэффициентов, начиная с $m = \pm 1$, чередуются. Коэффициенты нулевого и первого порядков могут иметь одинаковые или противоположные знаки в зависимости от соотношения между пропускаемостями α и β. В случае чисто фазовой решетки пропускасмости α и β имеют вид $e^{i\rho}$. Так как существенна только разность фаз между волнами, исходящими из участков a и b, то без ущерба для общности можно положить $\alpha = 1$, $\beta = e^{i\rho}$. Тогда из формулы (53.5) находим

$$D_m = (e^{i\rho} - 1)\,\frac{b}{d}\,\frac{\sin(m\pi b/d)}{m\pi b/d} \qquad (m \neq 0), \tag{53.6}$$

$$D_0 = (e^{i\rho} - 1)\,\frac{b}{d} + 1. \tag{53.7}$$

Как и в случае амплитудной решетки, коэффициенты Фурье D_m, начиная с $m = \pm 1$, попеременно меняют знаки. Никакого дополнительного сдвига фаз между этими коэффициентами нет.

Качественное отличие фазовой решетки от амплитудной состоит в том, что в случае фазовой решетки имеется дополнительный сдвиг фаз φ между спектром нулевого и спектрами всех прочих порядков. Чтобы его вычислить, найдем из формул (53.6) и (53.7) комплексное отношение D_m/D_0. Аргумент этого комплексного числа и будет φ. Простое вычисление дает

$$\operatorname{tg}\varphi = \frac{b+a}{b-a}\,\frac{\sin\rho}{1-\cos\rho}. \tag{53.8}$$

Сдвиг фаз φ один и тот же для всех порядков m. Так как после дифракции на решетке спектры различных порядков пространственно разделяются на независимые пучки, то можно оказывать воздействие на каждый из них, не меняя при этом амплитуды и фазы всех остальных пучков. Например, если на пути нулевого пучка поставить прозрачную пластинку, которая изменила бы его фазу на φ, то фазовые соотношения между дифрагированными пучками будут такими же, как и у амплитудной решетки. С введением такой пластинки фазовая решетка действует как амплитудная. На этом основан *метод фазового контраста*, используемый в микроскопии (см. § 59).

Отметим два частных случая. Во-первых, случай $a = b$. Тогда формула (53.8) дает $\operatorname{tg}\varphi = \infty$, т.е. $\varphi = \pi/2$. Во-вторых, случай малых значений ρ. Тогда

$$\operatorname{tg}\varphi \approx \frac{b+a}{b-a}\,\frac{2}{\rho},$$

т.е. $\operatorname{tg}\varphi$ очень велик, а самый угол φ практически равен $\pi/2$. В обоих случаях для превращения фазовой решетки в амплитудную на пути нулевого пучка или на пути всех прочих дифрагированных пучков достаточно ввести пластинку, вносящую дополнительную разность фаз $\pm\pi/2$.

4. В качестве последнего примера рассмотрим фазовую решетку, профиль штрихов которой показан на рис. 205. Поперечное сечение штриха имеет форму треугольника, одна сторона которого длинная и пологая, а другая — короткая и крутая. Такая решетка интересна тем, что при определенных условиях она может концентрировать основную часть света в спектре одного порядка. Решетка может действовать и как пропускающая, и как

отражательная. Ради определенности разберем действие пропускающей решетки.

Поместим начало координат O в середине длинной стороны AB. Пусть a — длина проекции длинной, а b — короткой стороны на направление оси X. Если $b \ll a$, то в интеграле (53.1) можно пренебречь вкладом, вносимым короткой стороной. В этом приближении можно положить $a = d$ и вычислить пропускаемость решетки только на участке a. Влияние пропус-

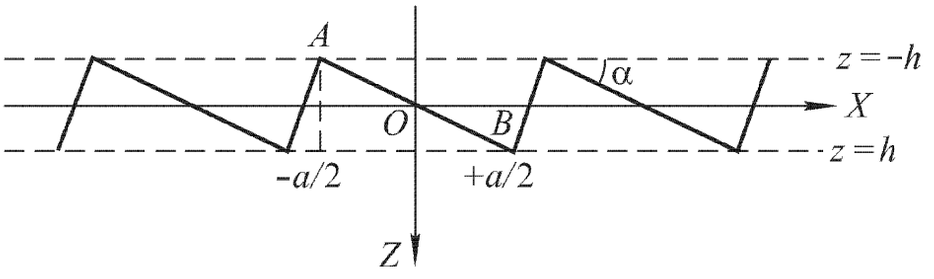


Рис. 205

каемости участка b практически не отразится на результатах. Пусть волна падает перпендикулярно к плоскости XY и в воздухе представляется выражением $E = E_0 e^{-\imath kz}$. На входе, т.е. в плоскости $z = -h$, поле представляется выражением $E_{\text{вх}} = E_0 e^{ikh}$. Чтобы вычислить поле на выходе при $z = +h$, можно, ввиду малости угла наклона α, пренебречь преломлением. Если z_0 — текущая координата точки на прямой AB, то поле на выходе в лежащей под ней точкой будет равно

$$E_{\text{вых}} = E_0 e^{-\imath kz_0} \cdot e^{-\imath kn(h-z_0)} = e^{\imath\delta} e^{\imath k(n-1)z_0} E_{\text{вх}},$$

где δ — некоторая постоянная. Постоянный фазовый множитель $e^{\imath\delta}$ не играет роли и может быть отброшен. Таким образом, пропускаемость решетки $D = e^{\imath k(n-1)z_0}$, или после подстановки $z_0 = x \operatorname{tg}\alpha \approx \alpha x$

$$D(x) = e^{\imath k(n-1)\alpha x}.$$

Коэффициенты Фурье D_m вычисляем по формуле (53.1) и находим

$$D_m = \frac{\sin \pi[m + (d/\lambda)(n-1)\alpha]}{\pi[m + (d/\lambda)(n-1)\alpha]}. \tag{53.9}$$

Если знаменатель этого выражения обращается в нуль, то почти весь свет сконцентрируется в спектре порядка m. Для этого должно быть $m + (d/\lambda)(n-1)\alpha = 0$. С другой стороны, $d \sin\vartheta = m\lambda$, или, ввиду малости угла дифракции, $d \cdot \vartheta = m\lambda$. Исключая m, получаем

$$\vartheta = -(n-1)\alpha. \tag{53.10}$$

Эта формула показывает, что угол ϑ равен углу отклонения луча при преломлении в призме с малым преломляющим углом α. Таким образом, почти весь свет может сконцентрироваться в одном направлении, если это направление совпадает с направлением преломленных лучей. Для концентрации необходимо, чтобы разность хода между пучками, преломленными на соседних ступеньках решетки, составляла целое число волн. С подобной концентрацией дифрагированного света в спектре одного порядка мы столкнулись также в § 48 при изучении эшелона Майкельсона.

§ 54. Голография

1. Применим метод Рэлея для уяснения идеи *голографии*, т.е. безлинзового получения оптических изображений путем так называемого *восстановления волнового фронта*. В принципе идея голографии была выдвинута и экспериментально проверена польским физиком М. Вольфке (1883–1947). Его работа была опубликована еще в 1920 г., но была забыта. Эту идею независимо от Вольфке вновь предложил и обосновал в 1947 г. английский инженер и физик Габор (р. 1900), который по праву считается изобретателем голографии. Однако понадобилось 15 лет, чтобы стало возможно практическое осуществление голографии. Причина столь длительной задержки заключается в том, что в голографии требуются источники света, обладающие *высокой степенью временной и пространственной когерентности*. Таких источников в 1947 г. еще не существовало. Положение изменилось в 1960 г. с изобретением лазеров и проникновением их в лабораторную технику. Первые изображения по методу голографии были получены американцами Лейтом и Упатниексом в 1962 г.

При освещении или просвечивании предмета от него распространяется рассеянная или прошедшая волна. (Для определенности в дальнейшем предполагается, что предмет освещается, а не просвечивается, хотя принципиального различия между этими случаями при нашем рассмотрении нет.) Отделившись от предмета, рассеянная волна сохраняет в дальнейшем *независимое существование* и несет полную информацию о форме и прочих свойствах предмета, какая может быть получена путем освещения его световыми лучами. Попадая в глаз или объектив фотоаппарата, эта волна образует на сетчатке или на фотопластинке изображение предмета. Если любым путем создать такую же волну, то, очевидно, она сможет вызвать в точности такие же эффекты, что и исходная волна, рассеянная предметом. На этом замечании и основана идея голографии.

Процесс получения изображения в голографии распадается на две стадии. На первой стадии изготовляется *голограмма*, т.е. фотопластинка, с помощью которой можно восстанавливать световую волну, рассеянную телом. Вторую стадию составляет само *восстановление* этой волны и получение оптического изображения.

2. Пусть какой-то предмет A (рис. 206) освещается пучком паралелльных лучей от лазера. Рассеянные лучи попадают на фотопластинку Γ. По степени почернения пластинки после проявления можно судить об амплитуде рассеянной волны во всех местах пластинки, которых эта волна достигла. В этом смысле экспонированная и проявленная пластинка сохраняет информацию об амплитуде волнового поля. Для восстановления волново-

го поля такой информации, конечно, недостаточно. Нужна еще дополнительная информация о *фазе*, которой пластинка не содержит, так как степень почернения зависит только от интенсивности, но не от фазы волны. Габор указал, что необходимую информацию о фазе можно получить и записать на той же фотопластинке $Г$, если осветить ее вторым пучком от того же лазера и заставить его интерферировать с пучком, рассеянным предметом. Практически этого можно достигнуть, расширив предварительно пучок от лазера, а затем разделив его на два пучка. Один из них (*предметный*) направляется на предмет A, другой (*опорный*) отражается от плоского зеркала S. Оба пучка направляются на фотопластинку $Г$ и там интерферируют между собой. Интерференционная картина фотографируется. Так полученная фотография и называется *голограммой.*

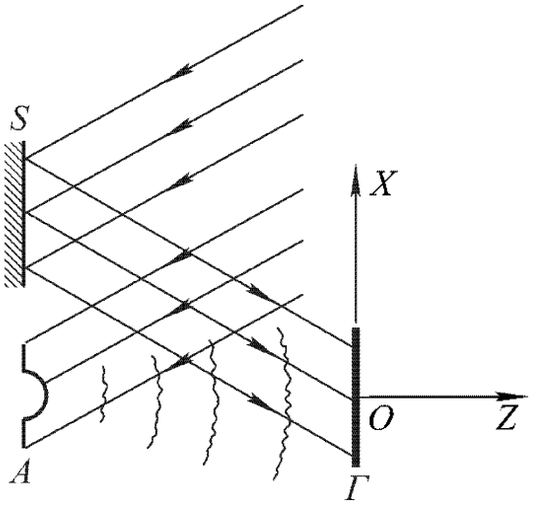


Рис. 206

Поскольку волна, рассеянная предметом, возникает при отражении и дифракции на *макроскопических* деталях предмета со сложной формой и взаимным расположением, реальная голограмма представляет собой очень сложную и запутанную интерференционную картину с очень мелкими деталями, которые невозможно различить невооруженным глазом. (На ней обычно видны и крупные дифракционные кольца. Но они не имеют никакого отношения к делу, так как возникают при дифракции на случайных пылинках, встретившихся на пути распространения света.)

Дифракционная картина на голограмме не имеет ни малейшего сходства с предметом. При рассматривании ее в микроскоп в ней трудно усмотреть следы каких-либо закономерностей. И тем не менее расположение, форма и интенсивность дифракционных пятен голограммы полностью определяются геометрической формой и физическими свойствами отражающей поверхности объекта. Голограмма *в закодированной форме* содержит *полную информацию* об амплитудах и фазах рассеянной волны, которая достаточна для ее восстановления и получения оптического изображения. Само название «голография» происходит от греческих слов «голог» — полный и «графе» — пишу и может быть переведено как «полная запись».

3. Разность хода между опорной волной и волнами, рассеянными предметом, очень велика и может составлять несколь-

ко метров. Поэтому при изготовлении голограммы время когерентности света должно быть очень велико (не меньше 10^{-5}–10^{-7} с). Длина когерентности также должна быть очень большой (не меньше 1–10 м). Никакие источники света, кроме лазеров, не могут обеспечить столь высокую степень временной и пространственной когерентности. Кроме того, необходима *неподвижность* (с точностью до долей длины световой волны) и *высокая контрастность* интерференционной картины, образующейся в области перекрытия предметного и опорного пучков во время экспозиции. Для этого также требуется высокая когерентность света, а также *механическая жесткость* всех элементов голографической установки.

Необходимая степень монохроматичности света определяется по общей формуле $\lambda/\delta l \gtrsim m$, где m — максимальный порядок интерференции, наблюдающейся при голографировании. При рациональном расположении источника света и прочей аппаратуры его можно оценить по формуле $m \sim L/\lambda$, где L — линейные размеры предмета. Таким образом, должно быть $\delta\lambda < \lambda^2/L$. Уже при небольших размерах предмета ~ 10 см и $\lambda = 500$ нм эта формула дает $\delta\lambda < 10^{-3}$ нм. Между тем спектральные линии ртутной лампы, даже с низким давлением ртутных паров, имеют ширину ~ 30 нм. Требования к размерам источника света, пожалуй, еще более жесткие. Для получения высокой контрастности интерференционной картины на голограмме поперечные размеры Δx источника света (т.е. размеры параллельно плоскости голограммы) должны быть малы по сравнению с шириной интерференционной полосы λ/α, где α — угол схождения крайних интерферирующих лучей (см. § 28). По порядку величины $\alpha = h/l$, где h — ширина опорного пучка, а l — расстояние между предметом и голограммой. Если $h = 3$ см, а $l = 30$ см, то должно быть $\Delta x < \lambda l/h = 5$ мкм.

С нелазерными источниками (например, ртутной дуговой лампой) этим требованиям пытались удовлетворить стандартными приемами, пропуская свет через различные монохроматоры, а затем фокусируя его на малом отверстии. При этом через отверстие проходила и достигала освещаемого объекта лишь ничтожная часть светового потока. Требующееся время экспозиции при этом будет достигать порядка года. Поэтому Габору удалось получить голограммы только простейших микроскопических объектов при экспозициях в несколько часов. Лазеры сделали описанные операции ненужными, а идея голографии получила практическое осуществление.

4. Представим поле волны, рассеянной предметом A, в виде

$$u = a(\mathbf{r})e^{i[\omega t - \Phi(\mathbf{r})]}, \tag{54.1}$$

а поле волны, отраженной плоским зеркалом S, — в виде

$$v = be^{i(\omega t - \mathbf{kr})} \tag{54.2}$$

(b и $\mathbf{k}$ — постоянные). Оба поля записаны в скалярной форме. Это упрощение не существенно, так как изменения поляризации волн в интересующих нас процессах не играют заметной роли. При таком упрощении величины $a(\mathbf{r})$, $\Phi(\mathbf{r})$ и b могут считаться вещественными. Интенсивность результирующего поля на пластинке Γ и перед ней представится выражением

$$I = (u + v)(u^* + v^*) = v^*u + vu^* + v^*v + u^*u.$$

Начало координат поместим в какой-либо точке O в плоскости пластинки Γ. С той же плоскостью совместим координатную плоскость XY, направив ось Y перпендикулярно к направлению опорного пучка, отраженного зеркалом S ($k_y = 0$). Тогда ось Z будет нормальна к плоскости пластинки Γ. Направим ее в сторону распространения опорного пучка. Интенсивность I_0 светового поля в плоскости пластинки найдется из предыдущего выражения, если все входящие в него функции заменить их значениями при $z = 0$:

$$I_0 = ba(x, y, 0)e^{i[k_x x - \Phi(x,y,0)]} + ba(x, y, 0)e^{-i[k_x x - \Phi(x,y,0)]} + \\ + b^2 + a^2(x, y, 0). \tag{54.3}$$

Допустим теперь, что контактным способом изготовлен позитив голограммы при таком режиме проявления, что амплитудная прозрачность (пропускаемость) D позитива оказалась пропорциональной I_0. Для краткости коэффициент пропорциональности можно принять равным единице, т.е. положить $D = I_0$. Этого всегда можно достигнуть надлежащим выбором единиц. Такая позитивная голограмма может быть использована для восстановления рассеянной волны $u(\mathbf{r}, t)$. Для этого, убрав предмет A, голограмму просвечивают таким же опорным пучком $v(\mathbf{r}, t)$, какой применялся при ее изготовлении. Этот пучок будет испытывать дифракцию на голограмме подобно тому, как дифрагирует свет на дифракционной решетке. Задача сводится к расчету дифракционной картины за голограммой. Для ее решения можно применить метод Рэлея. Обозначим через $E(\mathbf{r}, t)$ волновое поле за голограммой. На выходе голограммы оно представится выражением

$$E_{\text{вых}} = Dv(x, y, 0) = I_0 be^{i(\omega t - k_x x)}. \tag{54.4}$$

Надо найти решение волнового уравнения

$$\frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} + \frac{\partial^2 E}{\partial z^2} + k^2 E = 0, \tag{54.5}$$

удовлетворяющее краевому условию (54.4). Ввиду линейности и однородности такого уравнения и условия (54.4), эту задачу можно расчленить на четыре независимые задачи. Представим функцию E в виде суммы

$$E = E_1 + E_2 + E_3 + E_4 \tag{54.6}$$

и потребуем, чтобы слагаемые E_1, E_2, E_3, E_4 были решениями уравнения (54.5) и удовлетворяли следующим граничным условиям:

$$E_{1\text{ вых}} = b^2 a(x, y, 0) e^{i[\omega t - \Phi(x,y,0)]}, \tag{54.7}$$

$$E_{2\text{ вых}} = b^2 a(x, y, 0) e^{i[\omega t + \Phi(x,y,0) - 2k_x x]}, \tag{54.8}$$

$$E_{3\text{ вых}} = b^3 e^{i(\omega t - k_x x)}, \tag{54.9}$$

$$E_{4\text{ вых}} = b a^2(x, y, 0) e^{i(\omega t - k_x x)}. \tag{54.10}$$

Сюда необходимо добавить физическое условие, чтобы волны, дифрагированные на голограмме, были *уходящими*. Этим обеспечивается *единственность решения*.

5. Проще всего найти функцию E_3, т.е. решение задачи, удовлетворяющее граничному условию (54.9). Ввиду постоянства b таким решением, очевидно, будет

$$E_3 = b^3 e^{i(\omega t - \mathbf{kr})} = b^2 v(\mathbf{r}, t). \tag{54.9а}$$

Если отвлечься от амплитудного множителя b^2, то это есть просвечивающая (опорная) волна, распространившаяся за голограмму.

6. Основной интерес в голографии представляет решение E_1, удовлетворяющее граничному условию (54.7). Ввиду постоянства b, его можно также сразу указать. Именно:

$$E_1 = b^2 a(x, y, z) e^{i[\omega t - \Phi(x,y,z)]} = b^2 u(\mathbf{r}, t). \tag{54.7а}$$

Действительно, функция E_1 удовлетворяет волновому уравнению (54.5) и представляет волну, уходящую от голограммы, которая при $z = 0$ переходит в выражение (54.7). Но эта волна с точностью до несущественного постоянного множителя тождественна с волной u, рассеянной объектом. Она, таким образом, точно воспроизводит рассеянную волну и дает *мнимое изображение* (МИ) объекта, в том же самом месте, в каком он находился во время получения голограммы (рис. 207 *а*). Тем самым разъяснен основной принцип голографии.

7. Для нахождения волны E_2, т.е. решения волнового уравнения, удовлетворяющего граничному условию (54.8), рассмотрим сначала частный случай, когда опорный, а следовательно, и просвечивающий пучки перпендикулярны к плоскости голограммы. Тогда $k_x = 0$, и граничное условие (54.8) переходит в

$$E_{2\text{ вых}} = b^2 a(x, y, 0) e^{i[\omega t + \Phi(x,y,0)]}. \tag{54.11}$$

Волна

$$\widetilde{E}_2(x,y,z) = b^2 a(x,y,z) e^{i[\omega t + \Phi(x,y,z)]}, \qquad (54.12)$$

очевидно, удовлетворяет этому граничному условию. Это есть волна, получающаяся из волны (54.7а) обращением направления ее распространения, а потому она также удовлетворяет и волновому уравнению (54.5). Она давала бы мнимое изображение предмета перед голограммой в точности в том же месте, где

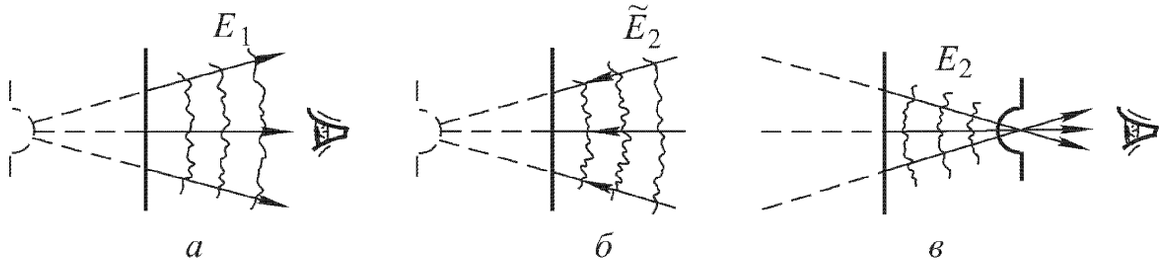


Рис. 207

находился сам предмет при получении голограммы (рис. 207 *б*). Однако волна $\widetilde{E}_2(\mathbf{r}, t)$ распространяется к голограмме, а не от нее, и по этой причине не может служить решением рассматриваемой нами задачи. Для нахождения нужного решения заметим, что волновое уравнение (54.5) не изменится, если изменить на противоположные знаки одной или нескольких координат. Так, при изменении знака у z уравнение (54.5) переходит в

$$\frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} + \frac{\partial^2 E}{\partial (-z)^2} + k^2 E = 0. \qquad (54.5а)$$

Это уравнение, очевидно, имеет решение $\widetilde{E}_2(x, y, -z)$. А так как оно тождественно с уравнением (54.5), то последнее имеет такое же решение, т.е.

$$E_2 = b^2 a(x,y,-z) e^{i[\omega t + \Phi(x,y,-z)]}. \qquad (54.11а)$$

Решение это, очевидно, удовлетворяет граничному условию (54.11). Оно получится из (54.12), если пространства перед и за голограммой поменять местами. Решение (54.11а) представляет волну, *идущую от голограммы*, а потому оно и будет искомым решением задачи. Полученная волна сходится за голограммой и дает там *действительное изображение* (ДИ) предмета. Положение его зеркально симметрично с положением самого предмета относительно плоскости голограммы (рис. 207 *в*). Однако глаз может видеть действительное изображение только тогда, когда оно рассматривается с задней стороны, т.е. когда изображение находится между голограммой и глазом. Поэтому действительное изображение будет *псевдоскопическим*, или вывернутым. На нем выпуклые места предмета глаз видит как

вогнутые и наоборот. (Псевдоскопичпость не будет проявляться, если действительное изображение спроецировать на экран.) Если смотреть на голограмму перпендикулярно к ее поверхности, то действительное изображение будет мешать рассматриванию мнимого изображения, так как оба изображения лежат на одной линии зрения. Такое же вредное влияние оказывают прямые лучи просвечивающего пучка, поскольку они также попадают в глаз. Эти недостатки устраняются применением наклонных опорных (а следовательно, и просвечивающих) пучков света.

В случае наклонных пучков $k_x \neq 0$, и надо найти решение E_2, удовлетворяющее общему граничному условию (54.8). Но этот случай легко сводится к частному случаю нормального падения, разобранному выше. Действительно, если бы граничное условие имело частный вид (54.11), то можно было бы перейти к общему граничному условию (54.8), наложив на голограмму дополнительную дифракционную решетку с амплитудной прозрачностью $D = e^{-2ik_x x}$. Это — фазовая решетка, дающая только спектр минус первого порядка (см. § 52, п. 5). Она действует как призма, т.е. только изменяет направление падающего пучка света. В результате действительное изображение смещается в

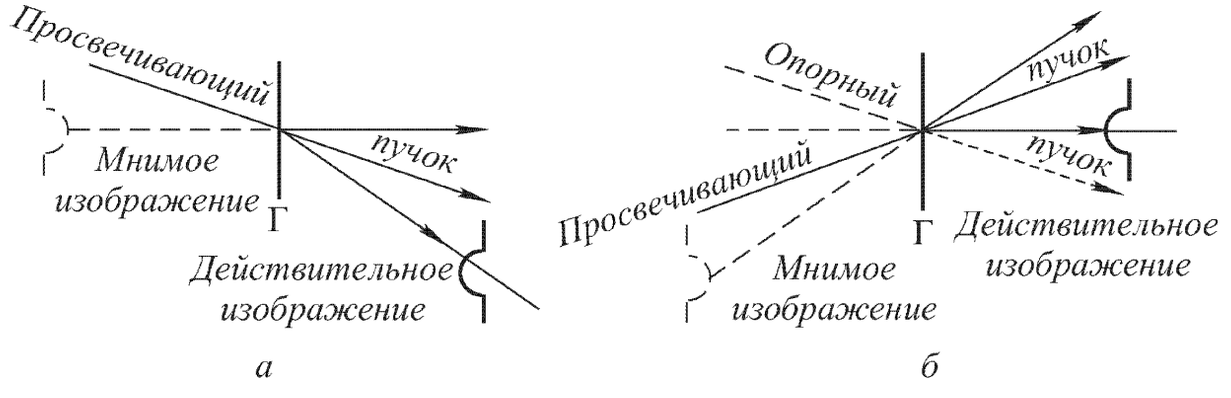


Рис. 208

сторону, как указано на рис. 208 *а*. Такое смещение, конечно, искажает изображение совершенно так же, как это происходит при рассматривании предмета через призму. При большом наклоне опорного (а следовательно, и просвечивающего) пучка действительного изображения может и не получиться.

8. Искажения действительного изображения можно избежать, если просвечивающий пучок направить на голограмму под тем же углом и в той же плоскости падения, что и опорный, но с другой стороны от нормали к плоскости голограммы. Тогда

$$v_{\text{прос}} = be^{i(\omega t + k_x x - k_z z)},$$

и вместо граничных условий (54.7) и (54.8) мы получим соответ-

ственно

$$E_{1\text{ вых}} = b^2 a(x,y,0) e^{i[\omega t - \Phi(x,y,0) + 2k_x x]}, \qquad (54.7б)$$

$$E_{2\text{ вых}} = b^2 a(x,y,0) e^{i[\omega t + \Phi(x,y,0)]}. \qquad (54.8б)$$

Теперь действительное изображение будет расположено зеркально симметрично (относительно плоскости голограммы) с самим предметом в том положении, какое он занимал при изготовлении голограммы. Оно по-прежнему будет псевдоскопическим, но не искаженным. Зато мнимое изображение сместится в сторону (рис. 208 *б*) и будет искажено, как при рассматривании предмета через призму.

9. Остается выяснить влияние волны E_4, удовлетворяющей граничному условию (54.10). Найти решение в этом случае без конкретизации вида функции $a(x,y,0)$ невозможно, так как в условие (54.10) входит не сама эта функция, а ее квадрат. Для отыскания E_4 в общем случае следует разложить функцию $a^2(x,y,0)$ в интеграл Фурье, а затем стандартным способом (см. § 52) представить волну E_4 в виде суперпозиции плоских волн, распространяющихся в различных направлениях. Таким образом, за голограммой возникает волновое поле, создающее фон, на котором получаются мнимое и действительное изображения. Интенсивность этого фона зависит от квадрата амплитуды $a(x,y,z)$ волны, рассеянной предметом при изготовлении голограммы. Поэтому для ослабления вредного влияния такого фона интенсивность опорной волны берут значительно большей интенсивности предметной волны.

Так как фотослой, на котором записана голограмма, тонкий, то при восстановлении изображения не возникает интерференция столь же высоких порядков, что и при изготовлении голограммы. По этой причине удовлетворительное изображение предмета можно получить с голограммой, освещаемой при восстановлении светом от ртутной лампы.

10. Чем больше углы, под которыми сходятся интерферирующие лучи на голограмме, а следовательно, чем мельче детали интерференционной картины, тем выше разрешающая способность голограммы (см. § 57). При предельно возможном разрешении эти детали порядка длины световой волны. Но и при углах схождения в несколько градусов интерференционная картина на голограмме получается все же довольно тонкой. Чтобы ее зафиксировать на фотопластинке, требуется фотографическая эмульсия весьма высокого качества. Наиболее высокие требования к фотопластинкам предъявляются в цветной фотографии Денисюка, где предметная и опорная волны идут навстречу друг к другу (см. п. 13). Современные мелкозернистые фотопластинки, применяемые в голографии, разрешают 1000–

10 000 линий на миллиметр (расстояние между разрешаемыми линиями 1000–100 нм). Следует, однако, заметить, что повышение разрешающей способности фотоэмульсии, как правило, сопровождается падением ее чувствительности.

11. Для получения изображения нет необходимости изготовлять позитивную голограмму; негатив столь же хорош, что и позитив. Это непосредственно следует из теоремы Бабине. Различие в дифракционных картинах касается только просвечивающего пучка, проникшего за голограмму, и совсем не затрагивает волновые поля E_1 и E_2, дающие мнимое и действительное изображения. В обоих случаях при восстановлении волнового фронта получаются одинаковые изображения.

Конечно, при изготовлении голограммы условие $D \sim I_0$, соблюдение которого предполагалось при нашем изложении, не может быть выполнено вполне точно. Его невыполнение приводит к появлению *дополнительных изображений*. Это проще всего пояснить на примере, в котором предметная и опорная волны обе плоские и падают на фотопластинку под разными углами (предметом является бесконечно удаленная светящаяся точка). Тогда при соблюдении условия $D \sim I_0$ голограмма будет представлять собой дифракционную решетку с синусоидальной амплитудной прозрачностью. При ее просвечивании, наряду со спектром нулевого порядка (т.е. прямого просвечивающего пучка, проникшего за голограмму), появятся спектры первого и минус первого порядков, один из которых будет мнимым, а другой действительным изображениями предмета. Если же условие $D \sim I_0$ не соблюдается, то разложение функции $D(x)$ в ряд Фурье будет содержать гармоники высших порядков. При просвечивании голограммы, наряду со спектрами первых порядков, появятся спектры высших порядков, т.е. дополнительные мнимые и действительные изображения предмета. Однако если такие дополнительные изображения слабы и получаются в стороне от основного изображения, то они не причиняют существенного вреда.

Для получения большей яркости изображения применяют так называемые *отбеленные голограммы*. Они получаются из обычных голограмм путем обработки их специальными отбеливающими растворами, образующими прозрачные соединения с серебром, отложившимся в эмульсии голограммы. Получается прозрачная фазовая голограмма, пропускаемость которой варьируется как за счет толщины, так и за счет показателя преломления отбеленного слоя. Такая голограмма сохраняет способность давать оптические изображения. В какой степени отбелка голограммы и отступления от условия $D \sim I_0$ сказываются на получаемых изображениях, эти вопросы представляют громадный интерес в практической голографии. Однако они

трудно поддаются теоретическому изучению и по своему содержанию не могут быть включены в курс общей физики.

12. По сравнению с обычными фотографиями изображения, получаемые по методу голографии, обладают тем преимуществом, что они *трехмерны.* Мнимые изображения полностью воспроизводят без каких бы то ни было искажений взаимное расположение реальных предметов в пространстве. Если при рассматривании голограммы одни предметы заслоняются другими, то достаточно сместить в сторону глаз, чтобы увидеть и заслоненные предметы. Часть голограммы действует как целая голограмма. Например, для воспроизведения изображения годится каждый кусочек разбитой голограммы. По мере уменьшения размеров голограммы ослабевают лишь четкость изображения (разрешающая способность) и ощущение объемности. Эта особенность голографического метода связана с тем, что при экспонировании, как правило, все части фотопластинки подвергаются действию света, рассеянного *всеми точками предмета.* Поэтому в каждой части голограммы хранится в закодированной форме изображение всего предмета. Вопрос о разрешающей способности голограмм будет разобран в § 57 (п. 5), как частный случай общего вопроса о разрешающей способности оптических приборов.

На одной фотопластинке можно последовательно зарегистрировать несколько изображений, причем каждое из них можно восстановить без помех со стороны других изображений. Можно, например, произвести несколько экспозиций при одном и том же опорном пучке, голографируя каждый раз только один предмет. Во избежание наложения восстановленных изображений друг на друга голографируемые объекты должны располагаться в различных местах. Можно также объекты помещать в одно и то же место, но менять направление опорного пучка. Можно получать цветные изображения, освещая предмет тремя источниками света с различными длинами волн, подобранными так, чтобы наиболее полно передать цвет предмета. От тех же трех источников формируются опорные пучки, создающие вместе с волнами, рассеянными предметом, единую голограмму на черно-белой фотопластинке. Для восстановления изображения такая «трехкомпонентная голограмма» помещается в то же место, какое она занимала при экспонировании, и одновременно освещается тремя просвечивающими пучками, тождественными с опорными. При просвечивании голограммы каждый из этих пучков создает три действительных и три мнимых изображения. Три из этих мнимых изображений, и притом разноцветных, пространственно совпадут друг с другом, образуя цветное изображение. Остальные изображения расположатся в разных местах, не мешая основному изображению.

13. Голограммы, о которых шла речь выше, можно назвать *двумерными*, так как в них применяются фотопластинки с тонкими слоями эмульсии. В 1962 г. советский физик Ю.Н. Денисюк (р. 1927) стал получать *трехмерные*, *объемные* голограммы на фотопластинках с *толстослойной эмульсией*. В его методе удачно сочетаются принципы голографии и цветной фотографии Липпмана. Толщина фотографического слоя составляет 15–20 мкм, т.е. 30–40 длин волн зеленого цвета. Фотослой настолько прозрачен, что через него можно освещать голографируемый объект. Опорная плоская монохроматическая волна от лазера падает на фотопластинку со стороны стекла (рис. 209). Пройдя через фотопластинку, она освещает голографируемый предмет. Волна, рассеянная предметом, распространяется навстречу опорной волне, интерферируя с ней в толще фотоэмульсии. Интерференционная картина представляет *стоячие волны*, на которые наложен довольно причудливый узор мелких деталей из максимумов и минимумов, так как среди интерферирующих волн только опорная волна является плоской. Проявленная и отфиксированная фотопластинка и будет объемной голограммой Денисюка. Она состоит как бы из нескольких десятков поверхностных голограмм, расположенных в толще эмульсии.

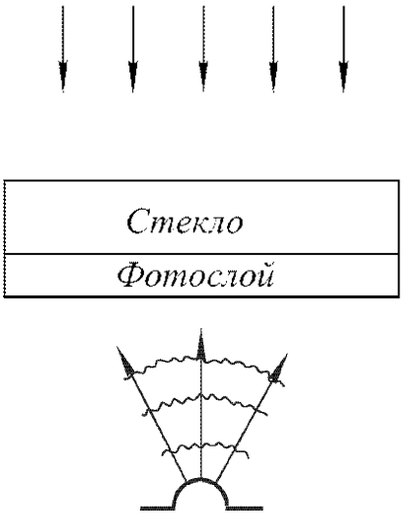


Рис. 209

Восстановление предметной волны производится расходящимся пучком *белого света*. Каждый слой выделившегося серебра, действуя подобно двумерной голограмме, дает слабые мнимое и действительное изображения предмета. При многолучевой интерференции происходит усиление тех волн, длина которых равна длине волны излучения лазера, в тех направлениях, в которых разность фаз между волнами от соседних слоев серебра равна 2π. В результате возникают изображения того же цвета, что и цвет луча лазера. Остальные изображения гасят друг друга при интерференции.

Таким образом, голограмма производит *монохроматизацию* белого света, которым она освещается. Конечно, такая монохроматизация сравнительно невысокая, из-за незначительного числа отложившихся слоев серебра и связанной с этим небольшой спектральной разрешающей способности голограммы. Кроме того, цвет изображения может существенно отличаться от цвета излучения лазера. Это связано с изменением расстояний

между слоями почернения при проявлении, фиксировании и сушке фотопластинки.

Метод Денисюка, подобно трехцветной фотографии, позволяет получать изображения предметов в *натуральных цветах.* Для этого на одной и той же фотопластинке получают голограмму предмета с помощью трех лазеров, излучения которых имеют различные длины волн. Последние подбираются так, чтобы при смешении они наиболее совершенно воспроизводили цвет предмета. Такая голограмма действует как три голограммы, дающие при освещении белым светом совмещенные изображения предмета в трех цветах. При этом цвет изображения кажется глазу таким же, как и цвет самого предмета.

Голография в настоящее время представляет самостоятельный, быстро развивающийся раздел науки, техники и искусства, возможно с большим будущим. Ей посвящены специальные руководства, к которым и следует обратиться всем, кто пожелает глубже и детальнее изучить голографию.

§ 55. Световое поле вблизи фокуса

1. Согласно геометрической оптике, волновое поле в центре сферической сходящейся волны (фокусе) обращается в бесконечность. Это указывает на неприменимость геометрической оптики в фокусе и его ближайшей окрестности.

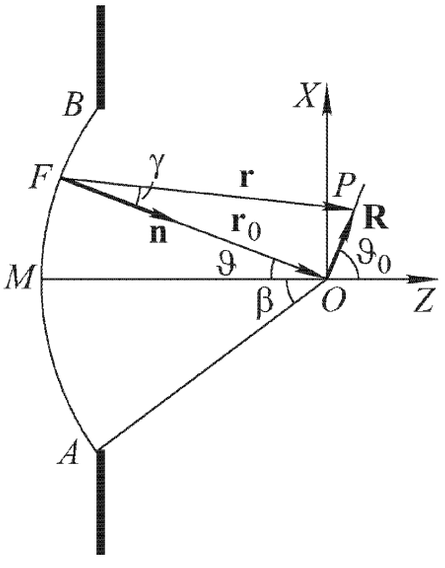


Рис. 210

Рассмотрим задачу о волновом поле в окрестности фокуса с точки зрения волновой оптики. Пусть на пути сходящейся сферической волны поставлена диафрагма с отверстием AB (рис. 210). Неприкрытую часть волнового фронта F примем за вспомогательную поверхность, из которой исходят вторичные волны Гюйгенса. Следуя приближенному методу Френеля (§ 41), поле на поверхности F запишем в виде

$$E_F = \frac{1}{r_0} e^{i(\omega t + k r_0)}.$$

Если пренебречь зависимостью амплитуд вторичных волн от (малого) угла γ, то поле в точке наблюдения P представится интегралом

$$E_P \sim e^{i\omega t} \int \frac{e^{ik(r_0 - r)}}{r_0 r}\, dF,$$

распространенным по всей неприкрытой поверхности волнового фронта. Если точка P лежит в малой окрестности фокуса O, то расстояния r_0 и r в знаменателе подынтегрального выражения могут считаться постоянными. Различие между ними надо учесть только в фазовом множителе $e^{ik(r_0-r)}$. Как видно из рис. 210, $\mathbf{r} = \mathbf{r}_0 + \mathbf{R}$. Возведем это равенство в квадрат, учтем, что $R \ll r_0$, а затем приближенно извлечем квадратный корень с точностью до квадратичных членов по R включительно. Получим

$$r - r_0 = \frac{(\mathbf{r}_0\mathbf{R})}{r_0} + \frac{R^2}{2r_0} - \frac{(\mathbf{r}_0\mathbf{R})^2}{r_0^3} = (\mathbf{nR}) + \frac{R^2}{2r_0}\sin^2(\mathbf{R},\mathbf{n}),$$

где $\mathbf{n} = \mathbf{r}_0/r_0$ — единичный вектор нормали к поверхности F, направленный к фокусу O. В дальнейшем в этом выражении сохраним только линейный член. Это можно делать, если поправка в фазе, вносимая квадратичным членом, много меньше π, т.е. когда

$$R^2/r_0 \ll \lambda. \tag{55.1}$$

С учетом всего изложенного для волнового поля (в условных единицах) в точке P получим

$$E_P = \int e^{i(\omega t - k\mathbf{nR})}\, d\Omega, \tag{55.2}$$

где $d\Omega = dF/r_0^2$ — телесный угол, под которым из точки O видна площадка dF.

Когда диафрагма удалена в бесконечность, а расстояние R остается конечным, то $R^2/r_0 = 0$. Тем самым ограничение (55.1) снимается, а (55.2) переходит в точное решение задачи. Это видно и из самой структуры выражения (55.2). Действительно, $d\Omega\, e^{i(\omega t - k\mathbf{nR})}$ есть плоская волна, распространяющаяся в направлении волнового вектора $\mathbf{k} = k\mathbf{n}$, а потому она удовлетворяет волновому уравнению (43.1). Интеграл (55.2) есть суперпозиция таких плоских волн различных направлений, а следовательно, является точным решением того же уравнения. Такое решение позволяет точно исследовать все явления, связанные с прохождением сферической волны через фокус.

2. Если точка P находится в фокусе O, т.е. $R = 0$, то все вторичные волны приходят в O в одинаковых фазах, а потому интенсивность света в этой точке максимальна. Формула (55.2) в этом случае дает $E_O = \Omega e^{i\omega t}$. Наибольший интерес представляет распределение интенсивности света в фокальной плоскости, т.е. в плоскости, проходящей через фокус O перпендикулярно к оси пучка MO. Чтобы интегрирование в (55.2) выполнялось элементарно, будем предполагать телесный угол Ω небольшим, так что при вычислении интеграла неприкрытую часть сфери-

ческого волнового фронта можно заменить плоской. Саму диафрагму возьмем прямоугольной формы со сторонами $2a$ и $2b$. Угловые размеры этих сторон при наблюдении из точки O будут $2\alpha = 2a/r_0$ и $2\beta = 2b/r_0$ соответственно. Начало координат поместим в точке O, направив оси X и Y параллельно сторонам прямоугольной диафрагмы. Координаты точки наблюдения P обозначим через X и Y, текущие координаты в плоскости диафрагмы — через x и y. Тогда $dF = dx\,dy = r_0^2\,d\varphi\,d\psi$, $d\Omega = d\varphi\,d\psi$, где φ и ψ — углы, под которыми из точки O видны отрезки x и y. Далее, $(\mathbf{nR}) = n_x X + n_y Y = -(x/r_0)X - (y/r_0)Y = -(\varphi X + \psi Y)$. В результате интеграл (55.2) перейдет в

$$E_P = e^{i\omega t} \int\limits_{-\alpha}^{+\alpha} \int\limits_{-\beta}^{+\beta} e^{ik(X\varphi + Y\psi)}\,d\varphi\,d\psi.$$

Это в точности такой же интеграл, какой встречался в § 45 при рассмотрении фраунгоферовой дифракции на прямоугольном отверстии. Поэтому распределение интенсивности света в фокальной плоскости в окрестности фокуса можно представить формулой

$$I = I_0 \left(\frac{\sin kx\alpha}{kx\alpha}\right)^2 \left(\frac{\sin ky\beta}{ky\beta}\right)^2. \tag{55.3}$$

В фокальной плоскости получается система светлых пятен прямоугольной формы со светлым центром (см. рис. 179). Расстояния между двумя соседними минимумами, а также от центра центрального максимума до первого минимума равны

$$\Delta x = \frac{\pi}{k\beta} = \frac{1}{2}\frac{\lambda}{\beta}. \tag{55.4}$$

Остается проверить, выполняется ли условие (55.1). Так как заметная интенсивность по формуле (55.3) получается при R порядка Δx, то формула (55.1) переходит в

$$\beta \gg \sqrt{\lambda/r_0}. \tag{55.5}$$

В оптических приборах под r_0 следует понимать расстояние от линзы (или сферического зеркала) до точки геометрического схождения лучей. Например, если лучи сходятся в главном фокусе, то r_0 равно фокусному расстоянию f линзы или зеркала. Ввиду малости длины волны, условие (55.5) очень хорошо выполняется во всех оптических приборах. Так, при $f = 10$ см, $\lambda = 500$ нм из (55.5) получаем $\beta > 4 \cdot 10^{-3}$ рад $\sim 15'$. Поэтому применимость выведенных здесь формул к оптическим приборам с линзами и зеркалами не вызывает сомнений.

Для теории оптических приборов наибольшее значение имеет случай круглой диафрагмы. Исследование этого случая, конечно, не встречает каких-либо затруднений. Качественно ясно, что

дифракционная картина в фокальной плоскости должна иметь вид светлых и темных концентрических кругов со светлым центром (см. рис. 180). Для определения размеров дифракционных кругов надо вычислить интеграл (55.2). В случае круглой диафрагмы результат вычисления выражается через бесселеву функцию первого порядка. Радиусы темных колец имеют следующие значения:

$$R = 0{,}61\lambda/\beta; \quad 1{,}12\lambda/\beta; \quad 1{,}62\lambda/\beta; \ldots, \tag{55.6}$$

где β — угловой размер радиуса диафрагмы, если его рассматривать из точки O.

3. Найдем теперь распределение интенсивности света на оси MO, предполагая, что пучок лучей ограничен круглой диафрагмой. В этом случае в (55.2) можно взять $d\Omega = 2\pi \sin\vartheta\, d\vartheta$ (см. рис. 210). Если точка P лежит перед фокусом, то $\mathbf{nR} = -R\cos\vartheta$. Если же она лежит за фокусом, то $\mathbf{nR} = R\cos\vartheta$. По формуле (55.2)

$$E_P = 2\pi e^{i\omega t} \int\limits_0^\beta e^{\pm ikR\cos\vartheta} \sin\vartheta\, d\vartheta = \frac{2\pi e^{i(\omega t \pm kR)}}{\pm ikR}[1 - e^{\pm ikR(\cos\beta - 1)}],$$

где верхний знак относится к случаю, когда точка наблюдения расположена левее, а нижний — правее фокуса. Вычислив модуль этого выражения, находим амплитуду колебаний поля на оси пучка:

$$a \sim \frac{1}{R} \sin\frac{kR(1-\cos\beta)}{2}. \tag{55.7}$$

Во всякой плоскости, перпендикулярной к оси пучка, дифракционная картина имеет вид концентрических колец. Однако в зависимости от положения этой плоскости центр пучка может быть и светлым, и темным. Амплитуда на оси обращается в нуль, когда $kR(1-\cos\beta) = m\pi$, т.е.

$$R = \frac{m\lambda}{1-\cos\beta} \qquad (m = 1, 2, 3, \ldots). \tag{55.8}$$

В этом случае, как легко убедиться, в неприкрытой части сферического волнового фронта содержится четное число зон Френеля, а потому центр колец и получается темным. Расстояние между ближайшими минимумами интенсивности, расположенными по разные стороны от фокуса, равно

$$\delta = \frac{2\lambda}{1-\cos\beta} \approx \frac{4\lambda}{\beta^2}. \tag{55.9}$$

Эта величина может служить мерой продольных размеров области, в которой концентрируется свет вблизи фокуса. Объем этой области порядка $\dfrac{4\lambda}{\beta^2} \cdot \left(\dfrac{\lambda}{\beta}\right)^2 = \dfrac{4\lambda^3}{\beta^4}$.

§ 56. Разрешающая способность телескопа и микроскопа

1. Будем предполагать, что оптические системы идеальны в смысле геометрической оптики, т.е. дают строго точечные изображения каждой точки объекта. По волновой оптике это не так. Как показано в предыдущем параграфе, изображением светящейся точки в сопряженной плоскости является дифракционная картина, состоящая из концентрических колец, окружающих центральный светлый дифракционный кружок. Распределение интенсивности в такой картине представлено сплошной кривой на рис. 181. Основная доля энергии света (около 84 %) приходится на центральный дифракционный кружок. Этот кружок и будет изображением светящейся точки, если пренебречь энергией, приходящейся на окружающие его дифракционные кольца. Он называется *кружком Эйри* по имени ученого, впервые рассчитавшего дифракционную картину Фраунгофера от круглого отверстия.

Конечный объект можно рассматривать как совокупность точечных источников, каждый из которых изображается кружком Эйри (1801–1892) с окружающими его дифракционными кольцами. Изображение объекта есть наложение таких кружков и дифракционных колец. Задача теории сводится к расчету распределения интенсивности света в такой картине. Следует различать два предельных случая: 1) *точечные источники некогерентны*; 2) *точечные источники когерентны*. В первом случае складываются *интенсивности* волновых полей, во втором — их *напряженности*. Приближенно первый случай реализуется для *самосветящихся*, второй — для освещаемых объектов. Первый случай имеет основное значение в теории телескопа, а второй — микроскопа.

В простейшем случае сложный объект состоит из двух точечных источников S_1 и S_2 (рис. 211). Если расстояние между центрами кружков Эйри от этих точечных источников мало по сравнению с размерами самих кружков, то распределение интенсивности света в изображении будет мало отличаться от того, которое получилось бы только от одного точечного источника. Глаз не сможет решить, является ли объект точечным или состоит из нескольких близко расположенных светящихся точек. Говорят, что в этом случае светящиеся точки *не разрешаются* оптическим прибором. При увеличении расстояния между S_1 и S_2 будет увеличиваться и рас-

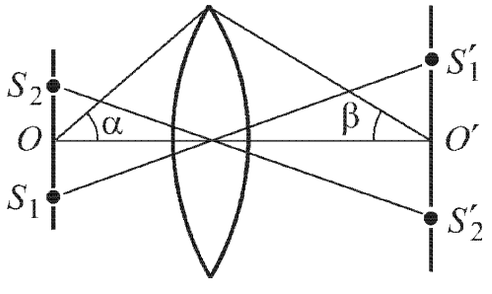


Рис. 211

стояние между центрами соответствующих им кружков Эйри S'_1 и S'_2, тогда как размеры самих кружков останутся неизменными. Начиная с некоторого минимального расстояния $l = l_{\text{мин}}$ на кривой распределения интенсивности в ее центре появится провал, который сможет зарегистрировать глаз или другой приемник света. Тогда говорят, что светящиеся точки S_1 и S_2 *разрешаются* оптическим прибором. Величина $l_{\text{мин}}$ называется *разрешаемым расстоянием прибора* (объектива), а обратная ему величина $1/l_{\text{мин}}$ — *разрешающей способностью.*

Точно указать величину $l_{\text{мин}}$ невозможно. Она в значительной степени зависит от индивидуальных особенностей глаза или другого приемника излучения. Речь может идти только о рациональной оценке $l_{\text{мин}}$. Согласно Рэлею, за $l_{\text{мин}}$ принимается такое расстояние между светящимися точками S_1 и S_2, при котором расстояние между центрами кружков Эйри равно радиусу одного кружка. В этом случае центр дифракционной картины от одной светящейся точки будет накладываться на первый дифракционный минимум дифракционной картины от второй светящейся точки. Такой критерий аналогичен критерию спектрального разрешения, предложенному также Рэлеем (см. § 47).

2. Для исследования вопроса о применимости критерия Рэлея к различным условиям наблюдения надо рассчитать распределение интенсивности света в плоскости изображения двух точечных источников света S_1 и S_2. С этой целью можно воспользоваться формулами предыдущего параграфа. Если источники S_1 и S_2 некогерентны, то надо сложить интенсивности создаваемых ими дифракционных картин. Если же они когерентны, то надо складывать волновые поля, а затем вычислять интенсивность результирующего волнового поля. Распределение интенсивности света в плоскости изображения в случае одной светящейся точки представлено сплошной кривой на рис. 181. Для нахождения распределения интенсивности света в случае двух одинаковых светящихся точек S_1 и S_2 применим графический метод.

Рассмотрим сначала случай, когда источники S_1 и S_2 некогерентны. На рис. 212 штриховые кривые представляют распределение интенсивности в плоскости изображения для каждого из точечных источников. Расстояние между центрами дифракционных картин предполагается равным $0{,}61\lambda/\beta$, т.е. радиусу кружка Эйри, как этого требует критерий Рэлея на пределе разрешения (угол β имеет тот же смысл, что и в предыдущем параграфе). Результирующая интенсивность получается сложением ординат обеих штриховых кривых и изображена сплошной линией. В центре C она имеет провал, где интенсивность составляет 74 % от максимальной интенсивности. Между тем средний

глаз уверенно обнаруживает наличие провала уже тогда, когда интенсивность в центре составляет около 85 %. Таким образом, можно считать, что при выбранном расстоянии между светящимися точками получится разрешение. С другой стороны, сближая штриховые кривые, легко установить, что провал исчезает, когда расстояние между центрами кружков Эйри уменьшает-

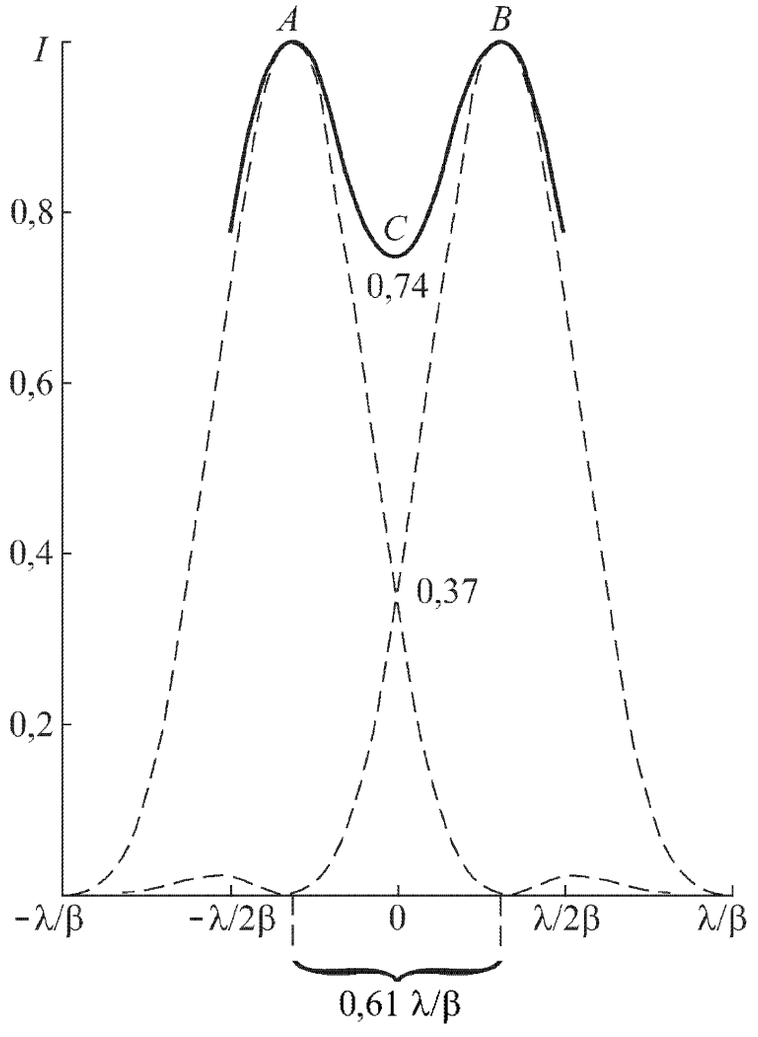


Рис. 212

ся до λ/β, что очень мало отличается от рэлеевского предела $0{,}61\lambda/\beta$. В этом случае разрешения нет. Таким образом, в случае некогерентных точечных источников света при визуальном наблюдении критерий Рэлея дает практически необходимое и достаточное условие для их разрешения.

3. Рассмотрим теперь случай, когда источники S_1 и S_2 когерентны. В качестве таких источников можно взять, например, два малых отверстия в непрозрачном экране, освещаемые когерентным светом (рис. 213). Такие отверстия ведут себя как точечные когерентные источники, излучающие вторичные волны Гюйгенса. Пусть отверстия освещаются пучком параллельных лучей, наклоненных под углом ϑ к нормали к плоскости

экрана. Если l — расстояние между центрами отверстий, то разность фаз между волнами, приходящими в эти отверстия, будет $\Delta = (2\pi/\lambda) l \sin\vartheta$. Такова же будет и разность фаз между колебаниями, приходящими в центр дифракционной картины, т.е. в точку, расположенную посередине между центрами кружков Эйри от источников S_1 и S_2. Интенсивность света в этой точке представится выражением

$$I_C = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos\Delta = \\ = 2I_1(1+\cos\Delta) = 4I_1\cos^2\frac{\Delta}{2}.$$

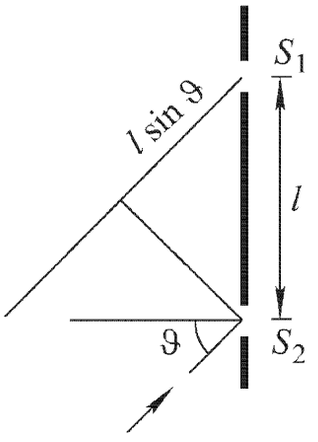


Рис. 213

Принципиальное отличие от случая некогерентных источников света состоит в появлении дополнительного — интерференционного — члена. Это сказывается на критерии разрешения и на разрешающей способности объектива. Меняя угол наклона ϑ, можно повысить разрешающую способность объектива.

Если свет, освещающий отверстия S_1 и S_2, падает нормально к плоскости экрана, то $\Delta = 0$ и, следовательно, $I_C = 4I_1$. Допустим, что в этом случае расстояние между центрами дифракционных кружков Эйри равно радиусу одного кружка, т.е. наименьшему расстоянию, которое требуется критерием Рэлея для разрешения светящихся точек. Тогда, как мы видели, интенсивность I_1 составляет 37 % от максимальной интенсивности, которую создает в центре кружка Эйри только один источник света. Поэтому результирующая интенсивность в центре дифракционной картины будет равна $4 \cdot 37 = 148\,\%$ от той же максимальной интенсивности. Распределение интенсивности света для рассматриваемого случая представлено на рис. 214. Штриховые кривые по-прежнему представляют интенсивности от каждого из источников S_1 и S_2 в отдельности, а сплошная кривая — результирующую интенсивность. На кривой результирующей интенсивности имеется всего лишь один максимум C. Поэтому никакого раздельного изображения светящихся точек в рассматриваемом случае получиться не может. Для этого необходимо развести центры кружков Эйри на расстояние, превышающее рэлеевский предел примерно в 1,2 раза.

Перейдем теперь к рассмотрению наклонного освещения. Если $\Delta = \pi/2$, т.е. разность хода между лучами, освещающими отверстия S_1 и S_2, равна $\lambda/4$, то $\cos\Delta = 0$. Интерференционный член пропадает, и получается в точности такое же распределение интенсивности света, а с ним и тот же критерий разрешения, что и в случае некогерентных светящихся точек. Если $\Delta = \pi$, а

следовательно, $l = \lambda/(2\sin\vartheta)$, что возможно лишь при $l \geqslant \lambda/2$, то в центре дифракционной картины всегда получается нулевая интенсивность, а потому разделение максимумов выражено осо-

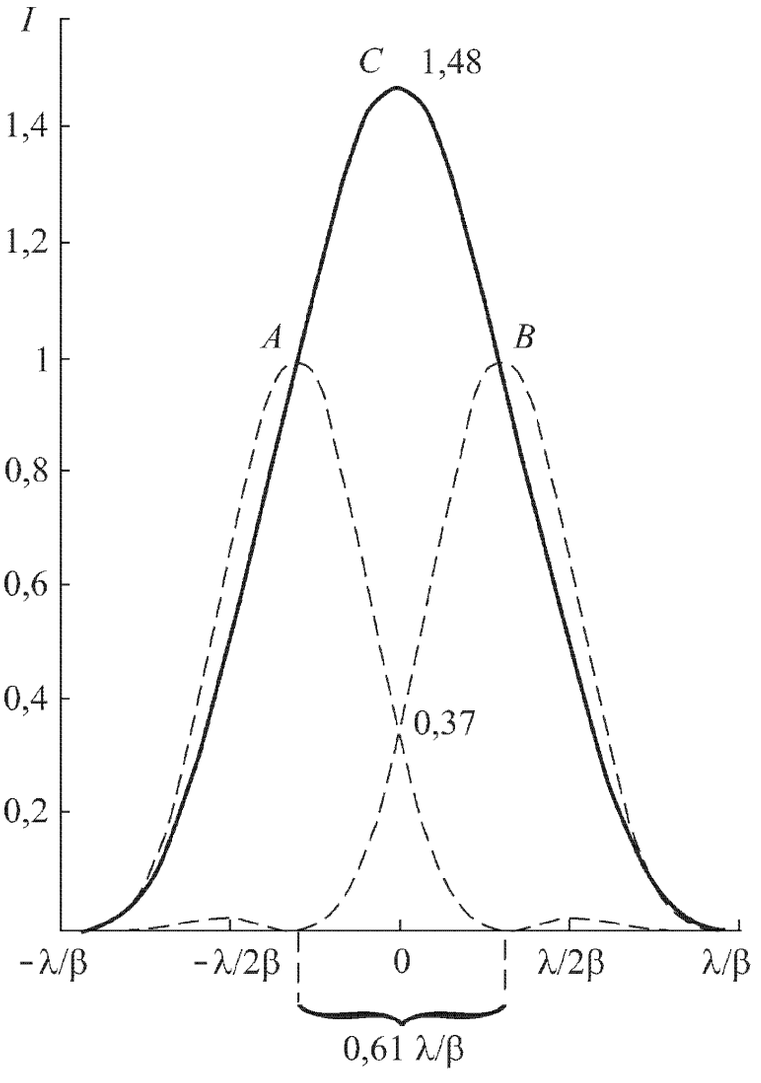


Рис. 214

бенно отчетливо. На рис. 215 представлено распределение интенсивностей для рассматриваемого случая в предположении, что центры кружков Эйри разведены до рэлеевского предела. При скользящем падении лучей эти центры можно сблизить примерно вдвое по сравнению с рэлеевским пределом, и все же провал на кривой интенсивности сохранится. Таким образом, применение наклонного освещения приводит к повышению разрешающей способности. При скользящем падении разрешающая способность возрастает почти вдвое по сравнению со случаем некогерентных светящихся точек.

Легко теперь сообразить, что следует ожидать при освещении объекта лучами всевозможных направлений от протяженного источника света. Нормально падающим лучам соответствует наименьшая разрешающая способность и наибольшее разрешаемое расстояние, превосходящее соответствующее расстояние для

некогерентных светящихся точек. При возрастании угла наклона лучей разрешаемое расстояние уменьшается. При угле наклона, когда разность фаз Δ становится равной $\pi/2$, оно такое же, что и для некогерентных источников. При дальнейшем увеличении наклона разрешаемое расстояние продолжает монотонно уменьшаться и при скользящем падении становится почти вдвое

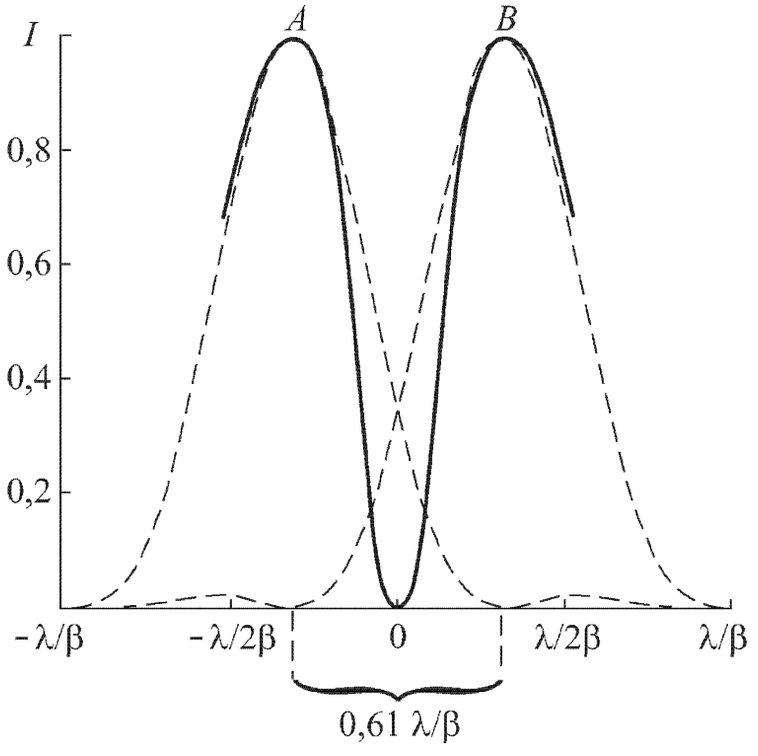


Рис. 215

меньше соответствующего расстояния для некогерентных источников. Естественно поэтому ожидать, и это подтверждается расчетом, что при освещении широкими пучками лучей получится такое же распределение интенсивности света в дифракционной картине и такой же критерий разрешения, что и при некогерентных источниках света.

Таким образом, разрешаемое расстояние меняется с изменением условий освещения и должно рассматриваться как ориентировочное. При этом надо еще раз подчеркнуть, что полученные результаты относятся к визуальным наблюдениям. Применение других физических методов исследования может понизить разрешаемый предел.

4. В случае телескопа рассматриваемые объекты, например компоненты двойной звезды, всегда излучают *некогерентно.* Для телескопа интерес представляет не линейное, а *угловое разрешаемое расстояние.* Если ϑ — угловое расстояние между рассматриваемыми звездами (рис. 216), то расстояние между центрами соответствующих кружков Эйри в фокальной плоскости будет $x = S_1'S_2' = f\vartheta$, где f — фокусное расстояние объектива.

Согласно критерию Рэлея, для разрешения звезд это расстояние должно быть не меньше $0{,}61\lambda/\beta$, т.е. $f\vartheta \geqslant 0{,}61\lambda/\beta$. Учитывая, что для небольших углов $f\beta = D/2$, где D — диаметр объектива, отсюда получим

$$\vartheta \geqslant 1{,}22\frac{\lambda}{D}. \qquad (56.1)$$

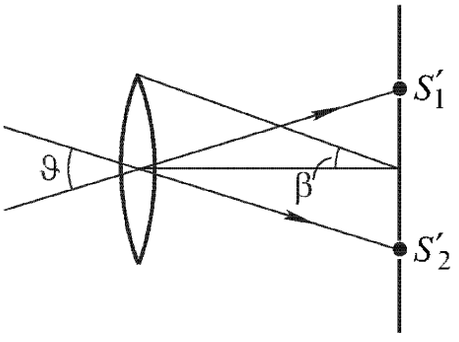


Рис. 216

Разрешаемое угловое расстояние равно

$$\vartheta_{\text{мин}} = 1{,}22\frac{\lambda}{D}. \qquad (56.2)$$

Глаз при рассматривании удаленных предметов действует принципиально так же, как объектив телескопа. Поэтому формулы (56.1) и (56.2) применимы и к глазу. Роль D играет диаметр зрачка глаза d. Полагая $d = 4$ мм, $\lambda = 550$ нм, находим для разрешаемого расстояния человеческого глаза

$$\vartheta_{\text{гл}} = 1{,}22\frac{\lambda}{d} = 1{,}67 \cdot 10^{-4} \text{ рад} = 35''. \qquad (56.3)$$

Этот результат удивительно хорошо согласуется с физиологической оценкой разрешающей способности глаза (см. § 21, п. 7). Достойна восхищения способность живого организма приспосабливаться к окружающим условиям и в процессе эволюции достигать максимума того, что принципиально допускается законами природы.

Таким образом, разрешающая способность телескопа пропорциональна диаметру его объектива. Крупнейший в мире действующий телескоп-рефлектор (см. § 24) имеет диаметр параболического зеркала $D = 6$ м. Теоретическая разрешающая сила его превосходит разрешающую силу глаза в $6000/4 = 1500$ раз. Теоретическое разрешаемое расстояние составляет $35''/1500 = 0{,}023''$. Для второго по величине телескопа-рефлектора Маунт-Паломарской обсерватории с диаметром зеркала 5 м теоретическое разрешаемое расстояние равно $0{,}028''$. Разрешающей способности таких гигантских телескопов достаточно, чтобы изображения звезд с наибольшими угловыми размерами получались уже в виде дисков, подобно изображениям планет.

Однако наличие земной атмосферы не позволяет полностью использовать разрешающую способность телескопа. Нерегулярные процессы в атмосфере, сопровождающиеся изменениями показателя преломления на пути световых лучей, портят изображения и снижают реальную разрешающую способность телескопов до величин порядка $1'' - 0{,}5''$, причем особенно сильно это снижение сказывается на больших телескопах. Такая разре-

шающая способность может быть достигнута с помощью много меньших объективов. Большие телескопы строятся не с целью повышения разрешающей способности, а для *увеличения количества света*, поступающего в телескоп от наблюдаемых небесных объектов. Оно пропорционально площади отверстия объектива. Поэтому с помощью больших телескопов можно обнаружить и сфотографировать более слабые небесные объекты, чем с помощью малых. Для повышения же разрешающей способности астрономических телескопов необходимо исключить вредное влияние атмосферы. Большие перспективы открываются перед телескопами, устанавливаемыми на космических кораблях, а еще лучше — на Луне.

5. Увеличение телескопа должно быть согласовано с разрешающей способностью его объектива. Допустим, что угловой размер предмета равен минимальному углу $\vartheta_{\text{мин}}$, разрешаемому объективом телескопа. Если угловое увеличение телескопа равно N, то в телескоп предмет будет виден под углом $N\vartheta_{\text{мин}}$. Для полного использования разрешающей способности телескопа необходимо, чтобы этот угол разрешался глазом, т.е. должно быть $N\vartheta_{\text{мин}} \geqslant \vartheta_{\text{гл}}$. Подставляя сюда значения $\vartheta_{\text{мин}}$ и $\vartheta_{\text{гл}}$ из формул (56.2) и (56.3), получим

$$N \geqslant \frac{D}{d} \equiv N_{\text{норм}}, \tag{56.4}$$

где $N_{\text{норм}}$ — нормальное увеличение телескопа (см. § 23, п. 4).

В § 11 (п. 10) было показано, что $N = D/h$, т.е. увеличение телескопа равно отношению ширины D падающего параллельного пучка лучей к ширине h выходящего пучка. Когда увеличение меньше нормального, т.е. $D/H < D/d$, то $h > d$. Значит, не весь пучок параллельных лучей, падающий на объектив, попадает в зрачок глаза, а только часть его. Действующей является только центральная часть объектива, диаметр которой меньше D. Телескоп действует так, как если бы его объектив и разрешающая способность уменьшились. При нормальном увеличении $h = d$, разрешающая способность объектива согласована с разрешающей способностью глаза. Наконец, когда увеличение больше нормального, т.е. $D/h > D/d$, то $h < d$. В этом случае глаз действует так, как если бы его зрачок сузился, а разрешающая способность понизилась.

Применение увеличений больше нормального не может выявить новые детали рассматриваемого объекта. Увеличения, значительно превосходящие нормальное, приносят только вред, поскольку чрезмерное сужение выходящего пучка может внести в изображение значительные дифракционные искажения. Однако по физиологическим причинам иногда бывает полезно брать увеличения, превосходящие нормальное в 2–4 раза. Рассматри-

вая детали на пределе разрешающей способности, глаз работает с напряжением и быстро утомляется, его чувствительность и разрешающая способность понижаются. Применение увеличений, несколько превышающих нормальное, не выявляет новых подробностей в строении рассматриваемых объектов, но позволяет рассматривать их с меньшим напряжением. К таким же выводам мы пришли в § 23 из фотометрических соображений.

6. Вопрос о разрешающей способности микроскопа решается так же, как и для телескопа. Критерий разрешения Рэлея в этом случае требует

$$l' \geqslant 0{,}61\frac{\lambda}{\beta}, \tag{56.5}$$

где l' — расстояние между центрами кружков Эйри для рассматриваемых двух светящихся точек (предполагается, что апертурный угол со стороны изображения 2β мал). Чтобы перейти к линейным размерам самого объекта, следует воспользоваться условием синусов Аббе, которое должно выполняться для объектива всякого микроскопа. При малых углах β (рис. 217) оно может быть написано в виде $ln\sin\alpha = l'\beta$, где n — показатель преломления пространства объектов, а 2α — апертурный угол со стороны этого пространства. Исключая β и l', приходим к соотношению

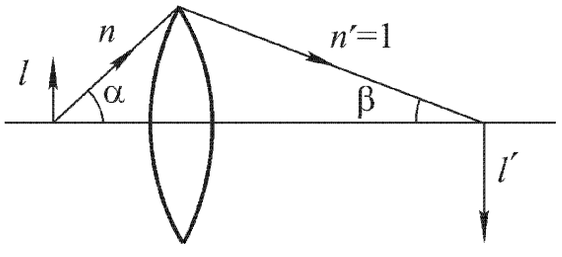


Рис. 217

$$l \geqslant 0{,}61\frac{\lambda}{n\sin\alpha}, \tag{56.6}$$

которое и определяет предел разрешения микроскопа. Минимальное разрешаемое расстояние (с точностью до небольшого численного коэффициента) равно

$$l_{\text{мин}} = 0{,}61\frac{\lambda}{n\sin\alpha}. \tag{56.7}$$

Этот предел разрешения определяется *волновой природой света* и не может быть превзойден никакими техническими усовершенствованиями микроскопа. При прочих равных условиях предел разрешения тем меньше, чем короче длина волны. Наличие в знаменателе показателя преломления n объясняется тем, что разрешаемое расстояние может непосредственно зависеть от длины волны не в вакууме, а в той среде, *где находится объект*, т.е. от величины λ/n.

Нетрудно проверить, что при малых апертурах формула (56.7) переходит в формулу (56.2), определяющую разрешающую способность телескопа. В этом случае обе формулы отличаются одна от другой только по форме. В формуле (56.7)

разрешаемое расстояние выражено в линейных единицах, а в формуле (56.2) — в угловых.

Для повышения разрешающей способности микроскопа можно идти двумя путями: 1) уменьшать длину волны (переход к ультрафиолету); 2) увеличивать числовую апертуру объектива микроскопа $n\sin\alpha$. Таким образом, угол α должен быть как можно больше. В лучших современных объективах он практически достиг своего теоретического предела $\alpha = \pi/2$. Для повышения числовой апертуры применяют также иммерсию, т.е. жидкость с возможно высоким показателем преломления, заполняющую пространство между покровным стеклом и фронтальной линзой объектива (см. § 18, п. 4). Если принять $n \approx 1{,}5$, то максимальное значение числовой апертуры будет $n\sin\alpha \approx 1{,}5$. Применение иммерсии, таким образом, позволяет снизить разрешаемый предел примерно в полтора раза, т.е. довести его до величины $l \approx 0{,}61\lambda/1{,}5 \approx 0{,}4\lambda$. Детали объекта, размеры которых меньше примерно $0{,}4\lambda$, принципиально не могут быть выявлены с помощью микроскопа. Ни при каком увеличении нельзя определить форму объекта, рассматривая его в микроскоп, если размеры самого объекта меньше приблизительно $0{,}4\lambda$. Разумеется, с помощью микроскопа можно обнаружить объекты и много меньших размеров, если только они светятся достаточно ярко (ультрамикроскоп). Нельзя определить только их форму.

Радикальное повышение разрешающей способности достигается в электронном микроскопе, где роль световых лучей выполняют электроны, фокусируемые электрическими и магнитными линзами. Электроны ведут себя подобно волнам с длиной волны

$$\lambda = \frac{h}{p} = \frac{h}{\sqrt{2m\mathcal{E}}},$$

где h — постоянная Планка, m — масса, p — импульс, $\mathcal{E}$ — энергия электрона. Эти волны называются *волнами де Бройля* (р. 1892). При $\mathcal{E} = 10\,000$ эВ приведенная формула дает $\lambda = 0{,}0122$ нм. Такая длина волны на порядок меньше размеров атома. Однако такое разрешение в электронном микроскопе не достигается, так как для уменьшения геометрических аберраций в электронных микроскопах приходится пользоваться узкими параксиальными пучками с малыми апертурами. Но даже в таких пучках электронный микроскоп позволяет разрешать детали, превосходящие размеры атомов всего в несколько раз.

7. Рассмотрим в заключение вопрос о рациональном увеличении микроскопа. Напомним, что увеличением микроскопа (лупы) называется отношение угла, под которым виден объект в микроскоп, к тому углу, под которым он был бы виден невооруженным глазом, если бы был помещен на расстоянии ясного зрения. Пусть $l_{\text{мин}}$ — минимальное расстояние, разрешаемое ми-

кроскопом. Невооруженным глазом с расстояния ясного зрения оно видно под углом $\vartheta = l_{\text{мин}}/L$. В микроскоп то же расстояние видно под углом $\vartheta' = N\vartheta$. Для разрешения необходимо, чтобы угол ϑ' был не меньше минимального угла, разрешаемого глазом, т.е. $\vartheta' = N\vartheta \geqslant \vartheta_{\text{гл}}$. Отсюда, используя выражения (56.3) и (56.7), найдем

$$N \geqslant \frac{2Ln\sin\alpha}{d}. \tag{56.8}$$

Знаку равенства соответствует нормальное увеличение

$$N_{\text{норм}} = \frac{2Ln\sin\alpha}{d}. \tag{56.9}$$

Как и в случае телескопа, нормальное увеличение микроскопа есть наименьшее увеличение, при котором может быть использована вся разрешающая способность объектива. О целесообразности работать при увеличениях больше нормального для микроскопа можно повторить без всяких изменений все, что выше было сказано относительно телескопа.

Максимальная числовая апертура, как указывалось выше, для сухих систем имеет своим пределом единицу. Для иммерсионных систем этот предел равен примерно 1,5. Диаметр зрачка глаза d примем равным 2 мм. Тогда для нормального глаза ($L = 25$ см) получатся следующие предельные значения нормальных увеличений:

для сухих систем $N_{\text{норм}} = 250$,

для иммерсионных систем $N_{\text{норм}} = 375$.

По физиологическим соображениям имеет смысл переходить к несколько большим увеличениям. Однако бессмысленно строить микроскопы с увеличением больше, чем в 1000–1500 раз.

§ 57. Теория и демонстрационные опыты Аббе

1. Впервые предел разрешения объектива микроскопа был найден Гельмгольцем в 1874 г. Гельмгольц рассматривал *самосветящиеся объекты*. Примерно в то же время и независимо от Гельмгольца вопрос о разрешающей способности микроскопа был разобран Аббе, но для случая *освещаемых объектов*. Именно этот случай встречается в практике микроскопии. От освещения очень сильно зависит качество изображения в микроскопе. Этот вопрос подвергся глубокому изучению Д.С. Рождественским (1876–1940) в последние годы его жизни. Ниже излагается метод Аббе с некоторыми изменениями и затрагиваются дополнительные вопросы, связанные с этим методом.

Допустим сначала, что объектом является достаточно большая (бесконечная) плоская дифракционная решетка, освещаемая параллельным пучком монохроматических лучей. Свет, прошедший через решетку, будет состоять из дискретного ряда плоских волн (дифракционных пучков или спектров различных порядков), распространяющихся от решетки в различных направлениях. Пучки низких порядков будут однородными волнами, а пучки, порядок которых превосходит определенное значение, — неоднородными (см. § 52). Неоднородные волны, затухая на расстояниях порядка длины волны, в объектив не попадают и потому не влияют на изображение, даваемое им.

2. Допустим, что перед объективом поставлена диафрагма, пропускающая спектры определенных порядков. Если диафрагма пропускает пучок одного лишь нулевого порядка, то прошедший свет будет состоять только из одной однородной плоской волны, как если бы на месте дифракционной решетки стояла прозрачная идеально плоскопараллельная пластинка. Никакой информации о структуре объекта такой свет не несет. В плоскости изображения, сопряженной с плоскостью решетки, получится равномерно освещенное поле.

Возьмем теперь такую диафрагму, чтобы она пропускала только два спектра соседних порядков, например m-го и $(m+1)$-го (рис. 218). За решеткой эффективными останутся только две однородные волны, распространяющиеся в различных направлениях:

$$E_m = a_m \cos(\omega t - \mathbf{k}_m \mathbf{r}),$$
$$E_{m+1} = a_{m+1} \cos(\omega t - \mathbf{k}_{m+1} \mathbf{r}),$$

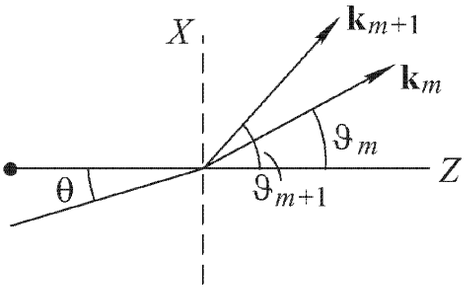


Рис. 218

интерферирующие между собой. Плоскость решетки примем за координатную плоскость XY, направив ось X перпендикулярно к ее штрихам, а ось Z — в сторону распространения падающего света. Интерференционные полосы во всякой плоскости $z = \text{const}$, параллельной плоскости решетки, будут параллельны ее штрихам. Найдем расстояние Δx между ними. Разность фаз между рассматриваемыми волнами в указанной плоскости представляется выражением

$$\varphi = (\mathbf{k}_{m+1} - \mathbf{k}_m)\mathbf{r} = (k_{m+1,x} - k_{m,x})x + (k_{m+1,z} - k_{m,z})z.$$

При изменении x на Δx она изменится на $\Delta\varphi = (k_{m+1,x} - k_{m,x})\Delta x$. Интенсивность света будет периодически повторяться, когда $\Delta\varphi = 2\pi, 4\pi, \ldots$ Ширина интерференционной полосы Δx найдется, если положить $\Delta\varphi = 2\pi$. Направления на дифракционные максимумы рассматриваемых порядков определяются

формулами

$$d(\sin\vartheta_m - \sin\theta) = m\lambda, \quad d(\sin\vartheta_{m+1} - \sin\theta) = (m+1)\lambda,$$

где θ — угол, под которым свет падает на решетку. Отсюда $d(\sin\vartheta_{m+1} - \sin\vartheta_m) = \lambda$. Но $k_{mx} = (2\pi/\lambda)\sin\vartheta_m$, $k_{m+1,x} = (2\pi/\lambda)\sin\vartheta_{m+1}$, так что

$$k_{m+1,x} - k_{mx} = (2\pi/\lambda)(\sin\vartheta_{m+1} - \sin\vartheta_m) = 2\pi/d.$$

Следовательно, ширина интерференционной полосы будет $\Delta x = 2\pi : (2\pi/d) = d$, независимо от направления освещающего пучка лучей. В частности, такая система интерференционных полос возникнет на выходе самой решетки, т.е. в плоскости $z = 0$. Объектив даст изображение этой системы полос в плоскости изображения, сопряженной с плоскостью решетки (см. § 27, п. 6). Такое изображение передает основной характер структуры решетки — ее периодичность с основным периодом d. Распределение интенсивности в плоскости изображения будет синусоидальным, т.е. таким же, какое получилось бы от синусоидальной решетки Рэлея. Чтобы изображение передавало более мелкие детали структуры объекта, необходимо расширить диафрагму. Тогда за решеткой появятся плоские дифрагированные волны, порядки которых отличаются друг от друга не только на единицу, но также на 2, 3, ... Рассуждая как выше, нетрудно показать, что интерференция двух волн, порядки которых отличаются на Δm, даст интерференционные полосы с шириной $\Delta x = d/\Delta m$.

Структуру решетки можно характеризовать разложением ее пропускаемости в ряд Фурье. Полностью подобное изображение со всеми деталями получилось бы, если бы интенсивность света в плоскости изображения представлялась рядом Фурье с теми же коэффициентами. Однако этого никогда не бывает. Из ряда Фурье выпадают компоненты, соответствующие неоднородным волнам и тем однородным волнам, которые задерживаются диафрагмой. Чем больше дифрагированных волн различных порядков проходит через диафрагму, тем совершеннее получается изображение.

3. Если при определении разрешающей способности довольствоваться наименее совершенным изображением, которое правильно передавало бы только периодическую структуру решетки с периодом d, то разрешающую способность объектива микроскопа можно определять следующим образом. Пусть решетка освещается нормально падающими лучами света (рис. 219 *а*). Тогда для разрешения необходимо, чтобы наряду с прямо прошедшим пучком света в объектив попали также дифракционные пучки первого и минус первого порядков. Если объектив иммерсионный с показателем иммерсии n, то минимальный период

решетки, при котором выполняется это условие, будет $d \sin \alpha = \lambda/n$, где λ — длина волны в вакууме. Отсюда и находится минимальное расстояние, разрешаемое объективом микроскопа:

$$l_{\text{мин}} = \frac{\lambda}{n \sin \alpha}. \tag{57.1}$$

Полученный результат отличается от (56.7) только несущественным числовым коэффициентом. Различие в коэффициентах связано с некоторым различием критериев разрешения, кото-

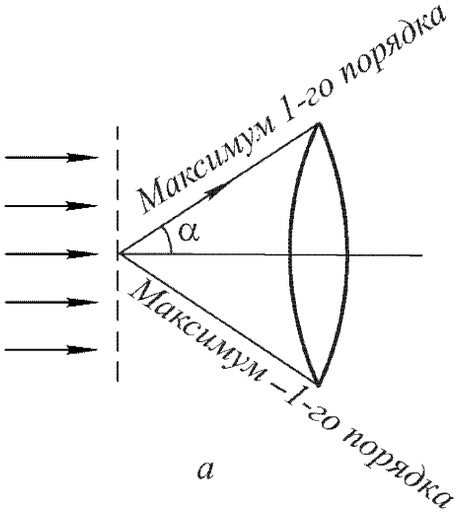


Рис. 219

рые использовались при выводе формул (56.7) и (57.1), а также с тем, что формула (56.7) относится к разрешению точечных, а (57.1) — линейных объектов.

Разрешаемый предел (57.1) можно понизить за счет применения наклонного освещения. При нормальном падении освещающего пучка за решеткой интерферируют три пучка. При интерференции пучков нулевого и одного из первых порядков в изображении выявляется структура с основным периодом d. Такая же картина интерференции получается при наложении нулевого пучка с другим пучком первого порядка. Второй пучок первого порядка просто усиливает действие другого пучка того же порядка. Кроме того, оба пучка первого порядка интерферируют между собой, в результате чего на интерференционную картину накладываются новые, более слабые, интерференционные полосы, расстояние между которыми вдвое меньше, т.е. равно $d/2$. В этом проявляется действие первых гармоник в разложении пропускаемости решетки в ряд Фурье. Не меняя основного периода в изображении решетки, они несколько усиливают его контраст. Но для выявления структуры с основным периодом d, как выяснено выше, достаточно *двух* соседних пучков, например, нулевого и одного из пучков первого порядка. Максимально возможное разрешение получится, если освещающий пучок направить

под углом α к решетке, как указано на рис. 219 *б*. Тогда будет $2d\sin\alpha = \lambda/n$ и, следовательно,

$$l_{\text{мин}} = \frac{1}{2}\frac{\lambda}{n\sin\alpha}, \tag{57.2}$$

что вдвое меньше величины (57.1).

4. Таким образом, разрешающую способность микроскопа можно повысить изменением направления освещающих лучей. Однако влияние способа освещения на изображение в микроскопе этим не ограничивается. Оно более существенно. Если бы освещение производилось пучком параллельных лучей, то, как было выяснено выше, интерференционная картина за решеткой в каждой параллельной ей плоскости П имела бы вид параллельных полос с расстоянием между ними d. Такие полосы *не локализованы*. Изображение, видимое в микроскоп, имело бы один и тот же вид независимо от того, куда сфокусирован микроскоп: на плоскость решетки или на любую плоскость, ей параллельную. Для устранения этого вредного обстоятельства в микроскопе объекты освещаются пучками лучей различных направлений с широкими апертурами. Интерференционные полосы, даваемые параллельными пучками различных направлений, смещены относительно друг друга вдоль оси X. Поэтому во всех плоскостях П происходит гашение интерференционных полос. Исключение составляет только случай, когда плоскость П совпадает с плоскостью самой решетки. В этом случае все интерференционные полосы, создаваемые освещающими пучками различных направлений, пространственно совпадают между собой. Поэтому изображение объекта в микроскопе может возникнуть только тогда, когда он сфокусирован на плоскость самого объекта. В этом основная причина использования широких апертур при освещении объектов в микроскопе.

5. В качестве объекта была взята дифракционная решетка лишь с целью упрощения рассуждений. Это упрощение не принципиальное. Если взять объект произвольной формы и осветить его параллельным пучком лучей, то за объектом возникнут дифрагированные пучки различных направлений с максимумами и минимумами. Угол дифракционной расходимости ϑ, определяющий направление на первый дифракционный минимум, определяется формулой $nl\sin\vartheta \sim \lambda$, где l — линейный размер объекта. Если диафрагма пропускает лишь незначительную часть центрального дифракционного пучка, то изображение в микроскопе получится таким же, как и в случае точечного объекта, излучающего сферические волны равномерно по всем направлениям. Это был бы дифракционный кружок Эйри, совершенно не зависящий от формы объекта, и никакого разрешения не получилось бы. Минимальные размеры объекта $l_{\text{мин}}$, при которых

возможно разрешение, определяются условием $\vartheta \sim \alpha$, т.е.

$$l_{\text{мин}} \sim \frac{\lambda}{n \sin \alpha}. \tag{57.3}$$

Мы снова пришли к формуле вида (57.1) или (57.2).

В приведенном рассуждении не столь существенно, куда попадает свет после рассеяния на объекте. Вместо объектива можно, например, взять фотопластинку и произвести на ней голографическую запись предмета. Из предыдущих рассуждений следует, что минимальный размер объекта, который может быть записан таким способом, определяется формулой

$$l_{\text{мин}} \sim \lambda/\alpha, \tag{57.4}$$

где α — угловой размер голограммы, если ее рассматривать из точки нахождения предмета во время голографирования (предполагается, что он невелик).

6. Сам Аббе представил изложенную теорию в следующей форме. Каждый пучок параллельных лучей, получившийся в результате дифракции на объекте AB (например, дифракционной решетке), собирается объективом в точке фокальной плоскости (рис. 220). Дифракционную картину в фокальной плоскости, состоящую из совокупности таких точек: C_0, C_1, C_{-1}, C_2, C_{-2}, ..., Аббе назвал *первичным изображением* объекта. Если известно первичное изображение, то с помощью принципа Гюйгенса–Френеля можно рассчитать световое поле во всем пространстве за фокальной плоскостью. В частности, изображение объекта есть не что иное, как интерференционная картина, получающаяся в плоскости изображения в результате интерференции вторичных волн Гюйгенса, исходящих из различных точек первичного изображения. Эту интерференционную картину Аббе назвал *вторичным изображением.*

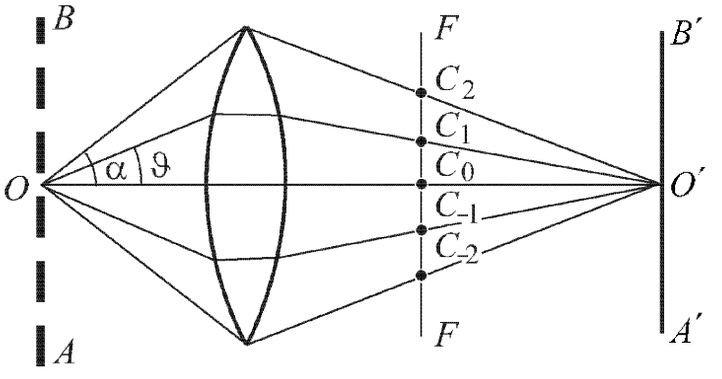


Рис. 220

Если диафрагма объектива пропускает один только центральный дифрагированный пучок, то первичное изображение будет состоять только из одного центрального точечного максимума. Получится лишь одна вторичная сферическая волна, исходящая из этого центрального максимума, которой не с чем будет интерферировать. Вторичное изображение получится бесструктурным и будет представлять собой более или менее равномерно освещенное поле. Для появления какой-то структуры во

вторичном изображении необходимо, чтобы первичное изображение состояло по крайней мере из двух точечных дифракционных максимумов. Но это условие лишь по форме отличается от того условия, из которого были выведены формулы (57.1) и (57.2).

Если закрыть некоторые дифракционные максимумы в фокальной плоскости, то вторичное изображение будет искажено. Характер искажения часто можно предсказать. Этим воспользовался Аббе для экспериментального подтверждения своей теории. Объектом служила грубая дифракционная решетка. Аббе помещал в фокальной плоскости другую решетку из проволок, которые закрывали дифракционные максимумы через один. Тогда расстояния между действующими (открытыми) максимумами возрастали вдвое. В изображении решетки наблюдалось удвоение штрихов. Все происходило так, как если бы экранирования не было, а объект был заменен более мелкой решеткой, расстояния между штрихами которой вдвое меньше.

Еще более интересные искажения наблюдаются в случае двумерной решетки, например, квадратной проволочной сетки. В этом случае дифракционная картина в фокальной плоскости состоит из светлых пятен, расположенных в узлах также квадратной сетки. Поместим в фокальной плоскости узкую щель, с помощью которой можно открывать прямолинейные ряды из таких пятен, а остальные пятна закрывать. Если щель горизонтальна и достаточно узка, то она выделит прямолинейный ряд максимумов, расположенных вдоль горизонтальной прямой

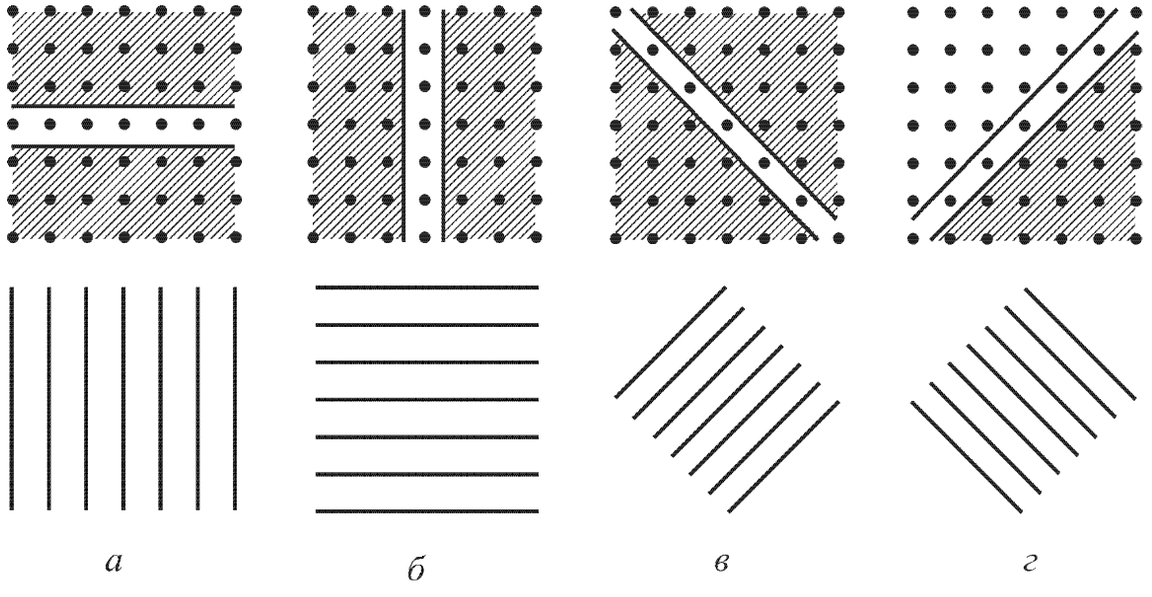


Рис. 221

(рис. 221 *а*). Такой ряд максимумов аналогичен дифракционной картине от одномерной решетки с вертикальными щелями. Поэтому оптическое изображение квадратной сетки при вве-

дении горизонтальной щели перейдет в систему вертикальных полос. Если щель повернуть на 90° в вертикальное положение (рис. 221 *б*), то полосы сделаются горизонтальными. Если щель повернуть параллельно диагонали сетки (рис. 221 *в* и *г*), то она выделит прямолинейный ряд максимумов, параллельный той же диагонали, причем расстояния между максимумами увеличатся в $\sqrt{2}$ раз. В результате оптическое изображение сетки перейдет в систему наклонных полос, перпендикулярных к щели, а сами полосы сделаются в $\sqrt{2}$ раз у́же.

Д.С. Рождественский указал, что непосредственной причиной появления ложных структур в опытах Аббе является *дифракция света на экранирующей сетке*. В отсутствие последней объектив дает геометрически подобное изображение объекта. Например, сетка из параллельных проволок изображается в виде системы параллельных полос. При введении экранирующей сетки эти полосы остаются на прежних местах. Но на них накладывается дифракционная картина, возникающая при дифракции света на экранирующей сетке. Если экранирующая сетка состоит из проволок, параллельных проволокам сетки-объекта, то возникнут дифракционные полосы, параллельные полосам в прежнем изображении сетки-объекта. При надлежащем периоде экранирующей сетки дифракционные полосы расположатся посередине между прежними полосами. Произойдет удвоение полос в изображении, как если бы проволоки в сетке-объекте были натянуты вдвое чаще. Так же, хотя и не столь просто, объясняется и появление диагоналей, когда объектом является квадратная проволочная сетка.

Дифракция на экранирующей сетке происходит независимо от того, освещается ли объект посторонним светом или является самосветящимся. Поэтому и в случае самосвечения должны появляться такие же ложные структуры, как и при освещении. Это было теоретически предсказано Л.И. Мандельштамом (1879–1944) еще в 1911 г. Опыты с накаленными сетками, поставленные самим Мандельштамом, подтвердили это предсказание.

§ 58. Телескоп без объектива. Получение изображений с помощью малых отверстий

1. В принципе можно построить телескоп без объектива сколь угодно высокой разрешающей способности. Роль объектива может играть круглое отверстие. Следующее рассуждение, принадлежащее Рэлею, разъясняет идею такого телескопа. Допустим сначала, что в отверстие вставлен объектив. Лучи, идущие от какой-либо точки объекта к ее изображению вдоль оптической оси и по периферии, имеют различные геометрические длины. При отсутствии объектива различны и их оптические длины. Объ-

ектив должен скомпенсировать своей толщиной различие оптических длин всех лучей, чтобы они приходили в точку-изображение в одинаковых фазах. Но небольшое расхождение в фазах лучей мало сказывается на результате их интерференции. Например, если интерферируют два одинаковых луча в одинаковых фазах, то результирующая интенсивность превосходит интенсивность отдельного луча в 4 раза. Если же разность фаз составляет $\pi/4$, то она будет превосходить в $2+\sqrt{2}=3{,}42$ раза, т.е. изменится незначительно. Поэтому разности фаз в $\pi/4$ или меньше можно не принимать во внимание. Допустим теперь, что фокусное расстояние объектива увеличивается все больше и больше. Различие в геометрических длинах центрального и крайнего периферийного лучей будет становиться все меньше и меньше. Когда это различие достигнет примерно $\lambda/8$, надобность в объективе отпадет. Действительно, в этом случае максимальная разность фаз между лучами, приходящими в точку-изображение от различных точек плоскости отверстия, не будет превышать $\pi/4$, и компенсация различия оптических длин становится практически не нужной. Отверстие будет действовать как объектив телескопа. При увеличении фокусного расстояния f разрешающая способность объектива не меняется и все время определяется формулой (56.2). Она останется неизменной и в предельном случае, когда $f=\infty$, т.е. когда объектив совсем удален.

Разность геометрических длин крайнего периферийного и центрального лучей составляет $D^2/(8l)$, где D — диаметр отверстия, а l — длина телескопа, которую можно считать равной фокусному расстоянию. Для оценки l приравняем эту разность $\lambda/8$ и получим

$$l=D^2/\lambda. \tag{58.1}$$

Эта формула показывает, что о практической реализации телескопа без объектива не может быть и речи ввиду колоссальных размеров, которые должен был бы иметь такой телескоп. Например, при $D=1$ м, $\lambda=500$ нм $=5\cdot10^{-7}$ м формула (58.1) дает $l=2\cdot10^6$ м $=2000$ км. При гигантских размерах телескоп обладал бы ничтожной светосилой: $(D/f)^2=l\lambda/f^2=\lambda/l\approx10^{-13}$.

2. К идее безлинзового телескопа примыкает способ получения изображений и фотографирования при помощи камеры с малым отверстием (камеры-обскуры). Она отличается от безлинзового телескопа только масштабами и возможностью практического осуществления. Поэтому приводимые ниже рассуждения относятся не только к камере с малым отверстием, но дополняют и уточняют идею безлинзового телескопа.

При уменьшении размеров отверстия четкость изображения в камере сначала улучшается, а затем начинает ухудшаться из-за дифракции. Дифракция несущественна при больших отверстиях, а при малых отверстиях становится основным фактором, определяющим четкость изображения. Оптимальные размеры отверстия, при которых достигается наибольшая четкость, легко оценить с помощью следующих соображений. Пусть отверстие имеет форму круга радиуса R. Расстояние до фотографируемого объекта может считаться бесконечно большим по сравнению с глубиной камеры l. Если бы была применима геометрическая оптика, то светящаяся точка изобразилась бы кружком того же радиуса R. Из-за дифракции точка изобразится дифракционным кружком, радиус которого порядка $\lambda l/R$. Уменьшать размеры отверстия имеет смысл лишь до тех пор, пока дифракционные ошибки не превзойдут геометрические. Наилучшая четкость изображения достигается при таких размерах отверстия, когда эти ошибки примерно

одинаковы, т.е. при выполнении условия $\lambda l/R \approx R$, или $R \approx \sqrt{l\lambda}$. Это значит, что размер отверстия должен быть порядка центральной френелевой зоны. Рэлей, более подробно исследовавший вопрос как теоретически, так и экспериментально, нашел для наивыгоднейшего радиуса отверстия:

$$R = 0{,}9\sqrt{\frac{ab\lambda}{a+b}}, \tag{58.2}$$

где a и b — расстояния предмета и его изображения от отверстия.

ЗАДАЧИ

1. При наблюдении в телескоп с нормальным увеличением освещенность изображения звезды на сетчатке глаза в $\alpha = 10$ раз меньше освещенности дневного неба, рассматриваемого в тот же телескоп. Во сколько раз надо увеличить диаметр объектива для того, чтобы освещенность изображения звезды на сетчатке стала в $\beta = 10$ раз больше освещенности изображения неба, если вместе с объективом телескопа заменен и окуляр таким образом, что увеличение телескопа осталось нормальным?

О т в е т. В $\sqrt{\alpha\beta} = 10$ раз.

2. В фокальной плоскости объектива телескопа помещена фотопластинка. Освещенность изображения звезды на фотопластинке в $\alpha = 10$ раз меньше освещенности дневного неба. Во сколько раз надо увеличить диаметр объектива, чтобы освещенность изображения звезды на фотопластинке стала в $\beta = 10$ раз больше освещенности изображения неба?

О т в е т. В $\sqrt{\alpha\beta} = 10$ раз.

3. Ракету, удаляющуюся от Земли, фотографируют, помещая фотопленку в фокальной плоскости объектива астрономического телескопа. При диаметре объектива $D_1 = 80$ мм дифракционное изображение ракеты становится неразличимым на фоне изображения неба, когда ракета находится на расстоянии $l_1 = 2 \cdot 10^4$ км от Земли. На каком расстоянии от Земли l_2 удастся заметить ракету на фотографии, полученной с помощью объектива с диаметром $D_2 = 200$ мм при той же контрастной чувствительности пленки?

О т в е т. $l_2 = l_1 D_2/D_1 = 5 \cdot 10^4$ км.

4. Как изменится разрешающая способность объектива телескопа, если его центральную часть закрыть круглым экраном, диаметр которого мало отличается от диаметра самого объектива?

У к а з а н и е. При решении задачи круглые объектив и экран можно заменить квадратными.

О т в е т. Разрешающая способность увеличится приблизительно в два раза.

5. Производится фотографирование удаленных предметов с помощью объектива телескопа на фотопластинке, помещенной в его фокальной плоскости. Полученный снимок с помощью окуляра того же телескопа проецируется на удаленный экран. Каково должно быть угловое увеличение телескопа, чтобы при этом была использована полностью разрешающая способность объектива телескопа? Изображение на экране рассматривается с того места, где установлен проекционный аппарат.

О т в е т. $N \geqslant D/d$, где D — диаметр объектива, а d — зрачка глаза.

6. С какого расстояния r можно увидеть невооруженным глазом свет лазера, генерирующего в непрерывном режиме $P = 10$ кВт на частоте $\omega = 4 \cdot 10^{15}$ с$^{-1}$, если для формирования луча используется параболическое

зеркало с диаметром $D = 5$ м? Глаз видит источник, если в зрачок диаметром $d = 5$ мм попадает в 1 с $n = 60$ квантов излучения, лежащего в зеленой части спектра.

О т в е т. $r \approx \dfrac{Dd}{2}\sqrt{\dfrac{P}{hc\lambda n}} \approx 0{,}1$ светового года (h — постоянная Планка.)

7. Излучение лазера непрерывного действия на волне $\lambda = 500$ нм мощностью $P = 1$ Вт направляется на спутник с помощью телескопа, диаметр объектива которого равен $D = 30$ см. Свет, рассеянный спутником, улавливается другим таким же телескопом и фокусируется на фотоумножитель с пороговой чувствительностью $P_{\text{пор}} = 10^{-14}$ Вт. При каких расстояниях l до спутника отраженный сигнал может быть обнаружен, если поверхность спутника равномерно рассеивает падающий на него свет (по закону Ламберта)? Диаметр спутника $d = 20$ см.

О т в е т. $l \lesssim \dfrac{D}{2}\sqrt[4]{\dfrac{Pd^2}{2P_{\text{пор}}\lambda^2}} \approx 250$ км.

8. В § 40 был описан способ Поля фотографирования предметов с помощью непрозрачного гладкого шара. Оценить минимальное угловое разрешаемое расстояние $\delta\varphi$ при фотографировании удаленных предметов этим способом.

У к а з а н и е. Размер светлого кружка в центре геометрической тени от точечного источника можно оценить из условия, что первый дифракционный минимум получается, когда разность хода лучей от противоположных точек шара порядка длины волны.

О т в е т. $\delta\varphi \approx \lambda/D$.

9. Существующие радиотелескопы и радиоустановки, предназначенные для изучения радиоизлучения Солнца и Галактики, обладают малой разрешающей способностью из-за больших длин волн радиоизлучения.

1) Найти минимальное разрешаемое угловое расстояние $\delta\varphi$ радиотелескопа с диаметром зеркала $d = 50$ м для длин волн $\lambda = 1$ м и $\lambda = 10$ см.

2) Для увеличения разрешающей способности предлагалось использовать дифракцию радиоволн от края Луны (см. задачу 2 к § 42). Оценить разрешающую способность этого метода для тех же волн в предположении, что край Луны действует как тонкий непрозрачный экран, ограниченный прямолинейным краем.

3) Оценить, каковы должны быть высоты h неровностей лунной поверхности, чтобы можно было пользоваться этим методом. Расстояние до Луны $b = 380\,000$ км.

О т в е т. 1) $\delta\varphi \approx \lambda/d$. При $\lambda = 1$ м $\delta\varphi \approx 1^\circ$; при $\lambda = 10$ см $\delta\varphi \approx 7''$.

2) $\delta\varphi \approx \sqrt{\lambda/b}$. При $\lambda = 1$ м $\delta\varphi \approx 2'$; при $\lambda = 10$ см $\delta\varphi \approx 40''$.

3) $h \ll \sqrt{b\lambda}$. При $\lambda = 1$ м $\sqrt{b\lambda} = 19{,}5$ км; при $\lambda = 10$ см $\sqrt{b\lambda} = 6{,}2$ км.

10. Один из принципиально возможных (но практически не осуществимых) способов повышения разрешающей способности радиоустановок для изучения радиоизлучения космических масс состоит в том, чтобы использовать дифракционный максимум интенсивности радиоизлучения, получающийся в центре геометрической тени Луны от точечного источника. Оценить разрешающую способность этого метода, найти минимальное угловое разрешаемое расстояние для тех же длин волн, что и в предыдущей задаче.

Обсудить возможности использования рассматриваемого метода. Диаметр Луны $D = 3470$ км.

О т в е т. $\delta\varphi \approx \lambda/D$. При $\lambda = 1$ м $\delta\varphi \approx 0{,}06''$. При $\lambda = 10$ см $\delta\varphi \approx 0{,}006''$.

Метод требует, чтобы источник радиоизлучения находился на прямой, соединяющей точку наблюдения с центром Луны. Кроме того, он предъявляет весьма жесткие требования к гладкости лунной поверхности и к отступлениям формы Луны от сферической. Высота неровностей поверхности по всей границе лунного диска должна быть мала по сравнению с $h = b\lambda/D$. Разность наибольшего и наименьшего диаметров лунного диска также не должна превышать этой величины. При $\lambda = 1$ м $h \approx 100$ м; при $\lambda = 10$ см $h \approx 10$ м. Эти жесткие требования исключают возможность использования рассматриваемого метода по крайней мере для радиоволн с длиной волны $\lambda < 100$ м.

11. С помощью объектива микроскопа получена микрофотография малого объекта (например, растительных клеток или бактерий) с линейным увеличением N. Тот же объектив был использован для проецирования полученной микрофотографии на удаленный экран. Каково должно быть минимальное значение увеличения N, чтобы полностью была использована разрешающая способность микроскопа? Диаметр апертурной диафрагмы объектива равен D, диаметр зрачка глаза d. Изображение на экране рассматривается с места нахождения объектива.

О т в е т. $N \geqslant D/d$.

12. Каково должно быть фокусное расстояние f_2 окуляра микроскопа, чтобы была полностью использована разрешающая способность объектива? Числовая апертура объектива равна $n \sin\alpha$, фокусное расстояние объектива f_1, длина тубуса (трубы микроскопа) l. Длину тубуса можно считать равной расстоянию между объективом и плоскостью первого изображения (т.е. изображения, даваемого объективом).

О т в е т. $f_2 \leqslant \dfrac{ld}{2f_1 n \sin\alpha}$, где d — диаметр зрачка глаза.

13. С помощью объектива телескопа с диаметром D и фокусным расстоянием f производится фотографирование удаленных объектов на мелкозернистой пластинке, помещенной в фокальной плоскости объектива. Полученное изображение рассматривается в микроскоп с числовой апертурой $n \sin\alpha$ и увеличением N. Каким условиям должны удовлетворять числовая апертура и увеличение микроскопа, чтобы полностью использовать разрешающую способность объектива телескопа?

О т в е т. $n \sin\alpha \geqslant D/(2f)$, $N \geqslant DL/(fd)$, где L — расстояние ясного зрения, d — диаметр зрачка глаза.

14. Современные фотопластинки способны разрешать до $z = 10^4$ линий на сантиметр. Какую светосилу (т.е. отношение квадратов диаметра D и фокусного расстояния f) должен иметь объектив фотоаппарата, чтобы полностью использовать разрешающую способность пленки?

О т в е т. $(D/f)^2 \geqslant z^2\lambda^2 \approx 0{,}25$.

§ 59. Фазовый контраст

1. В зависимости от вида изучаемых объектов при работе микроскопа следует различать два предельных случая. Одни объекты, называемые *абсорбционными*, в различных местах обладают различной прозрачностью. Такие объекты в основном влияют на *амплитуду* проходящего света. Дру-

гие объекты, называемые *рефракционными*, практически не поглощают света. Имея в различных местах различные толщины и показатели преломления, они влияют не на интенсивность, а на *фазу* проходящего света. Типичными примерами абсорбционного и рефракционного объектов могут служить амплитудная и фазовая дифракционные решетки.

Абсорбционные объекты дают *контрастные изображения* с хорошо выраженными границами между темными и светлыми частями. На них можно обнаружить все детали, которые способен разрешить микроскоп при заданной разрешающей способности. Напротив, изображения рефракционных объектов почти *лишены контраста*. В таких изображениях трудно, а часто и практически невозможно разрешить детали изучаемого объекта, хотя бы разрешающей способности микроскопа и было достаточно для этой цели. Причина такого различия между абсорбционными и рефракционными структурами состоит в том, что объектив микроскопа воспроизводит в плоскости изображения, а следовательно и на сетчатке глаза, то же распределение интенсивности светового поля, которое существует в плоскости объекта, а светочувствительные нервные окончания сетчатки реагируют именно на *интенсивность* световой волны, а не на ее фазу.

С рефракционными объектами постоянно приходится иметь дело в биологии при изучении хотя бы микроорганизмов. Биологические объекты в подавляющем большинстве случаев практически совершенно прозрачны в видимой области спектра. Отсутствие контраста в изображении затрудняет изучение таких объектов. Поэтому проблема контрастности изображения стоит в биологии особенно остро. Один из методов ее решения состоит в превращении рефракционных объектов в абсорбционные путем дифференциального окрашивания объекта. Однако такой метод не всегда возможен. Кроме того, он убивает живые организмы или по крайней мере нарушает их нормальную жизнедеятельность. Единственный метод изучения биологических объектов в естественных условиях состоит в том, чтобы воздействовать не на самый объект, а на *его изображение*. Это достигается в *методе фазового контраста*, предложенном Цернике (1888–1966) в 1934 г.

2. Идею метода фазового контраста проще всего выяснить на примере периодической структуры — одномерной дифракционной решетки. Различие между амплитудной и фазовой решетками с интересующей нас точки зрения по существу уже было выяснено в § 53. Остановимся на этом вопросе более подробно.

Отвлекаясь от поляризации, будем рассматривать свет как скалярное волновое поле и представлять световые колебания векторами на векторной диаграмме. Пусть свет падает нормально на поверхность решетки. Допустим сначала, что решетка амплитудная и состоит из чередующихся участков различной прозрачности, причем на участках I прозрачность больше, а на участках II — меньше. Для простоты (это несущественно для выяснения существа вопроса) предположим, что участки I и II имеют одинаковую ширину. Колебание на выходе участка I изобразится более длинной стрелкой $\mathbf{a}$, чем колебание на выходе участка II, представляемое стрелкой $\mathbf{b}$ (рис. 222). Так как амплитудная решетка не вносит разности фаз между волнами, прошедшими через различные участки ее, то обе стрелки $\mathbf{a}$ и $\mathbf{b}$ будут направлены одинаково.

Допустим теперь, что решетка фазовая и геометрически подобна амплитудной, т.е. состоит из чередующихся участков той же ширины, влияющих на фазу, но не на амплитуду волны. Колебания на выходе всех участков изобразятся стрелками $\mathbf{A}$ и $\mathbf{B}$ одинаковой длины, но различно направленными (рис. 222, положение *а*). Поскольку существенна лишь относительная

разность фаз между обоими колебаниями, стрелки $\mathbf{A}$ и $\mathbf{B}$ можно повернуть на один и тот же угол, ничего не меняя в физических условиях задачи. Поэтому, не нарушая общности, можно предположить, что биссектриса угла между векторами $\mathbf{A}$ и $\mathbf{B}$ горизонтальна, т.е. параллельна поверхности решетки. Разложим каждый из векторов $\mathbf{A}$ и $\mathbf{B}$ на горизонтальную и вертикальную составляющие: $\mathbf{A} = \mathbf{D} + \mathbf{c}$, $\mathbf{B} = \mathbf{D} - \mathbf{c}$ (рис. 222, положение *б*). Допустим теперь, что оба вектора $\mathbf{c}$ и $-\mathbf{c}$ повернуты на 90° в одном и том

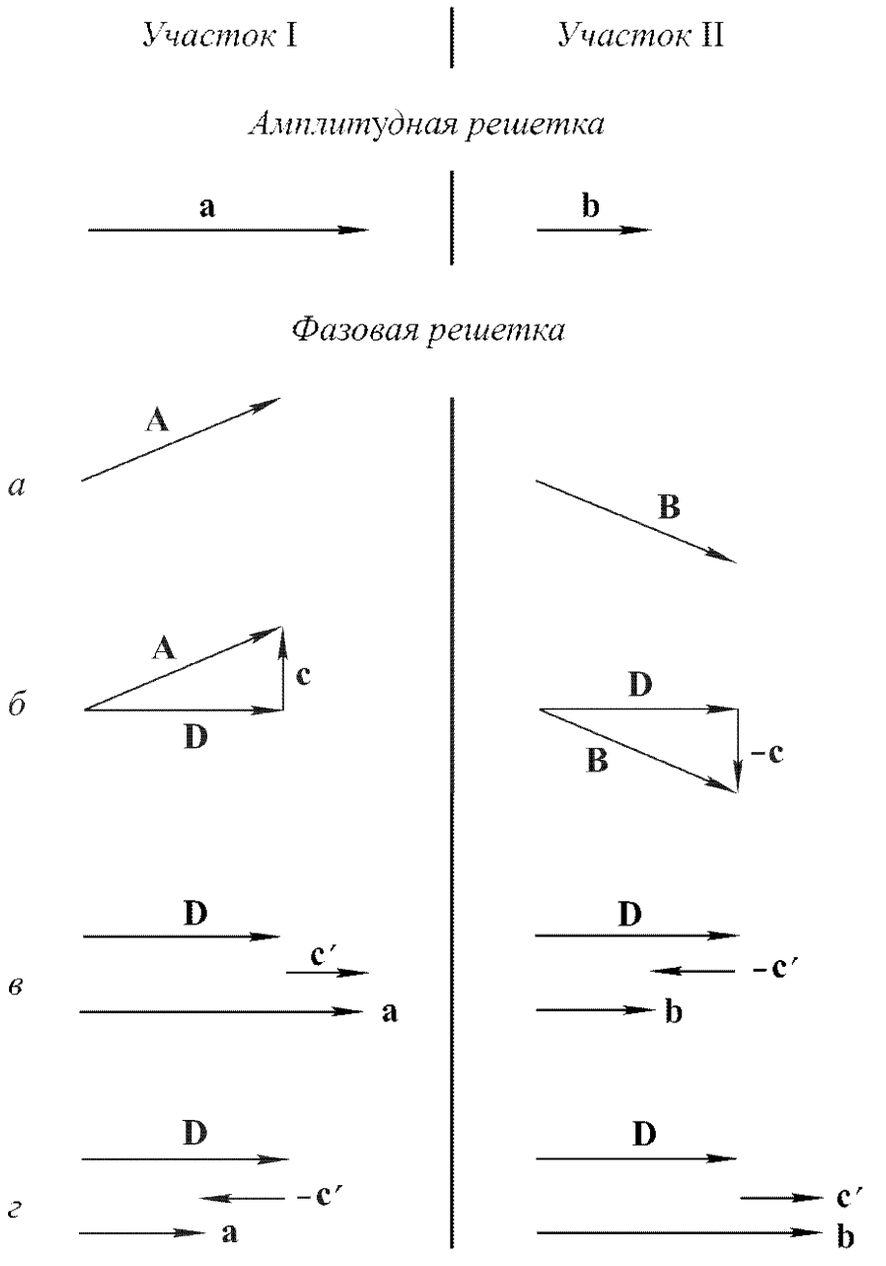


Рис. 222

же направлении в положении $\mathbf{c}'$ и $-\mathbf{c}'$ (рис. 222, положение *в*). Тогда на выходе участков I и II колебания представятся векторами $\mathbf{a} = \mathbf{D} + \mathbf{c}'$ и $\mathbf{b} = \mathbf{D} - \mathbf{c}'$, параллельными поверхности решетки. Поле на выходе фазовой решетки будет иметь такой же вид, что и на выходе амплитудной решетки. Поворот на 90° означает изменение фаз соответствующих колебаний на такой же угол. Таким образом, изменением фазы колебания на 90° можно превратить фазовую решетку в амплитудную. В этом и состоит идея метода фазового контраста.

При рассмотренном повороте по часовой стрелке вектор $\mathbf{a}$ получается длиннее вектора $\mathbf{b}$. Это значит, что светлым местам в изображении амплитудной решетки будут соответствовать светлые же места в изображении фазовой решетки, а темным — темные (*позитивный фазовый контраст*). Если же векторы $\mathbf{c}$ и $-\mathbf{c}$ повернуть на 90° в противоположном направлении (рис. 222, положение *г*), то соотношение между длинами векторов $\mathbf{a}$ и $\mathbf{b}$, а с ним и соответствие между светлыми и темными частями заменятся на противоположные (*негативный фазовый контраст*).

3. Чтобы повернуть векторы $\mathbf{c}$ и $-\mathbf{c}$, сохраняя неизменным направление вектора $\mathbf{D}$, надо прежде всего пространственно разделить волновые поля, представляемые этими векторами. Полное колебание на выходе решетки можно разложить на два колебания. Одно колебание имеет постоянную амплитуду на протяжении всей решетки и изображается постоянным вектором $\mathbf{D}$. Оно дает в фокальной плоскости объектива центральный максимум нулевого порядка и не влияет на все остальные максимумы. Другое колебание представляется периодической функцией, которая равна $+\mathbf{c}$ на одних участках решетки и $-\mathbf{c}$ на соседних участках. Так как среднее по периоду решетки значение такой функции равно нулю, то такое колебание будет возбуждать только боковые максимумы, не оказывая никакого влияния на центральный максимум нулевого порядка. Таким образом, в фокальной плоскости объектива оба колебания окажутся пространственно разделенными. Одно концентрируется в центральном максимуме, другое распределяется по всем остальным — боковым — максимумам. Поставив на пути либо центрального максимума, либо всех боковых максимумов прозрачную плоскопараллельную пластинку надлежащей толщины, можно внести необходимую разность фаз в 90° и тем самым осуществить фазовый контраст. Такая пластинка называется *фазовой*.

4. До внесения фазовой пластинки энергия на участке I в условных единицах представляется выражением $A^2 = D^2 + c^2$. Такова же энергия на участке II. Полная энергия на обоих участках равна $2(D^2 + c^2)$. После поворота векторов $\mathbf{c}$ и $-\mathbf{c}$ на 90° энергии на участках I и II будут равны соответственно $(D + c)^2$ и $(D - c)^2$, а их сумма $2(D^2 + c^2)$. Таким образом, энергия не изменяется, а лишь перераспределяется между участками I и II. Этим перераспределением и объясняется просветление участков I и потемнение участков II.

§ 60. Измерение угловых диаметров звезд

1. Пусть перед объективом телескопа помещен экран с двумя круглыми отверстиями, расстояние между центрами которых равно D (рис. 223). Направив телескоп на звезду, закроем сначала правое отверстие. В результате дифракции на открытом левом отверстии в фокальной плоскости объектива получится система дифракционных колец. Их положение и размеры будут зависеть только от диаметра отверстия, но не от его положения в плоскости экрана. Поэтому, если закрыть левое отверстие и открыть правое, то дифракционная картина в фокальной плоскости объектива, видимая глазом, не изменится. Если затем открыть оба отверстия, то одна система дифракционных колец точно совместится с другой. При этом не получится, однако, простого усиления яркости дифракционных колец. Дифракционные кольца будут пересечены параллельными интерференционными полосами, перпендикулярными к линии, соединяющей центры отверстий O_1 и O_2 (рис. 224).

Как и в опыте Юнга, эти полосы возникают из-за интерференции дифракционных пучков света, исходящих из первого и второго отверстий. Направления на соседние интерференционные максимумы (или минимумы) отличаются на угол $\vartheta = \lambda/D$. Это и есть угловое расстояние между соседними интерференционными полосами, если смотреть на фокальную плоскость из центра объектива.

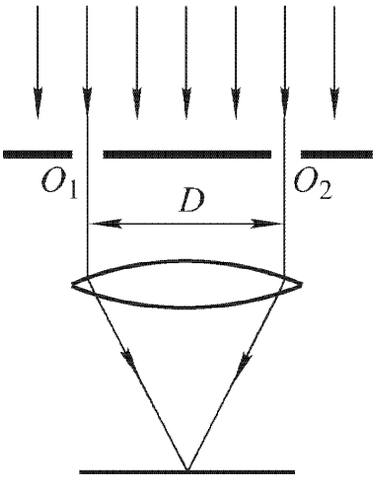


Рис. 223

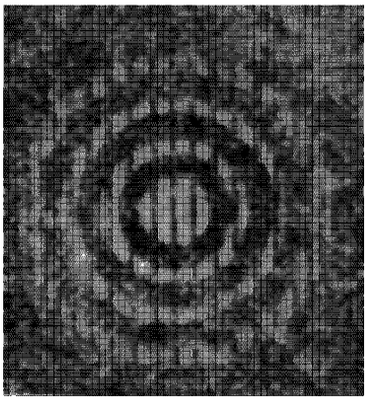


Рис. 224

Допустим, что звезда — двойная с угловым расстоянием между ее компонентами $\delta\varphi = \vartheta/2$. Тогда максимумы интерференционных полос от одной звезды наложатся на минимумы другой: интерференционные полосы либо пропадут, либо их видимость сделается наименьшей. На этом основан интерференционный метод измерения угловых расстояний между компонентами двойных звезд, идея которого была предложена Физо. Надо менять расстояние между отверстиями O_1 и O_2, пока не пропадут интерференционные полосы или их видимость сделается наименьшей. Если D — расстояние между отверстиями в этот момент, то угловое расстояние между компонентами двойной звезды найдется по формуле

$$\delta\varphi = \lambda/(2D). \tag{60.1}$$

Тот же метод применим и для измерения угловых диаметров одиночных звезд. Допустим для простоты вычислений, что звезда излучает как равномерно светящийся квадрат, плоскость которого параллельна фокальной плоскости объектива, а одна пара противоположных сторон параллельна прямой O_1O_2, соединяющей центры отверстий O_1 и O_2. Пусть угловой размер стороны квадрата будет $\delta\varphi' = \lambda/D$. Тогда можно мысленно разбить весь квадрат на пары узких одинаковых полосок, угловое расстояние между которыми равно $\lambda/(2D)$. Согласно формуле (60.1), каждая пара таких полосок не даст интерференционных полос. Следовательно, интерференционные полосы не получатся и от всей звезды. Таким образом, увеличивая расстояние между отверстиями O_1 и O_2, можно в этом случае добиться исчезновения интерференционных полос. Если D — расстояние между отверстиями в момент исчезновения, то угловой размер звезды найдется по формуле

$$\delta\varphi' = \lambda/D. \tag{60.2}$$

Если предположить, что звезда излучает как равномерно светящийся диск (что ближе соответствует действительности), то идейно простые, но

громоздкие вычисления показывают, что интерференционные полосы пропадут, когда

$$\delta\varphi' = 1{,}22\frac{\lambda}{D}. \tag{60.3}$$

2. Угловые размеры $\delta\varphi'$, даваемые формулами (60.2) и (60.3), совпадают с разрешаемым расстоянием телескопа. Однако если угловой размер звезды порядка (60.2) или (60.3), то ее изображение в телескопе настолько мало отличается от изображения точечного источника, что непосредственное измерение диаметра звезд с помощью телескопа становится практически непригодным. Интерференционный же метод дает в этом случае еще хорошую точность. Однако чтобы интерференционные полосы исчезли, а это необходимо по идее самого интерференционного метода, нужен телескоп с большим диаметром объектива. Физо указал способ, как преодолеть эту трудность. Идея Физо была использована Майкельсоном, соединившим телескоп с интерферометром.

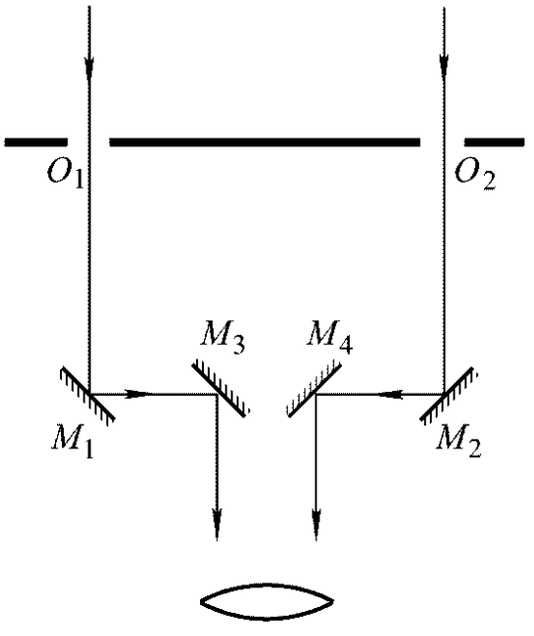


Рис. 225

Принципиальная схема *звездного интерферометра* Майкельсона показана на рис 225. Лучи от звезды падают на два круглых отверстия O_1 и O_2 и после отражения от зеркал M_1, M_3 и M_2, M_4 попадают в объектив телескопа. Если закрыть отверстие O_2, то из-за дифракции на краях отверстия O_1 в телескоп будут видны дифракционные кольца. Если открыть O_2 и закрыть O_1, то получится такая же система колец, но сдвинутая относительно первой. Поворотом зеркала M_2 ее можно совместить с первой системой. Тогда дифракционные кольца снова будут пересечены интерференционными полосами (рис. 224). Изменяя расстояние между отверстиями O_1 и O_2 с одновременным перемещением зеркал M_1 и M_2, можно добиться либо полного исчезновения полос (в случае одиночной звезды), либо их наименьшей видимости (в случае двойной звезды, состоящей из двух различных компонентов). Зная расстояние между центрами отверстий O_1 и O_2 в этот момент, можно вычислить угловой размер звезды по формулам (60.1) или (60.3).

3. Для уяснения действия установки Майкельсона отвлечемся от наличия звезды. Ее действие эквивалентно действию вторичных источников Гюйгенса, распределенных в плоскостях отверстий O_1 и O_2. Не теряя общности, можно для простоты рассуждений взять отверстия O_1 и O_2 бесконечно малыми. Тогда действие звезды сведется к действию вторичных точечных источников S_1 и S_2, помещенных в центрах отверстий O_1 и O_2 (рисунки 226 и 225). Если от звезды идет пучок параллельных лучей, перпендикулярных к плоскости экрана O_1O_2, то фазы вторичных источников S_1 и S_2 будут одинаковы. То же справедливо и для мнимых вторичных источников S_1', S_2' и S_1'', S_2'', являющихся изображениями S_1 и S_2 в плоских зеркалах. Таким образом, зеркала как бы сближают источники S_1 и S_2 и тем самым делают интерференционные полосы шире.

Допустим теперь, что вблизи первой звезды на угловом расстоянии $\delta\varphi$ находится вторая звезда. Волновой фронт от второй звезды будет дости-

гать отверстий O_1 и O_2 не одновременно. Разность хода между лучами, приходящими от второй звезды в S_1 и S_2, будет $D \cdot \delta\varphi$, где D — расстояние между S_1 и S_2. Если эта разность хода равна $\lambda/2$, то вторичные источники S_1 и S_2, заменяющие действие второй звезды, будут находиться в противоположных фазах. Следовательно, максимумы интерференционных полос, даваемые второй звездой, наложатся на минимумы интерференционных полос от первой звезды. Интерференционные полосы либо пропадут, либо их видимость станет наименьшей. Это произойдет, когда $D \cdot \delta\varphi = \lambda/2$. Таким образом, мы снова приходим к формуле (60.1), причем D означает расстояние между центрами отверстий O_1 и O_2. Аналогично, в случае одиночной звезды, излучающей как равномерно светящийся квадрат или диск, получаются формулы (60.2) и (60.3).

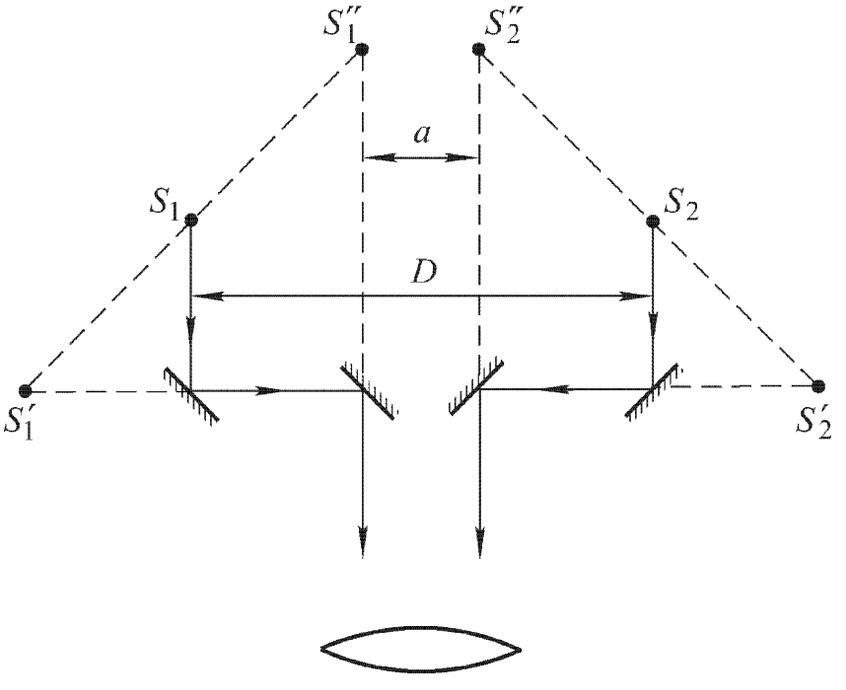


Рис. 226

Таким образом, нет надобности в телескопе с большим объективом. (Интерферометр Майкельсона был смонтирован на большом рефлекторе с диаметром зеркала 2,5 м обсерватории Маунт Вильсон, выбранном только из-за прочности своей механической конструкции. При расстоянии между зеркалами M_3 и M_4 114 см расстояние между интерференционными полосами в фокальной плоскости составляло около 0,02 мм.) Необходимо только иметь возможность достаточно далеко раздвигать отверстия O_1 и O_2 с зеркалами M_1 и M_2. Это предъявляет весьма жесткие требования к механическим качествам установки. Случайные колебания зеркал с амплитудами, составляющими незначительные доли длины волны, сделали бы измерения по этому методу невозможными. Допустим ради простоты, что центр зеркала все время остается неподвижным, а его концы испытывают беспорядочные смещения порядка h. Если бы зеркало было абсолютно твердым, то такие смещения вызвали бы беспорядочные вращения плоскости зеркала на углы порядка $2h/d$, где d — диаметр зеркала. Благодаря этому отраженные от зеркала лучи беспорядочно меняли бы свое направление на углы порядка $4h/d$. Это повело бы к дрожанию дифракционных колец. Для устойчивости интерференционных полос необходимо, чтобы угол $4h/d$ был мал по срав-

нению с угловым расстоянием между звездами $\lambda/(2D)$, т.е. должно быть $h \ll d\lambda/(8D)$. Хотя действительные беспорядочные колебания зеркал сложнее рассмотренных беспорядочных вращений, приведенный пример дает правильное представление о трудностях, которые должны быть преодолены при конструировании прибора. Майкельсон успешно справился с этой задачей.

4. В интерферометре Маикельсона зеркала M_1 и M_2 могли быть раздвинуты до расстояния $D = 6{,}1$ метра. Интерференционные полосы от звезды Бетельгейзе исчезали при $D = 306{,}5$ см, хотя при том же расстоянии интерференция от других звезд прекрасно наблюдалась. Считая эффективную длину волны для Бетельгейзе равной $\lambda = 5{,}75 \cdot 10^{-5}$ см, получим $\delta\varphi' = 0{,}047''$. Параллакс для этой звезды составляет $0{,}003''$. Это дает для линейного диаметра звезды около $9 \cdot 10^8$ км, что превосходит диаметр орбиты Марса.

Измерения Майкельсона были выполнены в двадцатых годах нашего столетия, когда астрофизика в современном ее понимании еще только зарождалась. Результаты этих измерений произвели сильное впечатление на современников. Майкельсон начал строить интерферометр с базой $D = 15{,}24$ м, но смерть помешала ему осуществить это намерение.

Звездным интерферометром Майкельсона были измерены угловые диаметры только нескольких звезд. Все они, как и Бетельгейзе, — гиганты, линейные диаметры которых во много раз превосходят диаметр Солнца. Одна из причин небольшого числа измеренных диаметров звезд связана с вредным влиянием турбулентности атмосферы. Правда, это влияние значительно сильнее проявляется при наблюдении в телескоп, чем при работе с интерферометром. В случае интерферометра изменение показателя преломления воздуха перед небольшими отверстиями зеркал лишь смещает интерференционную картину в целом, так что интерференционные полосы остаются различимыми, если только это смещение происходит достаточно медленно. При наблюдении же в тех же условиях в телескоп с недиафрагмированным объективом нерегулярные изменения показателя преломления воздуха сильно искажают изображение звезды.

Другая причина заключается в том, что линейные диаметры подавляющего большинства звезд мало отличаются от диаметра Солнца. На расстоянии же ближайшей звезды диск Солнца был бы виден под углом всего $0{,}007''$. Для измерения таких углов расстояние между внешними зеркалами интерферометра M_1 и M_2 (база) должно быть не меньше примерно 20 м. Постройка такого большого интерферометра трудна из-за высоких требований к механическим параметрам конструкции всей установки, о которых говорилось выше.

5. Принципы, изложенные выше, применимы, конечно, и в радиодиапазоне. Радиотелескопы, благодаря большим значениям длин радиоволн, характеризуются низкой разрешающей способностью. Даже для самых крупных из них с диаметром зеркала $D \sim 100$ м, работающих на длине волны $\lambda \sim 10$ см, разрешаемое угловое расстояние $\sim \lambda/D \approx 10^{-3}$ рад, т.е. всего около $3'$. Иное дело — радиоинтерферометр, работающий по принципу звездного интерферометра Майкельсона. Он состоит из двух антенн, установленных в разных местах, сигналы с которых подаются на общий детектор. Высокая разрешающая способность достигается за счет увеличения расстояния D между антеннами (базы радиоинтерферометра). Рекордное угловое разрешение достигается, когда антенны удалены друг от друга на межконтинентальные расстояния. Были осуществлены радиоинтерферометры

с базой Крым–США, США–Австралия. На самой короткой длине волны было достигнуто угловое разрешение $\sim 10^{-4}$ угловой секунды, т.е. разрешающая способность была примерно в 100 раз больше, чем у звездного интерферометра Майкельсона.

Надо отметить, что с увеличением базы возрастают технические трудности, обусловленные внесением нерегулярно меняющейся разности фаз на пути от антенн к детектору. Для устранения этого недостатка в радиоинтерферометре Брауна и Твисса сигналы, принятые антеннами, детектируются независимо. Измеряется корреляция флуктуирующих интенсивностей этих сигналов в зависимости от расстояния между антеннами. По этим данным находятся угловые размеры источника. Браун и Твисс показали, что аналогичное устройство может работать также и в видимой области спектра. Свет от звезды фокусируется двумя вогнутыми зеркалами на два фотоэлемента. Измеряется корреляция флуктуирующих фототоков, возбуждающихся в фотоэлементах, в зависимости от расстояния между зеркалами. Без особых осложнений расстояние между зеркалами может быть сделано большим и достигнута высокая разрешающая способность прибора.

§ 61. Дифракция на двумерных и трехмерных решетках. Дифракция рентгеновских лучей

1. *Двумерной решеткой* называется всякая структура, свойства которой периодически меняются в двух различных направлениях. Примером могут служить две скрещенные одномерные решетки, т.е. решетки, наложенные одна на другую таким образом, что штрихи одной решетки пересекаются со штрихами другой под некоторым углом. Дифракционная картина от такой структуры, как нетрудно показать, может быть получена путем наложения дифракционных картин от соответствующих одномерных решеток.

Трехмерные, пространственные решетки обладают периодичностью в трех различных направлениях. Они играют важную роль в физике рентгеновских лучей. Долгое время не удавалось получить дифракцию рентгеновских лучей, так как дифракционные решетки, применявшиеся в оптической области спектра, были слишком грубыми для рентгеновских лучей с длиной волны порядка 0,1 нм и меньше (см., впрочем, § 46, п. 8). Лауэ (1879–1960) в 1912 г. предложил для этой цели воспользоваться кристаллом, который можно рассматривать как *естественную периодическую пространственную решетку*, построенную из атомов или ионов. По указанию Лауэ его сотрудники Фридрих (1883–1968) и Книппинг (1883–1935) направили узкий пучок сплошного рентгеновского излучения на монокристалл и получили на фотопластинке картину дискретных, правильно расположенных пятен, возникших в результате дифракции рентгеновских лучей на кристаллической решетке. Этот основополагающий опыт установил *волновую природу* рентгеновского

излучения и заложил начала современной *физики твердого тела*. Он впервые принес экспериментальное доказательство давно высказывавшейся кристаллографами гипотезы о структуре кристаллов из частиц, правильно и периодически расположенных в пространстве.

Из этого опыта возникли два направления в физике: *рентгеновская спектроскопия* и *рентгеноструктурный анализ*. Рентгеновская спектроскопия использует естественные кристаллы известной кристаллической структуры для анализа рентгеновского излучения и измерения длин волн. Рентгеноструктурный анализ, напротив, использует рентгеновское излучение известной длины волны для выяснения кристаллической структуры кристаллов и измерения параметров этой структуры. По сравнению с оптической дифракцией дифракция рентгеновских лучей обладает той особенностью, что она осуществляется *без линз и зеркал*, ввиду отсутствия материалов, из которых можно было бы изготовить эти приспособления.

2. Качественное рассмотрение дифракции рентгеновских лучей удобно начать с *дифракции на прямолинейной цепочке*, состоящей из одинаковых равноотстоящих частиц (атомов). Расстояние между соседними частицами обозначим через a. Пусть на такую цепочку под углом скольжения α_0 падает параллельный пучок рентгеновских лучей (рис. 227). Разность хода между лучами, рассеянными соседними атомами под углом α, будет $AD - CB = a(\cos\alpha - \cos\alpha_0)$. Условие интерференционного усиления этих лучей имеет вид

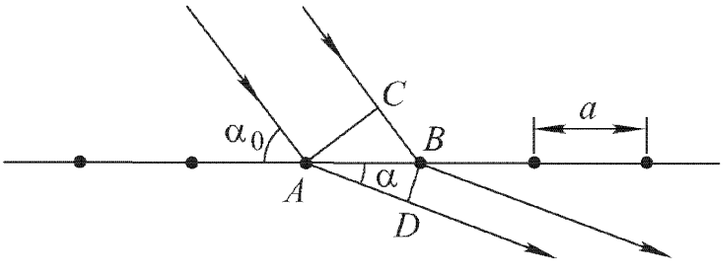


Рис. 227

$$a(\cos\alpha - \cos\alpha_0) = m\lambda \qquad (61.1)$$

($m = 0, \pm1, \pm2, \ldots$). Оно определяет *положения дифракционных максимумов в фраунгоферовой дифракционной картине*, т.е. на бесконечных расстояниях от цепочки. Бесконечными могут считаться расстояния r, удовлетворяющие условию:

$$r \gg l^2/\lambda, \qquad (61.2)$$

где l — длина цепочки. Ввиду малости длин рентгеновских волн, эти расстояния всегда очень велики. Так, при $l = 1$ мм, $\lambda = 0{,}1$ нм $= 10^{-7}$ мм условие (61.2) дает $r \gg 10^7$ мм $= 10$ км. Таким образом, для осуществления фраунгоферовой дифракции потребовались бы очень большие расстояния. В реальных опытах фотографическая пластинка, регистрирующая дифрак-

ционную картину, ставится на расстоянии в несколько десятков сантиметров от рассеивающего кристалла или меньше, т.е. не в волновой зоне, а в десятки и сотни тысяч раз ближе. Это замечание относится, конечно, не только к линейным цепочкам, но и к поверхностным и объемным решеткам.

Однако, как было выяснено в § 52, *волновое поле, возникающее при дифракции на плоской решетке, представляется на всех расстояниях суперпозицией плоских волн различных направлений.* В эту суперпозицию входят и *неоднородные волны.* Но на расстояниях от решетки, превышающих ее период, неоднородные волны не играют роли, так как на таких расстояниях они практически полностью затухают. Так же обстоит дело и в случае дифракции на кристаллической решетке. (В случае линейной цепочки вместо плоских удобнее пользоваться *цилиндрическими волнами.*) Фотопластинка ставится на малых расстояниях от кристалла, где применима геометрическая оптика, но все же достаточно далеко, чтобы различные плоские волны, исходящие от кристалла, успели разделиться пространственно. Для определения положения максимумов интенсивности на фотопластинке достаточно знать *направления* распространения этих волн. Но эти направления не зависят от расстояния до кристалла. В области применимости геометрической оптики они такие же, что и в волновой зоне, а потому для определения положения максимумов интенсивности на фотопластинке можно пользоваться формулой (61.1). Однако такие максимумы отличаются от интерференционных максимумов в волновой зоне. В каждый максимум в волновой зоне колебания от всех атомов решетки приходят *либо в одинаковых фазах, либо в фазах, отличающихся на* $2m\pi$ (m — целые числа). Для максимумов же интенсивности в области применимости геометрической оптики это не имеет места.

При фиксированном угле α_0 условие (61.1) определяет дискретный набор углов α, удовлетворяющих этому условию. Оно выделяет в пространстве *дискретное семейство конусов*, вдоль образующих которых могут распространяться дифрагированные пучки лучей. В сечении таких конусов плоскостью фотопластинки получается *дискретное семейство эллипсов или гипербол* в зависимости от направления этой плоскости. В частности, когда плоскость пластинки перпендикулярна к направлению цепочки, возникает *семейство концентрических кругов.*

3. Двумерные и трехмерные решетки могут быть *простыми* (*примитивными*) и *составными* (см. т. II, § 130). Решетка называется простой, если она построена из одинаковых атомов, причем элементарная ячейка решетки состоит из восьми атомов, расположенных в вершинах параллелепипедов. Все остальные решетки называются составными. Составная решетка состо-

ит из нескольких простых решеток, вставленных друг в друга. Дифракционная картина, возникающая при дифракции рентгеновских волн на составной решетке, получается в результате интерференции дифракционных картин от простых решеток, из которых она состоит. Поэтому достаточно ограничиться рассмотрением дифракции на простых решетках.

Если какие-либо два атома простой решетки соединить прямой линией, то, ввиду периодичности распределения атомов в пространстве, на этой прямой окажется бесконечно много атомов, находящихся на одинаковых расстояниях друг от друга. Такие прямые мы будем называть *атомными прямыми*, а плоскости, в которых располагаются атомы, — *атомными плоскостями*. Всю неограниченную решетку можно рассматривать как бесконечную двоякопериодическую систему параллельных атомных прямых или как бесконечную однократно периодическую систему параллельных атомных плоскостей. Оба эти представления не единственны, а могут быть выполнены бесконечным множеством способов. Три произвольные атомные прямые, не лежащие в одной плоскости и пересекающиеся в каком-либо атоме, можно принять за координатные оси X, Y, Z прямолинейной (вообще говоря, косоугольной) системы координат. Тогда координаты атомов простой решетки представятся выражениями

$$\begin{gathered} x_{lmn} = la_1, \quad y_{lmn} = ma_2, \quad z_{lmn} = na_3 \\ (l, m, n = 0, \pm 1, \pm, 2, \ldots), \end{gathered} \tag{61.3}$$

где a_1, a_2, a_3 — постоянные, называемые *периодами решетки*. Элементарной ячейкой такой решетки является параллелепипед с ребрами a_1, a_2, a_3, в вершинах которого находятся атомы.

Пусть на простую решетку падает параллельный пучок рентгеновских лучей, образующий углы α_0, β_0, γ_0 с координатными осями X, Y, Z. Чтобы волны, рассеянные всеми атомами в направлении прямой, составляющей углы α, β, γ с координатными осями, при интерференции в волновой зоне усиливали друг друга, должны выполняться условия

$$\begin{gathered} a_1(\cos\alpha - \cos\alpha_0) = m_1\lambda, \\ a_2(\cos\beta - \cos\beta_0) = m_2\lambda, \\ a_3(\cos\gamma - \cos\gamma_0) = m_3\lambda \\ (m_1, m_2, m_3 = 0, \pm 1, \pm 2, \ldots), \end{gathered} \tag{61.4}$$

называемые *условиями Лауэ*. Необходимость первого условия станет очевидной, если заметить, что оно является условием интерференционного усиления волн, рассеянных под углом α к оси X атомами каждой атомной прямой, параллельной этой оси. Аналогичный смысл имеют и остальные два условия.

Но условия (61.4) являются и достаточными для интерференционного усиления волн, рассеянных в рассматриваемом направлении всеми атомами решетки. Действительно, проведем через произвольный атом 1 атомную прямую, параллельную оси X. При выполнении первого условия (61.4) в направлении под углом α к этой прямой получится интерференционный максимум. Проведем теперь через тот же атом 1 атомную прямую, параллельную оси Y. При выполнении второго условия (61.4) все атомы этой прямой рассеивают волны в рассматриваемом направлении в той же фазе, что и атом 1. Значит, все атомы обеих атомных прямых, а с ними и все атомы, лежащие в их плоскости, будут посылать волны в том же направлении также в одинаковых фазах. Таким образом, выполнение первых двух условий (61.4) приводит к интерференционному усилению волн, рассеиваемых в рассматриваемом направлении всеми атомами любой атомной плоскости, параллельной координатной плоскости XY. Аналогично убедимся, что при выполнении еще третьего условия (61.4) будет иметь место интерференционное усиление волн, рассеянных всеми такими атомными плоскостями. Тем самым достаточность условий (61.4) доказана.

4. Дифракционная картина, возникающая на фотопластинке, поставленной на пути рентгеновских пучков, рассеянных монокристаллом в опытах типа Лауэ, называется *лауэграммой*. Об использовании условий Лауэ в области применимости геометрической оптики можно повторить все, что было сказано выше в связи с формулой (61.1). *Формулы Лауэ* (61.4) *указывают направления пучков, возникающих при дифракции на кристалле*. Физический смысл лауэграммы хорошо иллюстрируется аналогией с отражением светового пучка от многогранного зеркала. Здесь возникают отраженные пучки, распространяющиеся в различных направлениях. При падении на экран они дают систему правильно расположенных светлых пятен, аналогичную лауэграмме, возникающей при дифракции рентгеновских лучей.

Рассмотрим сначала лауэграмму от плоской двумерной кристаллической решетки. В этом случае три условия Лауэ (61.4) сводятся к двум. Если плоскость решетки принять за атомную плоскость XY, то останутся только два первых условия (61.4). Первое условие (61.4) означает, что максимумы лежат на поверхности конуса, образующего угол α с осью X, а второе — на поверхности другого конуса, образующего угол β с осью Y. Прямые, по которым пересекаются поверхности обоих конусов, указывают направления дифрагированных пучков. При пересечении таких пучков плоскостью фотопластинки возникают дискретные дифракционные точечные максимумы, расположенные вдоль эллипсов, гипербол или кругов в зависимости от направления этой плоскости. Максимумы могут и не получиться, если

$\cos\alpha$ или $\cos\beta$, или обе эти величины, вычисленные по формуле (61.4), окажутся по модулю больше единицы.

Рассмотрим теперь трехмерную решетку. К двум конусам, выделяемым первыми двумя условиями (61.4), теперь добавляется еще третий конус, образующие которого составляют угол γ с осью Z. Дифрагированные пучки должны одновременно лежать на поверхности всех трех конусов. Но три конуса, вообще говоря, не пересекаются вдоль общей прямой. Отсюда следует, что при падении на монокристалл монохроматического рентгеновского луча дискретные пучки рассеянного излучения, вообще говоря, не возникают, а рассеяние происходит более или менее равномерно во все стороны. Исключение составляет только прямой луч, проходящий через кристалл без изменения направления. Но для избранных длин волн три конуса могут иметь общие образующие. Поэтому *для получения лауэграмм от трехмерных решеток необходимо сплошное рентгеновское излучение*, так как в таком излучении могут присутствовать такие длины волн, для которых выполняются все три условия Лауэ (61.4).

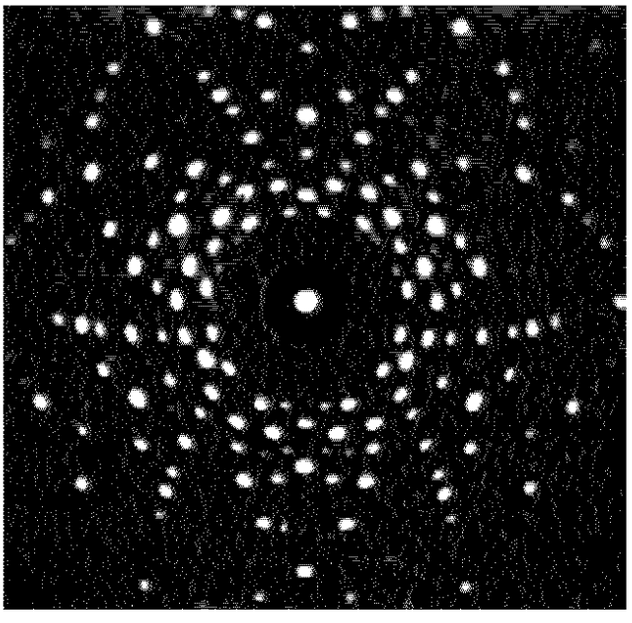


Рис. 228

Таким образом, если бы глаз обладал способностью воспринимать рентгеновские лучи и различать их цвета, то лауэграмма, наблюдаемая на экране, представлялась бы для него цветной, т.е. состоящей из пятен разного цвета.

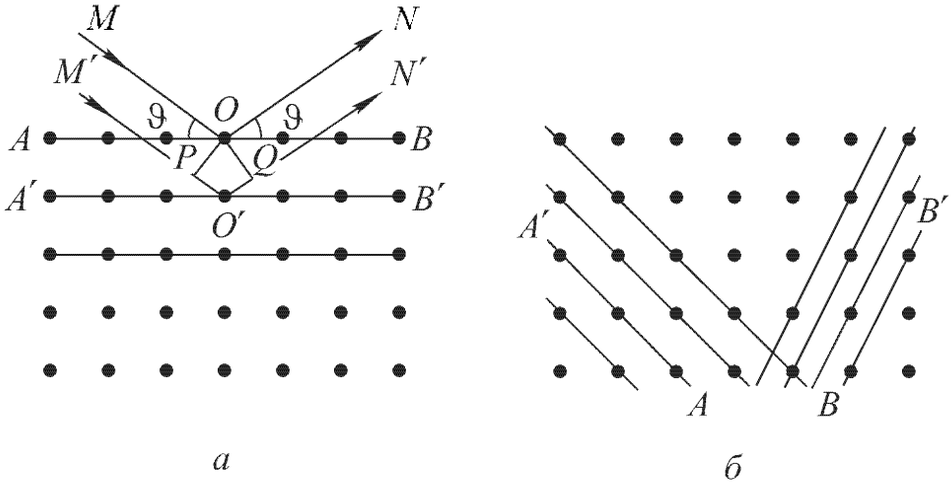
Подтвердим полученные результаты простым вычислением. Предположим, что простая кристаллическая решетка принадлежит к ромбической системе, так что элементарная ячейка будет прямоугольным параллелепипедом. Пусть для некоторой длины волны λ условия Лауэ (61.4) выполняются. Разрешим их относительно $\cos\alpha$, $\cos\beta$, $\cos\gamma$, возведем в квадрат и сложим почленно. Тогда, принимая во внимание соотношения $\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1$, $\cos^2\alpha_0 + \cos^2\beta_0 + \cos^2\gamma_0 = 1$, после простых преобразований получим

$$\lambda = -\frac{2(m_1\cos\alpha_0 + m_2\cos\beta_0 + m_3\cos\gamma_0)}{m_1^2 + m_2^2 + m_3^2}. \qquad (61.5)$$

Только при этом условии уравнения (61.4) совместны. Оно однозначно определяет длину волны, при которой может получиться дифракционный максимум, выделяемый целыми числами m_1, m_2, m_3 и углами α, β, γ.

На рис. 228 приведена лауэграмма для монокристалла кварца. Закономерное расположение пятен на фотопластинке указывает на правильность расположения атомов, из которых построен кристалл.

5. Дифракцию рентгеновских лучей в кристаллах можно трактовать несколько иначе. Такая трактовка была дана английским физиком Лоуренсом Брэггом (1890–1971) и независимо от него русским кристаллографом Ю.В. Вульфом (1863–1925). Проведем в кристалле произвольную атомную плоскость AB, достаточно густо усеянную атомами (рис. 229 *а*). Если на



а *б*

Рис. 229

нее падает рентгеновский луч MO, то под тем же углом возникнет отраженный луч ON. В том же направлении возникнут лучи, отраженные атомными плоскостями, параллельными плоскости AB. Интенсивность луча, отраженного отдельной атомной плоскостью, слишком мала, чтобы произвести заметное действие. Последнее может возникнуть лишь в результате интерференционного усиления всех лучей, отраженных рассматриваемыми атомными плоскостями. Разность хода $PO'Q$ между лучами, отраженными соседними плоскостями, равна $2d\sin\vartheta$, где d — расстояние между этими плоскостями. Для интерференционного усиления должно выполняться условие

$$2d\sin\vartheta = m\lambda, \tag{61.6}$$

называемое *условием Брэгга–Вульфа* ($m = 0, \pm1, \pm2, \ldots$). Каков бы ни был угол скольжения ϑ и порядок отражения m, всегда найдутся длины волн λ, удовлетворяющие условию (61.6). Толь-

ко волны с такими длинами могут отражаться от рассматриваемых атомных плоскостей.

В кристалле можно провести бесконечное множество систем параллельных атомных плоскостей в различных направлениях. Таковы, например, плоскости, параллельные атомной плоскости AA' или атомной плоскости BB' (рис. 229 *б*). Эффективными являются только такие плоскости, на которых атомы расположены достаточно часто. От всех этих плоскостей возможно интерференционное отражение. И дифракционную картину можно рассматривать как совокупность рентгеновских пучков, претерпевших отражения на таких атомных плоскостях.

Против последнего заключения можно, однако, выдвинуть следующее возражение. Рентгеновский пучок MO, падающий на атомную плоскость AB (рис. 229 *а*), дает не только отраженный пучок ON, но и боковые дифрагированные пучки, которые при определенных условиях могут усиливаться пучками того же направления, дифрагировавшими на параллельных атомных плоскостях. Такие пучки в наших рассуждениях не были учтены. Поэтому может показаться, что метод Брэгга–Вульфа дает не все возможные дифракционные пучки в дифракционной картине. Следующее простое рассуждение, устанавливающее *эквивалентность условий Лауэ и условия Брэгга–Вульфа*, показывает, что это не так.

6. Примем векторы $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$, являющиеся ребрами элементарного параллелепипеда кристаллической решетки, за базисные векторы косоугольной системы координат. Тогда радиус-вектор каждого атома решетки представится выражением

$$\mathbf{r} = x\mathbf{a}_1 + y\mathbf{a}_2 + z\mathbf{a}_3, \qquad (61.7)$$

в котором координаты x, y, z принимают целочисленные значения. Пусть $\mathbf{s}_0$ — единичный вектор, проведенный в направлении падающего луча, а $\mathbf{s}$ — единичный вектор, указывающий направление одного из дифрагированных пучков (рис. 230). Тогда условия Лауэ (61.4) можно записать в следующей векторной форме:

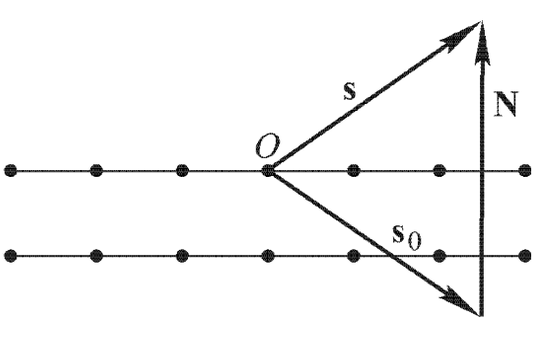


Рис. 230

$$\begin{aligned} (\mathbf{s}-\mathbf{s}_0)\mathbf{a}_1 &= m_1\lambda, \\ (\mathbf{s}-\mathbf{s}_0)\mathbf{a}_2 &= m_2\lambda, \\ (\mathbf{s}-\mathbf{s}_0)\mathbf{a}_3 &= m_3\lambda. \end{aligned} \qquad (61.8)$$

Вектор $\mathbf{N} = \mathbf{s} - \mathbf{s}_0$ направлен параллельно биссектрисе угла, образованного падающим и отраженным лучами. Введя этот век-

тор, получим

$$(\mathbf{Na}_1) = m_1\lambda, \quad (\mathbf{Na}_2) = m_2\lambda, \quad (\mathbf{Na}_3) = m_3\lambda. \tag{61.9}$$

Через атом, находящийся в начале координат O, проведем плоскость, перпендикулярную к вектору $\mathbf{N}$. Докажем, что она является атомной плоскостью. Уравнение рассматриваемой плоскости имеет вид $(\mathbf{Nr}) = 0$. Чтобы атом с координатами (61.7) лежал в этой плоскости, необходимо и достаточно, чтобы его координаты удовлетворяли уравнению

$$(\mathbf{Na}_1)x + (\mathbf{Na}_2)y + (\mathbf{Na}_3)z = 0,$$

или, ввиду соотношений (61.9), $m_1x + m_2y + m_3z = 0$. Каковы бы ни были целые числа m_1, m_2, m_3, существует двухпараметрическое семейство целочисленных решений этого уравнения. Тем самым доказано, что плоскость $(\mathbf{Nr}) = 0$ является атомной плоскостью.

Из доказанного следует, что для любого дифрагированного луча $\mathbf{s}$ можно указать атомную плоскость, а следовательно, и бесконечное семейство параллельных ей атомных плоскостей, при зеркальном отражении от которых возникают лучи того же направления, что и рассматриваемый дифрагированный луч. Тем самым доказано, что условием Брэгга–Вульфа охватываются все направления, по которым могут распространяться дифрагированные рентгеновские пучки. Значит, *каждый боковой дифракционный пучок, возникший при дифракции на той или иной атомной плоскости, совпадает по направлению с пучком, зеркально отразившимся какой-то другой атомной плоскостью. Направлениями зеркально отраженных лучей исчерпываются все возможные направления на дифракционные максимумы.*

Конечно, не всякие атомные плоскости эффективно отражают и дают максимумы, действительно наблюдающиеся на опыте. Необходимо, чтобы атомные плоскости были усеяны атомами *достаточно густо*. Иначе интенсивность отраженных лучей может оказаться настолько малой, что они не проявят никакого действия на опыте.

7. В связи с изложенным уточним смысл условия Брэгга–Вульфа. Выделим какое-либо семейство параллельных атомных плоскостей и рассмотрим лучи, возникшие при зеркальном отражении от каждой из этих плоскостей в отсутствие остальных. Условие Брэгга–Вульфа вовсе не означает, что при интерференции таких лучей между собой и с падающим лучом возникнет истинная картина распределения волнового поля в кристалле. Действительно, луч, который при выводе и интерпретации условия Брэгга–Вульфа принято называть лучом, отраженным отдельной атомной плоскостью, в действительности не является таковым. Он возникает в результате сложного процесса, в котором участвуют *атомы всего кристалла*, а не только атомы

рассматриваемой атомной плоскости. В частности, в формировании этого луча участвуют боковые пучки того же направления, возникающие при дифракции на других атомных плоскостях. Однако окончательная дифракционная картина будет такой, как если бы отдельные атомные плоскости только зеркально отражали рентгеновские лучи с некоторыми надлежаще выбранными эффективными коэффициентами отражения и не давали никаких боковых дифракционных пучков.

8. Изложенным по существу завершается доказательство эквивалентности методов Лауэ и Брэгга–Вульфа. Не лишено, однако, интереса показать, что из формул Лауэ без привлечения каких бы то ни было дополнительных соображений вытекает условие Брэгга–Вульфа. Вычислим прежде всего длину вектора $\mathbf{N} = \mathbf{s} - \mathbf{s}_0$, нормального к плоскости, от которой происходит рассматриваемое брэгговское отражение. Угол между единичными векторами $\mathbf{s}_0$ и $\mathbf{s}$ равен 2ϑ, т.е. удвоенному углу скольжения. Следовательно, $\mathbf{N}^2 = 2 - 2(\mathbf{s}\mathbf{s}_0) = 2(1 - \cos 2\vartheta) = 4\sin^2\vartheta$, а потому $N = 2\sin\vartheta$. Примем атомную плоскость, от которой происходит рассматриваемое брэгговское отражение, за координатную плоскость XY. Тогда межплоскостное расстояние d представится выражением

$$d = \left(\mathbf{a}_3 \frac{\mathbf{N}}{N}\right) = \frac{(\mathbf{a}_3\mathbf{N})}{2\sin\vartheta}.$$

Используя последнее соотношение (61.9), получим

$$d = m_3\lambda/(2\sin\vartheta),$$

что совпадает с условием Брэгга–Вульфа (61.6).

9. Рассмотрим два примера на применение условия (61.6).

а) Рентгеновский спектрограф с вращающимся кристаллом (рис. 231). Рентгеновский пучок после диафрагмирования на щелях направляется на монокристалл K известной кристаллической структуры. Кристалл может вращаться вокруг оси, перпендикулярной к падающему лучу, или совершать вращательные колебания вокруг этой оси. Дифрагированный пучок попадает на фотопластинку FF' или в ионизационную камеру. Таким образом, направления падающего и дифрагированного пучков, а с ними и величина угла скольжения ϑ фиксированы. Если падающий пучок монохроматичен или состоит из отдельных монохроматических линий, то при произвольном положении кристалла условие Брэгга–Вульфа, вообще говоря, не будет выполняться. Однако при повороте кристалл может занять такое положение, при котором это условие выполняется. Тогда возникнет отраженный максимум, регистрируемый по почернению фотопластинки или по току в ионизационной камере. Найдя такое положение, можно опреде-

лить угол скольжения ϑ, а затем по формуле (61.6) вычислить длину волны λ.

Разумеется, такой метод дает не абсолютное значение длины волны, а только ее отношение к постоянной решетки d. Для абсолютных измерений надо знать величину d. Для нахожде-

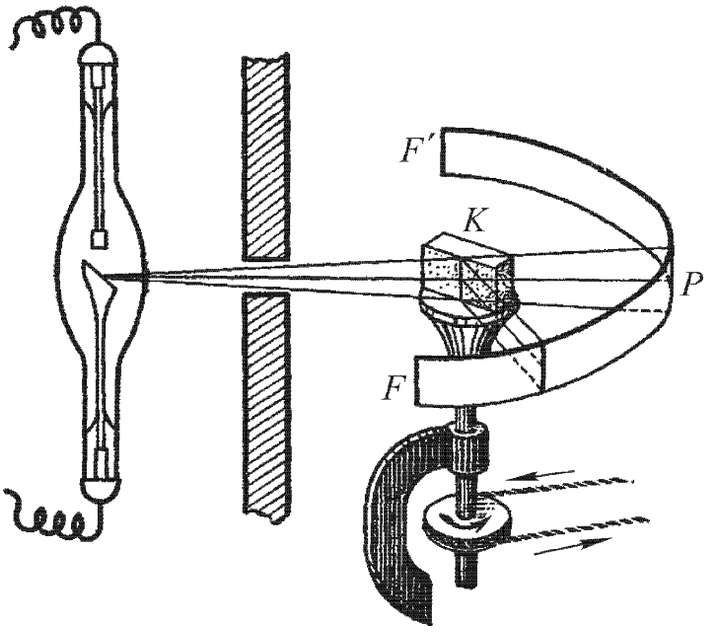


Рис. 231

ния d достаточно независимым способом измерить длину волны какой-либо строго определенной спектральной линии. Это можно сделать, например, с помощью обычной отражательной дифракционной решетки с известным периодом, измерив угол дифракции при скользящем падении луча (см. § 46, п. 8).

б) Метод Дебая–Шеррера–Хелла. Этот метод применяется в рентгеноструктурном анализе для исследования кристаллической структуры металлов и других кристаллических материалов в порошкообразном состоянии. Исследуемый образец (поликристалл) обычно имеет форму цилиндрика и состоит из множества мельчайших кристалликов, беспорядочно ориентированных во всевозможных направлениях. Установка аналогична приведенной на рис. 231. Однако столик, на котором помещается исследуемый образец, неподвижен. На образец направляется монохроматический рентгеновский луч с известной длиной волны λ. Дифракционная картина, называемая *дебаеграммой*, фотографируется.

Происхождение этой картины объясняется следующим образом. Среди множества беспорядочно ориентированных кристалликов найдется еще очень много кристалликов с такими ориентациями, что при заданной длине волны λ будет выполнено

условие Брэгга–Вульфа. Лучи, испытавшие брэгговские отражения от таких кристалликов, образуют поверхность конуса, ось которого направлена вдоль падающего луча, а угол раствора определяется межплоскостным расстоянием d (см. рис. 232, на котором показано отражение от отдельного микрокристаллика; сам микрокристаллик изображен в виде зеркальца). Так как эти расстояния образуют дискретный набор, то за образцом возникнет дискретное семейство конусов с общими вершиной и осью.

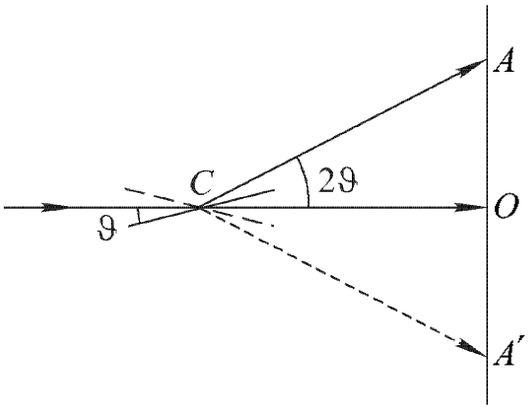


Рис. 232

Если бы фотопластинка была установлена перпендикулярно к этой общей оси, то дебаеграмма состояла бы из концентрических кругов. Измерив радиусы этих кругов, можно определить возможные значения угла ϑ, а затем по формуле Брэгга–Вульфа вычислить соответствующие межплоскостные расстояния и воспользоваться этими данными для воспроизведения кристаллической структуры образца. Чтобы получить действительно все межплоскостные расстояния, фотопластинке придают форму полоски, опоясывающей по окружности исследуемый образец. Дебаеграмма, полученная на кристаллическом порошке NaCl, приведена на рис. 233.

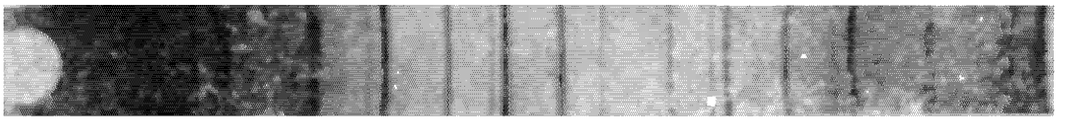


Рис. 233

Нелишне подчеркнуть, что, в отличие от лауэграмм, для получения которых требуется сплошное рентгеновское излучение, дебаеграммы получаются в *монохроматическом свете*. В белом свете никаких дебаеграмм с резкими дифракционными кольцами получиться не может.

10. Остановимся в заключение на вопросе о разрешающей способности дифракционной решетки (одномерной) в рентгеновской области спектра. Формула (47.3) для разрешающей способности решетки относится к случаю, когда решетка дает фраунгоферову дифракционную картину. Как было выяснено выше, в рентгеноскопии реализуется противоположный случай. Фокусирующего устройства нет, а фотопластинка ставится на малых расстояниях от решетки. В таких условиях дифракционные пучки подчиняются геометрической оптике, и формулой (47.3) пользоваться нельзя.

Пусть на решетку падает пучок параллельных лучей с длиной волны λ под углом скольжения α_0. Направление распространения дифрагированного пучка m-го порядка определится из условия $d(\cos\alpha_0 - \cos\alpha) = m\lambda$. Для такого же пучка с близкой длиной волны λ': $d(\cos\alpha_0 - \cos\alpha') = m\lambda'$. Отсюда $d(\cos\alpha' - \cos\alpha) = m(\lambda - \lambda')$, или

$$d\sin\alpha \cdot \delta\alpha = m\,\delta\lambda,$$

где введены обозначения: $\delta\alpha = |\alpha' - \alpha|$, $\delta\lambda = |\lambda' - \lambda|$. Для спектрального разрешения необходимо, чтобы оба пучка пространственно разделились. Если L — расстояние до фотопластинки, измеренное вдоль направления дифрагированного луча, то боковое смещение одного пучка относительно другого равно $x = L\delta\alpha$. Условие разрешения состоит в том, чтобы это смещение было не меньше ширины дифрагированного пучка: $x \geqslant h$. Последняя определяется выражением $h = D\sin\alpha$, где D — длина дифракционной решетки. В результате условие разрешения примет вид

$$\frac{Lm\,\delta\lambda}{d\sin\alpha} \geqslant D\sin\alpha.$$

Минимальному разрешаемому расстоянию соответствует знак равенства. Для него получаем

$$\delta\lambda = \frac{Dd\sin^2\alpha}{Lm}, \tag{61.10}$$

а для разрешающей способности

$$\frac{\lambda}{\delta\lambda} = \frac{Lm\lambda}{Dd\sin^2\alpha}. \tag{61.11}$$

Запишем эту формулу в виде

$$\frac{\lambda}{\delta\lambda} = Nm = \frac{L\lambda}{h^2}, \tag{61.12}$$

или

$$\frac{\lambda}{\delta\lambda} = \frac{Lm\lambda}{hd\sin\alpha}. \tag{61.13}$$

Для повышения разрешающей способности надо применять узкие пучки, а фотопластинку помещать как можно дальше. По сравнению с рэлеевской формулой формула (61.12) дает меньшее значение для разрешающей способности, поскольку она относится к области, в которой $L \ll h^2/\lambda$.

ГЛАВА V

ОТРАЖЕНИЕ И ПРЕЛОМЛЕНИЕ СВЕТА

§ 62. Поляризованный и естественный свет

1. При изучении интерференции и дифракции мы отвлеклись от поперечности световых колебаний, предполагая, что складываемые колебания совершаются в одном направлении. Обратимся теперь к изучению явлений *поляризации света*, типичных именно для *поперечных колебаний*. Плоская волна называется *линейно поляризованной* или *плоскополяризованной*, если электрический вектор $\mathbf{E}$ все время лежит в одной плоскости, в которой расположена также нормаль $\mathbf{N}$ к фронту волны (рис. 234). Эта плоскость называется *плоскостью колебаний* или *плоскостью поляризации* [1]). От поляризованного света следует отличать *естественный свет*. В нем в каждый момент времени векторы $\mathbf{E}$, $\mathbf{H}$, $\mathbf{N}$, хотя и остаются взаимно перпендикулярными, но направления векторов $\mathbf{E}$ и $\mathbf{H}$ беспорядочно изменяются с течением времени. Поэтому естественный свет обладает (статистически) *осевой симметрией* относительно направления его распространения. Для линейно поляризованного света такой симметрии нет. Его свойства в различных плоскостях, проходящих через направление нормали $\mathbf{N}$, различны. Среди этих плоскостей есть две избранные плоскости, в одной из которых лежит вектор $\mathbf{E}$, а в другой — вектор $\mathbf{H}$. Осевая асимметрия сохраняется и для смеси естественного света с линейно поляризованным. Такой свет называется *частично поляризованным*.

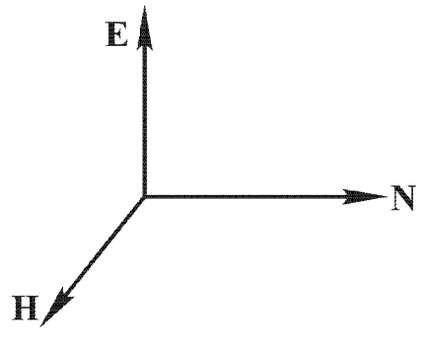


Рис. 234

[1]) По старой терминологии плоскостью поляризации называлась плоскость $(\mathbf{N}, \mathbf{H})$, содержащая магнитный вектор $\mathbf{H}$.

2. Линейно поляризованный свет легко получить, пропустив естественный свет через пластинку турмалина, вырезанную параллельно его кристаллографической (оптической) оси. Турмалин сильно поглощает световые лучи, в которых электрический вектор перпендикулярен к оптической оси. Если же электрический вектор параллелен оси, то такие лучи проходят через турмалин почти без поглощения. Поэтому естественный свет, пройдя через пластинку турмалина, наполовину поглощается и становится линейно поляризованным с электрическим вектором, ориентированным параллельно оптической оси турмалина.

Таким же свойством обладают *поляроиды*, более удобные в обращении. Они представляют собой искусственно приготовляемые коллоидные пленки, служащие для получения поляризованного света. Наиболее распространенным материалом для приготовления поляроидов является *герапатит*, представляющий собой соединение йода с хинином. Этот материал вводят в целлулоидную или желатиновую пленку. В ней ультрамикроскопические кристаллики герапатита каким-либо способом (обычно механически, например протаскиванием вязкой массы через узкую щель) ориентируются своими осями в одном и том же направлении. Полученная масса, подобно турмалину, действует как один кристалл и поглощает световые колебания, электрический вектор которых перпендикулярен к оптической оси. С другими способами получения поляризованного света мы познакомимся в дальнейшем.

Всякий прибор, служащий для получения поляризованного света, называется *поляризатором*. Тот же прибор, применяемый для исследования поляризации света, называется *анализатором*. Таким образом, кристаллы турмалина или поляроиды могут служить и поляризаторами, и анализаторами.

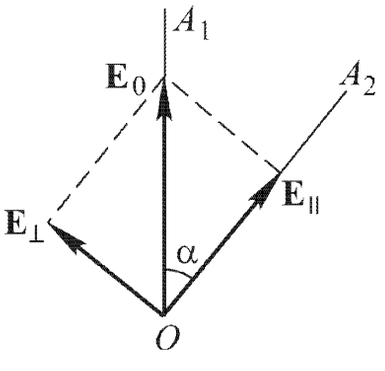


Рис. 235

3. Допустим, что два кристалла турмалина или два поляроида поставлены друг за другом, так что их оси OA_1 и OA_2 образуют между собой некоторый угол α (рис. 235). Первый поляроид пропустит свет, электрический вектор $\mathbf{E}_0$ которого параллелен его оси OA_1. Обозначим через I_0 интенсивность этого света. Разложим $\mathbf{E}_0$ на вектор $\mathbf{E}_\parallel$, параллельный оси OA_2 второго поляроида, и вектор $\mathbf{E}_\perp$, перпендикулярный к ней ($\mathbf{E}_0 = \mathbf{E}_\parallel + \mathbf{E}_\perp$). Составляющая $\mathbf{E}_\perp$ будет задержана вторым поляроидом. Через оба поляроида пройдет свет с электрическим вектором $\mathbf{E} \equiv \mathbf{E}_\parallel$, длина которого $E = E_0 \cos\alpha$. Интенсивность света, прошедшего через оба поляроида, будет

$$I = I_0 \cos^2 \alpha. \qquad (62.1)$$

Такое соотношение справедливо для любого поляризатора и анализатора. Оно называется *законом Малюса* (1775–1812).

4. Важные состояния поляризации возникают при наложении монохроматических волн. Их общий характер одинаков для векторных волн любой физической природы. Для наглядности начнем с механического примера, когда частица совершает два гармонических колебания с одной и той же частотой ω: одно колебание происходит вдоль оси X, другое — вдоль оси Y. Координаты частицы представляются выражениями

$$x = a\cos\omega t, \quad y = b\cos(\omega t + \delta) \qquad (a, b > 0). \tag{62.2}$$

Исключив из этих выражений время t, найдем уравнение траектории частицы:

$$\frac{x^2}{a^2} - \frac{2xy}{ab}\cos\delta + \frac{y^2}{b^2} = \sin^2\delta. \tag{62.3}$$

Это — кривая второго порядка, а именно *эллипс*, так как координаты x и y, как это видно из (62.2), не могут обращаться в бесконечность. Таким образом, *от сложения двух взаимно перпендикулярных колебаний одной и той же частоты возникает движение по эллипсу.*

Для определения направления движения частицы по эллипсу заметим, что начальную фазу δ всегда можно выбрать так, чтобы она была заключена между $-\pi$ и $+\pi$. Тогда колебание с большей фазой называют *опережающим*, а с меньшей фазой — *запаздывающим.* Напишем далее компоненты скорости частицы вдоль осей X и Y:

$$\dot{x} = -\omega a\sin\omega t, \quad \dot{y} = -\omega b\sin(\omega t + \delta).$$

При $t = 0$ получим $x = a$, $\dot{y} = -\omega b\sin\delta$. Таким образом, при $t = 0$ частица находится на вертикальной прямой $x = a$. Если $-\pi < \delta < 0$, то $\dot{y} > 0$, т.е. частица движется вверх, описывая эллипс против часовой стрелки. Если же $0 < \delta < \pi$, то $\dot{y} < 0$, частица движется вниз, описывая эллипс по часовой стрелке (рис. 236). В обоих случаях *движение по эллипсу совершается от положительного конца оси опережающего колебания к положительному концу оси запаздывающего колебания.*

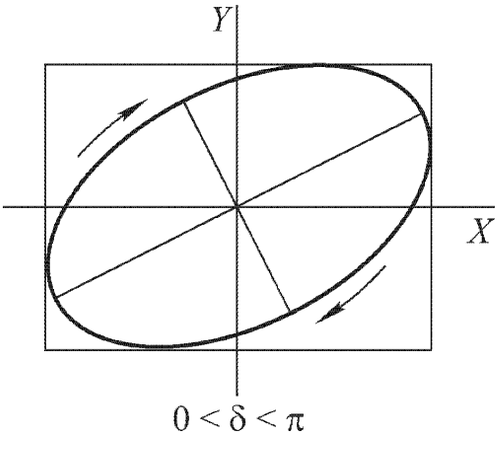


Рис. 236

В зависимости от значений параметров a, b, δ эллипс может вырождаться в отрезки прямой или в окружность. Из (62.3) видно, что для движения по окружности должны выполняться два условия: 1) $\cos\delta = 0$, т.е. $\delta = \pm\pi/2$; 2) $a = b$.

5. Все изложенное относится и к сложению *любых* векторных колебаний, в частности электромагнитных. Две электромагнитные волны, линейно поляризованные во взаимно перпендикулярных плоскостях, при сложении в общем случае дают волну, *поляризованную эллиптически*. В такой волне конец электрического (и магнитного) вектора в каждой точке пространства движется по эллипсу. Если эллипс вырождается в круг, то говорят, что волна *поляризована по кругу*.

Монохроматическое векторное поле всегда поляризовано, в общем случае эллиптически. Векторное поле называется монохроматическим, если все три его проекции на координатные оси совершают гармонические колебания с одной и той же частотой, т.е. представляются формулами вида

$$E_j = C_j(\mathbf{r})\cos[\omega t + \delta_j(\mathbf{r})] \qquad (j = x, y, z). \tag{62.4}$$

Умножая эти выражения на координатные орты $\mathbf{e}_j$ и суммируя по всем j, запишем монохроматическое поле в векторной форме:

$$\mathbf{E} = \mathbf{A}_1(\mathbf{r})\cos\omega t + \mathbf{A}_2(\mathbf{r})\sin\omega t. \tag{62.5}$$

Если векторы $\mathbf{A}_1(\mathbf{r})$ и $\mathbf{A}_2(\mathbf{r})$ везде или в некоторых точках коллинеарны, то в таких точках вектор $\mathbf{E}$ параллелен этим векторам, т.е. поле $\mathbf{E}$ поляризовано линейно. Если же $\mathbf{A}_1$ и $\mathbf{A}_2$ не коллинеарны, то, как видно из формулы (62.5), вектор $\mathbf{E}$ в любой момент времени лежит в плоскости векторов $\mathbf{A}_1(\mathbf{r})$ и $\mathbf{A}_2(\mathbf{r})$. Следовательно, конец вектора $\mathbf{E}$ описывает плоскую кривую. Чтобы найти ее форму, примем направление $\mathbf{A}_1$ за ось X, а перпендикулярное к нему направление, лежащее в плоскости $(\mathbf{A}_1, \mathbf{A}_2)$, — за ось Y. Тогда проекции E_x и E_y представятся в виде

$$E_x = a_x(\mathbf{r})\cos(\omega t + \delta_x), \quad E_y = a_y(\mathbf{r})\cos(\omega t + \delta_y).$$

Задача свелась к сложению двух взаимно перпендикулярных гармонических колебаний одной и той же частоты, сдвинутых по фазе относительно друг друга. От такого сложения получается движение по эллипсу.

Свет, испускаемый реальными источниками, *всегда не поляризован* или *поляризован частично*. Это является лучшим доказательством того, что он не монохроматичен.

§ 63. Число независимых граничных условий в электромагнитной теории света

1. *Формальная теория отражения и преломления света* строится на основе *граничных условий*, которым удовлетворяют векторы электромагнитного поля на границе раздела двух сред. Она определяет величины, характеризующие отраженную

и преломленную волны, но ничего не говорит о *механизме* возникновения этих волн. На последний вопрос, а также на более тонкие вопросы дает ответ *молекулярная теория*. Сначала мы изложим формальную, а затем дадим краткое представление о молекулярной теории отражения и преломления света.

Будем рассматривать все тела как сплошные среды и предположим, что на границах раздела таких сред нет (в сущности, искусственно вводимых) поверхностных зарядов и токов. Тогда на границах раздела должны быть непрерывны тангенциальные составляющие векторов $\mathbf{E}$ и $\mathbf{H}$ и нормальные составляющие векторов $\mathbf{D}$ и $\mathbf{B}$:

$$E_t^{(1)} = E_t^{(2)}, \quad H_t^{(1)} = H_t^{(2)}, \tag{63.1}$$

$$D_n^{(1)} = D_n^{(2)}, \quad B_n^{(1)} = B_n^{(2)}. \tag{63.2}$$

Все эти условия являются следствиями макроскопических уравнений Максвелла в интегральной форме, а потому они верны для *всяких сред*, пока последние можно рассматривать как сплошные. Условия (63.1) вытекают из уравнений

$$\begin{aligned} \oint \mathbf{E}\, d\mathbf{l} &= -\frac{1}{c}\int \dot{\mathbf{B}}\, d\mathbf{F}, \\ \oint \mathbf{H}\, d\mathbf{l} &= \frac{1}{c}\int \dot{\mathbf{D}}\, d\mathbf{F} + \frac{2\pi}{c}\mathcal{I}, \end{aligned} \tag{63.3}$$

а условия (63.2) — из уравнений

$$\oint \mathbf{D}\, d\mathbf{F} = 0, \quad \oint \mathbf{B}\, d\mathbf{F} = 0 \tag{63.4}$$

(см. т. III, § 82).

2. Уравнения (63.1) и (63.2) не совсем независимы. Для исследования этого вопроса построим бесконечно короткий цилиндр, образующие которого перпендикулярны к границе раздела, а основания F_1 и F_2 лежат по разные стороны от нее (рис. 237). На основании теоремы о циркуляции вектора $\mathbf{H}$ получим

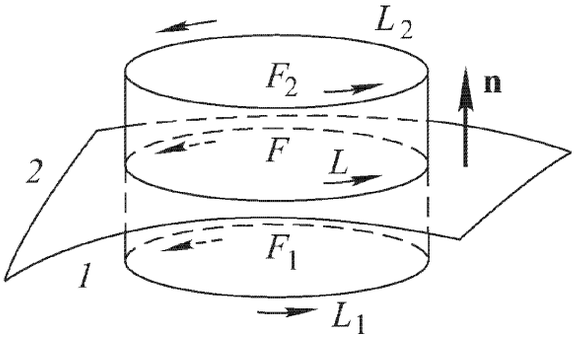


Рис. 237

$$\oint_{L_1} H_l^{(1)}\, dl = \frac{1}{c}\int \dot{D}_n^{(1)}\, dF_1,$$

$$\oint_{L_2} H_l^{(2)}\, dl = \frac{1}{c}\int \dot{D}_n^{(2)}\, dF_2.$$

В пределе, когда высота цилиндра обратится в нуль, контуры L_1 и L_2 сольются в общий контур L, а основания F_1 и F_2 — в общую площадку F, ограниченную контуром L. При

этом, ввиду непрерывности тангенциальных составляющих вектора $\mathbf{H}$, контурные интегралы совпадут между собой, а потому

$$\int\limits_F \dot{D}_n^{(1)}\,dF = \int\limits_F \dot{D}_n^{(2)}\,dF.$$

Аналогично,

$$\int\limits_F \dot{B}_n^{(1)}\,dF = \int\limits_F \dot{B}_n^{(2)}\,dF.$$

Отсюда, ввиду произвольности области интегрирования,

$$\dot{D}_n^{(1)} = \dot{D}_n^{(2)}, \quad \dot{B}_n^{(1)} = \dot{B}_n^{(2)}. \tag{63.5}$$

Таким образом, граничные условия (63.5), как видно из их вывода, являются следствиями граничных условий (63.1) и уравнений Максвелла для циркуляций векторов $\mathbf{E}$ и $\mathbf{H}$. В случае монохроматического поля $\dot{\mathbf{D}} = i\omega\mathbf{D}$, $\dot{\mathbf{B}} = i\omega\mathbf{B}$, так что условия (63.5) переходят в (63.2). Отсюда следует, что для монохроматических полей граничные условия (63.2) выполняются автоматически, если только выполняются условия (63.1). Поэтому в дальнейшем можно пользоваться только условиями (63.1), не заботясь о выполнении условий (63.2).

Если условия (63.1) записать в координатной форме, то получатся четыре уравнения, так как каждый из векторов $\mathbf{E}$ или $\mathbf{H}$ можно разложить на две тангенциальные и одну нормальную составляющие. Таким образом, электродинамика приводит к *четырем независимым граничным условиям*.

В старых теориях упругого эфира число независимых граничных условий было *шесть*: равенство трех составляющих смещений и трех составляющих сил упругих напряжений по обе стороны границы раздела. Чтобы удовлетворить этим шести граничным условиям, вообще говоря, необходимо, чтобы кроме поперечных волн существовали также и *продольные*. Но опыт говорил против существования продольных волн. Возникшую трудность теория пыталась устранить, наделяя эфир такими свойствами, чтобы продольные волны в нем никогда не возникали (несжимаемый или бесконечно сжимаемый эфир). Однако удовлетворительного решения проблемы таким путем получено не было. Электромагнитная теория не знает этой трудности, поскольку число независимых граничных условий в ней равно четырем. Им можно удовлетворить с помощью двух поперечных составляющих отраженной и двух поперечных составляющих преломленной волн.

§ 64. Геометрические законы отражения и преломления волн

1. Необходимость отражения и преломления света на границе раздела двух сред следует уже из граничных условий. Действительно, как будет видно из дальнейшего, граничные условия могут быть удовлетворены, вообще говоря, лишь при наличии отраженной и преломленной волн. Будем обозначать падающую волну индексом e (entfallende), отраженную — индексом r (reflektierte), проходящую — индексом d (durchgehende).

Пусть на плоскую неподвижную границу раздела падает плоская монохроматическая волна

$$\mathbf{E}^{(e)} = \boldsymbol{\mathcal{E}} e^{i(\omega t - \mathbf{k}_1 \mathbf{r})}. \tag{64.1}$$

Из соображений симметрии следует, что отраженная и прошедшая волны

$$\mathbf{E}^{(r)} = \mathbf{R} e^{i(\omega t - \mathbf{k}_1' \mathbf{r})}, \tag{64.2}$$

$$\mathbf{E}^{(d)} = \mathbf{D} e^{i(\omega t - \mathbf{k}_2 \mathbf{r})} \tag{64.3}$$

будут также плоскими и притом той же частоты ω. Равенство частот следует из линейности и однородности граничных условий. Если среды неподвижны, то коэффициенты при напряженностях полей в граничных условиях могут зависеть от координат, но не от времени. Пусть ω_r и ω_d — частоты отраженной и прошедшей волн. Тогда любое из граничных условий (63.1) принимает вид

$$A(\mathbf{r}) e^{i\omega t} + B(\mathbf{r}) e^{i\omega_r t} + C(\mathbf{r}) e^{i\omega_d t} = 0.$$

Коэффициенты $A(\mathbf{r})$, $B(\mathbf{r})$, $C(\mathbf{r})$ отличны от нуля, если только отраженная и прошедшая волны действительно существуют. Следовательно, функции $e^{i\omega t}$, $e^{i\omega_r t}$, $e^{i\omega_d t}$ линейно зависимы, а это возможно лишь при $\omega = \omega_r = \omega_d$. Если граница движется, то A, B и C зависят не только от $\mathbf{r}$, но и от времени. Тогда имеет место изменение частоты (эффект Доплера). В этой главе всюду предполагается, что среды неподвижны.

2. Найдем теперь волновые векторы отраженной и прошедшей волн. Формулы, определяющие эти векторы, называются *геометрическими законами отражения и преломления волн.* Они определяют *направления распространения отраженной и прошедшей волн*, а в случае их неоднородности также и *затухание в пространстве.*

Примем границу раздела сред за координатную плоскость XY. За ось X возьмем линию пересечения плоскости раздела сред с плоскостью падения. Ось Z направим вниз, т.е. в сторону второй среды. Тогда ось Y окажется перпендикулярной к плоскости падения и будет лежать в плоскости раздела сред. Так как

по доказанному частоты падающей, отраженной и прошедшей волн одинаковы, то любое из граничных условий (63.1) примет вид

$$Ae^{-i(k_{1x}x+k_{1y}y)} + Be^{-i(k'_{1x}x+k'_{1y}y)} + Ce^{-i(k_{2x}x+k_{2y}y)} = 0,$$

где A, B, C — постоянные и притом отличные от нуля, если только отраженная и прошедшая волны действительно существуют. Полагая $y = 0$, получаем линейную зависимость между функциями $e^{-ik_{1x}x}$, $e^{-ik'_{1x}x}$, $e^{-ik_{2x}x}$ и поэтому заключаем, что

$$k_{1x} = k'_{1x} = k_{2x}. \tag{64.4}$$

Аналогично,

$$k_{1y} = k'_{1y} = k_{2y}. \tag{64.5}$$

Таким образом, тангенциальные составляющие волновых векторов отраженной и прошедшей волн равны тангенциальной составляющей волнового вектора падающей волны. Остается найти нормальные составляющие этих векторов. Согласно соотношению (5.14),

$$k'^2_1 = k^2_1 = \frac{\omega^2}{c^2}\varepsilon_1, \tag{64.6}$$

$$k^2_2 = \frac{\omega^2}{c^2}\varepsilon_2, \tag{64.7}$$

где ε_1 и ε_2 — диэлектрические проницаемости первой и второй сред. Далее

$$k'_{1z} = -\sqrt{k^2_1 - k^2_{1x}}, \tag{64.8}$$

$$k_{2z} = \sqrt{k^2_2 - k^2_{1x}}. \tag{64.9}$$

Знак минус перед корнем в формуле (64.8) взят потому, что плюс соответствует падающей волне. Что касается знака перед корнем в (64.9), то он будет определен в дальнейшем из физических соображений.

Если падающая волна однородна, то из (64.4), (64.5) и (64.8) следует, что отраженная волна также однородна. Ее волновая нормаль лежит в плоскости падения, а угол отражения равен углу падения. Для проходящей волны надо различать два случая.

П е р в ы й с л у ч а й. $k^2_2 > k^2_{1x}$, т.е. преломленная волна однородна. Определим, какой знак следует выбрать в этом случае перед квадратным корнем в (64.9). Знаку плюс соответствует волна, распространяющаяся от границы раздела, — направление ее распространения обозначено на рис. 238 сплошной стрелкой. Знаку минус соответствует волна, идущая к границе раздела, — ее направление обозначено штриховой стрелкой.

Эти стрелки указывают направления распространения волновых фронтов, т.е. плоскостей равных фаз. Ясно, что отраженная и преломленная волны должны быть *уходящими* от границы раздела. Этим требованием обеспечивается однозначность решения задачи. Однако требование ухода относится не к фазе, а к энергии волны. Можно показать, что в случае электромагнитных волн в изотропных средах направления распространения фазы и энергии волны совпадают. Поэтому знак минус перед корнем в (64.9) следует отбросить; условиям задачи удовлетворяет только знак плюс.

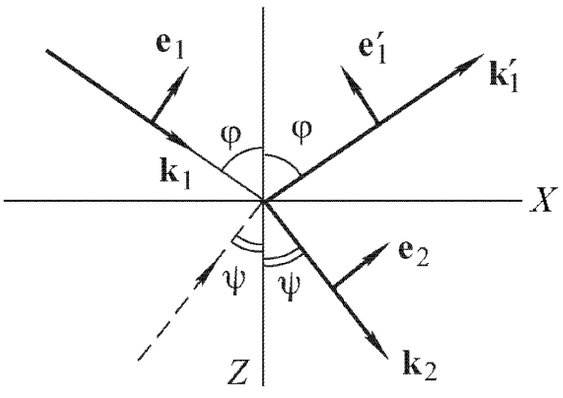


Рис. 238

Как видно из (64.4), нормали к падающей и преломленной волнам лежат в плоскости падения. Если φ — угол падения, а ψ — угол преломления, то

$$k_{1x} = k_1 \sin\varphi = \frac{\omega}{v_1}\sin\varphi, \quad k_{2x} = k_2 \sin\psi = \frac{\omega}{v_2}\sin\psi,$$

откуда на основании (64.4)

$$\frac{\sin\varphi}{\sin\psi} = \frac{v_1}{v_2} = \frac{n_2}{n_1} = \sqrt{\frac{\varepsilon_2}{\varepsilon_1}}. \tag{64.10}$$

В т о р о й с л у ч а й. $k_2^2 < k_{1x}^2$, или $\omega^2/v_2^2 < \omega^2 \sin^2\varphi/v_1^2$, откуда $\sin\varphi > v_1/v_2 = n$. Здесь n — относительный показатель преломления второй среды относительно первой. Так как $\sin\varphi < 1$, то рассматриваемый случай возможен только при $n < 1$. Составляющая k_{2z} чисто мнимая, а волна во второй среде, если она существует, неоднородная. Знак корня в (64.9) определится из требования, чтобы при удалении от границы раздела амплитуда волны затухала. Этому требованию удовлетворяет только выражение

$$k_{2z} = -i\sqrt{k_{1x}^2 - k_2^2} = -\frac{i}{2h}. \tag{64.11}$$

В самом деле, тогда (64.3) принимает вид

$$\mathbf{E}^{(d)} = \mathbf{D}e^{-z/2h}e^{i(\omega t - k_{1x}x)}, \tag{64.12}$$

т.е. волна во второй среде будет затухать в направлении оси Z, чего не получилось бы, если бы в (64.11) вместо минуса взять плюс.

Плоскости равных фаз волны (64.12) перпендикулярны к оси X и распространяются вдоль нее с фазовой скоростью $v_x =$

$= \omega/k_{1x}$. Плоскости равных амплитуд параллельны границе раздела. При смещении вглубь среды на h интенсивность волны (пропорциональная квадрату амплитуды) убывает в e раз. Величина h называется *глубиной проникновения* волны во вторую среду. Она равна

$$h = \frac{1}{\sqrt{k_{1x}^2 - k_2^2}} = \frac{\lambda_1}{4\pi\sqrt{\sin^2\varphi - n^2}}, \tag{64.13}$$

где λ_1 — длина волны в первой среде.

Из (64.12) видно, что на больших (по сравнению с глубиной проникновения) расстояниях от границы раздела волна во второй среде практически полностью затухает. А так как поглощения света нет, то энергия падающей волны, проникшая во вторую среду, должна снова целиком возвратиться в первую среду. Иными словами, при $\sin\varphi > n$ отражение света должно быть *полным*. Угол φ_0, определяемый соотношением $\sin\varphi_0 = n$, называется *предельным углом полного отражения*.

3. В случае обыкновенного отражения

$$\begin{aligned} &\sin\varphi = k_{1x}/k_1, \quad \cos\varphi = k_{1z}/k_1, \\ &\sin\psi = k_{2x}/k_2, \quad \cos\psi = k_{2z}/k_2. \end{aligned} \tag{64.14}$$

В случае полного отражения не существует вещественного угла ψ, удовлетворяющего соотношениям (64.14), так как они дают для $\sin\psi$ значения, превосходящие единицу, а для $\cos\psi$ — мнимые значения. Однако в целях сохранения единой формы записи при обыкновенном и полном отражениях целесообразно сохранить формулы (64.14) как простые определения $\sin\psi$ и $\cos\psi$. Поскольку эти величины удовлетворяют соотношению $\sin^2\psi + \cos^2\psi = 1$, они могут рассматриваться как синус и косинус комплексного аргумента ψ в смысле теории функций комплексного переменного:

$$\sin\psi = \frac{e^{i\psi} - e^{-i\psi}}{2i}, \quad \cos\psi = \frac{e^{i\psi} + e^{-i\psi}}{2}.$$

Если $\sin\psi$ и $\cos\psi$ известны, то этими формулами аргумент ψ определяется с точностью до целого кратного от 2π. Это не может сказаться на однозначности физических выводов, так как во все формулы будет входить не сам комплексный угол ψ, а его синус и косинус. К так определенным функциям $\sin\psi$ и $\cos\psi$ применимы все формальные соотношения обычной тригонометрии. Поэтому над комплексными $\sin\psi$ и $\cos\psi$ можно выполнять все преобразования, как если бы они были обыкновенными синусом и косинусом.

Заметим, наконец, что вместо (64.14) можно написать

$$\sin\psi = \sin\varphi/n, \quad \cos\psi = -\frac{i}{n}\sqrt{\sin^2\varphi - n^2}. \tag{64.15}$$

§ 65. Формулы Френеля

1. При выводе геометрических законов отражения и преломления волн явный вид граничных условий не использовался. Для определения амплитуд отраженной и проходящей волн необходимо использовать граничные условия *в явном виде.*

Разложим электрическое поле каждой волны на две составляющие. Одна из них лежит в плоскости падения, другая перпендикулярна к этой плоскости. Часто эти составляющие называют *главными составляющими* соответствующих волн. Они обозначаются значками $\parallel$ и $\perp$ соответственно. Пусть $\mathbf{e}_x$, $\mathbf{e}_y$, $\mathbf{e}_z$ — единичные векторы вдоль координатных осей, а $\mathbf{e}_1$, $\mathbf{e}'_1$, $\mathbf{e}_2$ — единичные векторы, лежащие в плоскости падения и перпендикулярные соответственно к падающему, отраженному и преломленному лучам (рис. 238). Тогда

$$\mathbf{e}_1 = \frac{[\mathbf{e}_y \mathbf{k}_1]}{k_1}, \quad \mathbf{e}'_1 = \frac{[\mathbf{e}_y \mathbf{k}'_1]}{k_1}, \quad \mathbf{e}_2 = \frac{[\mathbf{e}_y \mathbf{k}_2]}{k_2}. \tag{65.1}$$

В случае полного отражения вектор $\mathbf{e}_2$ — комплексный, его геометрическая интерпретация как единичного вектора, перпендикулярного к преломленному лучу, теряет смысл. Поэтому, чтобы охватить не только обыкновенное, но и полное отражение, дальнейшие вычисления проводятся чисто аналитически, не прибегая к чертежу. Введем разложения

$$\begin{aligned} \boldsymbol{\mathcal{E}} &= \mathcal{E}_\perp \mathbf{e}_y + \mathcal{E}_\parallel \mathbf{e}_1, \\ \mathbf{R} &= R_\perp \mathbf{e}_y + R_\parallel \mathbf{e}'_1, \\ \mathbf{D} &= D_\perp \mathbf{e}_y + D_\parallel \mathbf{e}_2 \end{aligned} \tag{65.2}$$

и определим амплитуды $R_\perp$, $R_\parallel$, $D_\perp$, $D_\parallel$. Умножая скалярно первое из уравнений (65.2) на $\mathbf{e}_x$, находим

$$\mathcal{E}_x = \mathcal{E}_\parallel (\mathbf{e}_1 \mathbf{e}_x) = \frac{\mathcal{E}_\parallel}{k_1} (\mathbf{e}_x [\mathbf{e}_y \mathbf{k}_1]) = \frac{\mathcal{E}_\parallel}{k_1} ([\mathbf{e}_x \mathbf{e}_y] \mathbf{k}_1) = \frac{\mathcal{E}_\parallel}{k_1} (\mathbf{e}_z \mathbf{k}_1) = \mathcal{E}_\parallel \cos\varphi.$$

Аналогично, $\mathcal{E}_y = \mathcal{E}_\perp$, $\mathcal{E}_z = -\mathcal{E}_\parallel \sin\varphi$. Таким образом, x-, y-, z-составляющие электрического поля на границе раздела сред (т.е. при $z = 0$) будут

$$E_x^{(e)} = \cos\varphi \cdot \mathcal{E}_\parallel, \quad E_y^{(e)} = \mathcal{E}_\perp, \quad E_z^{(e)} = -\sin\varphi \cdot \mathcal{E}_\parallel. \tag{65.3}$$

При этом мы всюду опустили фазовые множители, так как в любой точке границы раздела они одинаковы для всех трех волн: падающей, отраженной и прошедшей. Магнитные свойства сред учитывать не будем, считая, что $\mathbf{H} \equiv \mathbf{B}$. Магнитное поле вычисляем по формуле (5.6) и находим

$$H_x^{(e)} = -n_1 \cos\varphi \cdot \mathcal{E}_\perp, \quad H_y^{(e)} = n_1 \mathcal{E}_\parallel, \quad H_z^{(e)} = n_1 \sin\varphi \cdot \mathcal{E}_\perp. \tag{65.3а}$$

Для отраженной волны:

$$\begin{aligned} E_x^{(r)} &= -\cos\varphi\cdot R_{\|}, & E_y^{(r)} &= R_{\perp}, & E_z^{(r)} &= -\sin\varphi\cdot R_{\|},\\ H_x^{(r)} &= n_1\cos\varphi\cdot R_{\perp}, & H_y^{(r)} &= n_1 R_{\|}, & H_z^{(r)} &= n_1\sin\varphi\cdot R_{\perp}. \end{aligned} \tag{65.4}$$

Для прошедшей волны:

$$\begin{aligned} E_x^{(d)} &= \cos\psi\cdot D_{\|}, & E_y^{(d)} &= D_{\perp}, & E_z^{(d)} &= -\sin\psi\cdot D_{\|},\\ H_x^{(d)} &= -n_2\cos\psi\cdot D_{\perp}, & H_y^{(d)} &= n_2 D_{\|}, & H_z^{(d)} &= D_{\|} n_2\sin\psi. \end{aligned} \tag{65.5}$$

Для определения четырех неизвестных $R_{\perp}$, $R_{\|}$, $D_{\perp}$, $D_{\|}$ электродинамика дает четыре независимых граничных условия:

$$\begin{aligned} E_x^{(e)} + E_x^{(r)} &= E_x^{(d)}, & E_y^{(e)} + E_y^{(r)} &= E_y^{(d)},\\ H_x^{(e)} + H_x^{(r)} &= H_x^{(d)}, & H_y^{(e)} + H_y^{(r)} &= H_y^{(d)}. \end{aligned} \tag{65.6}$$

Подставляя в них найденные выше значения, получим

$$\begin{aligned} \cos\varphi\cdot(\mathcal{E}_{\|} - R_{\|}) &= \cos\psi\cdot D_{\|}, & \mathcal{E}_{\perp} + R_{\perp} &= D_{\perp},\\ n_1\cos\varphi\cdot(\mathcal{E}_{\perp} - R_{\perp}) &= n_2\cos\psi\cdot D_{\perp}, & n_1(\mathcal{E}_{\|} + R_{\|}) &= n_2 D_{\|}. \end{aligned}$$

Отсюда

$$\begin{aligned} r_{\perp} \equiv \frac{R_{\perp}}{\mathcal{E}_{\perp}} &= \frac{n_1\cos\varphi - n_2\cos\psi}{n_1\cos\varphi + n_2\cos\psi}, & d_{\perp} \equiv \frac{D_{\perp}}{\mathcal{E}_{\perp}} &= \frac{2n_1\cos\varphi}{n_1\cos\varphi + n_2\cos\psi},\\ r_{\|} \equiv \frac{R_{\|}}{\mathcal{E}_{\|}} &= \frac{n_2\cos\varphi - n_1\cos\psi}{n_2\cos\varphi + n_1\cos\psi}, & d_{\|} \equiv \frac{D_{\|}}{\mathcal{E}_{\|}} &= \frac{2n_1\cos\varphi}{n_2\cos\varphi + n_1\cos\psi}. \end{aligned} \tag{65.7}$$

Отношения $R_{\perp}/\mathcal{E}_{\perp}$, $D_{\perp}/\mathcal{E}_{\perp}$, ... называются *коэффициентами Френеля*. С помощью соотношений (64.10) формулы (65.7) легко привести к виду

$$\begin{aligned} \frac{R_{\perp}}{\mathcal{E}_{\perp}} &= -\frac{\sin(\varphi-\psi)}{\sin(\varphi+\psi)}, & \frac{D_{\perp}}{\mathcal{E}_{\perp}} &= \frac{2\cos\varphi\sin\varphi}{\sin(\varphi+\psi)},\\ \frac{R_{\|}}{\mathcal{E}_{\|}} &= \frac{\operatorname{tg}(\varphi-\psi)}{\operatorname{tg}(\varphi+\psi)}, & \frac{D_{\|}}{\mathcal{E}_{\|}} &= \frac{2\cos\varphi\sin\varphi}{\sin(\varphi+\psi)\cos(\varphi-\psi)}. \end{aligned} \tag{65.8}$$

Эти формулы и решают поставленную задачу. Они были впервые получены в 1821 г. Френелем из механической теории упругого эфира с помощью весьма простых, но нестрогих и противоречивых рассуждений. Последующие попытки строгого решения задачи с помощью уравнений упругости ухудшили дело: в случае волн, поляризованных в плоскости падения, они приводили к формулам, не согласующимся с опытом. Только электромагнитная теория света впервые дала строгое и удовлетворительное решение задачи.

2. При обыкновенном отражении угол ψ, а с ним и все коэффициенты Френеля *вещественны.* Отсюда следует, что отражение и преломление *не сопровождаются изменением фаз*, за исключением, быть может, изменения фазы отраженной волны на 180°. *Если падающая волна была линейно поляризована, то отраженная и преломленная волны будут также линейно поляризованы.*

При нормальном падении волны формулы (65.7) дают

$$\frac{R_\perp}{\mathcal{E}_\perp} = -\frac{R_\parallel}{\mathcal{E}_\parallel} = \frac{n_1 - n_2}{n_1 + n_2}. \tag{65.9}$$

В этом случае исчезает разница между волнами, поляризованными в плоскости падения и перпендикулярно к ней. Различие в знаках коэффициентов Френеля $R_\perp/\mathcal{E}_\perp$ и $R_\parallel/\mathcal{E}_\parallel$ связано с условным выбором положительных направлений, в которых проведены векторы $\mathbf{e}_1$ и $\mathbf{e}_1'$: в предельном случае нормального падения эти векторы направлены прямо противоположно (см. рис. 238). Если $n_1 < n_2$, то $R_\perp/\mathcal{E}_\perp < 0$; если $n_1 > n_2$, то $R_\perp/\mathcal{E}_\perp > 0$. *В первом случае отражение света сопровождается скачкообразным изменением фазы на* 180°. *Во втором оно происходит без изменения фазы.*

При скользящем падении, когда угол φ близок к 90°,

$$\frac{R_\perp}{\mathcal{E}_\perp} = \frac{R_\parallel}{\mathcal{E}_\parallel} \to -1,$$

т.е. отражение практически полное. С этим связано происхождение ярких и красивых изображений противоположных берегов в спокойных водоемах. Этим же объясняется, почему изображение заходящего солнца в тех же условиях по яркости почти не уступает самому солнцу.

Как видно из формул (65.7) или (65.8), отношение $R_\perp/\mathcal{E}_\perp$ никогда не обращается в нуль, за исключением тривиального случая $n_1 = n_2$. Напротив, при $\varphi + \psi = \pi/2$ знаменатель $\operatorname{tg}(\varphi + \psi)$ обращается в бесконечность. В этом случае $R_\parallel = 0$, т.е. отражение отсутствует. Таким образом, если электрический вектор падающей волны лежит в плоскости падения, то при некотором угле падения отражение света исчезает. В этом состоит *закон Брюстера* (1781–1868), открытый им экспериментально в 1815 г. Соответствующий угол падения φ_B называется *углом Брюстера.* Для его вычисления заметим, что при $\varphi + \psi = \pi/2$ направления прошедшего и отраженного лучей взаимно перпендикулярны (рис. 239). Следовательно, $\cos\varphi_B = \sin\psi_B = \frac{1}{n}\sin\varphi_B$, т.е.

$$\operatorname{tg}\varphi_B = n. \tag{65.10}$$

Если неполяризованный свет падает под углом Брюстера, то

составляющая с электрическим вектором $\mathbf{E}_{\parallel}$ *отражаться не будет. Отраженный свет окажется линейно поляризованным и притом перпендикулярно к плоскости падения* [1]).

Отражение под углом Брюстера дает простейший способ получения поляризованного света. Недостаток этого способа — малая интенсивность отраженного света. Для его устранения применяют несколько стеклянных пластинок, сложенных в стопу (стопу Столетова). При большом числе пластинок отражается почти половина падающего света. Если свет падает под углом Брюстера, то отраженный свет оказывается полностью поляризованным перпендикулярно к плоскости падения. Прошедший свет поляризован в плоскости падения, но эта поляризация не совсем полная. В настоящее время стопа для получения поляризованного света почти не употребляется, так как существуют более удобные и совершенные способы (*поляризационные призмы, полароиды*).

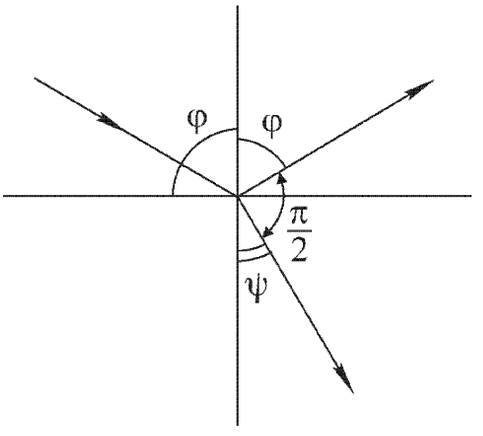


Рис. 239

3. Отношение отраженной энергии к энергии падающей называется *коэффициентом отражения*. Так как энергия пропорциональна квадрату амплитуды, то формулы (65.7) дают для коэффициентов отражения главных составляющих падающей волны следующие выражения:

$$\rho_{\perp} = \left(\frac{\cos\varphi - n\cos\psi}{\cos\varphi + n\cos\psi}\right)^2, \quad \rho_{\parallel} = \left(\frac{n\cos\varphi - \cos\psi}{n\cos\varphi + \cos\psi}\right)^2. \tag{65.11}$$

Коэффициент отражения при нормальном падении R называется *отражательной способностью*:

$$R = \left(\frac{n-1}{n+1}\right)^2. \tag{65.12}$$

Для стекла ($n = 1{,}5$) $R = 0{,}04 = 4\,\%$; для воды ($n = 1{,}33$) $R = 2\,\%$. При отражении от воды длинных электромагнитных волн ($n = \sqrt{\varepsilon} = 9$) $R = 64\,\%$. Эти цифры показывают, что ни вода, ни стекло при нормальном падении не могут служить зеркалом. Обычные зеркала используют отражение света от металлических поверхностей. Стекло служит только для защиты их

[1]) Малюс, открывший поляризацию света при отражении, назвал плоскость падения плоскостью поляризации отраженного света. Здесь под *плоскостью поляризации* понимается плоскость, перпендикулярная к плоскости поляризации Малюса.

задних посеребренных поверхностей. Однако наличие даже слабого отражения от передней стороны стекла делает такие посеребренные с задней стороны зеркала непригодными для оптических целей. Для этих целей необходимо покрывать металлом (лучше всего родием) переднюю поверхность стекла.

Отношение b прошедшей энергии к энергии падающей называется *коэффициентом пропускания*. Коэффициент пропускания можно также определить как отношение нормальных компонент усредненного по времени вектора Пойнтинга прошедшей и падающей волн. На основании (65.3) и (65.3а) находим для падающей волны

$$S_z^{(e)} = \frac{c}{16\pi}[E_x^{(e)} H_y^{*(e)} - E_y^{(e)} H_x^{*(e)}] + \text{компл. сопр.} = \\ = \frac{cn_1}{8\pi}\cos\varphi(|\mathcal{E}_\perp|^2 + |\mathcal{E}_\parallel|^2),$$

а для прошедшей

$$S_z^{(d)} = \frac{cn_2}{8\pi}\cos\psi(|D_\perp|^2 + |D_\parallel|^2).$$

Предполагая сначала, что вектор $\mathbf{E}$ перпендикулярен к плоскости падения, а затем параллелен ей, находим

$$b_\perp = \frac{n\cos\psi}{\cos\varphi}\left(\frac{2\cos\varphi}{\cos\varphi + n\cos\psi}\right)^2, \quad b_\parallel = \frac{n\cos\psi}{\cos\varphi}\left(\frac{2\cos\varphi}{n\cos\varphi + \cos\psi}\right)^2. \tag{65.13}$$

Нетрудно проверить, что $\rho_\perp + b_\perp = \rho_\parallel + b_\parallel = 1$, как и должно быть согласно закону сохранения энергии.

Коэффициент пропускания при нормальном падении

$$B = \frac{4n}{(n+1)^2} \tag{65.14}$$

называется *поверхностной прозрачностью*. Измерения отражательной способности и поверхностной прозрачности дают удобный метод измерения показателей преломления в инфракрасной части спектра.

Найдем теперь коэффициент отражения для случая, когда плоскость поляризации падающей волны составляет с плоскостью падения угол α. Этот угол называется *азимутом поляризации* падающей волны. Очевидно, $\mathcal{E}_\parallel = A\cos\alpha$, $\mathcal{E}_\perp = A\sin\alpha$, где A — амплитуда падающей волны. Полная отраженная энергия пропорциональна $\rho_\parallel\mathcal{E}_\parallel^2 + \rho_\perp\mathcal{E}_\perp^2 = A^2(\rho_\parallel\cos^2\alpha + \rho_\perp\sin^2\alpha)$. Следовательно, коэффициент отражения равен

$$\rho_\alpha = \rho_\parallel\cos^2\alpha + \rho_\perp\sin^2\alpha. \tag{65.15}$$

Коэффициент отражения неполяризованного света ρ получается отсюда путем усреднения по α. Так как в неполяризованном свете все направления электрического (и магнитного) вектора

представлены с одинаковой вероятностью, то $\overline{\cos^2\alpha} = \overline{\sin^2\alpha} = 1/2$. Поэтому

$$\rho = \frac{1}{2}(\rho_{\parallel} + \rho_{\perp}). \qquad (65.16)$$

На рис. 240 приведены теоретические кривые для коэффициентов отражения видимого света от стекла ($n = 1{,}5$), а на

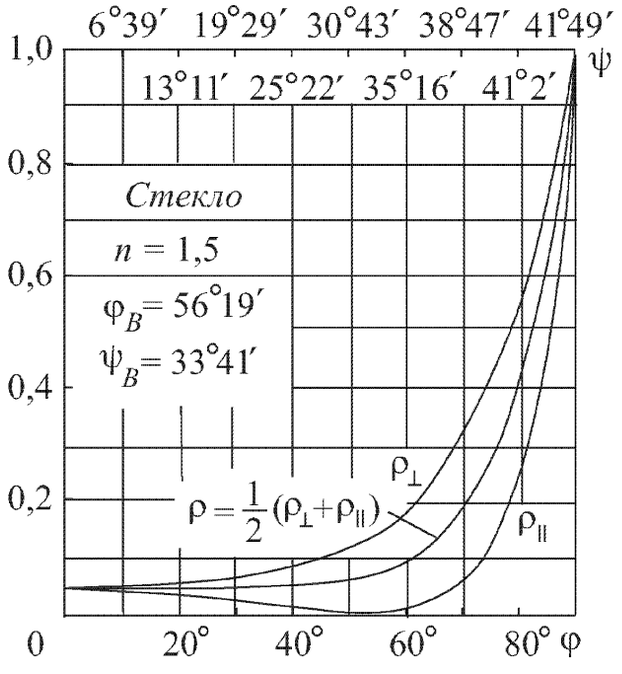


Рис. 240

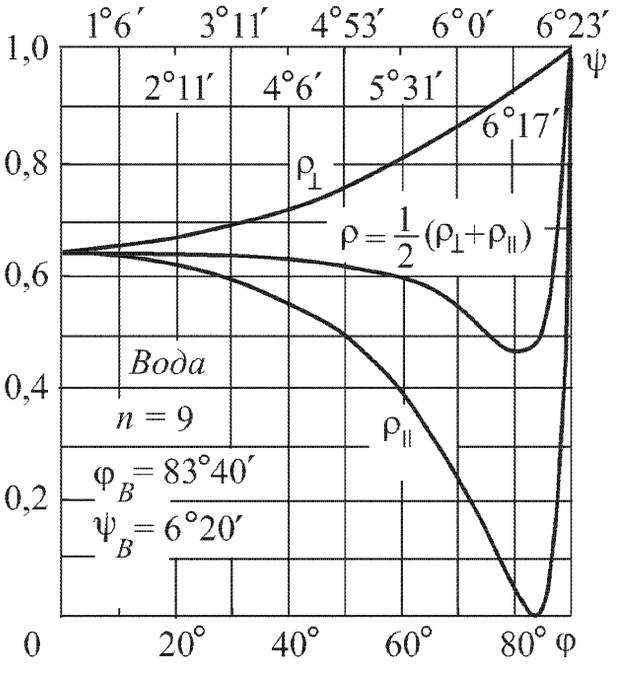


Рис. 241

рис. 241 — для радиоволн при их отражении от поверхности воды ($n = \sqrt{\varepsilon_0} = 9$)

Многочисленные измерения коэффициентов отражения при различных углах падения и при различных поляризациях падающей волны, предпринимавшиеся с целью проверки формул Френеля, находятся в очень хорошем согласии с этими формулами как в случае видимого света, так и в случае радиоволн. Исключение составляет случай отражения под углом Брюстера и в его ближайшей окрестности, где наблюдаются незначительные отступления от формул Френеля (см. § 70).

4. Формулы Френеля мы вывели в предположении, что свет *монохроматический.* Однако в случае обыкновенного отражения в эти формулы не входит длина волны, а отражение не сопровождается изменением фазы. Поэтому в случае недиспергирующих сред и обыкновенного отражения формулы Френеля справедливы и для *немонохроматических волн.* Надо только под $\mathcal{E}_\perp$, $R_\perp$, $D_\perp$ и т.д. понимать соответствующие компоненты напряженностей электрического поля падающей, отраженной и прошедшей волн *на границе раздела.* Это непосредственно следует из теоремы Фурье и принципа суперпозиции.

§ 66. Полное отражение

1. В случае полного отражения $\cos\psi$ — чисто мнимая величина, при этом формулы Френеля удобно записать в виде

$$\frac{R_\perp}{\mathcal{E}_\perp}=\frac{\cos\varphi+i\sqrt{\sin^2\varphi-n^2}}{\cos\varphi-i\sqrt{\sin^2\varphi-n^2}},\quad \frac{D_\perp}{\mathcal{E}_\perp}=\frac{2\cos\varphi}{\cos\varphi-i\sqrt{\sin^2\varphi-n^2}},$$
$$\frac{R_\parallel}{\mathcal{E}_\parallel}=\frac{n^2\cos\varphi+i\sqrt{\sin^2\varphi-n^2}}{n^2\cos\varphi-i\sqrt{\sin^2\varphi-n^2}},\quad \frac{D_\parallel}{\mathcal{E}_\parallel}=\frac{2n\cos\varphi}{n^2\cos\varphi-i\sqrt{\sin^2\varphi-n^2}}. \tag{66.1}$$

Из этих формул видно, что $|R_\perp|=|\mathcal{E}_\perp|$ и $|R_\parallel|=|\mathcal{E}_\parallel|$, т.е. отражение действительно является «полным». В то же время неоднородная волна во второй среде, вообще говоря, не исчезает. Например, если $\varphi=\varphi_0=\arcsin n$, то $D_\perp=2\mathcal{E}_\perp$. Здесь нет нарушения закона сохранения энергии. Дело в том, что формулы Френеля относятся к *монохроматическому полю*, т.е. к некоторому *установившемуся процессу*. А в этом случае закон сохранения энергии требует только, чтобы количество энергии во второй среде менялось во времени *периодически*. Иными словами, *среднее за период изменение энергии во второй среде должно быть равно нулю*. Что это условие действительно выполняется, показывает следующее простое вычисление.

Вычислим усредненную по времени нормальную составляющую вектора Пойнтинга во второй среде. Пусть электрический вектор лежит в плоскости падения. Тогда

$$\overline{S}_z=\frac{c}{16\pi}E_x^*H_y+\text{компл. сопр.}$$

Поле во второй среде определяется формулами (65.5). Так как $\cos\psi$ — величина чисто мнимая, то из (65.5) видно, что между электрическим и магнитным полями существует сдвиг фаз в $90°$. При этом величина $E_x^*H_y$ также чисто мнимая, а потому $\bar{S}_z=0$. То же справедливо и в случае, когда электрический вектор перпендикулярен к плоскости падения. Этим наше утверждение доказано.

Вычисление остальных компонент усредненного вектора Пойнтинга дает

$$\bar{S}_x=\frac{cn_1}{8\pi}(|D_\perp|^2+|D_\parallel|^2)\sin\varphi e^{-z/h},\quad \bar{S}_y=0, \tag{66.2}$$

где h — глубина проникновения, определяемая выражением (64.13). Таким образом, во второй среде действительно имеется энергия, средний запас которой со временем не изменяется, а средний поток параллелен оси X. Существование такой энергии при установившемся режиме, очевидно, не противоречит полному отражению падающей волны.

Чтобы ответить на вопрос, как появилась энергия во второй среде, надо было бы исследовать *процесс установления колебаний*. Можно, например, рассмотреть квазимонохроматическую волну с передовым фронтом, перед которым нет никакого волнового возмущения. Пока фронт волны не достиг границы раздела, во второй среде нет поля. Как только волна дойдет до границы раздела, она сначала будет почти целиком проникать во вторую среду и лишь частично отражаться. По мере установления колебаний коэффициент отражения будет быстро нарастать и стремиться к своему предельному значению — единице. Полное отражение имеет место лишь для установившегося режима. Пока процесс не установился, отражение *всегда частичное*.

2. Пусть две одинаковые среды разделены тонким воздушным промежутком, на который падает волна под углом, бóльшим предельного. Можно ожидать (см. § 67), что она проникнет в воздушный зазор в виде неоднородной волны. При достаточно малой толщине зазора неоднородная волна достигнет его второй границы еще не очень сильно ослабленной. Вступив из воздушного зазора во вторую среду, волна снова превратится в однородную и может быть обнаружена обычными средствами. Опыт такого рода был выполнен еще Ньютоном и рассматривался как экспериментальное доказательство проникновения света в оптически менее плотную среду при полном отражении. (Строго говоря, здесь нет полного отражения: оно возмущено наличием второй среды.) Ньютон прижимал к гипотенузной грани прямоугольной призмы другую призму, сошлифованную сферически (рис. 242). Оказалось, что свет проходит во вторую призму не только в месте соприкосновения (оптический контакт), но и в небольшом кольце вокруг него — там, где толщина воздушного промежутка сравнима с длиной волны. При наблюдении в белом свете внешний край кольца имел красноватую окраску. Этого и следовало ожидать, так как глубина проникновения пропорциональна длине волны: для красных лучей она больше, чем для синих.

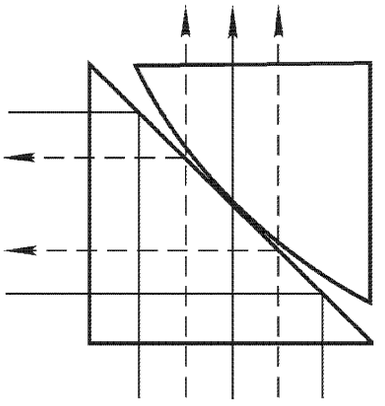


Рис. 242

Меняя толщину воздушного зазора, можно изменять интенсивность проходящего света. На этом основано устройство *светового телефона*, запатентованного фирмой Цейсс. Роль одной из сред играет прозрачная мембрана, колеблющаяся под действием падающего на нее звука. Свет, прошедший через зазор, меняет интенсивность в такт с изменениями силы звука. Падая на фотоэлемент, он возбуждает переменный ток, сила которо-

го изменяется также в соответствии с изменениями силы звука. Этот ток усиливается и используется в дальнейшем.

Явления прохождения волн через тонкие зазоры, аналогичные описанному, не специфичны для оптики. Они имеют место для волн *любой природы* и могут происходить, когда фазовая скорость в зазоре превосходит фазовую скорость в окружающих средах. Легче всего они реализуются для длинных волн (радиотехника, акустика). Особо важное значение эти явления приобрели в ядерной и атомной физике, а также в теории твердого тела. Об этом будет говориться в пятом томе нашего курса.

3. При обыкновенном отражении оба отношения $R_\perp/\mathcal{E}_\perp$ и $R_\parallel/\mathcal{E}_\parallel$ всегда вещественны. При полном отражении они, вообще говоря, *комплексны*. Это значит, что при полном отражении *фаза волны испытывает скачок*, и притом, как правило, отличный от нуля или π. Пусть волна поляризована перпендикулярно к плоскости падения. Можно написать

$$\cos\varphi + i\sqrt{\sin^2\varphi - n^2} = Ae^{i\delta_\perp/2},$$

где A и $\delta_\perp$ — величины вещественные. Тогда

$$\cos\varphi - i\sqrt{\sin^2\varphi - n^2} = Ae^{-i\delta_\perp/2}, \quad \frac{R_\perp}{\mathcal{E}_\perp} = e^{i\delta_\perp}.$$

Отсюда видно, что $\delta_\perp$ и есть интересующий нас скачок фазы. Приравнивая вещественные и мнимые части, получим

$$A\cos\frac{\delta_\perp}{2} = \cos\varphi, \quad A\sin\frac{\delta_\perp}{2} = \sqrt{\sin^2\varphi - n^2}.$$

Отсюда

$$\operatorname{tg}\frac{\delta_\perp}{2} = \frac{\sqrt{\sin^2\varphi - n^2}}{\cos\varphi}. \tag{66.3}$$

Аналогично, для волны, поляризованной в плоскости падения,

$$\operatorname{tg}\frac{\delta_\parallel}{2} = \frac{\sqrt{\sin^2\varphi - n^2}}{n^2\cos\varphi}. \tag{66.4}$$

Скачки фаз $\delta_\perp$ и $\delta_\parallel$ не одинаковы. Если падающая волна поляризована линейно, то в отраженной волне между $\mathbf{E}_\parallel$ и $\mathbf{E}_\perp$ возникнет определенная разность фаз. Поэтому *отраженная волна будет, вообще говоря, поляризована эллиптически*. Допустим, что между $\mathbf{E}_\parallel$ и $\mathbf{E}_\perp$ в падающей волне нет разности фаз. Тогда разность фаз между ними в отраженной волне будет $\delta = \delta_\parallel - \delta_\perp$. С помощью формул (66.3) и (66.4) найдем

$$\operatorname{tg}\frac{\delta}{2} = \frac{\cos\varphi\sqrt{\sin^2\varphi - n^2}}{\sin^2\varphi} \tag{66.5}$$

Отсюда видно, что $\operatorname{tg}(\delta/2) > 0$, т.е. δ лежит в пределах от нуля до π. Значит, колебание $\mathbf{E}_\parallel$ опережает по фазе колебание $\mathbf{E}_\perp$,

и эллиптическая поляризация отраженного света будет *левой*. Наоборот, если разность фаз между такими же колебаниями в падающей волне равна π, то эллиптическая поляризация отраженной волны будет *правой*.

Величина δ обращается в нуль при $\varphi = \varphi_0$ и $\varphi = \pi/2$. Между этими пределами она должна достигать максимума. Максимум достигается при $\cos\varphi = \sqrt{(1-n^2)/(1+n^2)}$, максимальная разность фаз δ_m определяется уравнением

$$\operatorname{tg}\frac{\delta_m}{2} = \frac{1-n^2}{2n}. \tag{66.6}$$

Для получения круговой поляризации отраженного света необходимо выполнение двух условий: 1) $\mathcal{E}_{\parallel} = \pm\mathcal{E}_{\perp}$; 2) $\delta = \pi/2$ (см. § 62, п. 4). Чтобы получить $\delta_m = \pi/2$, нужен показатель преломления, удовлетворяющий уравнению $1 = (1-n^2)/2n$. Отсюда $n = \sqrt{2} - 1 = 0{,}414$. Показатель преломления оптически более плотной среды относительно менее плотной будет $n' = 1/n = 2{,}41$. Если n' не меньше этого значения, то при однократном отражении можно получить разность фаз $\delta = 90°$. В этом случае при $\mathcal{E}_{\perp} = \mathcal{E}_{\parallel}$ поляризация отраженного света будет *круговая* и притом левая. В видимой области спектра такой большой показатель преломления имеет только алмаз ($n' = 2{,}42$). Для всех прочих сред $n' < 2{,}41$. Поэтому для видимого света получить разность фаз $\delta = 90°$ при однократном отражении невозможно. Но это возможно для коротких радиоволн. Для воды, например, $n' = \sqrt{\varepsilon_0} = 9$, $n = 1/9$; круговая поляризация, согласно (66.6), должна получиться при $\varphi = 6°29'$ и $\varphi = 44°38'$. Опыт подтвердил это предсказание.

Для стекла ($n' = 1{,}51$) максимальная разность фаз достигается при $\varphi = 51°20'$ и составляет $45°36'$. Разность фаз δ в точности равна $45°$ при $\varphi = 48°37'$ и при $\varphi = 54°37'$. Если применить двукратное отражение под одним из этих углов, то возникнет относительная разность фаз $90°$. Этим воспользовался Френель для получения круговой поляризации. Он изготовил стеклянный параллелепипед с углом $A = 54°37'$ (рис. 243). Луч SM, падающий перпендикулярно на грань AB, претерпевает двукратное полное отражение на гранях AD и BC, после чего выходит из параллелепипеда в направлении NQ. При каждом из отражений колебание $\mathbf{E}_{\parallel}$ опережает по фазе колебание $\mathbf{E}_{\perp}$ на $45°$. Если па-

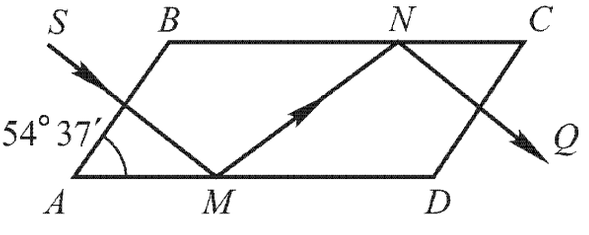


Рис. 243

дающий свет поляризован линейно и $\mathcal{E}_{||} = \mathcal{E}_{\perp}$, то выходящий свет будет поляризован по левому кругу. Если его заставить вторично пройти через параллелепипед Френеля, то он снова превратится в линейно поляризованный, но плоскость поляризации повернется на 90° относительно своего исходного положения.

4. Возникновение скачка фазы при полном отражении можно понять, не обращаясь к формулам Френеля. Допустим ради определенности, что электрический вектор лежит в плоскости падения. Если угол падения строго равен предельному углу полного отражения φ_0, то волна во второй среде будет еще однородной. Она должна распространяться параллельно границе раздела. Так как однородные волны поперечны, то отсюда следует, что электрический вектор во второй, а следовательно, в силу граничных условий, и в первой среде нормален к границе раздела. При том выборе положительного направления электрического поля, которого мы придерживаемся (рис. 244), это означает, что фазы падающей и отраженной волн на границе раздела совпадают.

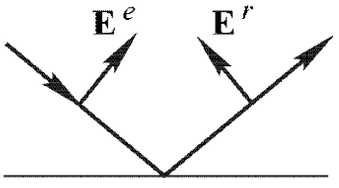


Рис. 244

Напротив, при $\varphi = 90°$ полное отражение сопровождается изменением фазы на 180°. В этом случае падающая и отраженная волны распространяются в одном и том же направлении — параллельно границе раздела. Граничные условия будут удовлетворены, если положить $R_{||} = -\mathcal{E}_{||}$. Падающая и отраженная волны гасят друг друга, в результате чего как в первой, так и во второй средах не будет никакого светового поля, и граничные условия выполняются тривиально. Но равенство $R_{||} = -\mathcal{E}_{||}$ означает, что при отражении появляется скачок фазы в 180°.

Следует заметить, что говорить об отражении света, когда угол падения точно равен 90°, не имеет смысла. Приведенное рассуждение показывает только, к какому пределу стремится амплитуда $R_{||}$, когда угол падения приближается к 90°. Если отражение полное, то в первой среде образуются стоячие волны с узловыми плоскостями, параллельными границе раздела. По мере приближения угла падения к 90° расстояние между узловыми плоскостями увеличивается и стремится к бесконечности. При этом ближайшая узловая плоскость в пределе сливается с границей раздела. Одновременно исчезает поле во второй (оптически менее плотной) среде.

Итак, при $\varphi = \varphi_0$ скачок фазы при отражении равен нулю, а при $\varphi = 90°$ он составляет 180°. При этом во всем интервале изменения угла φ от φ_0 до 90° отражение должно оставаться полным (§ 64). Изобразим колебание электрического поля в отраженной волне на векторной диаграмме. При $\varphi = \varphi_0$ ам-

плитуда изобразится вектором $\mathbf{A}(\varphi_0)$ (рис. 245), а само колебание — его проекцией на ось X (если заставить вектор $\mathbf{A}(\varphi_0)$ равномерно вращаться вокруг O с угловой скоростью ω). При $\varphi = \pi/2$ амплитуда изобразится вектором $\mathbf{A}(\pi/2)$, равным вектору $\mathbf{A}(\varphi_0)$, но противоположно направленным. Для промежуточного угла падения φ амплитуду, представляющую колебание, обозначим через $\mathbf{A}(\varphi)$.

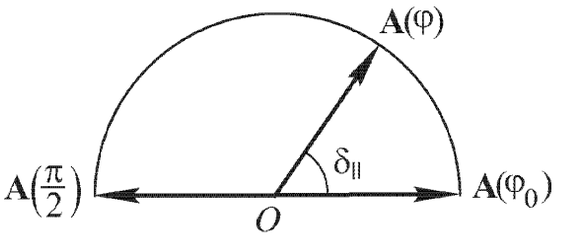


Рис. 245

При возрастании φ от φ_0 до $\pi/2$ длина вектора $\mathbf{A}(\varphi)$ не может изменяться, так как отражение полное. С другой стороны, вектор $\mathbf{A}(\varphi)$, изменяясь непрерывно, должен из исходного положения $\mathbf{A}(\varphi_0)$ перейти в конечное положение $\mathbf{A}(\pi/2)$. Это можно осуществить *только поворотом его вокруг точки* O. Значит, если угол падения φ лежит между φ_0 и 90°, то вектор $\mathbf{A}(\varphi)$ должен быть наклонен к оси X. Угол наклона $\delta_{\parallel}$ и есть скачок фазы при отражении.

5. Относительно перехода частичного отражения в полное сделаем следующее замечание. В области частичного отражения коэффициенты отражения $\rho_\perp$ и $\rho_{\parallel}$ меняются непрерывно и при $\varphi \to \varphi_0$ приближаются к единице. При $\varphi = \varphi_0$ они не испытывают разрыва. Однако их производные по φ претерпевают в этом месте разрыв. Вычислим производные по φ от коэффициентов Френеля $r_\perp = R_\perp/\mathcal{E}_\perp$ и $r_{\parallel} = R_{\parallel}/\mathcal{E}_{\parallel}$ в окрестности угла φ_0 в предположении, что $\varphi < \varphi_0$. Простое вычисление дает

$$\frac{dr_\perp}{d\varphi} = \frac{2}{\cos\psi}, \qquad \frac{dr_{\parallel}}{d\varphi} = \frac{2}{n^2\cos\psi}. \tag{66.7}$$

При $\varphi = \varphi_0$ $\cos\psi = 0$, и обе производные обращаются в бесконечность. Обращаются в бесконечность также $d\rho_\perp/d\varphi$ и $d\rho_{\parallel}/d\varphi$ ($\rho_\perp = r_\perp^2$, $\rho_{\parallel} = r_{\parallel}^2$). Поэтому вблизи предельного угла φ_0 коэффициенты частичного отражения $\rho_\perp$ и $\rho_{\parallel}$ должны очень резко меняться с изменением угла падения. Этим объясняется происхождение *резкой границы* между областями частичного и полного отражений, которая бывает видна в рефрактометрах, основанных на полном отражении света.

§ 67. Прохождение света через плоскопараллельную пластинку. Просветление оптики

1. Допустим, что на поверхность отражающего тела нанесена пленка толщины l с показателем преломления n (рис. 246).

Показатель преломления первой среды обозначим через n_1, а второй — через n_2. Пусть все показатели преломления n_1, n_2 и n постоянны (т.е. среды однородны), а падающая волна поляризована либо в плоскости падения, либо перпендикулярно к ней. Падающая волна

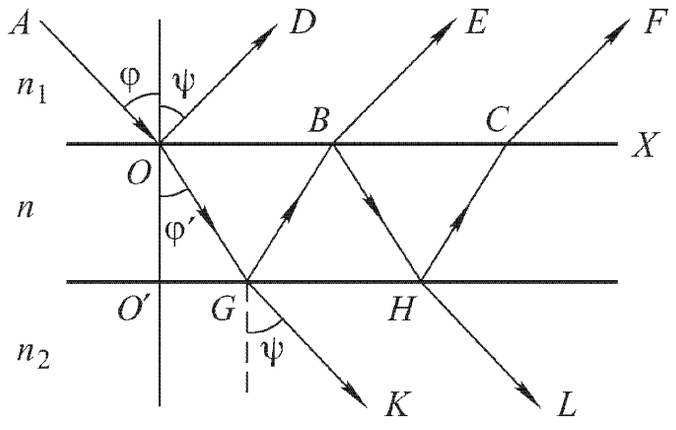


Рис. 246

$$E^{(e)} = \mathcal{E}e^{i(\omega t - \mathbf{k}_1\mathbf{r})} \quad (67.1)$$

(начало координат в O) претерпевает многократные отражения на границах пленки. В результате интерференции в первой среде возникает отраженная волна

$$E^{(r)} = Re^{i(\omega t - \mathbf{k}_1'\mathbf{r})}. \quad (67.2)$$

Поле внутри пленки будет состоять из двух плоских волн:

$$E = Ae^{i(\omega t - \mathbf{k}\mathbf{r})}, \quad E' = A'e^{i(\omega t - \mathbf{k}'\mathbf{r})}. \quad (67.3)$$

Наконец, во второй среде возникнет преломленная волна

$$E^{(d)} = De^{i(\omega t - \mathbf{k}_2\mathbf{r})}. \quad (67.4)$$

Ввиду граничных условий, тангенциальные компоненты всех волновых векторов должны быть одинаковы. Этим однозначно определяются направления всех волн (см. § 64). Для определения неизвестных амплитуд R, A, A', D электромагнитная теория дает четыре независимых граничных условия (два на верхней поверхности пленки и два на нижней), из которых эти неизвестные могут быть найдены.

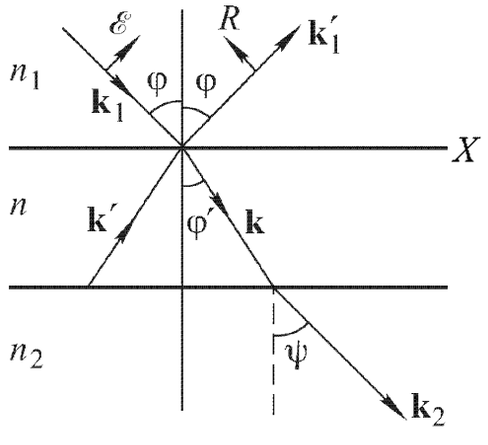


Рис. 247

Однако мы предпочитаем решить задачу более коротким способом. На верхней границе сходятся четыре волны (рис. 247), из которых две приходящие, а две другие — уходящие. При написании соотношений между их комплексными амплитудами *можно отвлечься от наличия нижней границы*. Тогда формально можно рассуждать так, *как если бы многократных отражений не было*. Все приходящие волны можно рассматривать как падающие, в результате (однократных) отражений и преломлений которых возникают уходящие волны. Падающая волна с амплитудой $\mathcal{E}$,

отражаясь от верхней границы, дает волну, идущую в направлении вектора $\mathbf{k}_1'$. Волна с амплитудой A', преломляясь на той же границе, дает волну, идущую в том же направлении. Результатом интерференции этих волн и является отраженная волна с амплитудой R. Это дает

$$R = r_1\mathcal{E} + d_1'A'. \tag{67.5}$$

Аналогично,

$$A = d_1\mathcal{E} + r_1'A'. \tag{67.6}$$

Здесь r_1 и d_1 означают коэффициенты Френеля на верхней границе пленки, а r_2 и d_2 — такие же коэффициенты на ее нижней границе (предполагается, что свет пересекает эти границы сверху вниз). Коэффициенты Френеля при обратном ходе лучей обозначаются теми же буквами, но со штрихами. С помощью формул Френеля нетрудно проверить, что эти величины удовлетворяют соотношениям

$$r' = -r, \quad r^2 + dd' = 1. \tag{67.7}$$

Такое же рассуждение можно провести и для нижней границы пленки. Надо только перенести начало координат в точку O', находящуюся на нижней поверхности пленки (см. рис. 246). Тогда волна (67.1) запишется в виде

$$E^{(e)} = \mathcal{E}e^{-ik_{1z}l}e^{i(\omega t - \mathbf{k}_1\mathbf{r}')},$$

откуда видно, что роль комплексной амплитуды будет играть величина $\mathcal{E}' = \mathcal{E}e^{-ik_{1z}l}$. Аналогично преобразуются комплексные амплитуды и остальных волн. В результате получим

$$A'E^{-ik_z'l} = r_2Ae^{-ik_zl}, \quad De^{-ik_{2z}l} = d_2Ae^{-ik_zl}. \tag{67.8}$$

Из полученных уравнений найдем

$$\begin{aligned} \frac{R}{\mathcal{E}} &= \frac{r_1 + r_2\exp\left(-2ik_0nl\cos\varphi\right)}{1 + r_1r_2\exp\left(-2ik_0nl\cos\varphi'\right)}, \\ \frac{D}{\mathcal{E}} &= \frac{d_1d_2\exp\left[ik_0l(n_2\cos\psi - n\cos\varphi)\right]}{1 + r_1r_2\exp\left(-2ik_0nl\cos\varphi'\right)}, \end{aligned} \tag{67.9}$$

где k_0 — волновое число в вакууме, а φ' и ψ — углы преломления на первой и второй границах. Для наглядности при выводе мы ссылались на чертеж, относящийся к однородным волнам. Но, очевидно, результаты (67.9) справедливы и в том случае, когда волны в пленке или во второй среде неоднородны.

2. При нормальном падении первая формула (67.9) переходит в

$$\frac{R}{\mathcal{E}} = \frac{r_1 + r_2e^{-2ik_0nl}}{1 + r_1r_2e^{-2ik_0nl}}. \tag{67.10}$$

Это выражение обращается в нуль, если выполняются два условия:

$$r_1 + r_2 \cos(2nk_0 l) = 0, \quad \sin(2nk_0 l) = 0.$$

Из второго условия получаем: $2nk_0 l = N\pi$, или

$$l = N\frac{\lambda_0}{4n} = N\frac{\lambda}{4}, \tag{67.11}$$

где N — целое число, а λ — длина волны в пленке. Теперь первое условие можно записать в виде: $r_1 + (-1)^N r_2 = 0$. Если N — нечетное, то $r_1 = r_2$; если же N четное, то $r_1 = -r_2$. Однако последнее равенство не может быть выполнено. Действительно,

$$r_1 = \frac{n - n_1}{n + n_1}, \quad r_2 = \frac{n_2 - n}{n_2 + n}.$$

Если бы $r_1 = -r_2$, то мы получили бы $n(n_2 - n_1) = 0$, откуда либо $n = 0$, что невозможно, либо $n_2 = n_1$, что не представляет интереса. Это показывает, что число N должно быть нечетным, а следовательно, $r_1 = r_2$. Отсюда

$$n = \sqrt{n_1 n_2}. \tag{67.12}$$

Таким образом, если $n = \sqrt{n_1 n_2}$, а толщина пленки равна $N\lambda/4$, где N — нечетное число, то отражательная способность обращается в нуль.

На этом результате основан один из методов увеличения поверхностной прозрачности стекол, применяемый в оптической промышленности (так называемое *просветление оптики*). Для стекла ($n = 1{,}5$) отражательная способность равна $R = (n-1)^2/(n+1)^2 = 0{,}04 = 4\,\%$, т.е. совсем невелика. Однако оптические приборы состоят из многих деталей, изготовленных из стекла. Отражение на границах их соприкосновения является главной причиной ослабления света при его прохождении через оптический прибор. Так, например, потери света в призменном бинокле составляют свыше 50 %, причем они почти целиком происходят за счет отражения света. Значительная доля отраженного света, благодаря последующим отражениям, доходит до глаза наблюдателя и, будучи в лучшем случае равномерно рассеянной, дает освещенный фон, ослабляющий контраст света и тени в изображении. Особенно вреден этот рассеянный свет в фотографических приборах. В лучшем случае он создает общую вуаль на эмульсии. При некоторых же неблагоприятных расположениях источников света могут получиться блики и дополнительные изображения.

Для увеличения поверхностной прозрачности стекла на его поверхности создается пленка с показателем преломления $n = \sqrt{n_1 n_2}$ и толщиной $\lambda/4$. Применение более толстых пленок (со-

ответствующих $N = 3, 5$ и т.д.) нецелесообразно, так как условие полного исчезновения отражения может быть точно выполнено только для одной определенной длины волны и одного определенного угла падения. При $N = 1$ возрастание коэффициента отражения с изменением длины волны, а также с изменением угла падения получается наиболее медленным. Поэтому в этом случае можно добиться почти полного исчезновения отражения для сравнительно большого участка спектра и сравнительно широкого интервала углов падения.

Просветляющие пленки создаются либо путем выщелачивания из поверхности стекла его компонентов, либо, что лучше, путем напыления на поверхность стекла слоя посторонних веществ. При этом возникают трудности, связанные с тем, что показатель преломления пленки должен быть значительно меньше показателя преломления стекла. Например, если $n_2 = 1{,}52$, $n_1 = 1$, то $n = 1{,}23$. Твердых веществ с таким малым показателем преломления в природе не встречается. Поэтому пленку приходится делать пористой, причем, во избежание заметного рассеяния света, размеры пор должны быть весьма малы по сравнению с длиной волны. Но пористая пленка не обладает достаточной механической прочностью. Трудности могут быть преодолены путем применения двуслойных покрытий. Сначала просветляемая поверхность покрывается пленкой, показатель преломления которой значительно превосходит показатель преломления стекла, а затем пленкой с меньшим показателем преломления.

3. Условия отсутствия отражения (67.11) и (67.12) могут быть также получены из простых интерференционных соображений.

При малых коэффициентах отражения можно пренебречь волнами, претерпевшими многократные отражения на границах пленки. Тогда останутся две отраженные волны, из которых одна отразилась от передней, а другая от задней поверхности пленки. Чтобы эти волны гасили друг друга, должны соблюдаться два условия: 1) фазы их должны быть противоположны; 2) интенсивности их должны быть равны. Первое условие при нормальном падении приводит к соотношению (67.11). Второе условие, если пренебречь ослаблением волны за счет отражения от передней поверхности пленки, сводится к $r_1 = r_2$. Отсюда, как показано выше, получается формула (67.12). Недостаток приведенного доказательства в том, что оно не учитывает многократные отражения.

ЗАДАЧИ

1. Вывести формулы для отражения и преломления электромагнитных волн в предположении, что магнитная проницаемость отлична от единицы.

О т в е т.

$$\frac{R_\perp}{\mathcal{E}_\perp}=\frac{\sqrt{\varepsilon_1/\mu_1}\cos\varphi-\sqrt{\varepsilon_2/\mu_2}\cos\psi}{\sqrt{\varepsilon_1/\mu_1}\cos\varphi+\sqrt{\varepsilon_2/\mu_2}\cos\psi},\quad \frac{D_\perp}{\mathcal{E}_\perp}=\frac{2\sqrt{\varepsilon_1/\mu_1}\cos\varphi}{\sqrt{\varepsilon_1/\mu_1}\cos\varphi+\sqrt{\varepsilon_2/\mu_2}\cos\psi},$$

$$\frac{R_\parallel}{\mathcal{E}_\parallel}=\frac{\sqrt{\varepsilon_2/\mu_2}\cos\varphi-\sqrt{\varepsilon_1/\mu_1}\cos\psi}{\sqrt{\varepsilon_2/\mu_2}\cos\varphi+\sqrt{\varepsilon_1/\mu_1}\cos\psi},\quad \frac{D_\parallel}{\mathcal{E}_\parallel}=\frac{2\sqrt{\varepsilon_1/\mu_1}\cos\varphi}{\sqrt{\varepsilon_2/\mu_2}\cos\varphi+\sqrt{\varepsilon_1/\mu_1}\cos\psi}. \tag{67.13}$$

2. Какой вид принимают формулы (67.13), если $\varepsilon_1=\varepsilon_2=1$?

О т в е т.

$$\frac{R_\perp}{\mathcal{E}_\perp}=\frac{n_2\cos\varphi-n_1\cos\psi}{n_2\cos\varphi+n_1\cos\psi},\quad \frac{D_\perp}{\mathcal{E}_\perp}=\frac{2n_2\cos\varphi}{n_2\cos\varphi+n_1\cos\psi},$$

$$\frac{R_\parallel}{\mathcal{E}_\parallel}=\frac{n_1\cos\varphi-n_2\cos\psi}{n_1\cos\varphi+n_2\cos\psi},\quad \frac{D_\parallel}{\mathcal{E}_\parallel}=\frac{2n_2\cos\varphi}{n_1\cos\varphi+n_2\cos\psi}. \tag{67.14}$$

Формулы для отражения в точности совпадают с формулами Френеля, если только поменять местами $\perp$ и $\parallel$-составляющие. Этот факт был отмечен Гельмгольцем.

3. Показать, что отражательная способность среды для радиоволн обращается в нуль, если $\varepsilon=\mu$.

4. Останется ли справедливым закон Брюстера для радиоволн, если магнитные проницаемости сред μ_1 и μ_2 отличны от единицы?

О т в е т. Если закон Брюстера имеет место, то угол Брюстера, при котором не отражается составляющая $E_\parallel$ электрического поля, определяется выражением

$$\operatorname{tg}\varphi_B=\sqrt{\frac{\varepsilon_2}{\varepsilon_1}\frac{\varepsilon_2\mu_1-\varepsilon_1\mu_2}{\varepsilon_2\mu_2-\varepsilon_1\mu_1}}. \tag{67.15}$$

Возможен случай, когда не будет отражаться составляющая $E_\perp$. Угол, при котором это имеет место, находится из уравнения

$$\operatorname{tg}\varphi'_B=\sqrt{\frac{\mu_2}{\mu_1}\frac{\varepsilon_2\mu_1-\varepsilon_1\mu_2}{\varepsilon_1\mu_1-\varepsilon_2\mu_2}}. \tag{67.16}$$

Оба случая взаимно исключают друг друга, так как знаки подкоренных выражений в (67.15) и (67.16) противоположны. (Предполагается, что ε и μ существенно положительны.) Таким образом, всегда существует угол, при котором не отражается либо составляющая $E_\parallel$, либо составляющая $E_\perp$.

5. Проверить, что коэффициенты Френеля удовлетворяют соотношениям (67.7).

6. Пользуясь формулами Френеля, показать, что линейно поляризованный свет остается линейно поляризованным после отражения на границе раздела двух прозрачных изотропных сред во всех случаях, за исключением случаев полного отражения.

7. Угол между плоскостью колебаний поляризованного света и плоскостью падения называется *азимутом колебания*. Найти азимут преломленной волны γ и азимут отраженной волны β, если азимут падающей волны α, а угол падения φ.

О т в е т. $\operatorname{tg}\gamma=\cos(\varphi-\psi)\operatorname{tg}\alpha$, $\operatorname{tg}\beta=-\dfrac{\cos(\varphi-\psi)}{\cos(\varphi+\psi)}\operatorname{tg}\alpha$.

8. 1) Найти угол Брюстера для света, отраженного от стекла с показателем преломления $n=1{,}5$; 2) найти для этого угла *степень поляризации*

преломленного света, т.е. величину $\Delta = \dfrac{I_\perp - I_\parallel}{I_\perp + I_\parallel}$, где $I_\parallel$ и $I_\perp$ — интенсивности отраженных волн, поляризованных соответственно в плоскости падения и перпендикулярно к ней. Падающий свет — естественный.

О т в е т. $\varphi_B = 56°19'$, $\Delta = \dfrac{4n^2 - (n^2+1)^2}{4n^2 + (n^2+1)^2} = -0{,}08$.

9. Естественный свет падает под углом Брюстера из воздуха на поверхность стекла с показателем преломления $n = 1{,}5$. Найти интенсивность I_r отраженного света, зная интенсивность падающего света I_e.

О т в е т. $\dfrac{I_r}{I_e} = \dfrac{1}{2}\left(\dfrac{n^2-1}{n^2+1}\right)^2 = 0{,}74$.

10. Естественный свет падает под углом Брюстера из воздуха на поверхность диэлектрика с показателем преломления n. Найти амплитуду $D_\parallel$ преломленной волны, если амплитуда падающей волны той же поляризации равна $\mathcal{E}_\parallel$.

О т в е т. $D_\parallel = \mathcal{E}_\parallel / n$.

11. На боковую грань призмы, изготовленной из стекла с показателем преломления $n = 1{,}5$, падает под углом Брюстера φ_B световой пучок, электрический вектор которого лежит в плоскости падения. Каким должен быть преломляющий угол A призмы, чтобы свет прошел через нее, не испытав потерь на отражение?

О т в е т. $A = \pi - 2\varphi_B = 68°$.

12. При каких условиях луч света, падающий на боковую грань прозрачной изотропной призмы с преломляющим углом $A = 60°$, проходит через нее без потерь на отражение?

О т в е т. Электрический вектор должен лежать в плоскости падения. Показатель преломления призмы должен быть равен

$$n = \operatorname{tg}\frac{A}{2} = \sqrt{3} = 1{,}73.$$

13. Свет падает из среды 1 на среду 2 под углом φ и преломляется под углом ψ. Доказать, что коэффициент отражения не изменится, если свет будет падать из среды 2 на среду 1 под углом ψ.

14. Пользуясь формулами Френеля, показать, что при отражении плоской электромагнитной волны от идеального зеркала, покрытого сверху слоем прозрачного диэлектрика, амплитуда отраженной волны равна амплитуде падающей. Изменяется лишь фаза волны, как это и должно быть согласно закону сохранения энергии.

Р е ш е н и е. При отражении от идеального зеркала может меняться только фаза, но не амплитуда волны. Поэтому коэффициент Френеля r_2 в формуле (67.10) должен иметь вид $r_2 = e^{i\alpha}$. Вводя еще обозначение $\delta = 2k_0 nl$, получим из той же формулы

$$\frac{R}{\mathcal{E}} = \frac{r_1 + e^{i(\alpha-\delta)}}{1 + r_1 e^{i(\alpha-\delta)}}.$$

Отсюда $|R| = |\mathcal{E}|$.

15. Свет, частично отражаясь, проходит через две параллельные полупрозрачные плоскости. Коэффициенты отражения и пропускания первой из них равны ρ_1 и b_1, а второй ρ_2 и b_2 соответственно. Степень монохроматичности падающего света невелика, так что интерференции не возникает, а имеет место сложение интенсивностей света. Найти коэффициенты отражения ρ и пропускания b для системы обеих плоскостей.

Р е ш е н и е.

$$\rho = \rho_1 + \rho_2 b_1^2 + \rho_2 b_1^2 \rho_1 \rho_2 + \ldots = \rho_1 + \frac{\rho_2 b_1^2}{1 - \rho_1 \rho_2},$$

$$b = b_1 b_2 + b_1 b_2 \rho_1 \rho_2 + b_1 b_2 (\rho_1 \rho_2)^2 + \ldots = \frac{b_1 b_2}{1 - \rho_1 \rho_2}.$$

Используя соотношения $\rho_1 + b_1 = 1$, $\rho_2 + b_2 = 1$, полученные результаты можно записать в виде

$$\rho = \frac{\rho_1 + \rho_2 - 2\rho_1 \rho_2}{1 - \rho_1 \rho_2}, \quad b = \frac{(1 - \rho_1)(1 - \rho_2)}{1 - \rho_1 \rho_2}. \tag{67.17}$$

При $\rho_1 = \rho_2$ отсюда получаем

$$\rho = \frac{2\rho_1}{1 + \rho_1}, \quad b = \frac{1 - \rho_1}{1 + \rho_1}. \tag{67.18}$$

16. Свет, частично отражаясь, проходит через систему m параллельных полупрозрачных плоскостей. Коэффициенты отражения и пропускания каждой из них равны ρ и b. Найти коэффициент отражения ρ_m и коэффициент пропускания b_m всей системы m плоскостей. (Относительно падающего света см. предыдущую задачу.)

Р е ш е н и е. Присоединим к системе m плоскостей одну такую же $(m+1)$-ю плоскость. Первые m плоскостей можно заменить одной плоскостью с коэффициентом отражения ρ_m. Тогда задача будет сведена к предыдущей, и мы получим для коэффициентов отражения и пропускания $(m+1)$ плоскостей

$$\rho_{m+1} = \frac{\rho_m + \rho - 2\rho_m \rho}{1 - \rho_m \rho},$$

$$b_{m+1} = \frac{(1 - \rho_m)(1 - \rho)}{1 - \rho_m \rho}.$$

Отсюда методом доказательства от m к $m+1$ нетрудно получить

$$\rho_m = \frac{m\rho}{1 + (m-1)\rho}, \quad b_m = \frac{1 - \rho}{1 + (m-1)\rho}. \tag{67.19}$$

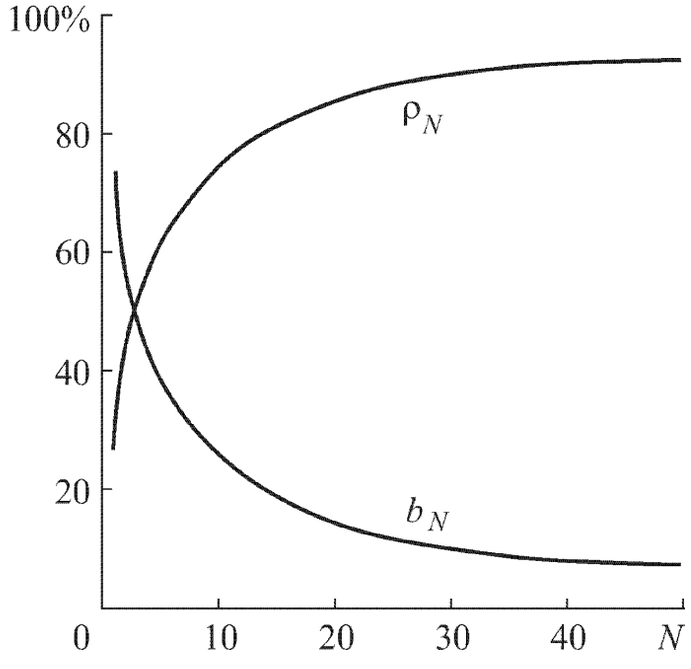


Рис. 248

17. Стопа Столетова (1839–1896) состоит из плоскопараллельных стеклянных пластинок с показателем преломления $n = 1{,}5$. На нее под углом Брюстера падает свет, поляризованный перпендикулярно к плоскости падения. Найти выражения и начертить график для коэффициентов отражения и пропускания стопы в зависимости от числа N пластинок.

О т в е т. $\rho_N = \dfrac{2N}{2N + 5{,}76}$, $b_N = \dfrac{5{,}76}{2N + 5{,}76}$ (рис. 248).

18. Естественный свет падает под углом φ на стопу Столетова, состоящую из N стеклянных пластинок с показателем преломления n. Найти степени поляризации Δ и Δ' прошедшего и отраженного света, а также отношение интенсивностей $I_\perp/I_\parallel$ главных компонент прошедшего света.

О т в е т.

$$\Delta = \frac{N(\rho_\perp - \rho_\parallel)}{(2N-1)\rho_\perp\rho_\parallel - (N-1)(\rho_\perp + \rho_\parallel) - 1},$$

$$\Delta' = \frac{\rho_\perp - \rho_\parallel}{2(N-1)\rho_\perp\rho_\parallel + (\rho_\perp + \rho_\parallel)}, \quad \frac{I_\perp}{I_\parallel} = \frac{1-\rho_\perp}{1-\rho_\parallel}\,\frac{(2N-1)\rho_\parallel + 1}{(2N-1)\rho_\perp + 1},$$

где $\rho_\parallel$ и $\rho_\perp$ — коэффициенты отражения волн, поляризованных в плоскости падения и перпендикулярно к ней, для одной отражающей поверхности:

$$\rho_\perp = \left(\frac{\cos\varphi - n\cos\psi}{\cos\varphi + n\cos\psi}\right)^2, \quad \rho_\parallel = \left(\frac{n\cos\varphi - \cos\psi}{n\cos\varphi + \cos\psi}\right)^2.$$

19. Падающий свет поляризован линейно с азимутом колебаний, равным $+45°$ [1]). Можно ли путем однократного отражения превратить его в свет, поляризованный по правому кругу?

О т в е т. Нельзя.

20. Какой должен быть минимальный показатель преломления параллелепипеда Френеля, чтобы при азимуте колебаний падающего света в $+45°$ выходящий свет был поляризован по правому кругу?

О т в е т. $n = \dfrac{1+\sin(3\pi/8)}{\cos(3\pi/8)} = 5{,}028$. В оптическом диапазоне спектра этот случай осуществить нельзя. Его можно было бы осуществить с более длинными электромагнитными волнами.

21. Линейно поляризованная электромагнитная волна с азимутом колебаний, равным $+135°$, отражается на границе вода–воздух. Диэлектрическая проницаемость воды $\varepsilon = 81$. Под каким углом должна падать эта волна, чтобы отраженная волна получилась поляризованной по кругу? Какая при этом будет поляризация: правая или левая?

О т в е т. $6°29'$ или $44°38'$. Правая.

22. Линейно поляризованный луч с азимутом колебаний $+135°$ падает перпендикулярно на грань AB стеклянной призмы $ABCD$ (рис. 249) и, испытав три раза полное отражение, выходит из нее. Каков должен быть преломляющий угол A призмы, чтобы вышедший свет был поляризован по кругу, если показатель преломления стекла призмы равен 1,52? Какая получится поляризация вышедшего света: правая или левая?

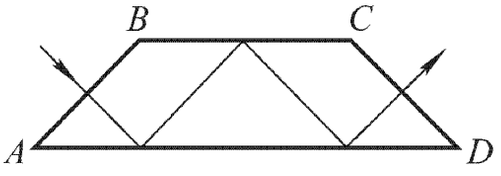


Рис. 249

О т в е т. $69°21'$ или $42°46'$. Правая.

23. Каким должен быть минимальный показатель преломления призмы, описанной в предыдущей задаче, чтобы при азимуте колебаний падающего света, равном $+45°$, выходящий свет был поляризован по правому кругу? Какой при этом должен быть угол A?

---

[1]) Азимут колебаний падающей волны может изменяться от $-\pi/2$ до $+\pi/2$. Он считается положительным, если $\mathcal{E}_\parallel/\mathcal{E}_\perp > 0$, и отрицательным, если $\mathcal{E}_\parallel/\mathcal{E}_\perp < 0$.

О т в е т. $n = 1/(\sqrt{2}-1) = 2{,}4143$; $A = 35°34'$.

24. С помощью векторной диаграммы показать, что скачок фазы при полном отражении превосходит вдвое скачок фазы, испытываемый преломленной (поверхностной) волной.

§ 68. Распространение света в среде с точки зрения молекулярной оптики

1. С точки зрения атомистических представлений всякую среду следует рассматривать как вакуум, в который вкраплены атомы вещества. (Молекулы можно рассматривать тоже как атомы.) Под действием падающей волны, а также излучений соседних атомов внутри каждого атома возбуждаются колебания электронов и атомных ядер. Вследствие этого атомы становятся источниками вторичных сферических волн, распространяющихся между этими частицами *со скоростью света в вакууме* c. Эти волны когерентны, поскольку они возбуждаются одной и той же падающей волной. Их интерференция между собой и с падающей волной определяет волновое поле во всем пространстве. В частности, отраженная волна возникает в результате интерференции вторичных волн, вышедших из среды в вакуум, с которым она граничит.

Почему же в среде свет распространяется с иной скоростью, чем в вакууме? Вопрос этот надо уточнить, указав, о какой скорости идет речь. В теории отражения и преломления света основной интерес представляет *фазовая скорость*, поскольку она определяет показатель преломления среды, а следовательно, и законы отражения и преломления волн на границе раздела сред. Отличие фазовой скорости света в среде от скорости света в вакууме вкратце объясняется тем, что в каждую точку пространства вторичные волны приходят не только от атомов, расположенных вдоль луча, проходящего через рассматриваемую точку, но и от множества других атомов, расположенных в стороне от него.

Более подробное рассмотрение приводится ниже для точечных атомов. В поле световой волны атомы приобретают меняющиеся во времени дипольные моменты и излучают как точечные диполи Герца. Для наших целей достаточно знать поле излучения такого диполя *в волновой зоне*. Оно определяется только составляющей дипольного момента $\mathbf{p}_\perp$, перпендикулярной к направлению излучения. Параллельная составляющая $\mathbf{p}_\parallel$ в волновой зоне на излучение диполя Герца не влияет.

2. Допустим, что в вакууме вдоль оси X распространяется плоская монохроматическая волна $\mathbf{E}_0 e^{i(\omega t - k_0 x)}$, на пути которой перпендикулярно к оси X поставлен бесконечно тонкий плоско-

параллельный слой толщины $d\xi$, состоящий из точечных неподвижных атомов, равномерно распределенных по объему слоя (рис. 250). Следуя Рэлею, выясним влияние такого слоя на фазу колебаний в какой-то удаленной точке $A(x)$, расположенной впереди слоя. Дипольные моменты атомов слоя, возбужденные падающей волной, можно представить в виде $\mathbf{p} = \mathbf{p}_0 e^{i(\omega t - k_0 \xi)}$, где ξ — абсцисса слоя. Предположим, что точка A расположена в волновой зоне ближайшего, а следовательно, и всех остальных диполей слоя. Тогда электрическое поле каждого диполя в точке A будет

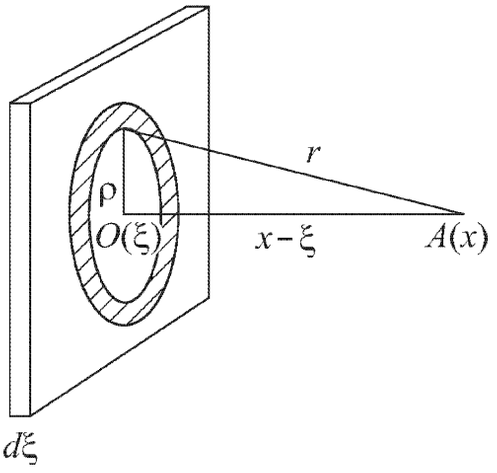


Рис. 250

$$\frac{k_0^2}{\mathbf{r}} \mathbf{p}_{0\perp} e^{i(\omega t - k_0 \xi - k_0 r)}, \qquad (68.1)$$

где r — расстояние от диполя (см. т. III, § 141). Такие выражения надо просуммировать по всем диполям слоя. Применим для этого метод кольцевых зон Френеля. Из теории таких зон известно, что результирующая напряженность $d\mathbf{E}_1$ всех диполей слоя в точке A будет равна половине напряженности поля, возбуждаемого в той же точке диполями одной только центральной зоны. Таким образом, для нахождения $d\mathbf{E}_1$ надо просуммировать выражение (68.1) по всем диполям центральной зоны и результат разделить на два. Вторичные волны, исходящие от края центральной зоны, отстают по фазе на π от волны, исходящей из ее центра O, а следовательно, и от падающей волны. Отставание по фазе вторичных волн, исходящих от остальных точек той же зоны, будет промежуточным. Таким образом, возникнет замедление скорости распространения фазы волны в результате прохождения ее через слой вещества.

Для фактического выполнения расчета заменим суммирование интегрированием. Возьмем кольцо с внутренним радиусом ρ и наружным $\rho + d\rho$, заштрихованное на рис. 250. В элементарном объеме $dV = 2\pi\rho\, d\rho\, d\xi$ находится $N\, dV$ диполей (N — число диполей в единице объема). Для возможности аппроксимации сумм интегралами и применимости метода зон Френеля предположим, что число $N\, dV$ еще достаточно велико. На это число надо умножить выражение (68.1), проинтегрировать по центральной зоне и результат разделить на два. Из соотношения $\rho^2 = r^2 - (x - \xi)^2$ при постоянном ξ получаем $\rho\, d\rho = r\, dr$ и вводим в качестве переменной интегрирования расстояние r. В пределах центральной зоны величину $\mathbf{p}_{0\perp}$ можно считать по-

стоянной и равной $\mathbf{p}_0$. Тогда интегрирование сведется к

$$\int\limits_{x-\xi}^{x-\xi+\lambda/2} e^{-ik_0 r}\,dr = -\frac{2}{k_0}i,$$

а после введения коэффициента перед интегралом получится

$$d\mathbf{E}_1 = -i\cdot 2\pi k_0 N\mathbf{p}_0\,d\xi\,e^{i(\omega t-k_0 x)}.$$

Интегрированием по остальным зонам убеждаемся, что из-за убывания $\mathbf{p}_{0\perp}$ их действия медленно убывают с возрастанием номера зоны n и при $n\to\infty$ стремятся к нулю. Это может служить обоснованием применимости метода зон Френеля к рассматриваемому случаю.

Добавив $d\mathbf{E}_1$ к падающей волне, найдем полное поле в точке A:

$$\mathbf{E} = (\mathbf{E}_0 - i\cdot 2\pi k_0 N\mathbf{p}_0\,d\xi)e^{i(\omega t-k_0 x)} = \mathbf{E}_0 e^{i(\omega t-k_0 x-d\Phi)},$$

где введено обозначение

$$d\Phi = \frac{2\pi k_0 N p_0}{E_0}\,d\xi.$$

Таким образом, наличие слоя вносит дополнительное отставание по фазе $d\Phi$. Если толщина слоя l конечна, то отставание по фазе будет равно

$$\Phi = \frac{2\pi k_0 N p_0}{E_0} l. \tag{68.2}$$

В этой формуле содержится *принципиальное объяснение* замедления фазовой скорости световой волны при ее распространении в веществе.

Для завершения расчета надо было бы найти связь между амплитудами $\mathbf{p}_0$ и $\mathbf{E}_0$. В общем случае это весьма сложная задача, так как дипольный момент атома $\mathbf{p}$ определяется не средним макроскопическим полем $\mathbf{E}$, а *микроскопическим полем, действующим на атомы среды*. Только для не очень плотных газов (когда $n-1\ll 1$) оба поля практически совпадают. Тогда $\mathbf{p}_0 = \beta\mathbf{E}_0$, где β — поляризуемость атома, связанная с диэлектрической проницаемостью ε и показателем преломления n соотношением

$$\varepsilon = n^2 = 1 + 4\pi N\beta = 1 + 4\pi N p_0/E_0.$$

Используя это соотношение, из (68.2) найдем

$$\Phi = \frac{n^2-1}{2}\,k_0 l \approx (n-1)k_0 l,$$

что совпадает с результатом феноменологического рассмотрения.

3. В феноменологической теории показатель преломления вводится с помощью *макроскопических уравнений Максвелла.* Последние предполагают, что в каждом элементарном объеме, линейные размеры которого малы по сравнению с длиной волны, содержится еще *очень много атомов.* Молекулярное рассмотрение, приведенное выше, показывает, что *это условие не обязательно.* Показатель преломления можно определить *через сдвиг фазы*, который вносит вещество, стоящее на пути световой волны. Такой сдвиг был вычислен выше в предположении, что велико число атомов во всяком элементе объема порядка $dV = 2\pi\rho\, d\rho\, d\xi$. А этому условию можно удовлетворить для сколь угодно разреженной среды, если только точку наблюдения A отодвинуть от слоя $d\xi$ достаточно далеко. Так, можно говорить о *показателе преломления рентгеновских лучей*, хотя макроскопические уравнения Максвелла на них не распространяются. Не лишено смысла говорить о *показателе преломления межпланетного и межзвездного пространства*, хотя плотность вещества в нем и ничтожна (не превышает примерно одного атома в кубическом сантиметре).

4. Если бы точку наблюдения A поместить перед слоем, то наше вычисление привело бы к волне, распространяющейся в противоположном направлении, т.е. к *отраженной волне.* Если средние межатомные расстояния меньше длины волны и атомы распределены в пространстве равномерно, то не возникнет никаких волн помимо прошедшей прямо и отраженной.

Не так будет, когда межатомные расстояния *больше длины волны.* Если атомы в среде распределены регулярно, например находятся в узлах кристаллической решетки, то вторичные волны, излучаемые атомами, когерентны, и будут складываться напряженности волновых полей. Условия интерференционного усиления вторичных волн могут выполняться не только в направлениях падающего и отраженного света, но и для некоторых других направлений. Возникнет дискретный ряд плоских волн, распространяющихся в различных направлениях (*интерференционное рассеяние*). Такой случай реализуется при дифракции коротких рентгеновских волн на кристаллической решетке. Если же атомы среды распределены в пространстве хаотически, то вторичные волны при рассмотрении бокового рассеяния ведут себя как некогерентные: складываются их интенсивности.

5. До сих пор не учитывалось *тепловое движение атомов.* Объясним теперь, как при наличии такового в среде может распространяться регулярная волна и как может возникнуть правильное отражение от зеркальных поверхностей твердых и жидких тел.

Рассмотрим сначала газы. Между столкновениями атомы газа движутся прямолинейно и равномерно. Из-за эффекта До-

плера атомы, движущиеся с различными скоростями, излучают свет с различными частотами. Казалось бы, что никакой интерференции при таких условиях возникнуть не может. На самом деле изменение частоты не имеет места, когда речь идет о вторичных волнах, идущих *в направлении распространения света.* Действительно, пусть в газе распространяется плоская монохроматическая волна с частотой ω. Речь идет о частоте в системе отсчета S, в которой газ покоится. Рассмотрим какой-либо движущийся атом. Перейдем в систему отсчета S', в которой атом неподвижен. В системе S' частота распространяющейся плоской волны изменится и будет равна, скажем, ω'. С той же частотой в системе S' возбудятся колебания атома и будут излучаться вторичные сферические волны. При обратном переходе в систему S частота ω' излучаемой сферической волны изменится и будет зависеть от направления излучения. Только для излучения, идущего в направлении первичной волны, получится прежняя частота ω, независимо от того, с какой скоростью и в каком направлении двигался атом.

Таким образом, в направлении распространения первичной волны все атомы будут излучать волны *с одной и той же частотой* ω. С этим и связана возможность регулярного распространения света в газах. Во всех других направлениях движущиеся атомы будут посылать волны с различными частотами. Например, если атом движется в направлении света, то в обратном направлении он будет излучать волны с меньшей частотой. Если же он движется навстречу свету, то частота излучаемой волны в направлении его движения увеличится.

В твердых и жидких телах тепловое движение носит иной характер. В этих случаях атомы движутся *ускоренно*, и рассуждение с переходом к движущейся системе отсчета здесь неприменимо. Атомы совершают колебания около положений равновесия и тем самым *модулируют* поле световой волны. В результате не только сохраняются вторичные волны с прежней частотой, но возникают и волны с новыми частотами. К излучениям с прежними частотами применимо все сказанное выше. С ними связана возможность регулярного распространения световых волн в твердых и жидких средах, а также правильного отражения и преломления их на зеркальных поверхностях тел. Излучения же с изменившимися частотами приводят к появлению в рассеянном свете новых частот.

6. Закончим этот параграф замечанием, которое понадобится нам при выводе формул Френеля с атомистической точки зрения. Если среда однородна и неограниченна, то в ней могут распространяться дипольные колебания в виде бегущей волны

$$\mathbf{p} = \mathbf{p}_0 e^{i(\omega t - \mathbf{kr})}, \tag{68.3}$$

где $\mathbf{p}$ — дипольный момент атома с радиусом-вектором $\mathbf{r}$. Каждый диполь, излучая, теряет энергию. Но эта убыль энергии восполняется за счет энергии, приходящей от других диполей. Излучение других диполей создает в месте нахождения рассматриваемого диполя электрическое поле, которое поддерживает установившиеся гармонические колебания этого диполя. Таким образом, вся среда ведет себя как *замкнутая система*, совершающая *свободные*, а не вынужденные колебания без каких бы то ни было внешних воздействий. Если длина волны велика по сравнению с межатомными расстояниями, то среду можно считать сплошной и характеризовать ее состояние вектором поляризации

$$\mathbf{P} = N\mathbf{p} = \mathbf{P}_0 e^{i(\omega t - \mathbf{kr})}, \tag{68.4}$$

где N — число атомов в единице объема.

§ 69. Вывод формул Френеля в молекулярной оптике

1. Допустим, что однородная изотропная среда граничит с вакуумом вдоль плоскости. Падающая на нее плоская электромагнитная волна возбудит в среде *дипольную волну* (68.3), которую при усреднении по физически бесконечно малым объемам можно рассматривать как *волну поляризации* (68.4). Направления распространения этой и отраженной волн найдутся из условия равенства фазовых скоростей всех волн параллельно границе раздела. Это приводит к равенству *тангенциальных слагающих* волновых векторов $\mathbf{k}_0$, $\mathbf{k}'_0$ и $\mathbf{k}$ падающей, отраженной и прошедшей волн. Кроме того, должны выполняться соотношения $k_0^2 = k_0'^2 = (\omega/c)^2$, так как волна падает на среду из вакуума.

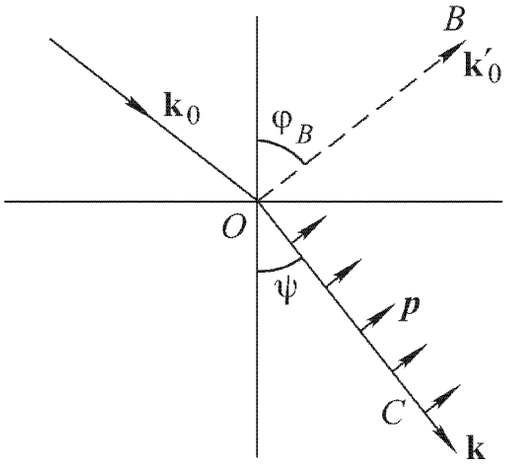


Рис. 251

Пусть теперь электрический вектор падающей волны лежит в плоскости падения (рис. 251). Если волна падает под углом Брюстера, то преломленный луч OC будет перпендикулярен к направлению отраженного луча OB. Ввиду поперечности световых волн электрический вектор в среде перпендикулярен к преломленному лучу OC. Возбуждаемые им дипольные моменты атомов будут также перпендикулярны к OC, а следовательно, параллельны OB. Но вдоль колебаний дипольного момента диполь Герца не излучает. Значит, в направлении OB атомы среды *не излу-*

чают и отраженная волна возникнуть не может. Так просто объясняется закон Брюстера. (См. также задачу 4 к следующему параграфу.)

2. При выводе формул Френеля для отраженного света среду можно считать сплошной, так как межатомные расстояния пренебрежимо малы по сравнению с расстояниями от границы среды до точки наблюдения, где определяется поле отраженной волны. При такой идеализации излучающими центрами будут не отдельные атомы, а элементы объема среды, малые по сравнению с λ^3, но содержащие еще очень много атомов.

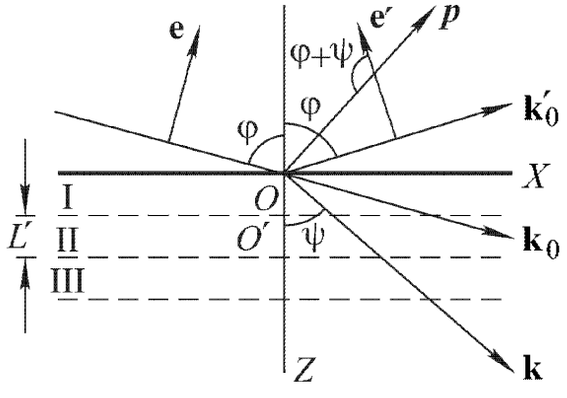


Рис. 252

Выберем координатные оси так же, как это было сделано на рис. 238. Разобьем среду на плоскопараллельные слои I, II, III, ... одинаковой толщины, называемые в дальнейшем *зонами*, по аналогии с кольцевыми зонами Френеля (рис. 252). Каждая плоскопараллельная зона будет излучать в верхнее полупространство плоскую волну с волновым вектором $\mathbf{k}_0'$.

Толщину зоны L' подберем так, чтобы каждая последующая волна запаздывала по фазе на π относительно предыдущей волны. В момент $t = 0$ фаза дипольных колебаний в точке O равна нулю, в точку O' дипольные колебания приходят с фазой $(-k_z L')$. Волна же, излученная в O', приходит в O с дополнительным запаздыванием по фазе на $k_{0z}L'$. Результирующее запаздывание по фазе $(k_z + k_{0z})L'$ по условию должно быть равно π, откуда и находится требуемое значение $L' = \pi/(k_z + k_{0z})$.

Теперь ясно, что электрическое поле $\mathbf{E}^{(r)}$ отраженной волны представится знакопеременным рядом

$$\mathbf{E}^{(r)} = \mathbf{E}_1 - \mathbf{E}_2 + \mathbf{E}_3 - \ldots, \tag{69.1}$$

члены которого представляют поля излучения соответствующих зон. При строгой однородности волны поляризации (68.4) члены ряда (69.1) одинаковы по абсолютной величине. Частичные суммы ряда колеблются между нулем и $\mathbf{E}_1$, а сумма самого ряда не имеет определенного значения.

В действительности реальная волна поляризации *не однородна.* Она имеет *передовой фронт*, перед которым волнового возмущения нет. На самом деле ряд (69.1) содержит конечное число членов, а потому сходится. Для нахождения его суммы мысленно удалим первую зону вместе с излучаемым ею полем и сдвинем

оставшуюся среду вверх на расстояние L'. Изменим далее фазы всех диполей среды на одну и ту же величину и притом такую, чтобы диполи, оказавшиеся после смещения на границе раздела, получили те же фазы, какие имели бы в тот же момент времени удаленные с этой границы диполи первой зоны. Ввиду медленности изменения членов ряда (69.1) и тождественности зон, поле излучения среды в верхнем полупространстве в результате этих операций практически останется без изменения. Но теперь оно может быть представлено рядом

$$\mathbf{E}^{(r)} = \mathbf{E}_2 - \mathbf{E}_3 + \mathbf{E}_4 - \ldots \tag{69.2}$$

Складывая его с предыдущим, найдем

$$\mathbf{E}^{(r)} = \frac{1}{2}\,\mathbf{E}_1. \tag{69.3}$$

Таким образом, *напряженность поля излучения среды в верхнем полупространстве равна половине напряженности поля, излучаемого в то же полупространство первой зоной.* Мы пришли к тому же конечному результату, что и в теории кольцевых зон Френеля (см. § 39). Однако здесь наши результаты более убедительны, поскольку соотношение (69.2) для плоских зон лучше обосновано, чем аналогичное соотношение для кольцевых зон Френеля.

3. Прежде чем идти дальше, необходимо сделать следующее замечание. Отраженная волна, излучаемая сплошной средой с непрерывно распределенным вектором поляризации (68.4), в верхнем полупространстве будет всюду плоской и однородной. В случае реальной среды, построенной из дискретных атомов, на плавно меняющееся поле волны вблизи границы среды накладывается «молекулярная рябь», т.е. дополнительное поле, резко меняющееся на расстояниях порядка межатомных расстояний. Это поле может быть представлено в виде суперпозиции волн, быстро затухающих при удалении от границы среды. Пространственный период вдоль оси X таких неоднородных или поверхностных волн порядка расстояния a между атомами. Соответствующая составляющая волнового вектора будет порядка $K_x \approx \approx 2\pi/a$. А так как длина самого вектора $\mathbf{K}$ должна быть равна ω/c, то для его z-составляющей получится $K_z = \sqrt{(\omega/c)^2 - K_x^2} \approx \approx i \cdot 2\pi/a$. Отсюда видно, что при удалении от границы раздела на расстояние a поле неоднородных волн убывает приблизительно в $e^{2\pi} \approx 500$ раз.

Выше указывалось, что замена дискретных излучающих центров (атомов) непрерывно распределенными источниками годится при вычислении поля излучения в верхнем полупространстве на таких расстояниях от границы, которые очень велики по сравнению с a. Приведенная оценка показывает, что это ограни-

чение в действительности должно быть сильно ослаблено. Поле излучения реальной среды не отличается от соответствующего поля ее идеализированной сплошной модели уже на расстояниях от границы порядка a.

4. Дальше при выводе формул Френеля можно было бы пойти двумя путями. Можно было бы рассчитать поле излучения диполей среды в занимаемом ею полупространстве. Это сложно. Мы обойдем вычисления, использовав замечание в конце предыдущего параграфа.

Представим падающую волну как *результат излучения некоторой среды*. Вообразим, что верхнее полупространство на рис. 252 заполнено такой же средой, что и нижнее, а во всем бесконечном пространстве распространяется однородная волна поляризации (68.4) Воображаемая среда оказывает влияние на среду в нижнем полупространстве посредством своего поля излучения. А так как в нижнем полупространстве ничего не изменилось, то поле, излучаемое воображаемой средой в нижнее полупространство, тождественно с полем падающей волны.

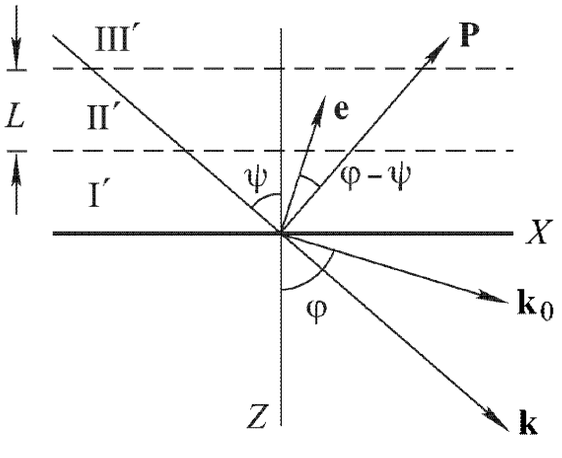


Рис. 253

Разобьем воображаемую среду на плоскопараллельные зоны I′, II′, III′, толщина каждой из которых равна $L = \pi/(k_z - k_{0z})$ (рис. 253). Рассуждая, как и раньше, легко показать, что напряженность поля, излучаемого воображаемой средой в нижнее полупространство, равна половине напряженности поля излучения зоны I'. Вычисление отношения амплитуд отраженной и падающей волн R и $\mathcal{E}$ свелось, таким образом, к сравнению полей излучения зон I и I' на границе раздела $z = 0$.

Если вектор $\mathbf{P}$ перпендикулярен к плоскости падения, то отношение амплитуд R и $\mathcal{E}$ будет просто равно отношению толщин зон I и I', взятому со знаком минус, поскольку фазы волн, излучаемых этими зонами, на границе раздела противоположны. Это дает первую формулу Френеля:

$$\frac{R_\perp}{\mathcal{E}_\perp} = -\frac{L'}{L} = -\frac{k_z - k_{0z}}{k_z + k_{0z}} = -\frac{\sin(\varphi - \psi)}{\sin(\varphi + \psi)}. \tag{69.4}$$

Если же вектор $\mathbf{P}$ параллелен плоскости падения, то необходимо еще учесть зависимость поля излучения диполя от угла между дипольным моментом и направлением излучения. Излучение дает только поперечная слагающая дипольного момен-

та. Поэтому величину $-L'/L$ необходимо умножить на отношение проекций вектора $\mathbf{P}$ на направления единичных векторов $\mathbf{e}'$ и $\mathbf{e}$, перпендикулярных к направлениям распространения отраженной и падающей волн. С учетом поперечности преломленной волны ($\mathbf{kP} = 0$) для этого отношения нетрудно найти $-\cos(\varphi+\psi)/\cos(\varphi-\psi)$ (см. рисунки 252 и 253), а для отношения самих амплитуд $R_{\parallel}$ и $\mathcal{E}_{\parallel}$

$$\frac{R_{\parallel}}{\mathcal{E}_{\parallel}} = -\frac{R_{\perp}}{\mathcal{E}_{\perp}}\frac{\cos(\varphi+\psi)}{\cos(\varphi-\psi)} = \frac{\operatorname{tg}(\varphi-\psi)}{\operatorname{tg}(\varphi+\psi)}. \tag{69.5}$$

Это — вторая формула Френеля.

Таким же путем может быть разобран случай обратного падения, когда волна падает из среды на ее границу с вакуумом, а также общий случай отражения на границе раздела двух материальных сред.

§ 70. Отступления от формул Френеля

1. Отражение и преломление света на чистой границе раздела двух прозрачных изотропных сред хорошо описывается формулами Френеля. Небольшие отступления от них заметны лишь при отражении под углом Брюстера и вблизи него. С особой отчетливостью они проявляются в существовании следующих двух явлений:

а) при отражении света, поляризованного в плоскости падения, *закон Брюстера соблюдается не строго*; коэффициент отражения не обращается в нуль ни для какого угла падения, хотя при угле Брюстера он и очень мал;

б) свет, линейно поляризованный под углом к плоскости падения, при отражении в малой окрестности угла Брюстера становится поляризованным *эллиптически*, тогда как при строгой справедливости формул Френеля он должен был бы оставаться поляризованным линейно во всех случаях, за исключением случаев полного отражения.

Эти два отступления *не независимы*, а взаимно обусловливают друг друга. Действительно, при точной справедливости формул Френеля

$$\frac{R_{\perp}}{\mathcal{E}_{\perp}} = -\frac{\sin(\varphi-\psi)}{\sin(\varphi+\psi)}, \quad \frac{R_{\parallel}}{\mathcal{E}_{\parallel}} = \frac{\operatorname{tg}(\varphi-\psi)}{\operatorname{tg}(\varphi+\psi)} \tag{70.1}$$

отношение $R_{\perp}/\mathcal{E}_{\perp}$ сохраняло бы знак при всех углах падения, тогда как знак отношения $R_{\parallel}/\mathcal{E}_{\parallel}$ менялся бы при переходе через угол Брюстера. Это значит, что во втором случае при переходе через угол Брюстера фаза отражающейся волны скачкообразно менялась бы на π. Никакого нарушения непрерывности $R_{\parallel}$, как

функции угла падения φ, при этом не произошло бы, так как, согласно второй формуле (70.1), эта функция при $\varphi = \varphi_B$ обращается в нуль. Однако функция $R_{||}(\varphi)$ сделалась бы разрывной при $\varphi = \varphi_B$, если бы, как это действительно имеет место, она при этом угле не была равна нулю. Для такого разрыва нет физических оснований. Поэтому следует заключить, что изменение фазы, претерпеваемое $R_{||}(\varphi)$, происходит не скачкообразно при угле $\varphi = \varphi_B$, а в какой-то малой окрестности вокруг угла Брюстера. В пределах этой окрестности фаза меняется непрерывно от 0 до π. В той же окрестности возникает сдвиг фаз между составляющими $\mathbf{E}_{||}$ и $\mathbf{E}_{\perp}$, а потому отраженный свет будет поляризован эллиптически, если только падающий был поляризован линейно. Это показывает, что *второе отступление является следствием первого.*

Обратное также справедливо: *первое отступление вытекает из второго*, если только среды не вращают плоскость поляризации. Это непосредственно очевидно, если заметить, что при соблюдении указанного условия волны $\mathbf{E}_{||}$ и $\mathbf{E}_{\perp}$ сохраняют характер своей поляризации и в соответствии с принципом суперпозиции распространяются независимо друг от друга.

2. Такие отступления от формул Френеля можно объяснить наличием на поверхности отражающей среды тонкого *переходного слоя*, оптические свойства которого отличны от оптических свойств самой среды. В правильности такого объяснения не приходится сомневаться, по крайней мере когда среда свободна от внутренних напряжений. Но на вопрос, как эти слои возникают и каковы их свойства, еще нельзя ответить с полной определенностью.

Простейшим является случай, когда на поверхности отражающей среды существует тонкая пленка постороннего вещества, нанесенная искусственно или возникшая в результате обработки и загрязнения поверхности. Когда толщина пленки велика по сравнению с межатомными расстояниями, ее можно характеризовать *макроскопическими величинами* — толщиной l и показателем преломления n. Пленки с толщиной порядка λ искусственно наносятся на отражающие поверхности для увеличения их поверхностной прозрачности, а также для изготовления интерференционных светофильтров. Если пленка однородна, т.е. величина n одинакова по всей толщине ее, то отражение и преломление света описывается формулами (67.9).

Нас, однако, сейчас интересуют *тонкие пленки*, толщина которых мала по сравнению с λ. В этом случае формулы (67.9) можно упростить, разложив их по степеням l/λ и сохранив только члены первой степени. Для таких пленок вычисление можно обобщить, не вводя предположения об однородности пленки, а

предполагая, что при переходе через пленку показатель преломления меняется *непрерывно*. Так поступили Друде (1863–1906) и многие другие авторы. Однако оценка толщин переходных слоев на поверхностях чистых жидкостей и свежих сколов кристаллов показала, что эти толщины *того же порядка*, что и атомные размеры или межатомные расстояния. Применять внутри таких слоев макроскопические уравнения Максвелла и характеризовать их показателем преломления не корректно. Мы получим нужные результаты методами молекулярной оптики, изложенными в предыдущем параграфе.

3. Допустим сначала, что никакого переходного слоя на отражающей поверхности нет. Как выяснено в предыдущем параграфе, для определения поля отраженной волны достаточно рассмотреть излучение первой зоны I (см. рис. 252). Разобьем эту зону на бесконечно тонкие слои толщины $d\zeta$ (рис. 254). Вектор поляризации в точке O' определяется выражением $\mathbf{P}_0 e^{i(\omega t - k_z \zeta)}$. На единицу площади слоя $d\zeta$ приходится дипольный момент $\mathbf{P}_0\, d\zeta\, e^{i(\omega t - k_z \zeta)}$. Колебания из O' приходят в O с запаздыванием по фазе на $k_{0z}\zeta$. При вычислениях слой $d\zeta$ можно мысленно перенести на границу среды, введя соответствующий сдвиг фаз. Дипольный момент, приходящийся на единицу площади так мысленно передвинутого слоя, будет

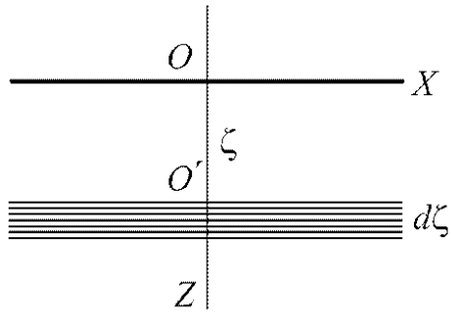


Рис. 254

$$d\mathbf{A}^{(r)} e^{i\omega t} = \mathbf{P}_0 e^{i\omega t} e^{-i(k_{0z}+k_z)\zeta}\, d\zeta.$$

Проинтегрируем это выражение по толщине первой зоны I и результат разделим на два. Тогда найдем, что отраженная волна может быть моделирована полем излучения «эквивалентной плоскости», расположенной на границе среды, на которой совершаются дипольные колебания с поверхностной плотностью амплитуды

$$\mathbf{A}_0^{(r)} = \frac{\mathbf{P}_0}{2} \int_0^{L'} e^{-i(k_{0z}+k_z)\zeta}\, d\zeta = i\, \frac{\mathbf{P}_0}{2(k_{0z}+k_z)} \left[e^{-i(k_{0z}+k_z)L'} - 1\right],$$

или, с учетом соотношения $(k_{0z}+k_z)L' = \pi$,

$$\mathbf{A}_0^{(r)} = -i\, \frac{\mathbf{P}_0}{k_{0z}+k_z}. \tag{70.2а}$$

Рассуждая так же применительно к зоне I′ (см. рис. 253), найдем, что падающая волна может быть моделирована полем излучения другой «эквивалентной плоскости», расположенной также на границе среды. На единицу площади такой плоскости приходится дипольный момент с амплитудой

$$\mathbf{A}_0^{(e)} = -i\, \frac{\mathbf{P}_0}{k_{0z}-k_z}. \tag{70.2б}$$

4. Учтем теперь наличие переходного слоя. Так как толщина слоя предполагается малой по сравнению с λ, его можно моделировать математиче-

ской плоскостью. Обозначим через $\boldsymbol{\tau} = \boldsymbol{\tau}_0 e^{i(\omega t - k_{0x} x)}$ дополнительный дипольный момент, вносимый переходным слоем на единицу его площади. Хотя $\boldsymbol{\tau}$ — величина макроскопическая, ею допустимо пользоваться для *сколь угодно тонких переходных слоев* (в том числе и мономолекулярных). Это связано с тем, что введение $\boldsymbol{\tau}$ предполагает усреднение не по малым объемам среды, а *по малым площадкам* отражающей поверхности. Последние же можно выбрать малыми по сравнению с длиной волны, но еще большими по сравнению с межатомными расстояниями.

Величину $\boldsymbol{\tau}_0$ надо добавить к амплитуде $\mathbf{A}_0^{(r)}$, чтобы моделировать поле отраженной волны с учетом излучения переходного слоя. Для моделирования падающей волны ту же величину $\boldsymbol{\tau}_0$ надо добавить к амплитуде $\mathbf{A}_0^{(e)}$, но с противоположным знаком. Действительно, после такого добавления волна, проникающая извне внутрь среды и воздействующая на нее, будет моделирована излучением двух плоскостей, расположенных на границе среды: плоскости с амплитудой дипольных колебаний $(\mathbf{A}_0^{(e)} - \boldsymbol{\tau}_0)$ и плоскости с амплитудой таких же колебаний $\boldsymbol{\tau}_0$. Амплитуды $\boldsymbol{\tau}_0$ и $-\boldsymbol{\tau}_0$ компенсируются. Остается *только одна амплитуда* $\mathbf{A}_0^{(e)}$, как если бы переходный слой не оказывал никакого влияния на среду вдали от ее границы. Но это как раз и необходимо, чтобы поляризация среды вне переходного слоя представлялась однородной волной (68.4).

Таким образом, вместо (70.2а) и (70.2б) получаются выражения

$$\mathbf{A}^{(r)} = -i\,\frac{\mathbf{P}_0}{k_z + k_{0z}} + \boldsymbol{\tau}_0, \quad \mathbf{A}^{(e)} = i\,\frac{\mathbf{P}_0}{k_z - k_{0z}} - \boldsymbol{\tau}_0, \tag{70.3}$$

которыми и надо пользоваться при вычислении поля падающей и отраженной волн. При этом, разумеется, надо учесть, что для излучения существенны не сами амплитуды (70.3), а только их составляющие, *перпендикулярные к направлению излучения*. Коэффициенты Френеля $R_\perp/\mathcal{E}_\perp$ и $R_\parallel/\mathcal{E}_\parallel$ найдутся как отношения таких составляющих. Ограничиваясь линейными членами по k_0, в результате простых, но несколько длинных вычислений получим

$$\frac{R_\perp}{\mathcal{E}_\perp} = -\frac{\sin(\varphi - \psi)}{\sin(\varphi + \psi)}(1 + 2ik_0\gamma_x \cos\varphi), \tag{70.4}$$

$$\frac{R_\parallel}{\mathcal{E}_\parallel} = \frac{\operatorname{tg}(\varphi - \psi)}{\operatorname{tg}(\varphi + \psi)}\left(1 + 2ik_0 \cos\varphi\,\frac{\gamma_x \cos^2\psi - \gamma_z \sin^2\varphi}{\cos^2\psi - \sin^2\varphi}\right). \tag{70.5}$$

Здесь введены обозначения:

$$\gamma_x = \frac{\tau_{0x}}{P_{0x}} \equiv \frac{\tau_{0y}}{P_{0y}}, \quad \gamma_z = \frac{\tau_{0z}}{P_{0z}}. \tag{70.6}$$

Параметры γ_x и γ_z имеют размерность длины. В изложенной теории ими и характеризуются свойства переходного слоя. Теория справедлива при выполнении условий

$$\gamma_x k_0 \ll 1, \quad \gamma_z k_0 \ll 1, \quad \text{или} \quad \gamma_x/\lambda \ll 1, \quad \gamma_z/\lambda \ll 1.$$

Квадратичные члены $(\gamma_x/\lambda)^2$ и $(\gamma_z/\lambda)^2$ в ней отброшены. В этом приближении все коэффициенты при малых величинах γ_x/λ и γ_z/λ не зависят от λ. Это значит, что вычисление параметров γ_x и γ_z есть *чисто электростатическая задача*. Для фактического вычисления самих параметров в молекулярной теории необходимо знать молекулярную структуру среды и переходного слоя.

5. Состояние поляризации отраженного света полностью характеризуется комплексным отношением $R_{\parallel}/R_{\perp}$. Если $\mathcal{E}_{\perp} = \mathcal{E}_{\parallel}$, то

$$\frac{R_{\parallel}}{R_{\perp}} = -\frac{\cos(\varphi+\psi)}{\cos(\varphi-\psi)} + 2ik_0(\gamma_z - \gamma_x)\frac{\cos\varphi\sin^2\varphi}{\cos^2(\varphi-\psi)}. \tag{70.7}$$

При угле Брюстера это отношение *чисто мнимое*, т.е. отраженный свет поляризован *эллиптически*. Одна из главных осей эллипса колебаний перпендикулярна к плоскости падения, а другая лежит в этой плоскости. Отношение этих осей называется *коэффициентом эллиптичности света*, отраженного под углом Брюстера. Он может быть и положительным, и отрицательным, так как разность фаз δ между составляющими $R_{\parallel}$ и $R_{\perp}$ при угле Брюстера принято считать равной $+\pi/2$. При таком соглашении $R_{\parallel}/R_{\perp} = \rho e^{i\pi/2} = i\rho$. Для ρ из формулы (70.7) нетрудно получить

$$\rho = \frac{\pi}{\lambda}\sqrt{n^2+1}\,(\gamma_z - \gamma_x), \tag{70.8}$$

где λ означает длину волны в вакууме. Положительному коэффициенту эллиптичности соответствует левая эллиптическая поляризация отраженного света, отрицательному — правая (предполагается, что падающий свет линейно поляризован под углом 45° к плоскости падения, т.е. $\mathcal{E}_{\parallel} = \mathcal{E}_{\perp}$). Эллиптическая поляризация называется *правой*, если для глаза, смотрящего против распространения света, движение по эллипсу совершается по часовой стрелке. В противоположном случае она называется *левой*.

Для всех исследованных жидкостей с чистой отражающей поверхностью коэффициент ρ положителен. Для обыкновенной воды он отрицателен. Однако, как показал Рэлей, при очищении поверхности воды величина ρ меняет знак, обращаясь в нуль при некоторой промежуточной степени очищения. Для предельно чистой поверхности воды Рэлей нашел $\rho = +0{,}00042$. В табл. 6 приведены значения ρ для различных жидкостей (при $\lambda = 546$ нм).

Таблица 6

Вещество	$\rho \cdot 10^5$	n	$l_{\text{мин}}/\lambda$
Вода	42	1,33	0,00057
Метиловый спирт	46	1,33	0,00062
Этиловый спирт	92	1,33	0,00124
Четыреххлористый углерод	107	1,46	0,00103
Бензол	135	1,51	0,00117
Толуол	144	1,50	0,00101
Ортоксилол	137	1,51	0,00119

В последнем столбце табл. 6 приведены значения минимальной толщины $l_{\text{мин}}$ переходного слоя, которые по теории Друде достаточны для получения наблюдаемой эллиптичности отраженного света (см. задачу 3 к этому параграфу). Хотя эта теория и макроскопическая, приведенная оценка дает правильный порядок величины $l_{\text{мин}}$. Из нее видно, что переходные слои порядка одного или нескольких межатомных расстояний уже достаточны, чтобы вызвать наблюдаемые отступления от формул Френеля. Это с несомненностью указывает на то, что, наряду с переходными слоями, вызванными загрязнением, обработкой, абсорбцией газов и пр., существуют и пере-

ходные слои, обусловленные *молекулярной структурой* самой отражающей среды вблизи ее поверхности.

Трудно указать причины возникновения таких переходных слоев. Вот некоторые из возможных причин. 1. Электрическое поле световой волны, действующее на приграничные молекулы среды, может отличаться от действующего поля для остальных молекул. 2. Анизотропные молекулы, если среда построена из таковых, могут ориентироваться вблизи поверхности в некоторых преимущественных направлениях. 3. Плотность среды в тонком приграничном слое может отличаться от плотности остальной среды.

ЗАДАЧИ

1. Считая переходный слой сплошным, вычислить для него параметры γ_x и γ_z.

Р е ш е н и е. Задача сводится к нахождению вектора поляризации $\mathbf{P}(z)$ среды и переходного слоя в однородном электрическом поле $\mathbf{E}$. Среду будем характеризовать диэлектрическими проницаемостями $\varepsilon_x(z)$ и $\varepsilon_z(z)$ в направлениях параллельно и перпендикулярно к слою. Координатные оси можно выбрать так, чтобы $E_y = D_y = 0$. В силу симметрии векторы $\mathbf{E}$ и $\mathbf{D}$ не могут зависеть от x и y. Поэтому уравнения $\operatorname{rot}\mathbf{E} = 0$ и $\operatorname{div}\mathbf{D} = 0$ переходят в

$$\frac{dE_x}{dz} = \frac{dD_z}{dz} = 0.$$

Отсюда следует, что во всем пространстве E_x и D_z постоянны, т.е. не зависят от z. Вектор поляризации в переходном слое будет

$$P_x = \frac{\varepsilon_x - 1}{4\pi} E_z, \quad P_z = \frac{\varepsilon_z - 1}{4\pi} E_z = \frac{1}{4\pi}\left(1 - \frac{1}{\varepsilon_z}\right) D_z.$$

Чтобы выделить дополнительную поляризацию $\Delta\mathbf{P}$ переходного слоя, вычтем отсюда вектор поляризации самой однородной среды:

$$\Delta P_x = \frac{\varepsilon_x - \varepsilon}{4\pi} E_x, \quad \Delta P_z = \frac{1}{r\pi}\left(\frac{1}{\varepsilon} - \frac{1}{\varepsilon_z}\right) D_z,$$

где ε — диэлектрическая проницаемость этой среды. Для нахождения τ_{0x} и τ_{oz} полученные выражения надо проинтегрировать по толщине l переходного слоя. Выполнив это и введя обозначения

$$p = \int_0^l \varepsilon_x(z)\, dz, \quad q = \int_0^l \frac{dz}{\varepsilon_z(z)},$$

найдем

$$\gamma_x = \frac{p - ln^2}{n^2 - 1}, \quad \gamma_z = \frac{l - qn^2}{n^2 - 1}. \tag{70.9}$$

Если подставить эти значения в (70.4) и (70.5), то получатся *формулы Друде.*

2. Считая переходный слой макроскопическим и однородным, получить выражение для коэффициента эллиптичности при угле Брюстера.

О т в е т.

$$\rho = \frac{\pi}{\lambda}\sqrt{n^2+1}\,\frac{(n^2 - n'^2)(n'^2 - 1)}{(n^2 - 1)n'^2}\, l, \tag{70.10}$$

где n' — показатель преломления переходного слоя.

3. Считая эллиптичность ρ положительной, найти значение n', чтобы при заданных n и l величина ρ получилась максимальной. Найти минимальное значение для l, чтобы при заданном n получилось заданное значение ρ.

О т в е т.

$$n' = \sqrt{n}, \quad l_{\text{мин}} = \frac{\rho}{\pi\sqrt{n^2+1}}\,\frac{n+1}{n-1}\,\lambda. \tag{70.11}$$

4. Изотропная среда построена из анизотропных молекул. Ее макроскопическая изотропия обусловлена беспорядочными изменениями направлений молекул в результате теплового движения. Дипольный момент анизотропной молекулы, возбуждаемый полем световой волны, вообще говоря, не совпадает по направлению с электрическим полем. Пусть электрический вектор падающей волны лежит в плоскости падения, а сама волна падает под углом Брюстера. Электрический вектор в среде хотя и перпендикулярен к преломленному лучу, но дипольные моменты молекул, вообще говоря, имеют составляющие вдоль луча. Такие молекулы, очевидно, будут излучать в направлении отраженного луча. Будет ли в этом случае выполняться закон Брюстера? (См. § 69, п. 1.)

Р е ш е н и е. Закон Брюстера останется справедливым. Для макроскопически изотропной среды составляющие дипольных моментов молекул, параллельные преломленному лучу, меняются хаотически во времени. В силу этого излучения отдельных молекул в направлении отраженного луча некогерентны. Они не могут привести к правильному отражению, а только к диффузному рассеянию света. Но если в тонком поверхностном слое анизотропные молекулы преимущественно ориентированы в некотором направлении, то излучения таких молекул становятся частично когерентными, и появляется отраженный свет (см. § 69).

ГЛАВА VI

ОПТИКА МЕТАЛЛОВ

§ 71. Уравнения Максвелла и волны в металлах

1. Как и в диэлектриках, распространение электромагнитных волн в металлах описывается уравнениями Максвелла:

$$\begin{aligned} \operatorname{rot}\mathbf{H} &= \frac{1}{c}\dot{\mathbf{D}} + \frac{4\pi}{c}\mathbf{j}, \\ \operatorname{rot}\mathbf{E} &= -\frac{1}{c}\dot{\mathbf{B}}, \\ \operatorname{div}\mathbf{D} &= 4\pi\rho, \\ \operatorname{div}\mathbf{B} &= 0. \end{aligned} \tag{71.1}$$

По сравнению с диэлектриками добавляется член с током проводимости $\mathbf{j}$. В статических и низкочастотных полях ток обусловлен движением практически одних только *свободных электронов*. Движение связанных электронов, а тем более атомных ядер в этой области спектра не играет никакой роли. Но уже в инфракрасной области, где лежат собственные частоты колебаний атомных ядер, движения последних начинают существенно влиять на оптическое поведение металлов (резонанс). В дальнейшем, в видимой и ультрафиолетовой областях спектра, в колебания вовлекаются и *связанные электроны*.

Смещения связанных зарядов вызывают *поляризацию металлов*, аналогичную поляризации диэлектриков. Появляется *ток поляризации*. Однако в высокочастотных полях нет существенной разницы между движениями свободных и связанных частиц. Нерационально связывать ток проводимости с движением только свободных, а ток поляризации — только связанных электронов. Этим понятиям надо дать точные определения, в соответствии с тем, как они используются в теории.

Можно поступить следующим образом. Вынужденное колебание заряженной частицы (электрона или атомного ядра) в по-

ле световой волны можно разложить на колебание, происходящее в фазе (или противофазе) с электрическим полем, и колебание, сдвинутое относительно него по фазе на 90°. Ток, обусловленный первыми колебаниями, условимся называть *током проводимости* и обозначать его плотность через $\mathbf{j}$. Токи проводимости вызывают поглощение электромагнитных волн. Токи же, обусловленные колебаниями, сдвинутыми по фазе на 90°, назовем *токами поляризации*. С ними не связано поглощение волн. Плотность тока поляризации можно представить в виде $\mathbf{j}_{\text{пол}} = \dot{\mathbf{P}}$, где $\mathbf{P}$ — вектор поляризации. На последнее соотношение можно смотреть как на определение вектора $\mathbf{P}$. Тем самым в металлах раскрывается и точный смысл вектора индукции $\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P}$, который входит в уравнения (71.1).

Магнитные свойства вещества, не играющие существенной роли в оптической области спектра, мы учитывать не будем. Металлы будем считать *оптически изотропными*, хотя все металлы (за исключением, конечно, жидких) и имеют кристаллическую структуру. Однако кристаллы кубической системы оптически изотропны. Другие металлы, как правило, *макроскопически изотропны*, так как они состоят из множества хаотически ориентированных кристалликов, размеры которых малы по сравнению с длиной волны. Для изотропных металлов материальные уравнения имеют вид

$$\mathbf{D} = \varepsilon'\mathbf{E}, \tag{71.2}$$

$$\mathbf{j} = \sigma\mathbf{E}. \tag{71.3}$$

Величина σ называется *электрической проводимостью*, а ε' — *диэлектрической проницаемостью* металла. (Мы пользуемся обозначением ε', сохраняя ε для обозначения комплексной величины, вводимой ниже.) Обе величины ε' и σ являются функциями частоты ω. Поэтому в уравнениях (71.2) и (71.3) поле $\mathbf{E}$ должно предполагаться монохроматическим. Немонохроматические поля надо разлагать на монохроматические составляющие и применять принцип суперпозиции.

С наличием проводимости σ связано *поглощение света* в металлах. С точки зрения излагаемой здесь формальной теории поглощение света есть не что иное, как превращение электромагнитной энергии в *джоулево тепло*.

Материальные уравнения (71.2) и (71.3) дают лишь грубое описание оптических свойств металлов (см. § 74). В ряде вопросов, в особенности для коротких волн (ультрафиолетовые, видимые и короткие инфракрасные лучи), они приводят к выводам, не совсем согласующимся с опытом. Более удовлетворительная теория должна основываться на *квантовой теории металлов*.

Однако изложение такой теории далеко выходит за рамки этой книги.

2. Для монохроматических полей $\dot{\mathbf{D}} = i\omega\mathbf{D} = i\omega\varepsilon'\mathbf{E}$, так что правая часть первого уравнения (71.1) преобразуется в $\frac{1}{c}(i\omega\varepsilon' + 4\pi\sigma)\mathbf{E}$. Если ввести величину

$$\varepsilon = \varepsilon' - i\frac{4\pi\sigma}{\omega} \equiv \varepsilon' - i\varepsilon'', \tag{71.4}$$

называемую *комплексной диэлектрической проницаемостью*, то система уравнений (71.1) примет вид

$$\operatorname{rot}\mathbf{H} = \frac{i\omega}{c}\varepsilon\mathbf{E}, \quad \operatorname{rot}\mathbf{E} = -\frac{i\omega}{c}\mathbf{H}, \tag{71.5}$$

т.е. формально совпадет с соответствующими уравнениями для диэлектриков. Граничные условия также имеют одинаковый вид: они требуют непрерывности тангенциальных компонент векторов $\mathbf{E}$ и $\mathbf{H}$ на границе раздела сред. Как следствие таких граничных условий и уравнений (71.5), получается непрерывность нормальных компонент вектора $\varepsilon\mathbf{E}$ (см. § 63). Поэтому *любое соотношение оптики прозрачных сред, полученное из уравнений* (71.5) *и граничных условий с помощью линейных вещественных операций, может быть формально перенесено в оптику металлов и других поглощающих сред простой заменой вещественной величины ε на комплексную.* Требуется только дополнительное исследование физического содержания и смысла полученного соотношения.

Для характеристики оптических свойств металлов применяется также *комплексный показатель преломления* ν. Он определяется соотношением

$$\nu^2 = \varepsilon = \varepsilon' - i\varepsilon''. \tag{71.6}$$

Полагая

$$\nu = n - i\varkappa, \tag{71.7}$$

где n и $\varkappa$ вещественны и существенно положительны, получим

$$n^2 - \varkappa^2 = \varepsilon', \quad 2n\varkappa = \varepsilon''. \tag{71.8}$$

Величина n называется *главным показателем преломления* металла. Величину $\varkappa$ называют *главным показателем затухания*. Этот термин надо предпочесть обычно употребляемому термину «показатель поглощения», так как *затухание волны может происходить и без поглощения*. Примером может служить плазма, когда частота ω меньше так называемой *плазменной частоты* (см. § 87). Для нее величина ε вещественна, но *отрицательна*, т.е. $\varepsilon'' = 0$. В этом случае поглощения нет, но есть затухание, так как $\sqrt{\varepsilon}$ — величина чисто мнимая, а потому $\varkappa \neq 0$.

Если среда однородна, то в ней могут распространяться плоские монохроматические волны вида (5.3). Для них должны выполняться соотношения

$$\mathbf{E} = -\frac{c}{\omega\varepsilon}[\mathbf{kB}], \quad \mathbf{B} = \frac{c}{\omega}[\mathbf{kE}], \tag{71.9}$$

$$\mathbf{k}^2 = \frac{\omega^2}{c^2}\varepsilon = \frac{\omega^2}{c^2}(\varepsilon' - i\varepsilon''). \tag{71.10}$$

В поглощающей среде волновой вектор $\mathbf{k}$ всегда комплексный, а соответствующая плоская волна *всегда неоднородна*. Это естественно, так как при наличии поглощения плоская волна не может распространяться без затухания. Положим

$$\mathbf{k} = \mathbf{k}' - i\mathbf{k}'', \tag{71.11}$$

где $\mathbf{k}'$ и $\mathbf{k}''$ — вещественные векторы. Тогда на основании (71.10)

$$k'^2 - k''^2 = \frac{\omega^2}{c^2}\varepsilon', \quad 2(\mathbf{k}'\mathbf{k}'') = \frac{\omega^2}{c^2}\varepsilon''. \tag{71.12}$$

Вектор $\mathbf{k}'$ указывает направление распространения *плоскостей равных фаз*. В направлении вектора $\mathbf{k}''$ убывает амплитуда волны. В общем случае *плоскости равных фаз и плоскости равных амплитуд не перпендикулярны между собой*. Перпендикулярность всегда имеет место только для непоглощающих сред, когда $\varepsilon'' = 0$.

§ 72. Геометрические законы отражения и преломления света на границе металла

1. Пусть из вакуума на плоскую границу металла падает плоская монохроматическая однородная волна, распространяющаяся вдоль волнового вектора $\mathbf{k}_1$ (рис. 255). При этом возникнет однородная отраженная волна с волновым вектором $\mathbf{k}'_1$ и неоднородная волна, прошедшая в металл. Комплексный волновой вектор прошедшей волны обозначим через $\mathbf{k}$ (без индекса). Как было показано ранее, из граничных условий получаются соотношения

$$k_{1x} = k'_{1x} = k_x. \tag{72.1}$$

Из них следует, что геометрические законы отражения света от металлов такие же, что и для непоглощающих сред. Различие есть лишь в законах преломления.

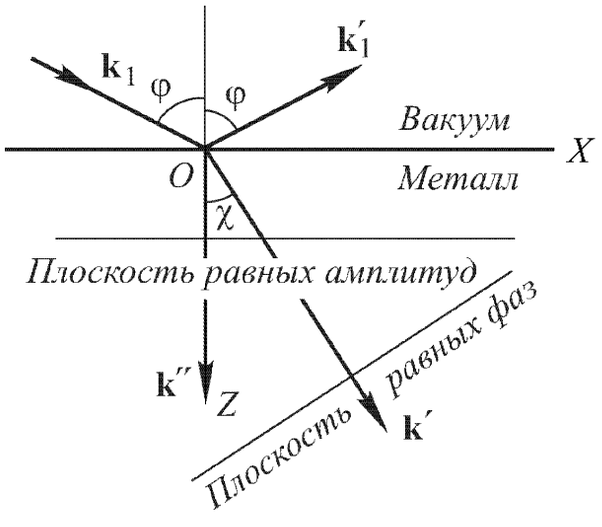


Рис. 255

Прежде всего отметим, что плоскости равных амплитуд прошедшей волны параллельны границе металла. Действительно, представим комплексный вектор $\mathbf{k}$ в виде (71.11). Из (72.1) следует, что тангенциальная составляющая вектора $\mathbf{k}$ вещественна, а потому вектор $\mathbf{k}''$ перпендикулярен к поверхности металла. Это и доказывает наше утверждение. Прошедшая волна затухает в направлении вектора $\mathbf{k}''$. Поэтому вектор $\mathbf{k}''$ надо направить вниз, т.е. в сторону металла, так как затухание волны в металле должно идти в этом, а не в противоположном направлении. В сторону металла должен быть направлен и вектор $\mathbf{k}'$, поскольку угол между векторами $\mathbf{k}'$ и $\mathbf{k}''$ острый, как это следует из второго соотношения (71.12).

Вектор $\mathbf{k}'$ перпендикулярен к плоскости равных фаз прошедшей волны. Угол χ, образуемый этим вектором с положительным направлением оси Z, называется *вещественным углом преломления*. Отношение

$$\frac{\sin\varphi}{\sin\chi} = n_\varphi, \tag{72.2}$$

вообще говоря, зависит от угла падения φ. Оно положительно и называется *показателем преломления*. Так как $k_x = k'_x = k'\sin\chi$, то из соотношения (72.1) следует

$$k' = k_1\frac{\sin\varphi}{\sin\chi} = \frac{\omega}{c}n_\varphi. \tag{72.3}$$

По аналогии с (72.3) введем другую положительную величину $\varkappa_\varphi$, определяемую соотношением

$$k'' = \frac{\omega}{c}\varkappa_\varphi. \tag{72.4}$$

Ее называют *показателем затухания*. Физический смысл показателя затухания легко установить, рассмотрев выражение для поля прошедшей волны. Таким путем не представляет труда установить, что на глубине

$$z = h_\varphi = \frac{\lambda_0}{4\pi\varkappa_\varphi}, \tag{72.5}$$

где λ_0 — длина волны в вакууме, интенсивность света (пропорциональная квадрату амплитуды) убывает в e раз. Величина h_φ называется *глубиной проникновения света в металл*. Таким образом, *показатель затухания $\varkappa_\varphi$ можно определить как отношение длины световой волны в вакууме к умноженной на 4π глубине проникновения света в металл.*

Для металлов показатель затухания в видимой области спектра обычно порядка единицы. Например, для золота при нормальном падении $\varkappa_\varphi = 2{,}82$, $h_\varphi = \lambda_0/(4\pi\varkappa_\varphi) = \lambda_0/35{,}4$. Отсюда находим, что на протяжении длины волны интенсивность света в золоте убывает в $\exp(\lambda_0/h_\varphi) \approx 2{,}4\cdot 10^{15}$ раз.

Пленки металлов с толщиной порядка длины волны, как правило, практически непрозрачны для света. Об оптических свойствах металлов обычно судят *по отраженному свету*. Тем не менее необходимо изучить законы проникновения света в металл, так как без этого нельзя понять и законы отражения. Свет, отраженный от металла (как и от диэлектрика), возникает в результате интерференции когерентных вторичных волн, излучаемых электронами и атомными ядрами металла. Но вторичные волны, очевидно, возбуждаются падающей волной, проникшей в металл. Если бы поле в металл совсем не проникало, то отражение света было бы невозможно.

2. Выразим теперь n_φ и $\varkappa_\varphi$ через оптические константы металла n, $\varkappa$ и угол падения φ. Подставляя выражения (72.3) и (72.4) в формулы (71.12), получим

$$n_\varphi^2 - \varkappa_\varphi^2 = \varepsilon', \quad 2n_\varphi\varkappa_\varphi\cos\chi = \varepsilon''. \tag{72.6}$$

Сравнивая эти выражения с (71.8), находим

$$n_\varphi^2 - \varkappa_\varphi^2 = n^2 - \varkappa^2, \quad n_\varphi\varkappa_\varphi\cos\chi = n\varkappa. \tag{72.7}$$

Кеттелер (1839–1900), впервые получивший эти соотношения, назвал их *главными уравнениями распространения световых волн в металлах и поглощающих средах*. Они показывают, что величины

$$a \equiv n_\varphi^2 - \varkappa_\varphi^2, \quad b \equiv 2n_\varphi\varkappa_\varphi\cos\chi \tag{72.8}$$

не зависят от угла падения φ, а только от рода металла, его физического состояния и от длины световой волны. Величины a и b называются *инвариантами Кеттелера* и могут служить для характеристики оптических свойств металлов (вместо n и $\varkappa$).

Если $\varphi = 0$, то $\chi = 0$, и уравнения (72.7) дают $n_\varphi = \pm n$, $\varkappa_\varphi = \pm\varkappa$. Здесь надо взять знак плюс, так как по определению величины n_φ и $\varkappa_\varphi$ существенно положительны. Таким образом, при нормальном падении показатели преломления и затухания n_φ и $\varkappa_\varphi$ принимают свои главные значения n и $\varkappa$.

Для нахождения n_φ и $\varkappa_\varphi$ как функций угла падения φ перепишем второе уравнение (72.7) в виде

$$n_\varphi^2\varkappa_\varphi^2\cos^2\chi = n_\varphi^2\varkappa_\varphi^2(1-\sin^2\chi) = n_\varphi^2\varkappa_\varphi^2 - \varkappa_\varphi^2\sin^2\varphi = \frac{1}{4}\,b^2.$$

Решая его совместно с первым уравнением, найдем

$$\begin{aligned} n_\varphi^2 &= \frac{1}{2}\left[\pm\sqrt{(a-\sin^2\varphi)^2+b^2} + (a+\sin^2\varphi)\right], \\ \varkappa_\varphi^2 &= \frac{1}{2}\left[\pm\sqrt{(a-\sin^2\varphi)^2+b^2} - (a-\sin^2\varphi)\right]. \end{aligned} \tag{72.9}$$

Знаки перед квадратными корнями должны быть одинаковы в обеих формулах, чтобы разность $n_\varphi^2 - \varkappa_\varphi^2$ была равна a. Кроме

того, n_φ и $\varkappa_\varphi$ должны быть непрерывными функциями угла φ. Для действительно поглощающих сред инвариант b не равен нулю, как это видно из выражения $b = 2n\varkappa = \varepsilon''$. Поэтому квадратный корень в формулах (72.9) не может обращаться в нуль и менять знак при изменении угла φ. Но при $\varphi = 0$ величины n_φ^2 и $\varkappa_\varphi^2$ могут быть существенно положительными тогда и только тогда, когда оба квадратных корня в (72.9) взяты со знаком плюс. Значит, знак плюс следует брать и при любых значениях угла φ. Таким образом, для дейсвительно поглощающих сред

$$\begin{aligned} n_\varphi^2 &= \frac{1}{2}\left[\sqrt{(a-\sin^2\varphi)^2+b^2}+(a+\sin^2\varphi)\right], \\ \varkappa_\varphi^2 &= \frac{1}{2}\left[\sqrt{(a-\sin^2\varphi)^2+b^2}-(a-\sin^2\varphi)\right]. \end{aligned} \tag{72.10}$$

В частности, при $\varphi = 0$

$$n^2 = \tfrac{1}{2}(\sqrt{a^2+b^2}+a), \quad \varkappa^2 = \tfrac{1}{2}(\sqrt{a^2+b^2}-a). \tag{72.11}$$

Особые случаи в выборе знака могут иметь место только при $b \equiv 2n\varkappa = 0$, т.е. либо при $n = 0$, $\varkappa \neq 0$, либо при $n \neq 0$, $\varkappa = 0$. В обоих случаях диэлектрическая проницаемость вещественна, т.е. среда непоглощающая. Однако величины n_φ и $\varkappa_\varphi$ сохраняют смысл и для таких случаев. Например, теорию полного отражения на границе прозрачных сред (см. § 66) можно представить как частный случай теории, изложенной в этом параграфе.

§ 73. Формулы Френеля. Измерение оптических констант металлов

1. Формулы Френеля (65.7) или (65.8) применимы и для металлов, если под $\sin\psi$ и $\cos\psi$ понимать комплексные величины

$$\sin\psi = \frac{\sin\varphi}{\nu}, \quad \cos\psi = \frac{1}{\nu}\sqrt{\nu^2-\sin^2\varphi}. \tag{73.1}$$

Здесь надо взять то значение квадратного корня, которое имеет отрицательную мнимую часть — только тогда неоднородная волна, проникшая в поглощающую среду, будет затухать при удалении от границы раздела.

При нормальном падении

$$\frac{R_\perp}{\mathcal{E}_\perp} = -\frac{R_\parallel}{\mathcal{E}_\parallel} = -\frac{\nu-1}{\nu+1} = -\frac{(n-1)-i\varkappa}{(n+1)-i\varkappa}. \tag{73.2}$$

Для отражательной способности металла получаем

$$R = \left|\frac{R_\perp}{\mathcal{E}_\perp}\right|^2 = \frac{(n-1)^2+\varkappa^2}{(n+1)^2+\varkappa^2}. \tag{73.3}$$

2. При отражении от металла оба отношения $R_\perp/\mathcal{E}_\perp$ и $R_\parallel/\mathcal{E}_\parallel$, вообще говоря, комплексны, т.е. появляются *скачки фаз*. Они, как правило, *различны* для составляющих $\mathbf{E}_\parallel$ и $\mathbf{E}_\perp$. *Если падающий свет поляризован линейно под углом к плоскости падения, то отраженный свет будет поляризован эллиптически.* Исследуя эллиптическую поляризацию отраженного света, можно определить оптические константы металла n и $\varkappa$. На этом основан *метод Друде*, излагаемый ниже.

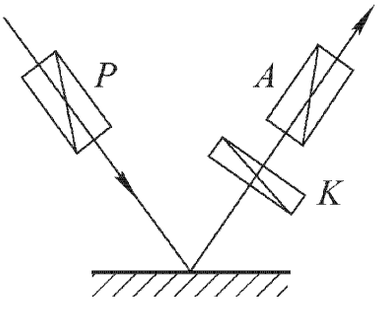


Рис. 256

Световой луч, пройдя через поляризатор P (рис. 256), поляризуется линейно. Для простоты расчета предположим, что азимут поляризации равен 45° ($\mathcal{E}_\perp = \mathcal{E}_\parallel$). Обобщение на случай произвольного азимута не встречает затруднений. Отраженный луч сначала проходит через компенсатор K, а затем через анализатор A. Изменяя установку компенсатора и вращая анализатор вокруг направления отраженного луча, можно погасить отраженный луч. В этом случае после прохождения через компенсатор свет становится поляризованным линейно. Азимут его поляризации называется *азимутом восстановленной линейной поляризации* отраженного света. Компенсатором K можно измерить разность фаз Δ между $\mathbf{E}_\parallel$ и $\mathbf{E}_\perp$ отраженной волны, а анализатором A — азимут β ее восстановленной линейной поляризации. По этим данным можно вычислить оптические константы металла n и $\varkappa$.

Действительно, из формул Френеля (65.8) при $\mathcal{E}_\perp = \mathcal{E}_\parallel$ получаем

$$\frac{R_\parallel}{R_\perp} = -\frac{\cos(\varphi+\psi)}{\cos(\varphi-\psi)}.$$

Очевидно,

$$\frac{R_\parallel}{R_\perp} = re^{i\Delta}, \quad r = \operatorname{tg}\beta.$$

Отсюда

$$\frac{1-re^{i\Delta}}{1+re^{i\Delta}} = \frac{\cos\varphi\cos\psi}{\sin\varphi\sin\psi} = \frac{\sqrt{\nu^2-\sin^2\varphi}}{\operatorname{tg}\varphi\sin\varphi}.$$

Далее,

$$\nu^2-\sin^2\varphi = (n-i\varkappa)^2-\sin^2\varphi = n^2-\varkappa^2-2in\varkappa-\sin^2\varphi.$$

На основании (72.7) и (72.2)

$$\nu^2-\sin^2\varphi = n_\varphi^2-\varkappa_\varphi^2-2in_\varphi\varkappa_\varphi\cos\chi-n_\varphi^2\sin^2\chi = (n_\varphi\cos\chi-i\varkappa_\varphi)^2.$$

на его поверхности. В табл. 7 приведены для ориентировки значения $\varkappa$ и n некоторых металлов, полученные по методу Друде ($\lambda = 589{,}3$ нм). Отражательная способность R вычислена по формуле (73.3).

3. Запишем формулу (73.3) в виде

$$R = \frac{(n^2 + \varkappa^2 + 1) - 2n}{(n^2 + \varkappa^2 + 1) + 2n}. \tag{73.8}$$

Как видно из табл. 7, почти для всех металлов $2n$ мало по сравнению с $n^2+\varkappa^2+1$. Благодаря этому отражательная способность металлов *очень велика* — для многих металлов она близка к единице. «Металлический блеск», присущий металлам, объясняется их большой отражательной способностью. Так как $\varkappa$ и n изменяются в зависимости от длины волны, некоторые металлы — в особенности медь и золото — обладают резко выраженным цветом. Металл кажется нам, например, красным, если он сильнее всего отражает красные лучи. Поэтому, грубо говоря, поверхностная окраска металла является дополнительной к цвету лучей, проходящих через тонкие металлические пленки. Например, очень тонкие золотые пленки в проходящем свете кажутся зелеными.

Таблица 7

Металл	$\varkappa$	n	R, %
Натрий	2,61	0,05	97,5
Серебро, цельное	3,64	0,18	95,1
Магний, цельный	4,42	0,37	93,1
Золото, цельное	2,82	0,37	84,9
Золото, электролитическое	2,83	0,47	81,5
Ртуть	4,41	1,62	75,4
Медь, цельная	2,62	0,64	73,2
Никель, цельный	3,32	1,79	62,0
Никель, электролитический	3,48	2,01	62,0
Никель, гальванически распыленный	1,97	1,30	43,3
Железо, гальванически распыленное	1,63	1,51	32,6

Для слабо поглощающих тел отражательная способность определяется почти исключительно главным показателем преломления n и практически не зависит от $\varkappa$. К таким телам относятся красящие вещества. В этом случае поверхностная окраска тел обусловлена не избирательным отражением, а *избирательным поглощением* в слое красящего вещества. Поэтому окраска тел кажется одинаковой в отраженном и проходящем свете. Слои лаковых красок свободны от неоднородностей, которые сделали бы эти слои мутными. Сквозь слои таких красок

видны детали поверхности тела. Свет от источника попадает на поверхность тела и на ней рассеивается. Таким образом, на пути к глазу он дважды проходит через поглощающий слой краски, претерпевая при этом избирательное поглощение. Свет, отраженный от передней поверхности лака, может дать слабый беловатый оттенок. Если поверхность слоя лака гладкая, то такой беловатый оттенок практически ограничивается областью зеркального отражения. Клеевые краски искусственно делаются мутными путем введения в них различных неоднородностей. Свет, падающий на слой краски, рассеивается на этих неоднородностях, не достигая поверхности тела. Таким образом, значительная доля света рассеивается верхними слоями краски, не претерпевая избирательного поглощения. Благодаря этому появляется заметный беловатый оттенок.

Различие в происхождении поверхностной окраски слабо и сильно поглощающих сред можно иллюстрировать и следующим примером. Покроем поверхность прозрачной стеклянной пластинки слоем фиолетовых чернил. Чернила, пока они не засохли, являются «слабо поглощающей средой». Они кажутся фиолетовыми как в отраженном, так и в проходящем свете: окраска обусловлена избирательным поглощением света, проходящего через слой чернил и испытывающего рассеяние в нем. Когда чернила засохнут, они превращаются в «сильно поглощающую среду». В проходящем свете они по-прежнему кажутся фиолетовыми — окраска обусловлена избирательным поглощением. В отраженном же свете засохший слой чернил приобретает дополнительный желтовато-зеленый металлический блеск, вызванный избирательным отражением.

Зависимость оптических констант многих металлов от длины волны выражена весьма резко. Так, серебро, характеризующееся в видимой области большим коэффициентом отражения (около 95 %), имеет в ультрафиолете резко выраженную область плохого отражения и большой прозрачности — вблизи $\lambda = 316$ нм отражательная способность серебра снижается до 4,2 %, т.е. становится такой же, как у стекла. Вуд показал, что тонкие пленки щелочных металлов прозрачны в ультрафиолетовой части спектра, но совершенно не пропускают видимых лучей. Ему удалось даже обнаружить угол Брюстера при отражении ультрафиолетовых лучей от этих металлов.

Для многих поглощающих тел отражательная способность имеет резко выраженный максимум в некоторой, иногда весьма узкой, области спектра. Путем многократных отражений от таких тел можно получить лучи с довольно высокой степенью монохроматичности. Такие лучи называются *остаточными*. Этот метод применяли Рубенс (1865–1922) и его сотрудники для получения монохроматических инфракрасных лучей.

4. В заключение остановимся на отражательной способности металлов для длинных волн (радиоволны, инфракрасные лучи). В этой области проводимость σ практически не зависит от частоты ω и равна своему статическому значению σ_0. Как показывает формула (71.4), мнимая часть ε'' диэлектрической проницаемости с уменьшением частоты растет. Для длинных волн вещественной частью ε' можно пренебречь, полагая

$$\varepsilon(\omega) = -i\frac{4\pi\sigma_0}{\omega}. \tag{73.9}$$

Тогда из соотношений (71.8) получаем

$$n = \varkappa = \sqrt{\frac{\varepsilon''}{2}} = \sqrt{\frac{2\pi\sigma_0}{\omega}} = \sqrt{\sigma_0 T} = \sqrt{\frac{\lambda\sigma_0}{c}}, \tag{73.10}$$

где T — период колебаний, λ — длина волны в вакууме.

Соотношения (73.10) впервые были получены Друде. Для видимого света они не оправдываются. Так, для меди $\sigma_0 = 5{,}14 \cdot 10^{17}$ с$^{-1}$. Полагая $\lambda = 589{,}3$ нм $= 5{,}893 \cdot 10^{-5}$ см, из формул (73.10) найдем $n = \varkappa \approx 33$, тогда как опыт дает $n = 0{,}64$, $\varkappa = 2{,}62$. Соотношениями (73.10) можно пользоваться только для длинных волн. В этом случае (73.8) переходит в

$$R = \frac{2n^2 - 2n + 1}{2n^2 + 2n + 1} \approx 1 - \frac{2}{n}\left(1 - \frac{1}{n} + \ldots\right),$$

или

$$R = 1 - \frac{2}{\sqrt{\sigma_0 T}}\left(1 - \frac{1}{\sqrt{\sigma_0 T}} + \ldots\right). \tag{73.11}$$

Эта формула проверялась Гагеном и Рубенсом для инфракрасных лучей ($\lambda = 4$; 8; 12; 25,5 мкм). Они нашли, что при $\lambda > 5$ мкм формула (73.11) начинает оправдываться количественно. Для радиоволн величина $1/\sqrt{\sigma_0 T}$ очень близка к нулю, а отражательная способность — к единице: металлы отражают радиоволны практически полностью.

ЗАДАЧА

В каких случаях возможно полное отражение света от среды?

Р е ш е н и е. Если свет поляризован перпендикулярно к плоскости падения, то условие полного отражения будет

$$\left|\frac{R_\perp}{\mathcal{E}_\perp}\right|^2 = \frac{\cos\varphi - \nu\cos\psi}{\cos\varphi + \nu\cos\psi} \cdot \frac{\cos\varphi - \nu^*\cos^*\psi}{\cos\varphi + \nu^*\cos^*\psi} = 1,$$

или $\cos\varphi(\nu\cos\varphi + \nu^*\cos^*\psi) = 0$. Пользуясь формулой (73.4), ему можно придать вид

$$n_\varphi \cos\varphi\cos\chi = 0. \tag{73.12}$$

Такое же условие получится и в случае света, поляризованного в плоскости падения. Полное отражение может наступить в одном из трех случаев:

1) $\cos\varphi = 0$; 2) $n_\varphi = 0$; 3) $\cos\chi = 0$. Первый случай соответствует скользящему падению. Во втором случае $n_\varphi = 0$, а потому, ввиду (72.7), $n\varkappa = 0$. В третьем случае $\cos\chi = 0$ и, следовательно, $n\varkappa$ также равно нулю. Итак, если $\varphi \neq \pi/2$, то для полного отражения необходимо, чтобы $n\varkappa = 0$. Это условие может осуществиться в двух случаях: 1) $n = 0$, $\varkappa \neq 0$; 2) $n \neq 0$, $\varkappa = 0$.

С л у ч а й 1. $n = 0$, $\varkappa \neq 0$. Отражение будет полным при любом угле падения. Действительно, в рассматриваемом случае $\nu = n - i\varkappa = -i\varkappa$, и формулы Френеля принимают вид

$$\frac{R_\perp}{\mathcal{E}_\perp} = \frac{\cos\varphi + i\varkappa\cos\psi}{\cos\varphi - i\varkappa\cos\psi}, \quad \frac{R_\parallel}{\mathcal{E}_\parallel} = -\frac{i\varkappa\cos\varphi + \cos\psi}{-i\varkappa\cos\varphi + \cos\psi}.$$

В них

$$\cos\psi = \sqrt{1 - \sin^2\psi} = \sqrt{1 - \frac{\sin^2\varphi}{\nu^2}} = \sqrt{1 + \left(\frac{\sin\varphi}{\varkappa}\right)^2},$$

откуда видно, что $\cos\psi$ — величина всегда вещественная. В знаменателях формул Френеля стоят величины, комплексно-сопряженные с числителями. Следовательно, $|R_\perp/\mathcal{E}_\perp|^2 = |R_\parallel/\mathcal{E}_\parallel|^2 = 1$. В рассматриваемом случае $\varepsilon = \nu^2 = -\varkappa^2 < 0$, т.е. среда непоглощающая с отрицательной диэлектрической проницаемостью. Такой случай осуществляется при отражении радиоволн от плазмы (см. § 87).

С л у ч а й 2. $n \neq 0$, $\varkappa = 0$. В этом случае $\nu = n$, $\varepsilon = n^2 > 0$, т.е. среда прозрачна. Полное отражение будет иметь место при $n < 1$, если φ превосходит предельный угол $\varphi_0 = \arcsin n$ (см. § 66).

§ 74. Аномальный скин-эффект и эффективная диэлектрическая проницаемость

1. Проникновение электромагнитной волны в тонкий поверхностный слой металла есть частный случай *скин-эффекта*, рассмотренного нами в т. III, § 144. Самый слой, в который проникает электромагнитное поле, называется *скин-слоем*. Толщина скин-слоя определяется формулой (72.5). Она выводится на основе макроскопических уравнений Максвелла (71.5). Из тех же уравнений следует, что напряженность поля в скин-слое убывает *экспоненциально*. Такой скин-слой называется *нормальным*.

Для применимости макроскопических уравнений Максвелла необходимо, чтобы *межатомные расстояния были малы по сравнению не только с длиной волны, но и с толщиной скин-слоя*. Это условие можно считать выполненным для всех металлов. Более жестким является условие применимости понятия диэлектрической проницаемости $\varepsilon(\omega)$, как оно было введено в § 71. Там было учтено, что электроны и ионы, движением которых создаются токи проводимости и поляризации, движутся в электрическом поле, которое меняется во времени, но не

было принято во внимание его изменение в пространстве. Это допустимо, когда *средняя длина свободного пробега электронов мала по сравнению с расстояниями, на которых заметно меняется напряженность электрического поля, т.е. по сравнению с длиной волны и толщиной скин-слоя.* Только тогда электрон от столкновения до столкновения движется практически в однородном поле. Если же средняя длина свободного пробега электрона порядка или больше толщины скин-слоя или длины волны, то результаты § 71 требуют пересмотра. Понятие диэлектрической проницаемости $\varepsilon(\omega)$ может потерять смысл. Тогда напряженность поля и ток будут убывать в глубь металла *не экспоненциально*, а по более сложному закону. Соответствующий скин-эффект называется *аномальным*.

Если воспользоваться значениями $\varkappa$ из табл. 7, то легко убедиться, что у всех металлов, приведенных в этой таблице, величина $h = \lambda/(4\pi\varkappa)$ для видимого света порядка 10^{-6} см. Того же порядка при комнатных температурах и средняя длина свободного пробега электронов. Это указывает на аномальный характер скин-эффекта. Только для плохих проводников, у которых длина свободного пробега меньше, скин-эффект при комнатных температурах нормальный. В области же низких температур, где средняя длина свободного пробега достигает для очень чистых образцов значений порядка $10^{-2} - 10^{-3}$ см, об использовании теории нормального скин-эффекта, основанной на понятии диэлектрической проницаемости, не может быть и речи. Поскольку характер скин-эффекта определяется соотношением между средней длиной свободного пробега электрона и толщиной скин-слоя, полная теория аномального скин-эффекта должна строиться на основе *кинетики электронов*, в ее квантовой форме.

2. Однако независимо от того, является ли скин-эффект нормальным или аномальным, отражение света возникает в результате излучения электромагнитных волн токами, текущими в поверхностном слое металла и возбуждаемыми падающей волной. Механизм отражения света от металлов вполне аналогичен соответствующему механизму для диэлектриков, разобранному в § 68 и 69.

В случае нормального скин-эффекта плотность полного тока убывает в глубь металла по экспоненциальному закону. В случае аномального скин-эффекта это не так. Однако если толщина скин-слоя много меньше длины волны, конкретный закон изменения плотности тока в поверхностном слое может лишь слабо сказаться на отражении света, так как в этом случае фазы источников вторичных волн, распределенных в поверхностном слое, практически одинаковы по всей его толщине. Поэтому при вычислении поля отраженной волны действительное распреде-

ление полного тока в скин-слое может быть заменено распределением, в котором плотность тока убывает экспоненциально. Такая замена эквивалентна введению вместо ε *эффективной диэлектрической проницаемости металла* $\varepsilon_{эф}$.

Пользуясь величиной $\varepsilon_{эф}$, можно вычислять поле отраженной волны так, как если бы скин-эффект был нормальным, а металл имел диэлектрическую проницаемость $\varepsilon = \varepsilon_{эф}$. Однако если скин-эффект аномальный, пользоваться $\varepsilon_{эф}$ для вычисления поля внутри металла нельзя: термин «эффективная» понимается в смысле: *эффективная в отношении отражения.* Эффективная диэлектрическая проницаемость, очевидно, может быть введена и в случае более толстых скин-слоев. Однако в этих случаях она, вообще говоря, зависит от состояния поляризации падающей волны и от угла падения. Для теоретического вычисления $\varepsilon_{эф}$ надо было бы решить задачу об отражении света методами кинетики и статистики, что далеко выходит за рамки этой книги.

ГЛАВА VII

КРИСТАЛЛООПТИКА

§ 75. Плоские волны в кристаллах

1. Большинство кристаллов *оптически анизотропны*, т.е. их оптические свойства в разных направлениях не одинаковы. Наиболее важным проявлением этой анизотропии является *двойное лучепреломление в кристаллах*. Изучение этого явления и связанной с ним поляризации света составит основное содержание настоящей главы.

Фундаментальные уравнения Максвелла справедливы без всяких изменений и в кристаллических средах. В отсутствие электрических зарядов и токов они имеют вид

$$\operatorname{rot}\mathbf{H}=\frac{1}{c}\dot{\mathbf{D}},\quad \operatorname{rot}\mathbf{E}=-\frac{1}{c}\dot{\mathbf{H}}. \tag{75.1}$$

Но материальные уравнения усложняются. Изотропные среды характеризуются скалярной диэлектрической проницаемостью $\varepsilon(\omega)$. Для характеристики оптических свойств кристаллов требуются девять величин $\varepsilon_{ik}(\omega)$, образующих *тензор диэлектрической проницаемости*, или *диэлектрический тензор*. Он вводится посредством соотношений

$$D_j=\sum_k \varepsilon_{jk}E_k \qquad (j,k=x,y,z). \tag{75.2}$$

Для прозрачных кристаллов, как можно показать, исходя из закона сохранения энергии, диэлектрический тензор *симметричен*, т.е. $\varepsilon_{ij}=\varepsilon_{ji}$ (см. § 80). Разумеется, в различных системах координат компоненты диэлектрического тензора имеют разные значения. При переходе от одной системы координат к другой они преобразуются как компоненты всякого тензора. Благодаря тензорной связи между $\mathbf{D}$ и $\mathbf{E}$ направления этих векторов в кристаллах, вообще говоря, не совпадают.

2. Если среды прозрачны и однородны, то в них могут распространяться плоские монохроматические волны. Каждую из них можно записать в виде

$$\mathbf{E} = \mathbf{E}_0 e^{i(\omega t - \mathbf{kr})}, \quad \mathbf{H} = \mathbf{H}_0 e^{i(\omega t - \mathbf{kr})}, \quad \mathbf{D} = \mathbf{D}_0 e^{i(\omega t - \mathbf{kr})}. \tag{75.3}$$

Рассмотрим сначала свойства таких волн, которые вытекают из одних только фундаментальных уравнений (75.1) без использования материальных уравнений (75.2). Как и в случае изотропных сред,

$$\operatorname{rot} \mathbf{H} = -i\,[\mathbf{kH}], \quad \dot{\mathbf{D}} = -i\omega \mathbf{D}, \ldots \tag{75.4}$$

Подставив эти выражения в (75.1) и вводя единичный вектор волновой нормали $\mathbf{N}$ по формуле $\mathbf{k} = \frac{\omega}{v}\mathbf{N}$, получим

$$\mathbf{D} = -\frac{c}{v}\,[\mathbf{NH}], \quad \mathbf{H} = \frac{c}{v}\,[\mathbf{NE}], \tag{75.5}$$

где v — нормальная скорость волны, т.е. фазовая скорость, с которой распространяется волновой фронт в направлении волновой нормали $\mathbf{N}$. Присоединим к этим формулам еще выражение для вектора Пойнтинга

$$\mathbf{S} = \frac{c}{4\pi}\,[\mathbf{EH}]. \tag{75.6}$$

Фундаментальное значение этого вектора в кристаллооптике состоит в том, что он определяет *направление световых лучей*, т.е. линий, вдоль которых происходит распространение энергии света. Чтобы убедиться в этом, достаточно показать, что направление $\mathbf{S}$ совпадает с направлением вектора групповой скорости $\mathbf{u} = d\omega/d\mathbf{k}$ (см. § 8). Это будет сделано в § 81. *В кристаллах векторы* $\mathbf{S}$ *и* $\mathbf{N}$, *вообще говоря, не совпадают по направлению.* Именно с этим связано двойное лучепреломление, а также коническая рефракция.

Из формул (75.5) видно, что *векторы* $\mathbf{D}$ *и* $\mathbf{H}$ *взаимно перпендикулярны.* Кроме того, *они перпендикулярны к волновой нормали* $\mathbf{N}$, т.е. параллельны фронту волны. Значит, *плоские волны в кристалле поперечны в отношении векторов* $\mathbf{D}$ *и* $\mathbf{H}$. Однако *в общем случае они не поперечны в отношении вектора* $\mathbf{E}$. *Четыре вектора* $\mathbf{E}$, $\mathbf{D}$, $\mathbf{N}$, $\mathbf{S}$ *лежат в одной плоскости, перпендикулярной к вектору* $\mathbf{H}$. Взаимное расположение этих векторов показано на рис. 257. Из него видно, что *заданием в кристалле направления вектора* $\mathbf{E}$ (*или* $\mathbf{D}$) *однозначно определяется направление вектора* $\mathbf{D}$ (*или* $\mathbf{E}$) *и с точностью до* 180° *— направления всех остальных векторов* $\mathbf{H}$, $\mathbf{N}$, $\mathbf{S}$. *Опре-*

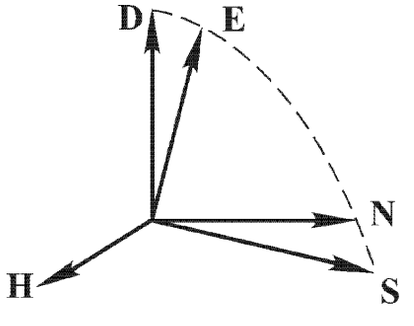


Рис. 257

деляется также величина нормальной скорости волны v. Действительно, если задано направление вектора $\mathbf{E}$, то уравнение (75.2) однозначно определит направление вектора $\mathbf{D}$, а с ним и плоскость $(\mathbf{E},\mathbf{D})$, к которой перпендикулярен вектор $\mathbf{H}$. Тем самым с точностью до 180° определится направление $\mathbf{H}$, а следовательно, и направления векторов $\mathbf{N}$ и $\mathbf{S}$. Исключение составляет случай, когда направления векторов $\mathbf{E}$ и $\mathbf{D}$ совпадают. Тогда всякая прямая, перпендикулярная к вектору $\mathbf{E}$, может служить направлением магнитного поля. Те же рассуждения применимы и к случаю, когда задано направление вектора $\mathbf{D}$. Для определения v исключим из (75.5) вектор $\mathbf{H}$:

$$\mathbf{D} = \frac{c^2}{v^2}\mathbf{E} - \frac{c^2}{v^2}(\mathbf{NE})\mathbf{N}. \tag{75.7}$$

Так как $(\mathbf{DN}) = 0$, то отсюда скалярным умножением на $\mathbf{D}$ находим

$$v^2 = c^2\frac{(\mathbf{DE})}{D^2}. \tag{75.8}$$

Таким образом, электрический вектор ($\mathbf{E}$ или $\mathbf{D}$) в кристалле в известном смысле является *главным*. Это и понятно, так как именно он определяет электрическую поляризацию среды, а возбуждение последней составляет сущность процесса распространения электромагнитных волн в материальных средах.

§ 76. Оптически одноосные кристаллы

1. Простейшими оптическими свойствами обладают *оптически одноосные кристаллы*, которые к тому же имеют наибольшее практическое значение. Поэтому имеет смысл особо выделить этот простейший частный случай. Оптически одноосными называются кристаллы, свойства которых обладают симметрией вращения относительно некоторого направления, называемого *оптической осью кристалла*.

Разложим электрические векторы $\mathbf{E}$ и $\mathbf{D}$ на составляющие $\mathbf{E}_{\|}$, $\mathbf{D}_{\|}$ вдоль оптической оси и составляющие $\mathbf{E}_{\perp}$, $\mathbf{D}_{\perp}$, перпендикулярные к ней. Тогда

$$\mathbf{D}_{\|} = \varepsilon_{\|}\mathbf{E}_{\|}, \quad \mathbf{D}_{\perp} = \varepsilon_{\perp}\mathbf{E}_{\perp}, \tag{76.1}$$

где $\varepsilon_{\|}$ и $\varepsilon_{\perp}$ — постоянные, называемые *продольной* и *поперечной диэлектрическими проницаемостями кристалла*. К этим двум величинам и сводится диэлектрический тензор одноосного кристалла. К оптически одноосным кристаллам относятся все кристаллы тетрагональной, гексагональной и ромбоэдрической систем. Кристаллы кубической системы являются вырожденным случаем их. Для них $\varepsilon_{\|} = \varepsilon_{\perp}$. Поэтому *кристаллы кубической*

системы в оптическом отношении ведут себя как оптически изотропные тела. Плоскость, в которой лежат оптическая ось кристалла и нормаль $\mathbf{N}$ к фронту волны, называется *главным сечением кристалла.* Главное сечение — это не какая-то определенная плоскость, а целое семейство параллельных плоскостей.

Рассмотрим теперь два частных случая.

С л у ч а й 1. Вектор $\mathbf{D}$ перпендикулярен к главному сечению кристалла. В этом случае $\mathbf{D} \equiv \mathbf{D}_\perp$, а потому $\mathbf{D} = \varepsilon_\perp \mathbf{E}$. Кристалл ведет себя как изотропная среда с диэлектрической проницаемостью $\varepsilon_\perp$. Для нее

$$\mathbf{D} = \varepsilon_\perp \mathbf{E}, \tag{76.2}$$

и из уравнений (75.5) получаем

$$D = \frac{c}{v} H, \quad H = \frac{c}{v} E, \tag{76.3}$$

или

$$\varepsilon_\perp E = \frac{c}{v} H, \quad H = \frac{c}{v} E, \tag{76.4}$$

откуда

$$v = v_\perp \equiv v_o = \frac{c}{\sqrt{\varepsilon_\perp}}. \tag{76.5}$$

Таким образом, *если электрический вектор перпендикулярен к главному сечению, то скорость волны не зависит от направления ее распространения.* Такая волна называется *обыкновенной.*

С л у ч а й 2. Вектор $\mathbf{D}$ лежит в главном сечении. Так как вектор $\mathbf{E}$ лежит также в главном сечении (см. рис. 257), то $\mathbf{E} = \mathbf{E}_N + \mathbf{E}_D$, где $\mathbf{E}_N$ — составляющая этого вектора вдоль $\mathbf{N}$, а $\mathbf{E}_D$ — вдоль $\mathbf{D}$. Из векторного произведения $[\mathbf{NE}]$ составляющая $\mathbf{E}_N$ выпадает. Поэтому вторую формулу (75.5) можно записать в виде

$$\mathbf{H} = \frac{c}{v} [\mathbf{N}\mathbf{E}_D].$$

Очевидно,

$$E_D = \frac{\mathbf{ED}}{D} = \frac{E_\parallel D_\parallel + E_\perp D_\perp}{D} = \frac{(D_\parallel^2/\varepsilon_\parallel) + (D_\perp^2/\varepsilon_\perp)}{D},$$

или

$$E_D = D\left(\frac{\sin^2\alpha}{\varepsilon_\parallel} + \frac{\cos^2\alpha}{\varepsilon_\perp}\right) = D\left(\frac{N_\perp^2}{\varepsilon_\parallel} + \frac{N_\parallel^2}{\varepsilon_\perp}\right),$$

где α — угол между оптической осью и волновой нормалью (рис. 258). Если ввести обозначение

$$\frac{1}{\varepsilon} = \frac{N_\perp^2}{\varepsilon_\parallel} + \frac{N_\parallel^2}{\varepsilon_\perp}, \tag{76.6}$$

то получится $D = \varepsilon E_D$, и мы придем к соотношениям

$$\varepsilon E_D = \frac{c}{v} H, \quad H = \frac{c}{v} E_D, \tag{76.7}$$

формально тождественным с соотношениями (76.4). Роль величины $\varepsilon_\perp$ теперь играет величина ε, определяемая выражением (76.6). Поэтому нормальная скорость волны будет определяться выражением

$$v = \frac{c}{\sqrt{\varepsilon}} = c\sqrt{\frac{N_\perp^2}{\varepsilon_\parallel} + \frac{N_\parallel^2}{\varepsilon_\perp}} \equiv v_\parallel. \tag{76.8}$$

Она меняется с изменением направления волновой нормали **N**. По этой причине волну, электрический вектор которой лежит в главном сечении кристалла, называют *необыкновенной. Зависимость ее скорости* $v_\parallel$ *от направления распространения обусловлена тем, что с изменением направления волновой нормали* **N** *меняется угол между электрическим вектором и оптической осью кристалла.* Когда $N_\perp = 0$, т.е. необыкновенная волна распространяется вдоль оптической оси кристалла, то из формулы (76.8) получаем $v = c/\sqrt{\varepsilon_\perp} = v_\perp \equiv v_o$. В этом случае, как это очевидно заранее, нет разницы между обыкновенной и необыкновенной волнами. Если же $N_\parallel = 0$, т.е. необыкновенная волна распространяется перпендикулярно к оптической оси, то скорость волны будет равна

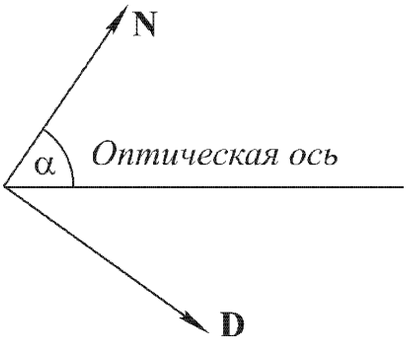


Рис. 258

$$v = v_e \equiv c/\sqrt{\varepsilon_\parallel}. \tag{76.9}$$

Термин «оптическая ось» был введен для обозначения такой прямой, вдоль которой обе волны в кристалле распространяются с одинаковыми скоростями. Таких прямых в общем случае две (см. § 80 и 81). Поэтому в этом случае кристалл называется *оптически двуосным.* В рассматриваемом нами частном случае оптические оси совпадают между собой, сливаясь в одну прямую, а потому кристалл и называется *оптически одноосным.*

2. Так как уравнения Максвелла в кристаллах линейны и однородны, то *в общем случае волна, вступающая в кристалл из изотропной среды, разделяется внутри кристалла на две линейно поляризованные волны: обыкновенную, вектор электрической индукции которой перпендикулярен к главному сечению, и необыкновенную с вектором электрической индукции, лежащим в главном сечении.* Эти волны распространяются в кристалле в различных направлениях и с различными скоростями $v_\perp$ и $v_\parallel$. В направлении оптической оси скорости обеих волн сов-

падают, так что в этом направлении может распространяться волна любой поляризации.

К обеим волнам применимы все рассуждения, которыми мы пользовались при выводе геометрических законов отражения и преломления (см. § 64). Но в кристаллах они относятся *к волновым нормалям, а не к световым лучам.* Волновые нормали отраженной и обеих преломленных волн лежат в плоскости падения. Их направления формально подчиняются закону Снеллиуса

$$\frac{\sin\varphi}{\sin\psi_\perp} = n_\perp, \quad \frac{\sin\varphi}{\sin\psi_\parallel} = n_\parallel, \tag{76.10}$$

где $n_\perp$ и $n_\parallel$ — *показатели преломления обыкновенной и необыкновенной волн*, т.е.

$$n_\perp = \frac{c}{v_\perp} = n_o, \quad n_\parallel = \frac{c}{v_\parallel} = \left(\frac{N_\perp^2}{\varepsilon_\parallel} + \frac{N_\parallel^2}{\varepsilon_\perp}\right)^{-1/2}. \tag{76.11}$$

Из них $n_\perp \equiv n_o$ не зависит, а $n_\parallel$ зависит от угла падения. Постоянная n_o называется *обыкновенным показателем преломления кристалла.* Когда необыкновенная волна распространяется перпендикулярно к оптической оси ($N_\perp = 1$, $N_\parallel = 0$),

$$n_\parallel = \sqrt{\varepsilon_\parallel} \equiv n_e. \tag{76.12}$$

Величину n_e называют *необыкновенным показателем преломления кристалла.* Ее нельзя смешивать с показателем преломления $n_\parallel$ необыкновенной волны. Величина n_e есть *постоянная*, а $n_\parallel$ — *функция направления распространения волны.* Обе величины совпадают только тогда, когда волна распространяется перпендикулярно к оптической оси.

3. Теперь легко понять происхождение *двойного лучепреломления.* Допустим, что плоская волна падает на плоскопараллельную пластинку из одноосного кристалла. При преломлении на первой поверхности пластинки волна внутри кристалла разделится на обыкновенную и необыкновенную. Эти волны поляризованы во взаимно перпендикулярных плоскостях и распространяются внутри пластинки *в разных направлениях и с разными скоростями. Волновые нормали обеих волн лежат в плоскости падения. Обыкновенный луч*, поскольку его направление совпадает с направлением волновой нормали, *также лежит в плоскости падения. Но необыкновенный луч, вообще говоря, выходит из этой плоскости.* (В случае двуосных кристаллов деление на обыкновенную и необыкновенную волны теряет смысл — внутри кристалла обе волны «необыкновенные». При преломлении волновые нормали обеих волн, конечно, остаются в плоскости падения, однако оба луча, вообще говоря, выходят из нее.)

Если падающая волна ограничена диафрагмой, то в пластинке получатся два пучка света, которые при достаточной толщине пластинки окажутся *разделенными пространственно*. При преломлении на второй границе пластинки из нее выйдут *два пучка света*, параллельные падающему лучу. Они будут линейно поляризованы во взаимно перпендикулярных плоскостях. Если падающий свет естественный, то всегда выйдут два пучка. Если же падающий свет линейно поляризован в плоскости главного сечения или перпендикулярно к ней, то двойного преломления не получится — из пластинки выйдет *только один пучок* с сохранением исходной поляризации.

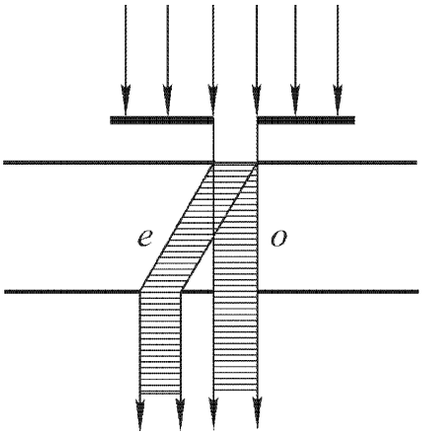


Рис. 259

Двойное преломление возникает и при нормальном падении света на пластинку. В этом случае преломление испытывает необыкновенный луч, хотя волновые нормали и волновые фронты не преломляются. Примером может служить кристаллическая пластинка, вырезанная параллельно оптической оси кристалла (рис. 259). Параллельный пучок света, ограниченный диафрагмой, падает нормально к поверхности пластинки. Волновые фронты, обозначенные на рисунке тонкими горизонтальными линиями, всюду параллельны той же поверхности. Обыкновенный пучок лучей *o* преломления не испытывает. Необыкновенный *e* в пластинке отклоняется вбок, но по выходе из нее снова идет в первоначальном направлении.

4. Двойное лучепреломление было открыто в 1669 г. Бартолинусом (1625–1698) на кристаллах исландского (известкового) шпата. Гюйгенс дал объяснение этого явления, введя гипотезу, что элементарная (вторичная) волна в кристалле распадается на две волны: сферическую (обыкновенную) и эллипсоидальную (необыкновенную). Гюйгенс же открыл и поляризацию света при двойном лучепреломлении.

Исландский шпат есть разновидность углекислого кальция ($CaCO_3$). Он встречается в природе в виде довольно больших и оптически чистых кристаллов. Его обыкновенный показатель преломления $n_o = 1{,}6585$, необыкновенный $n_e = 1{,}4863$ (для желтой линии). Благодаря большому различию n_o и n_e двойное преломление в исландском шпате выражено очень отчетливо. И до сих пор кристаллы исландского шпата наиболее удобны для демонстрации двойного лучепреломления и являются наилучшим материалом для изготовления поляризационных призм и других поляризационных приборов, хотя теперь известно мно-

го естественных и искусственных кристаллов с аналогичными свойствами.

Кристаллы исландского шпата принадлежат к гексагональной системе, но встречаются в различных формах. Каждый кристалл раскалыванием легко привести к форме ромбоэдра, ограниченного шестью подобными параллелограммами с углами $78°08'$ и $101°52'$ (рис. 260). В двух противоположных вершинах A и B сходятся стороны трех тупых углов, в остальных — стороны одного тупого и двух острых. Прямая, проходящая через точку A или B и одинаково наклоненная к ребрам, сходящимся в этих точках, а также всякая прямая, ей параллельная, есть оптическая ось кристалла. Если отшлифовать кристалл так, чтобы все ребра его имели одинаковую длину, то линия AB и будет оптической осью.

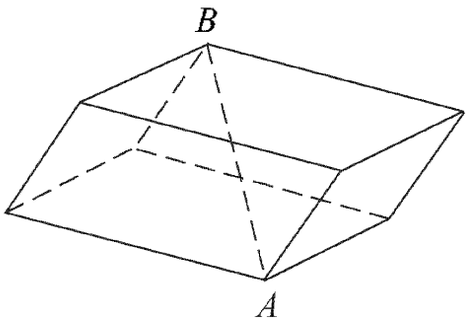


Рис. 260

5. Для демонстрации двойного преломления берут ромбоэдр исландского шпата, вставленный в круглую вращающуюся оправу, устанавливаемую на подставке оптической скамьи. Луч света должен проходить перпендикулярно к противоположным граням ромбоэдра. Перед конденсором проекционного фонаря на той же оптической скамье устанавливают ирисовую диафрагму. В отсутствие исландского шпата длиннофокусный объектив фонаря дает изображение отверстия диафрагмы на удаленном экране. При введении непосредственно за диафрагмой исландского шпата изображение раздваивается. Уменьшая диаметр диафрагмы, можно добиться, чтобы оба изображения не накладывались друг на друга. При освещении естественным светом изображения получаются одинаково яркими. Если вращать исландский шпат вокруг главной оптической оси установки, то при правильной юстировке одно изображение остается неподвижным, а другое движется вокруг него по кругу. Когда кристалл исландского шпата делает полоборота, второе изображение совершает полный оборот. С помощью поляроида, поставленного за объективом, легко убедиться, что оба пучка света линейно поляризованы во взаимно перпендикулярных плоскостях. Неподвижное изображение дает обыкновенный, а вращающееся — необыкновенный пучок света. Увеличив диафрагму, можно частично наложить одно изображение на другое. При вращении поляроида неперекрывающиеся части изображения периодически становятся светлыми и темными: когда максимальна освещенность одной части, освещенность другой обращается в нуль. Освещенность же перекрывающейся части при вращении поля-

роида все время остается неизменной — эта часть освещается неполяризованным светом.

ЗАДАЧИ

1. Если пластинку исландского шпата положить на страницу печатного текста, то происходит раздвоение букв. Будет ли происходить удвоение изображения, если через ту же пластинку смотреть на удаленный предмет?

О т в е т. Не будет.

2. Показатели преломления кристалла можно измерить с помощью кристалл-рефрактометра, действующего по следующему принципу. Пластинка исследуемого кристалла кладется на плоскую поверхность стеклянного полушария с очень высоким (до 2) показателем преломления N. Свет падает со стороны стеклянного полушария вдоль его радиуса и отражается от пластинки. Показатель преломления n исследуемого вещества вычисляется по предельному углу полного отражения по формуле $n = N \sin \varphi$. В случае отражения от кристалла существуют два предельных угла, соответствующих двум преломленным лучам в кристалле.

Как надо ориентировать на кристалл-рефрактометре плоскопараллельную пластинку, вырезанную произвольным образом из одноосного кристалла, чтобы получить оба главных показателя преломления кристалла n_o и n_e?

Р е ш е н и е. Для определения обыкновенного показателя преломления n_o пластинка, очевидно, может быть ориентирована как угодно. Чтобы получить необыкновенный показатель преломления n_e (т.е. максимальное или минимальное значение $n_\parallel$), пластинку надо ориентировать так, чтобы плоскость, проходящая через оптическую ось кристалла и нормаль к границе раздела пластинки со стеклом кристалл-рефрактометра, была перпендикулярна к плоскости падения.

Действительно, при полном отражении световое поле проникает во «вторую» среду, вообще говоря, в виде неоднородной (поверхностной) волны. Но если свет падает строго под предельным углом полного отражения, то волна во второй среде будет еще однородна. Ее волновая нормаль параллельна линии пересечения плоскости падения с плоскостью раздела сред. Повернем кристалл так, чтобы его оптическая ось стала перпендикулярна к этой линии. Тогда волна в кристалле будет распространяться перпендикулярно к оптической оси. При такой ориентировке кристалла электрический вектор необыкновенной волны будет параллелен оптической оси. Значит, на кристалл-рефрактометре в этом случае будет измерен главный показатель преломления необыкновенной волны, т.е. величина n_e.

3. При измерении показателей преломления кристаллической пластинки на кристалл-рефрактометре оказалось, что один показатель преломления сохраняет постоянное значение n_o при всех поворотах стеклянного полушария. Другой показатель преломления изменяется так, что значение n_o для него 1) максимально; 2) минимально. Определить оптический знак кристалла [1]) и ориентировку пластинки относительно оптической оси.

О т в е т. Пластинка вырезана параллельно оптической оси. 1) Кристалл отрицателен. 2) Кристалл положителен.

---

[1]) Одноосный кристалл называется *положительным*, если обыкновенный показатель преломления n_o меньше необыкновенного n_e ($n_o < n_e$). В противоположном случае ($n_o > n_e$) кристалл называется *отрицательным*.

4. Как будут меняться при вращении полушария кристалл-рефрактометра оба показателя преломления пластинки, если она вырезана перпендикулярно к оптической оси?

О т в е т. Оба показателя преломления будут оставаться постоянными и равными соответственно n_o и n_e.

5. Как надо вырезать призму из одноосного кристалла, чтобы методом наименьшего отклонения преломленного луча измерить обыкновенный n_o и необыкновенный n_e показатели преломления кристалла?

О т в е т. Оптическая ось должна быть параллельна преломляющему ребру призмы.

6. Узкий параллельный пучок неполяризованного света падает нормально на пластинку исландского шпата, вырезанную не перпендикулярно к оптической оси. После этого пучок падает на вторую такую же пластинку, параллельную первой, а затем попадает на экран. Главные сечения пластинок образуют между собой угол 30°. Какая картина будет наблюдаться на экране?

О т в е т. На экране образуются четыре пятна, интенсивности которых относятся как $1:3:1:3$. Два из них с интенсивностями $1:3$ поляризованы линейно, два других с такими же интенсивностями поляризованы также линейно, но в перпендикулярной плоскости.

7. Ветровое стекло и фары автомашины прикрыты поляроидами, главные сечения которых параллельны между собой и составляют угол 45° с горизонтом. У всех машин они должны быть повернуты в одну и ту же сторону (считая по ходу машины). Показать, что при таком устройстве водитель машины будет видеть дорогу, освещенную светом его фар, но ему не будет мешать слепящее действие фар встречных машин.

8. Определить максимальный угол $\delta_{\text{макс}}$ между направлениями луча и волновой нормали в кристалле исландского шпата.

Р е ш е н и е. Векторы $\mathbf{N}$, $\mathbf{S}$, $\mathbf{D}$, $\mathbf{E}$ в необыкновенной волне лежат в одной плоскости, причем угол δ между $\mathbf{N}$ и $\mathbf{S}$ равен углу между $\mathbf{D}$ и $\mathbf{E}$. Обозначим через β угол между $\mathbf{D}$ и оптической осью кристалла, а через γ — между $\mathbf{E}$ и той же осью. Тогда

$$D_\parallel = D\cos\beta = \varepsilon_\parallel E_\parallel = \varepsilon_\parallel E\cos\gamma,$$
$$D_\perp = D\sin\beta = \varepsilon_\perp E_\perp = \varepsilon_\perp E\sin\gamma.$$

Отсюда $\operatorname{tg}\beta = (\varepsilon_\perp/\varepsilon_\parallel)\operatorname{tg}\gamma = (n_o/n_e)^2\operatorname{tg}\gamma$. Искомый угол $\delta = \beta - \gamma$. Поэтому

$$\operatorname{tg}\delta = \operatorname{tg}(\beta-\gamma) = \frac{n_o^2 - n_e^2}{n_o^2 + n_e^2\operatorname{tg}^2\beta}. \tag{76.13}$$

Максимум достигается, когда $\operatorname{tg}\beta = n_o/n_e = 1{,}1157$, т.е. при $\beta = 48^\circ 7' 53''$. Максимальный угол $\delta_{\text{макс}}$ определяется формулой

$$\operatorname{tg}\delta_{\text{макс}} = \frac{n_o^2 - n_e^2}{2n_o n_e} = 0{,}1097. \tag{76.14}$$

Из нее находим $\delta_{\text{макс}} = 6^\circ 15' 46''$.

9. Определить показатель преломления $n_\parallel$ необыкновенной волны в ромбоэдре исландского шпата, если волновая нормаль параллельна боковому ребру кристалла. Для исландского шпата угол между боковым ребром и оптической осью равен $\alpha = 64^\circ$. Найти для этого случая угол δ между направлениями луча и волновой нормали в кристалле.

О т в е т. $n_\parallel = \sqrt{(\sin\alpha/n_e)^2 + (\cos\alpha/n_o)^2} = 1{,}515$; $\delta = 4^\circ 36' 20''$.

§ 77. Поляризационные устройства

1. Двупреломляющие кристаллы (лучше всего исландского шпата) можно использовать для получения поляризованного света. Для этого в опыте типа рис. 259 достаточно задержать один из двух линейно поляризованных пучков света, вышедших из кристалла. Однако гораздо удобнее пользоваться не простыми кристаллами, а их комбинациями, называемыми *поляризационными призмами*. Обычно для изготовления поляризационных призм применяется исландский шпат, иногда кварц и натронная селитра, что значительно удешевляет стоимость поляризационных приборов.

Поляризационная призма состоит из двух или более трехгранных призм из одноосного кристалла с одинаковой или различной ориентацией оптических осей, склеенных между собой прозрачными веществами или разделенных воздушной прослойкой. Для склейки применяются канадский бальзам ($n_D = 1{,}550$), льняное масло ($n_D = 1{,}49$), глицерин ($n_D = 1{,}474$) и другие материалы. Для работы в ультрафиолетовой части спектра применяют призмы, склеенные глицерином, касторовым маслом, а также призмы с воздушной прослойкой.

Различают *однолучевые поляризационные призмы*, из которых выходит только один пучок линейно поляризованного света, и *двухлучевые поляризационные призмы*, из которых выходят два пучка света, поляризованных во взаимно перпендикулярных плоскостях. Призмы первого типа действуют по принципу *полного отражения*. Падающий пучок естественного света, проникая в призму, расщепляется в ней на два пучка, поляризованных линейно во взаимно перпендикулярных плоскостях. Один из этих пучков претерпевает полное отражение на границе склейки призм и отклоняется вбок, другой проходит прямо и используется в дальнейшем.

2. Первая поляризационная призма была изобретена в 1828 г. шотландским физиком Николем (1768–1851). Ее сокращенно называют *николем*. То же название часто применяют и для других призм, действующих по тому же принципу. Для изготовления призмы Николя у продолговатого ромбоэдра, полученного скалыванием из куска исландского шпата, сошлифовывают основания так, чтобы новые основания составляли с боковыми ребрами угол 68° (вместо 71° у естественного кристалла). Затем кристалл разрезают вдоль плоскости, перпендикулярной к новым основаниям и к главному сечению кристалла (последнее нормально к этим основаниям и проходит через оптическую ось кристалла). Отполировав плоскости разреза, оба куска склеивают в прежнем положении тонким слоем канадского бальзама.

Сечение призмы Николя плоскостью главного сечения показано на рис. 261. Двойная стрелка, наклоненная под углом 64° к длинному ребру, указывает направление оптической оси. Такое обозначение применяется в дальнейшем и для других призм.

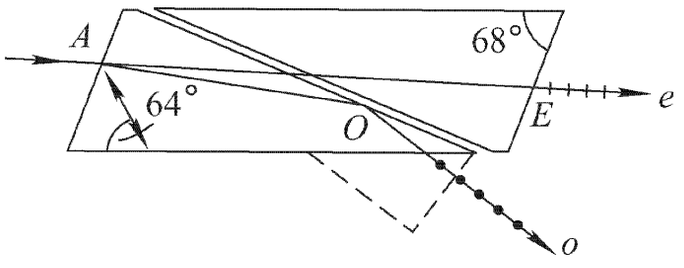


Рис. 261

Луч света, падая на искусственное основание кристалла, разделяется внутри кристалла на обыкновенный AO и необыкновенный AE. Показатель преломления канадского бальзама (1,550) имеет промежуточное значение между обыкновенным (n_o = = 1,658) и необыкновенным ($n_e = 1{,}486$) показателями преломления исландского шпата. Углы в призме Николя рассчитаны так, чтобы необыкновенный луч прошел через слой канадского бальзама, а обыкновенный претерпел на нем полное отражение и поглотился зачерненной боковой гранью. (В больших призмах во избежание их нагревания обыкновенный луч выводится наружу специальной призмочкой, наклеенной на боковую грань и изображенной на рис. 261 штриховой линией.) В результате свет, вышедший из призмы, окажется линейно поляризованным.

В пучке лучей, падающих на входную грань призмы Николя, обычно встречаются лучи разных направлений. Если луч падает на плоскость разреза слишком наклонно, то претерпит полное отражение не только обыкновенный луч, но и необыкновенный. Если же наклон недостаточен, то через призму пройдет и обыкновенный луч. Через призму может пройти один только необыкновенный луч тогда и только тогда, когда углы падения на входную грань лежат в определенных пределах. Разность углов наклона между крайними лучами падающего пучка, удовлетворяющими этому условию, определяет так называемую *апертуру полной поляризации призмы*. Для призмы Николя она составляет 29°, а отношение длины призмы к ее ширине — 3,28 (см. задачу 1 к этому параграфу).

В *призме Фуко* (рис. 262), устроенной как и николь, канадский бальзам заменен тонким слоем воздуха. Благодаря этому она может применяться для ультрафиолета, тогда как призма Николя непригодна для этой цели, так как канадский бальзам поглощает ультрафиолетовый свет. От николя призма Фуко отличается еще тем, что показатель преломления воздушной прослойки меньше обоих показателей n_o и n_e, а не равен промежуточному значению между ними, как это было в случае канадского бальзама. Поэтому полное отражение от плоскости разреза в призме Фуко осуществляется под меньшими углами, а сама

призма Фуко (при заданном поперечном сечении) гораздо короче и, следовательно, дешевле николя. Зато апертура полной поляризации ее составляет всего 8°.

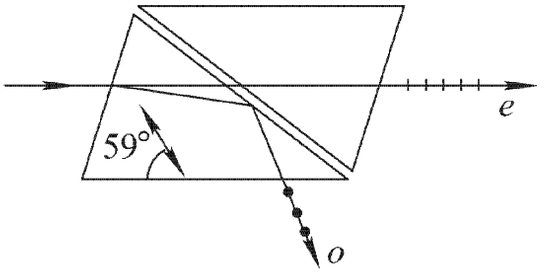


Рис. 262

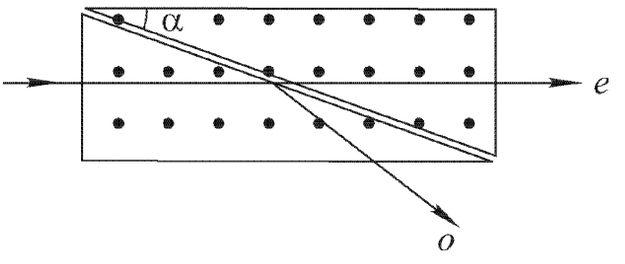


Рис. 263

3. Призмы Николя и Фуко имеют скошенные основания. Это вызывает параллельное боковое смещение падающего луча при прохождении его через призму. Следствием этого является кругообразное перемещение выходящего луча при вращении призмы вокруг ее оси. От этого недостатка свободны призмы, имеющие форму прямоугольного параллелепипеда: Глазебрука (рис. 263), Глана, Глана–Томпсона, Гартнака–Празмовского, Франка–Риттера (рис. 264), Аренса (рис. 265) и др. Сечение склеенной поляризационной призмы Глазебрука показано на

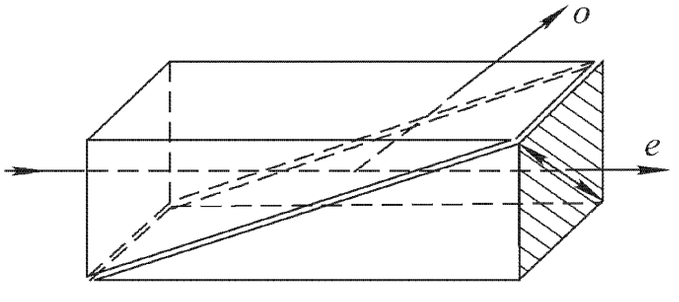


Рис. 264

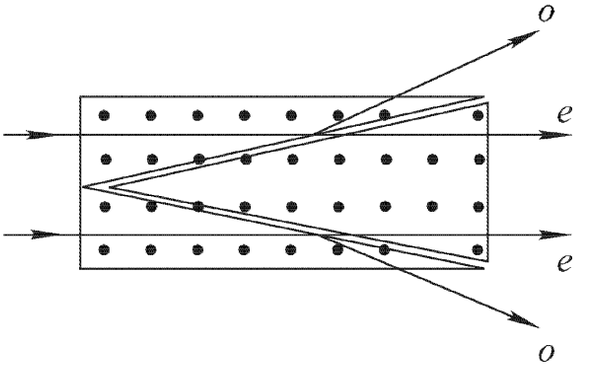


Рис. 265

рис. 263. Оптическая ось перпендикулярна к плоскости рисунка, что отмечено точками. Аналогично устроены призмы Глана, Глана–Томпсона и Гартнака–Празмовского. Их отличия от призмы Глазебрука заключаются в следующем. Призма Глана — воздушная. Призмы Глана–Томпсона имеют другие параметры. В призме Гартнака–Празмовского оптическая ось нормальна к плоскости распила, а в призме Франка–Риттера (рис. 264) она параллельна одной из диагоналей основания призмы. Призма Аренса — тройная призма, устройство и действие которой понятно из рис. 265. Параметры поляризационных призм указаны в табл. 8.

Таблица 8

Вид призмы	Апертурный угол полной поляризации	Угол плоскости разреза с боковой гранью	Отношение длины к ширине
Призма Николя	29°	22°	3,28
Укороченная призма Николя, склеенная:			
а) канадским бальзамом,	13°	25°	2,83
б) копайским бальзамом	24°	25°	2,83
Призма Фуко	8°	40°	1,53
Призма Глазебрука, склеенная:			
а) канадским бальзамом,	35°	12,1°	4,65
б) льняным маслом,	42°	14°	4,02
в) глицерином	32,1°	17,3°	3,2
Призма Глана с воздушной прослойкой	7,9°	50,3°	0,83
Призма Глана–Томпсона, склеенная:			
а) канадским бальзамом,	27,5°	13,5°	4,15
б) льняным маслом	41°	13,5°	4,15
Призма Гартнака–Празмовского	41,9°	13,9°	4,04
Призма Франка–Риттера, склеенная канадским бальзамом	19°	17,5°	3,17
Призма Аренса, склеенная канадским бальзамом	35°	12,1°	2,32

4. Различные типы *двухлучевых поляризационных призм* представлены на рис. 266. Первая призма состоит из комби-

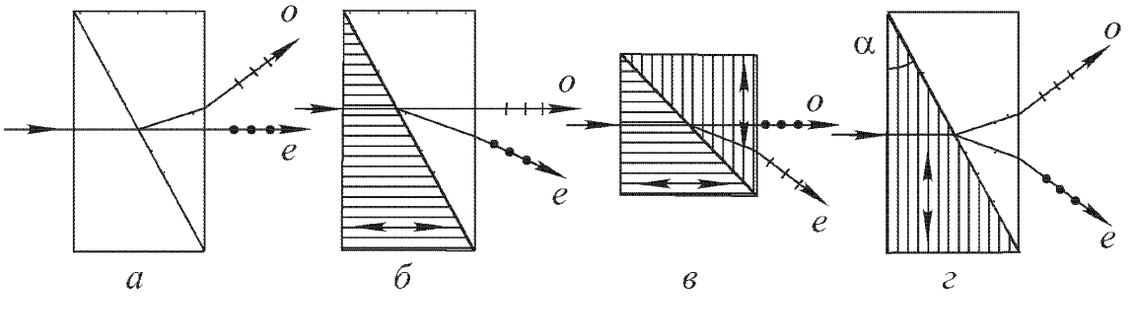


Рис. 266

нации стеклянной призмы с кристаллической из исландского шпата, оптическая ось которой параллельна преломляющему ребру. Призмы соприкасаются или склеиваются, как показано на рис. 266 *а*. Показатель преломления стекла (1,49) почти точно совпадает с необыкновенным показателем преломления n_e

исландского шпата. Падающий пучок неполяризованного света в кристаллической призме разделяется на обыкновенный и необыкновенный. Необыкновенный проходит без преломления. Обыкновенный сильно отклоняется к основанию кристаллической призмы в результате двукратного преломления на ее гранях. Аналогично действуют *призмы Рошона* (рис. 266 *б*) и *Сенармона* (рис. 266 *в*). В *призме Волластона* (рис. 266 *г*) выходящие лучи линейно поляризованного света отклоняются в разные стороны.

5. У многих кристаллов поглощение света зависит от направления электрического вектора в световой волне. Это явление также используется для получения линейно поляризованного света в так называемых *дихроичных пластинках*. К ним относятся, например, пластинки турмалина и поляроиды. В турмалине обыкновенный луч поглощается сильнее необыкновенного. Поэтому после прохождения через пластинку турмалина естественный свет становится частично поляризованным в плоскости главного сечения. Если пластинка достаточно толстая (около 1 мм), то в области видимого света обыкновенный луч поглощается практически полностью, так что прошедший свет окажется полностью линейно поляризованным. Для некоторых участков видимого спектра и необыкновенный луч испытывает заметное поглощение. Турмалин при надлежащей толщине действует не только как *поляризатор*, но и как *светофильтр*, пропускающий преимущественно желто-зеленую область спектра. Это — существенный недостаток турмалина как поляризатора. Зато апертура полной поляризации у него очень велика и ограничена одними только конструктивными условиями.

Заметим еще, что с изменением направления волновой нормали меняется и направление электрического вектора в обеих линейно поляризованных волнах, на которые разделяется естественный свет, распространяющийся вдоль этой нормали. Вследствие этого меняется и поглощение таких волн, которое к тому же обычно зависит от длины волны. Это приводит к зависимости окраски кристалла от направления распространения света. Такое явление называется *дихроизмом* (двуцветностью) или, лучше, *плеохроизмом* (многоцветностью) и в большей или меньшей степени присуще большинству двупреломляющих кристаллов.

ЗАДАЧИ

1. В поляризационной призме Николя плоскость разреза BC (рис. 267) перпендикулярна к основаниям AB и CD. Разрез произведен по диагонали призмы. Определить предельные углы полного отражения от слоя канадского бальзама для обыкновенной и необыкновенной волн. Найти соответ-

ствующие им углы падения на основание AB и апертуру полной поляризации призмы. Найти показатель преломления $n_{\|}$ необыкновенной волны в кристалле. Найти также отношение длины l призмы к ширине h ее поперечного сечения. Вычисления произвести для двух случаев: 1) основания призмы сошлифованы ($\gamma = 68°$); 2) для естественного кристалла ($\gamma = 71°$).

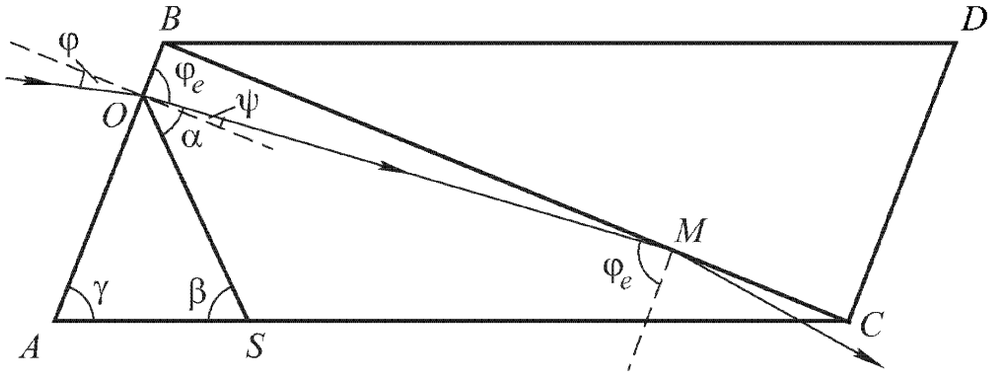


Рис. 267

Р е ш е н и е. Проведем расчет для необыкновенной волны в предположении, что $\gamma = 68°$. Пусть $\beta = 64°$ — угол между оптической осью кристалла и продольным ребром AC, OM — волновая нормаль для предельного случая полного отражения, φ_e — соответствующий угол падения на слой канадского бальзама BC. Тогда $\sin\varphi_e = n/n_{\|}$, где n — показатель преломления канадского бальзама. Показатель преломления $n_{\|}$ необыкновенной волны в кристалле для рассматриваемого направления волновой нормали определяется уравнением

$$\frac{1}{n_{\|}^2} = \frac{\sin^2\alpha}{n_e^2} + \frac{\cos^2\alpha}{n_o^2}.$$

Как видно из рис. 267, $\alpha = \beta + \gamma - \varphi_e$. Подставив числовые значения в предыдущее соотношение, придем к уравнению

$$\operatorname{tg}^2\varphi - 7{,}046608\operatorname{tg}\varphi - 32{,}84593 = 0.$$

Из него находим

$$\operatorname{tg}\varphi = 10{,}250829, \quad \varphi = 84{,}422825°, \quad n_{\|} = n/\sin\varphi_e = 1{,}5573726.$$

Угол преломления на основании AB равен $\psi = \pi/2 - \varphi_e = 5{,}5771175°$. По закону Снеллиуса находим угол падения

$$\sin\varphi = n_{\|}\sin\psi = 0{,}1513532, \quad \varphi = 8{,}705349°.$$

Чтобы необыкновенная волна не испытывала полного отражения, угол падения φ должен удовлетворять условию $\varphi > 8{,}705349°$.

Аналогичный расчет для обыкновенной волны, разумеется, проще, так как показатель преломления такой волны не зависит от направления ее распространения. Для предельного угла полного отражения в этом случае получаем $\varphi_o = 69{,}20573°$, а для соответствующего угла падения $\varphi = 36{,}05854°$. Чтобы обыкновенная волна испытывала полное отражение, угол падения должен быть меньше этого значения. При угле падения, удовлетворяющем условию $8{,}705349° < \varphi < 36{,}05854°$, прошедший свет будет поляризован полностью. Апертура полной поляризации $\varepsilon = 36{,}06 - 8{,}71 \approx 27{,}35°$. (При

вычислениях использовано неточное значение угла β. Десятичные знаки сохранялись для того, чтобы не потерять точность во время расчета.)

Для отношения l/h получаем

$$\frac{l}{h} = \frac{1}{\sin\gamma\cos\gamma} = \frac{2}{\sin 2\gamma} = 2{,}88.$$

Практически призму Николя распиливают не по диагонали, а как указано на рис. 261, чтобы легче монтировать ее. В этом случае отношение l/h несколько больше.

Рассмотрим теперь случай естественного кристалла, когда $\gamma = 71°$. Вычисления для обыкновенной волны остаются без всяких изменений. Вычисления же для необыкновенной волны сильно упрощаются благодаря случайному соотношению: $\beta + \gamma = 135°$, и следовательно, $\sin(\beta+\gamma) = -\cos(\beta+\gamma)$. Приводим окончательные результаты: $n_{\|} = 1{,}564955$, $\varphi_e = 82{,}0728°$, $\varphi = 12{,}46454°$, $\varepsilon = 23{,}59°$, $l/h = 3{,}25$. По сравнению с предыдущим случаем параметры поляризационной призмы ухудшились.

2. Призма Волластона (рис. 266 *г*) сделана из исландского шпата. Угол $\alpha = 15°$. На какой угол δ будут разведены обыкновенный и необыкновенный лучи?

О т в е т. $\delta = 5°17'$.

3. Две поляризационные призмы с воздушной прослойкой изготовлены из исландского шпата. В одной призме оптическая ось перпендикулярна, в другой — параллельна плоскости падения (рис. 268). Опишите действие

Рис. 268

каждой призмы. Как будет поляризован проходящий свет? Какая призма будет пропускать больше света? В каких пределах должен быть заключен угол α, чтобы из призмы выходил линейно поляризованный свет? Свет падает на грань призмы перпендикулярно.

О т в е т. Через призмы проходит необыкновенный луч. Вторая призма пропускает больше света. В обоих случаях должно выполняться условие $1/n_o < \sin\alpha < 1/n_e$, т.е. $37°6' < \alpha < 42°18'$.

4. В призме Глазебрука (рис. 263) склеивающим веществом является либо канадский бальзам ($n = 1{,}550$), либо льняное масло ($n = 1{,}494$). Найти наивыгоднейшую апертуру полной поляризации призмы и соответствующее ей отношение длины a к ширине b для обоих склеивающих материалов. Пучок падающих лучей, дающих полную поляризацию, должен быть симметричен относительно оси, параллельной боковому ребру призмы.

Р е ш е н и е. В рассматриваемом случае электрический вектор необыкновенной волны всегда параллелен оптической оси, а обыкновенной — перпендикулярен к ней. Поэтому направления лучей и волновых нормалей совпадают между собой. Необыкновенный луч не может претерпеть полное отражение от плоскости разреза призмы, так как n_e меньше показателя преломления склеивающего вещества. Наибольший наклон необыкновенного луча, идущего сверху вниз направо (рис. 269), определяется

направлением диагонали BC. Угол преломления для такого луча $\psi_e = \arcsin\left(1/\sqrt{1+(a/b)^2}\right)$, а угол падения

$$\varphi = \arcsin \frac{n_e}{\sqrt{1+(a/b)^2}}. \tag{77.1}$$

Если угол падения превышает это значение, то необыкновенный луч на плоскость разреза не попадет, а поглотится боковой гранью AC. Подберем теперь отношение a/b так, чтобы обыкновенный луч SO, падающий под углом φ в направлении снизу вверх направо, встретил плоскость разреза BC под предельным углом полного отражения $\beta = \arcsin(n/n_o)$. Тогда всякий обыкновенный луч, падающий снизу вверх под меньшим углом, а

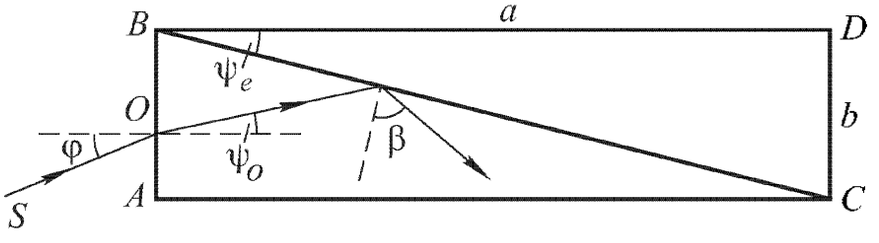


Рис. 269

также приходящий сверху вниз полностью отразится от плоскости разреза BC. Апертура полной поляризации будет $\varepsilon = 2\varphi$. Как видно из рис. 269, $\beta = \dfrac{\pi}{2} - (\psi_e + \psi_o)$, и следовательно, $\cos(\psi_e + \psi_o) = n/n_o$. Добавив сюда закон преломления $\sin\varphi = n_o \sin\psi_o = n_e \sin\psi_e$, получим полную систему уравнений для нахождения искомых величин: отношения a/b и апертуры 2φ. Из нее находим удобные расчетные формулы:

$$(n_o^2 - n^2)\left(\frac{a^2}{b^2}\right)^2 + [n_o^2 - n_e^2 - 2n(n+n_e)]\frac{a^2}{b^2} - (n+n_e)^2 = 0, \tag{77.2}$$

$$\sin\frac{\varepsilon}{2} = \frac{n_e}{\sqrt{1+(a/b)^2}}. \tag{77.3}$$

Они дают для канадского бальзама $a/b = 5{,}32$, $\varepsilon = 31°51'$, для льняного масла $a/b = 4{,}25$, $\varepsilon = 39°50'$.

5. Решить предыдущую задачу в предположении, что слой склеивающего вещества заменен прослойкой воздуха.

Р е ш е н и е. Для того чтобы обыкновенный луч, падающий на грань AB под углом φ снизу вверх направо (рис. 270), испытал полное отражение от плоскости разреза BC, необходимо, чтобы угол падения на эту плоскость превосходил предельный угол полного отражения:

$$\alpha - \psi_o \geqslant \beta_o, \quad \sin\beta_o = 1/n_o. \tag{77.4}$$

Для того чтобы необыкновенный луч, падающий на грань AB под углом φ' сверху вниз направо, прошел через плоскость разреза BC, необходимо, чтобы он падал на BC под углом, не превосходящим предельного угла полного отражения:

$$\alpha + \psi_e' \leqslant \beta_e, \quad \sin\beta_e = 1/n_e. \tag{77.5}$$

Наибольшей апертуре призмы соответствуют в (77.4) и (77.5) знаки равенства. В этом случае угол $\varphi + \varphi'$ равен апертуре ε полной поляризации призмы. Потребовав дополнительно $\varphi = \varphi'$, будем иметь

$$\alpha - \psi_o = \beta_o, \quad \alpha + \psi'_e = \beta_e, \quad (77.6)$$

причем

$$\sin\varphi = n_o \sin\psi_o = n_e \sin\psi'_e, \quad (77.7)$$

$$\operatorname{tg}\alpha = a/b. \quad (77.8)$$

Отсюда могут быть найдены все интересующие нас величины. Из (77.6) получаем

$$\psi_o + \psi'_e = \beta_e - \beta_o = 5^\circ 12'. \quad (77.9)$$

Таким образом, углы ψ_o, ψ'_e, φ малы и их синусы можно заменить самими углами. Это дает для апертуры

$$\varepsilon = 2\varphi = \frac{2n_o n_e}{n_o + n_e}(\beta_e - \beta_o) = 8^\circ 9' \quad (77.10)$$

и для отношения сторон

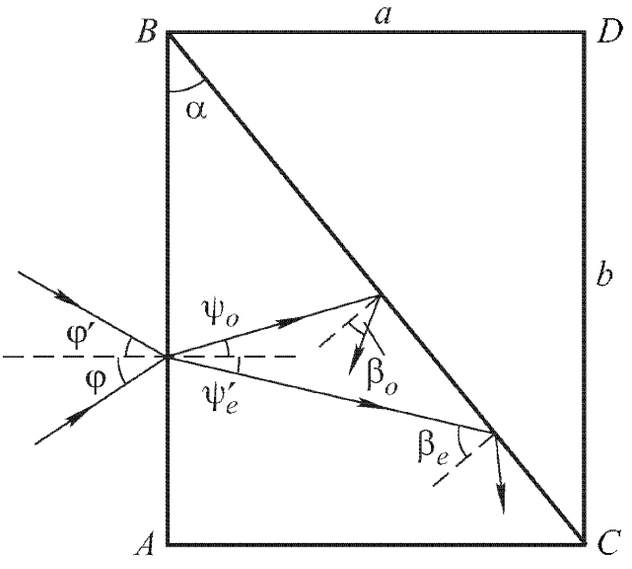


Рис. 270

$$\frac{a}{b} = \operatorname{tg}(\beta_o + \psi_o) = \operatorname{tg}\left(\beta_o + \frac{\varphi}{n_o}\right) = 0{,}826. \quad (77.11)$$

§ 78. Анализ поляризованного света

1. В линейной поляризации света можно убедиться с помощью поляризационной призмы (николя) или любого поляризатора, способного давать полностью линейно поляризованный свет. (С целью сокращения всякий поляризатор в дальнейшем называется *николем*.) Для этого николь ставят на пути исследуемого света. При вращении николя вокруг направления луча интенсивность проходящего света, вообще говоря, будет изменяться. Если при некотором положении николя проходящий свет полностью гасится, то падающий свет был поляризован *линейно*.

Если падающий свет естественный или поляризован по кругу, то при вращении николя интенсивность проходящего света меняться не будет. Для отличия одного случая от другого применяется *пластинка в четверть волны* (короче, $\lambda/4$) или *компенсатор*. Пластинка в четверть волны есть кристаллическая пластинка, которая вносит дополнительную разность фаз в $\pi/2$ между проходящими через нее лучами, поляризованными во взаимно перпендикулярных плоскостях. Эти плоскости определяют в плоскости пластинки два направления, называемые *главными направлениями пластинки*. Обычно пластинка

$\lambda/4$ вырезается из одноосного кристалла (например, кварца) параллельно его оптической оси. Тогда дополнительная разность фаз в $\pi/2$ вносится между обыкновенным и необыкновенным лучами. Но пластинку $\lambda/4$ можно изготовить и из двуосного кристалла, например слюды. В дальнейшем для определенности предполагается, что пластинка $\lambda/4$ вырезана из одноосного кристалла. В свете, поляризованном по кругу, разность фаз между любыми двумя взаимно перпендикулярными колебаниями равна $\pm\pi/2$. Если на пути такого света поставить пластинку $\lambda/4$, то она внесет дополнительную разность фаз $\pm\pi/2$. Результирующая разность фаз получится 0 или π, и свет станет поляризованным *линейно*. Его можно полностью погасить поворотом николя. Если же падающий свет естественный, то он останется таковым и после прохождения через пластинку $\lambda/4$. В этом случае гашения не будет.

2. Для отличия правой круговой поляризации от левой можно воспользоваться тем же приспособлением. Допустим, что свет распространяется вдоль оси Z по направлению к читателю, а оптическая ось пластинки $\lambda/4$ ориентирована вдоль оси Y. Тогда вектор $\mathbf{E}$ в необыкновенной волне будет параллелен оси Y, а в обыкновенной — оси X. Для определенности допустим, что по-

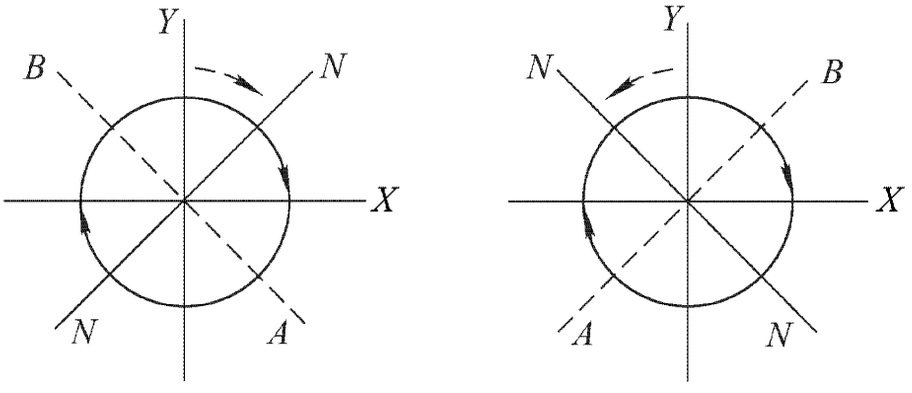


Рис. 271

ляризация правая (рис. 271), т.е. вращение вектора $\mathbf{E}$ совершается по часовой стрелке. В этом случае перед пластинкой необыкновенная волна опережает по фазе обыкновенную на $\pi/2$ (см. § 62, п. 4). Если z — толщина пластинки, то после прохождения через нее между этими волнами возникнет дополнительная разность фаз

$$\Delta_{\text{доп}} = (\omega t - kn_e z) - (\omega t - kn_o z) = k(n_o - n_e)z,$$

где k — волновое число в вакууме. Для отрицательного кристалла ($n_o > n_e$, исландский шпат) $\Delta_{\text{доп}} = +\pi/2$, для положительного ($n_o < n_e$, кварц) $\Delta_{\text{доп}} = -\pi/2$. В первом случае

полная разность хода равна $\Delta = \pi$, и плоскость колебаний прошедшего света на рис. 271 *а* будет направлена вдоль штриховой биссектрисы AB. Если первоначально главное сечение николя проходило через оптическую ось Y пластинки $\lambda/4$, то его надо повернуть в положение NN в направлении штриховой стрелки, т.е. *в ту же сторону*, куда вращается вектор $\mathbf{E}$ в падающей волне, чтобы полностью погасить проходящий свет. Во втором случае $\Delta = 0$ (рис. 271 *б*) и николь надо повернуть в противоположную сторону, т.е. *против направления вращения вектора* $\mathbf{E}$ в падающей волне. Если поляризация падающей волны *левая*, то все вращения николя надо производить в *противоположную сторону*.

Результаты можно резюмировать в виде следующего правила. *Пусть в исходном положении плоскость главного сечения николя проходит через оптическую ось пластинки* $\lambda/4$. *Положительным считается вращение николя вправо, отрицательным — влево (если смотреть против направления распространения света, выходящего из николя). Тогда если знаки кристаллической пластинки* $\lambda/4$ *и вращения николя одинаковы, то круговая поляризация левая, если же они разные, то правая.*

3. Рассмотрим более общий вопрос: как отличить естественный свет от света, поляризованного по кругу, и от смеси естественного света с поляризованным по кругу? Для этого поставим снова на пути света пластинку в четверть волны и николь. *Если при вращении николя при любом положении пластинки интенсивность не меняется, то свет естественный. Если интенсивность меняется и падает до нуля, то он поляризован по кругу. Если же интенсивность меняется, но не падает до нуля, то падающий свет состоит из смеси естественного с поляризованным по кругу.*

4. Допустим теперь, что падающая волна *поляризована эллиптически*. Если поставить николь, то при его вращении интенсивность проходящего света в двух положениях (отличающихся друг от друга на 180°) будет максимальна, а в перпендикулярных к ним положениях минимальна. Эти положения определят направления главных осей эллипса колебаний. После этого на пути падающего света поставим пластинку $\lambda/4$, оптическая ось которой ориентирована параллельно одной из главных осей эллипса. Тогда после прохождения через пластинку свет станет поляризован линейно и может быть погашен поворотом николя. При этом николь надо будет повернуть на некоторый угол относительно исходного положения, когда интенсивность проходящего через него света была минимальна или максимальна. Действительно, в исходном положении главное сечение николя было ориентировано параллельно одной из главных осей эллипса колебаний. После же прохождения через пластинку $\lambda/4$ плоскость

колебаний линейно поляризованного света будет проходить через одну из диагоналей прямоугольника на рис. 236.

5. Нетрудно сообразить, как поступить, чтобы отличить друг от друга: 1) эллиптически поляризованный свет; 2) смесь естественного света с линейно поляризованным светом (*отчасти линейно поляризованный свет*); 3) смесь естественного света с эллиптически поляризованным светом (*отчасти эллиптически поляризованный свет*). Опять надо поместить на пути распространения света пластинку в четверть волны, а за ней николь. Если вращением пластинки вокруг направления луча можно найти такое положение, при котором свет, прошедший через нее, можно погасить последующим вращением николя, то падающий свет был *эллиптически поляризован.* Если это сделать не удается, то мы имеем дело либо со смесью естественного света с линейно поляризованным, либо со смесью естественного света с эллиптически поляризованным. Чтобы отличить друг от друга эти два последних случая, на пути света ставят сначала только один николь и устанавливают его на минимум интенсивности проходящего света. Затем перед николем помещают пластинку в четверть волны. Вращением пластинки и николя снова добиваются минимума интенсивности. Если этот минимум интенсивности получается при прежнем положении николя (или при повороте его на 180°), то мы имеем *смесь естественного света с линейно поляризованным.* Если же для получения минимума требуется повернуть николь на некоторый угол, — то *смесь естественного света с эллиптически поляризованным.*

6. Вместо пластинки $\lambda/4$ (а также пластинки $\lambda/2$, вносящей дополнительную разность хода в полволны) применяются более совершенные приспособления, называемые *компенсаторами.* Они также могут служить для анализа поляризации света.

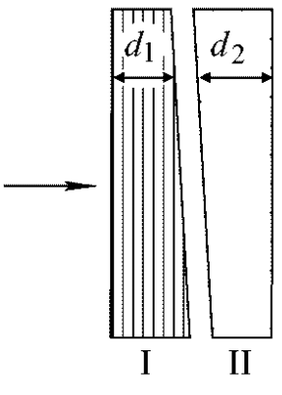


Рис. 272

Компенсатор Бабине (рис. 272) состоит из двух слабо скошенных кварцевых клиньев I и II, оптические оси которых взаимно перпендикулярны, что показано на рисунках точками и направлением штриховки. Клин I неподвижен, клин II может относительно него перемещаться вверх и вниз с помощью микрометрического винта. При таких перемещениях обращенные друг к другу поверхности клиньев остаются параллельными. Пусть луч падает на компенсатор слева направо перпендикулярно к его поверхности. В компенсаторе он разделяется на обыкновенный и необыкновенный лучи, идущие в одном направлении. Обыкновенный луч в первом клине при вступлении во второй, очевидно, становится необыкновенным и наоборот. В результате между колебаниями, параллельными и

перпендикулярными к плоскости рисунка, возникнет дополнительная разность фаз

$$\Delta_{\text{доп}} = (\omega t - kn_e d_1 - kn_o d_2) - (\omega t - kn_o d_1 - kn_e d_2) = \\ = k(n_e - n_o)(d_2 - d_1),$$

где d_1 и d_2 — толщины первого и второго клиньев вдоль пересекающего их луча, а $k = 2\pi/\lambda$ — волновое число в вакууме.

Для кварца (положительный кристалл) $n_e > n_o$. В том месте, где $d_1 = d_2$, дополнительная разность фаз $\Delta_{\text{доп}}$ равна нулю. Выше этого места величина $\Delta_{\text{доп}}$ положительна, ниже — отрицательна, причем она линейно возрастает с расстоянием при смещении луча вверх. Если на компенсатор падает параллельный пучок линейно или эллиптически поляризованного света, то на выходе компенсатора в разных местах полная разность фаз Δ будет разной. Линии равной разности фаз имеют форму прямых, параллельных ребрам клиньев. Выходящий свет останется поляризованным эллиптически, но форма эллипса колебаний будет другой, меняясь при переходе от одной линии равной разности фаз к другой. Когда полная разность фаз равна $\Delta = 2m\pi$ $(m = 0, 1, 2, \ldots)$, свет будет поляризован линейно. Если его рассматривать через надлежаще ориентированный николь, то на поверхности компенсатора появятся равноотстоящие параллельные темные полосы. Если николь повернуть на 90°, то полосы сместятся на полполосы, т.е. появятся там, где $\Delta = (2m + 1)\pi$. Если микрометрическим винтом перемещать клин II, то при неизменном положении луча будет меняться и дополнительная разность фаз $\Delta_{\text{доп}}$. Шкалу на барабане микрометрического винта можно проградуировать таким образом, чтобы сразу получить значение величины $\Delta_{\text{доп}}$.

Разность фаз $\Delta_{\text{доп}}$, вносимая компенсатором Бабине, зависит от точки вхождения падающего луча. В усовершенствованной форме, приданной компенсатору Солейлем (1798–1878), этот недостаток устранен: все падающие лучи приобретают одну и ту же дополнительную разность фаз. Как и компенсатор Бабине, *компенсатор Бабине–Солейля* состоит из таких же кварцевых клиньев, но ориентированных оптическими осями параллельно друг другу (рис. 273). Перед клиньями располагается плоско-параллельная кварцевая пластинка, оптическая ось которой перпендикулярна к оптической оси клиньев. Толщина пластинки равна сумме толщин клиньев в нулевом положении, когда они по высоте не смещены относительно друг друга. Оба клина вместе действуют как пло-

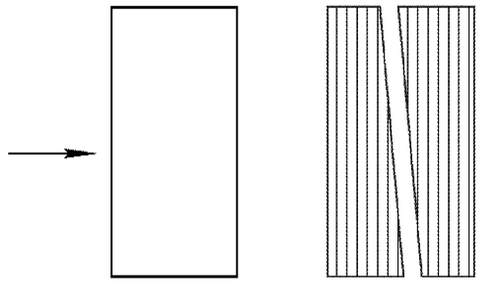


Рис. 273

скопараллельная пластинка, толщину которой можно менять микрометрическим винтом. Тем самым действие компенсатора Бабине–Солейля сведено к действию двух кристаллических пластинок с взаимно перпендикулярными оптическими осями.

ЗАДАЧИ

1. Определить наименьшую толщину d пластинки слюды, чтобы она могла служить пластинкой в четверть волны для желтого света натрия. Для такого света показатели преломления волн, идущих внутри пластинки перпендикулярно к ней, соответственно равны $n_1 = 1{,}5941$, $n_2 = 1{,}5887$. Сделать то же самое для кварца ($n_o = 1{,}5442$, $n_e = 1{,}5533$).

О т в е т. 1) $d = \dfrac{\lambda}{4(n_1 - n_2)} = 27$ мкм; 2) $d = \dfrac{\lambda}{4(n_e - n_o)} = 16{,}2$ мкм.

2. Смесь света, поляризованного по кругу, и естественного рассматривается через кристаллическую пластинку в четверть волны и николь. При вращении николя вокруг оси светового пучка найдено, что максимальная интенсивность света, прошедшего через систему, в $m = 3$ раза превосходит минимальную интенсивность. Найти отношение интенсивности света $I_к$, поляризованного по кругу, к интенсивности естественного света I_e.

О т в е т. $I_к/I_e = {}^1\!/_2(m-1) = 1$.

3. На пути линейно поляризованного света поставлена пластинка в полволны. Плоскость колебаний падающего света составляет угол α с оптической осью пластинки. Определить поляризацию света, прошедшего через пластинку.

О т в е т. Линейная поляризация сохранится, но плоскость колебаний будет наклонена к оптической оси под углом $\beta = -\alpha$, т.е. повернется на угол 2α. В частности, когда $\alpha = 45°$, плоскость колебаний повернется на $90°$.

4. Параллельный пучок монохроматического света, поляризованный по правому кругу, падает нормально на пластинку в полволны. Найти состояние поляризации света, прошедшего через эту пластинку.

О т в е т. Свет будет поляризован по левому кругу.

5. Параллельный пучок монохроматического света падает нормально на поляроид, а затем на пластинку в полволны. Главная плоскость поляроида (в которой лежит электрический вектор пропускаемой им волны) составляет угол α с осью этой пластинки. Найти состояние поляризации прошедшего света на выходе из пластинки в полволны.

О т в е т. Свет останется линейно поляризованным, но плоскость колебаний электрического вектора повернется на угол 2α и станет симметрично расположенной со своим исходным положением относительно оси пластинки в полволны.

6. Параллельный пучок монохроматического света проходит через два николя, главные сечения которых повернуты друг относительно друга на угол $\alpha = 20°$. Между николями ставится пластинка в полволны из одноосного кристалла, вырезанная параллельно оптической оси. Какой угол β должна составлять оптическая ось пластинки с главным направлением первого николя, чтобы свет через эту систему не прошел?

Р е ш е н и е. Свет не пройдет через второй николь, если электрический вектор перпендикулярен к главному сечению этого николя, т.е. параллелен

прямой AB, перпендикулярной к тому же сечению (рис. 274). Ось пластинки должна быть ориентирована по биссектрисе угла AON_1 или угла N_1OB, ему дополнительного до π (см. предыдущую задачу). Это дает два значения угла β:

$$\beta_1 = -\frac{\pi}{4} + \frac{\alpha}{2} = -35^\circ, \quad \beta_2 = \frac{\pi}{4} + \frac{\alpha}{2} = +55^\circ.$$

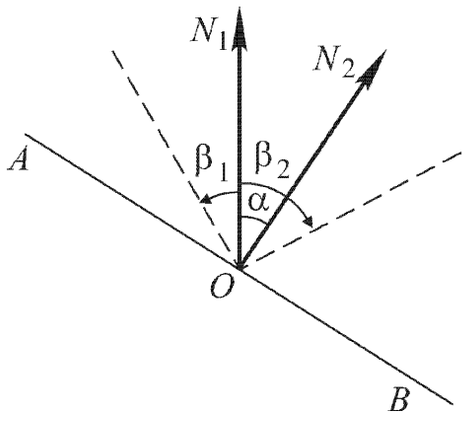


Рис. 274

7. Линейно поляризованная волна проходит через кристаллическую пластинку, одно из главных направлений которой составляет с главным сечением поляризатора угол φ. Разность фаз, сообщаемая пластинкой, равна δ. Найти: 1) отношение полуосей эллипса колебаний полученного эллиптически поляризованного света; 2) угол между главными направлениями пластинки и полуосями эллипса.

Р е ш е н и е. Линейно поляризованный луч по выходе из пластинки превращается в эллиптически поляризованный:

$$x = a\cos\omega t, \quad y = b\cos(\omega t + \delta),$$

где x и y — составляющие электрического вектора вдоль координатных осей, совпадающих с главными направлениями пластинки. Для определения полуосей полученного эллипса найдем проекции светового вектора x' и y' на оси координат, повернутые на угол θ относительно первоначальной системы:

$$x' = a\cos\theta\cos\omega t + b\sin\theta\cos(\omega t + \delta) = A\cos(\omega t + \alpha),$$
$$y' = -a\sin\theta\cos\omega t + b\cos\theta\cos(\omega t + \delta) = B\cos(\omega t + \beta),$$

где A и B — полуоси эллипса, определяющиеся уравнениями

$$\begin{aligned} A\cos\alpha &= a\cos\theta + b\sin\theta\cos\delta, & A\sin\alpha &= b\sin\theta\sin\delta, \\ B\cos\beta &= -a\sin\theta + b\cos\theta\cos\delta, & B\sin\beta &= b\cos\theta\sin\delta. \end{aligned} \tag{78.1}$$

Отсюда

$$\begin{aligned} A^2 &= a^2\cos^2\theta + b^2\sin^2\theta + ab\sin 2\theta\cos\delta, \\ B^2 &= a^2\sin^2\theta + b^2\cos^2\theta - ab\sin 2\theta\cos\delta. \end{aligned} \tag{78.2}$$

Складывая и вычитая эти соотношения, получим

$$A^2 + B^2 = a^2 + b^2 = R^2, \tag{78.3}$$

$$A^2 - B^2 = (a^2 - b^2)\cos 2\theta + 2ab\sin 2\theta\cos\delta, \tag{78.4}$$

где R — амплитуда волны, прошедшей через поляризатор. Чтобы амплитуды A и B соответствовали колебаниям вдоль осей эллипса, надо выбрать θ так, чтобы разность $A^2 - B^2$ была максимальной или минимальной. Приравнивая ее производную нулю, из этого условия найдем

$$\operatorname{tg} 2\theta = \frac{2ab}{a^2 - b^2}\cos\delta. \tag{78.5}$$

Из (78.1) находим

$$\begin{aligned} AB\sin(\alpha - \beta) &= -ab\sin\delta, \\ AB\cos(\alpha - \beta) &= ab\cos 2\theta\cos\delta - \frac{a^2 - b^2}{2}\sin 2\theta, \end{aligned} \tag{78.6}$$

или

$$\frac{AB\cos(\alpha-\beta)}{ab\cos 2\theta\cos\delta}=1-\frac{a^2-b^2}{2ab\cos\delta}\operatorname{tg}2\theta.$$

Подставляя вместо $\operatorname{tg}2\theta$ его значение из (78.5), получаем $\dfrac{AB\cos(\alpha-\beta)}{ab\cos 2\theta\cos\delta}=0$, откуда $\cos(\alpha-\beta)=0$, $(\alpha-\beta)=\pm\pi/2$. Следовательно, из первого уравнения (78.6) имеем $AB=\pm ab\sin\delta$. Так как $b/a=\operatorname{tg}\varphi$, то

$$\sin 2\varphi=\frac{2ab}{a^2+b^2},\quad \cos 2\varphi=\frac{a^2-b^2}{a^2+b^2},\quad \operatorname{tg}2\varphi=\frac{2ab}{a^2-b^2}.$$

Таким образом, уравнение (78.5) получит следующий вид:

$$\operatorname{tg}2\theta=\operatorname{tg}2\varphi\cos\delta.$$

Этим уравнением и определяется угол θ между главными направлениями пластинки и осями эллипса. Для второго неизвестного введем обозначение $B/A=\operatorname{tg}I$ и воспользуемся формулой $\sin 2I=2\operatorname{tg}I/(1+\operatorname{tg}^2 I)$. Тогда получим

$$\sin 2I=\frac{2AB}{A^2+B^2}=\frac{2ab\sin\delta}{a^2+b^2}=\sin 2\varphi\sin\delta.$$

8. На кристаллическую пластинку, вырезанную параллельно оптической оси, падает нормально свет, поляризованный по кругу. Прошедший свет рассматривается через анализатор.

1) Пренебрегая потерями света на отражение, определить интенсивность прошедшего света, если главное сечение анализатора составляет угол α с одним из главных направлений пластинки. 2) Под каким углом надо поставить анализатор, чтобы получить максимальную и минимальную интенсивности?

Р е ш е н и е. 1) Если свет поляризован по кругу, то слагающие колебания по координатным осям могут быть представлены в виде

$$x=a\cos\omega t,\quad y=a\sin\omega t.$$

После прохождения через кристаллическую пластинку, сообщающую разность фаз δ, уравнения колебаний перейдут в

$$x=a\cos\omega t,\quad y=a\sin(\omega t+\delta).$$

При выходе из анализатора результирующее колебание будет

$$\xi=a\cos\alpha\cos\omega t+a\sin\alpha\sin(\omega t+\delta)=$$
$$=a(\cos\alpha+\sin\alpha\sin\delta)\cos\omega t+a\sin\alpha\cos\delta\sin\omega t.$$

Отсюда получаем для интенсивности

$$I=a^2\{(\cos\alpha+\sin\alpha\sin\delta)^2+(\sin\alpha\cos\delta)^2\}=a^2(1+\sin 2\alpha\sin\delta).$$

2) При постоянном δ интенсивность достигает максимума или минимума, когда $\cos 2\alpha=0$, т.е. при $\alpha={}^1\!/_4\pi,{}^3\!/_4\pi$. Если $\sin\delta>0$, то первому значению соответствует максимум, а второму — минимум; при $\sin\delta<0$ — наоборот.

9. Клин из двоякопреломляющего вещества помещен на пути монохроматического света, поляризованного по кругу. Оптическая ось параллельна ребру клина. Описать картину, наблюдаемую через николь, когда клин

неподвижен и когда он поворачивается вокруг направления распространения света.

Р е ш е н и е. Согласно предыдущей задаче интенсивность света, прошедшего через анализатор,

$$I = a^2(1 + \sin 2\alpha \sin \delta).$$

Если $\sin 2\alpha > 0$, то при постоянном α интенсивность минимальна, когда

$$\sin \delta = -1, \quad \text{т.е. при} \quad \delta = 3\pi/2, \quad 7\pi/2, \ldots,$$

и максимальна, когда

$$\sin \delta = 1, \quad \text{т.е. при} \quad \delta = \pi/2, \quad 5\pi/2, \quad 9\pi/2, \ldots$$

Если же $\sin 2\alpha < 0$, то в первом случае будет максимум, а во втором — минимум интенсивности. Во всех случаях в поле зрения будут видны чередующиеся светлые и темные полосы. При вращении клина будет меняться угол α, а с ним и интенсивность в каждой точке клина. При углах $\alpha = 90$, 180 и 270° весь клин будет освещен равномерно, а при углах $\alpha = 45$, 135, 225, 315° будет наблюдаться наибольший контраст темных и светлых полос. При переходе через углы $\alpha = 90$, 180, 270° темные полосы будут переходить в светлые, а светлые — в темные.

10. Два когерентных пучка квазимонохроматического неполяризованного света равной интенсивности дают на экране интерференционные полосы. Какой толщины кристаллическую пластинку надо ввести на пути одного из этих пучков, чтобы интерференционные полосы исчезли и притом так, чтобы их нельзя было восстановить никакой стеклянной пластинкой, вводимой в другой пучок? Как изменится картина, если за кристаллической пластинкой поставить поляроид? При каком положении поляроида интерференционных полос не будет?

Р е ш е н и е. Разложим мысленно световую волну на две составляющие, электрические векторы которых взаимно перпендикулярны и параллельны главным осям пластинки. При введении пластинки интерференционные полосы от каждой составляющей сместятся. Если введенная пластинка является пластинкой в полволны, то разность смещений составит половину ширины полосы. В этом случае при введении пластинки интерференционные полосы пропадут. При введении поляроида они появятся вновь. Исключение составляет случай, когда оси поляроида наклонены под углом 45° к осям пластинки. В этом случае интерференционные полосы наблюдаться не будут.

11. Плоская световая волна эллиптически поляризована. Длины полуосей эллипса колебаний равны соответственно a и b. Какую кристаллическую пластинку надо поставить на пути распространения волны и как надо ориентировать эту пластинку, чтобы получить свет, поляризованный по кругу: 1) с тем же направлением вращения; 2) с противоположным направлением вращения?

Р е ш е н и е. В системе главных осей X, Y эллиптическое колебание представляется уравнениями $E_X = a\cos\omega t$, $E_Y = b\sin\omega t$ (рис. 275). Перейдем к новой системе ξ, η, оси которой являются биссектрисами прежних координатных углов. В этой системе то же колебание представится в виде

$$E_\xi = \frac{1}{\sqrt{2}}(a\cos\omega t + b\sin\omega t) = \sqrt{\frac{a^2+b^2}{2}}\cos(\omega t - \varphi),$$

$$E_\eta = \frac{1}{\sqrt{2}}(-a\cos\omega t + b\sin\omega t) = \sqrt{\frac{a^2+b^2}{2}}\cos[\omega t - (\pi - \varphi)],$$

где φ — острый угол, определяемый уравнением $\operatorname{tg}\varphi = b/a$. Колебания вдоль осей ξ и η совершаются с одинаковыми амплитудами $\sqrt{(a^2+b^2)/2}$, причем колебание вдоль оси ξ опережает по фазе колебание вдоль оси η на угол $\delta = \pi - 2\varphi$. Внесем кристаллическую пластинку, чтобы ее оси были ориентированы вдоль ξ и η и чтобы она изменила разность фаз до $\pm\pi/2$. Для этого должно быть выполнено соотношение

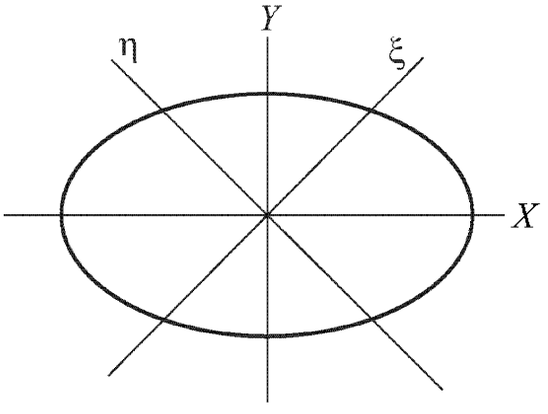


Рис. 275

$$(\omega t - \varphi - k_\xi l) - (\omega t - \pi + \varphi - k_\eta l) = \pm\pi/2,$$

откуда

$$l = \frac{2\varphi - \pi \pm \pi/2}{k_\eta - k_\xi} = \lambda \frac{\varphi/\pi - {}^1\!/_2 \pm {}^1\!/_4}{n_\eta - n_\xi}.$$

Тогда волна перейдет в волну, поляризованную по кругу. Знаку плюс соответствует то же направление вращения, что и в исходной эллиптически поляризованной волне, а знаку минус — противоположное. Такой же результат получится, если толщину пластинки изменить на $m\lambda/(n_\eta - n_\xi)$, где m — целое число.

§ 79. Интерференция поляризованных лучей

1. Явления *интерференции поляризованных лучей* в истории оптики имели большое значение для выяснения фундаментального вопроса о *природе световых колебаний*. Они исследовались в классических опытах Френеля и Араго (1816 г.). Конечно, лучи от независимых источников света интерферировать не будут, даже если они предварительно пропущены через поляризационное приспособление. Для интерференции необходима *когерентность*. Однако, как видно из формулы (26.2), результат интерференции линейно поляризованных лучей зависит от угла между плоскостями световых колебаний. Интерференционные полосы наиболее контрастны, когда плоскости колебаний параллельны. Интерференция никогда не наблюдается, если волны поляризованы во взаимно перпендикулярных плоскостях. Это впервые было установлено в упомянутых выше опытах Френеля и Араго. Отсюда Френель пришел к заключению о *поперечности световых колебаний* (см. § 26, п. 5).

2. Одна из возможных схем для исследования интерференции поляризованных лучей изображена на рис. 276. Лучи от первичного источника света S проходят через поляроид П или другое поляризационное приспособление. Вторичные когерентные источники S_1 и S_2 получаются одним из способов, применяемых для осуществления двухлучевой интерференции. Исходящие из них пучки поляризованы в параллельных плоскостях.

На пути одного из пучков вводится полуволновая кристаллическая пластинка K. В другом пучке для компенсации возникшей разности хода помещается стеклянная пластинка P надлежащей толщины. Стеклянная пластинка, конечно, не меняет направления колебаний проходящей через нее линейно поляризованной волны. Кристаллическая пластинка действует так же только в том случае, когда ее оптическая ось параллельна или перпендикулярна к плоскости колебаний. В этом случае из пластинок K и P выходят одинаково поляризованные когерентные пучки света. Направления колебаний в этих пучках изображены на рис. 276 совпадающими векторами $\mathbf{E}_1$ и $\mathbf{E}_2$.

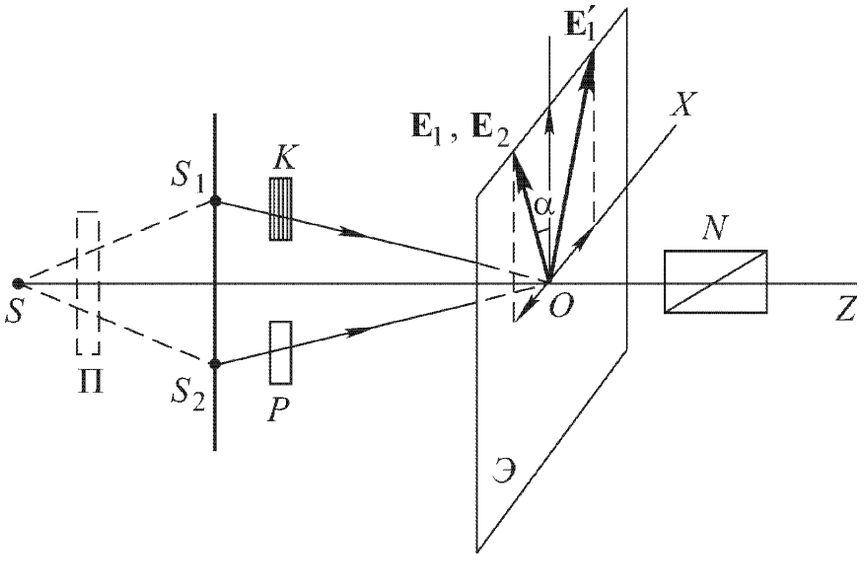


Рис. 276

Если повернуть пластинку K вокруг ее нормали на угол α, то вектор $\mathbf{E}_1$ повернется на угол 2α и перейдет в положение $\mathbf{E}'_1$ (см. задачу 5 к предыдущему параграфу). Таким образом, можно исследовать интерференцию при различных значениях угла 2α между плоскостями колебаний интерферирующих волн.

Интерференционную картину можно получать на экране $Э$, а также наблюдать без экрана с помощью лупы или зрительной трубы. В дальнейших пояснениях плоскость экрана принимается за координатную плоскость XY, направление оптической оси пластинки K — за ось Y, а нормаль к экрану, проведенная в сторону распространения света, — за ось Z. Следовательно, при повороте пластинки K будет поворачиваться и ось Y.

От наложения линейно поляризованных монохроматических волн везде, в том числе и на экране $Э$, возникнет эллиптическая поляризация света. При неизменном угле между плоскостями колебаний форма и ориентация эллипса колебаний будут изменяться с изменением разности хода между интерферирующими лучами. При определенных значениях разности хода эллипти-

ческие колебания могут вырождаться в линейные. Поверхности (а на экране — линии) равной разности хода являются в то же время поверхностями (линиями) одинаковых и одинаково ориентированных эллипсов колебаний.

Разложим, например, в описанной выше установке все колебания по осям X и Y. В центре экрана O y-составляющие векторов $\mathbf{E}_1'$ и $\mathbf{E}_2$ направлены в одну сторону, т.е. колебания вдоль оси Y совершаются в фазе. Составляющие тех же векторов вдоль оси X направлены противоположно — вдоль этой оси колебания противофазны и взаимно гасят друг друга. В результате в точке O, а также во всякой точке оси X получится линейное колебание параллельно оси Y. Если сместиться параллельно оси Y на половину ширины интерференционной полосы, то x-колебание станет синфазным, а y-колебание противофазным. Результирующее колебание также будет линейным, но уже вдоль оси X. В промежуточных точках из-за разности хода между лучами полного гашения не будет ни для x-, ни для y-колебания. Однако одно из этих колебаний всегда будет сдвинуто относительно другого по фазе на $\pi/2$: результирующее колебание — эллиптическое с главными осями эллипса, параллельными координатным осям X и Y.

3. Эллиптический характер поляризации проявляется при наблюдении через николь. Если николь пропускает колебания, параллельные оси Y, то на экране получаются интерференционные полосы, параллельные оси X. Центральная полоса, проходящая через центр O экрана (где разность хода равна нулю), будет светлой. При повороте николя на 90° эти полосы гасятся. Зато появляется другая система интерференционных полос, сдвинутая относительно предыдущей на половину ширины полосы. В ней, следовательно, центральная полоса будет темной. Таким образом, поворот николя на 90° из указанного исходного положения приводит к замене светлых полос темными и наоборот.

То же, в частности, будет, когда поляризация круговая (т.е. когда в нашей интерференционной установке кристаллическая пластинка K повернута из своего исходного положения на угол $\alpha = 45°$). Тогда интерференционные полосы, как они воспринимаются глазом, будут иметь один и тот же вид при любой ориентации николя. При вращении николя полосы непрерывно перемещаются параллельно самим себе. При повороте на 90° это перемещение составляет половину ширины полосы. Направление колебаний изменяется также на 90°. Если убрать николь, то произойдет наложение двух сдвинутых на полполосы некогерентных интерференционных картин со взаимно перпендикулярными колебаниями. Это даст равномерную освещенность экрана, в соответствии с тем, что волны со взаимно перпенди-

кулярными колебаниями никогда не интерферируют. Таким образом, поворот плоскости колебаний одного из линейно поляризованных лучей на 90° (осуществляемый, например, вращением кристаллической пластинки K) существенно меняет распределение освещенности на экране. При отсутствии такого поворота на экране видны полосы интерференции. При наблюдении через николь их интенсивность изменяется; полосы пропадают, когда колебания перпендикулярны к главному сечению николя. При наличии же поворота плоскости колебаний одного из лучей на 90° и отсутствии николя никаких интерференционных полос не наблюдается. Зато они появляются при наблюдении через николь и непрерывно перемещаются параллельно самим себе при его вращении, причем при повороте николя на 90° это перемещение равно половине ширины полосы.

Если опыт производить в неполяризованном свете (т.е. убрать поляроид П), то введение кристаллической пластинки, повернутой на любой угол, конечно, оставляет естественным свет, проходящий через нее. Однако, как указал и подтвердил свое заключение на опыте С.И. Вавилов, при этом также должна претерпеть изменения интерференционная картина на экране. Это объясняется тем, что и после введения кристаллической пластинки лучи не остаются независимыми, а определенным образом скоррелированы между собой.

В самом деле, падающий естественный свет можно представить в виде суперпозиции линейно поляризованных некогерентных волн с всевозможными направлениями плоскости колебаний. Возьмем одну из таких волн. При наличии кристаллической пластинки K и при наблюдении через николь N на экране получатся интерференционные полосы, описанные выше. Существенно, что положение полос не зависит от угла между плоскостями колебаний интерферирующих волн. Значит, все линейно поляризованные компоненты, на которые разложена падающая волна, дадут совпадающие по положению интерференционные картины. А так как эти картины некогерентны, то их наложение приведет только к взаимному усилению интерференционных полос. При вращении николя, ввиду статистической изотропии естественного света относительно всех направлений, перпендикулярных к лучу, вид интерференционных полос изменяться не будет. Будет происходить лишь перемещение полос параллельно самим себе, как при круговой поляризации, описанной выше. При повороте николя на 90° светлые полосы заменятся темными и наоборот. Не то будет, когда нет кристаллической пластинки. В этом случае в отсутствие николя интерференционные полосы наблюдаться будут. Однако при введении николя интенсивность полос уменьшится вдвое. Вращение же николя вообще не сказывается на положении и интенсивности полос.

4. Важный случай интерференции осуществляется при прохождении через кристалл поляризованного света. Пусть на плоскопараллельную кристаллическую пластинку падает волна под углом φ (рис. 277). В кристалле она разделяется на две волны, распространяющиеся в разных направлениях и с различными скоростями. Пусть AB и AC — волновые нормали этих волн, а ψ_1 и ψ_2 — соответствующие им углы преломления. В кристалле направления волновых нормалей не совпадают. Вне кристалла различие между этими направлениями пропадает. Из кристалла выходят два луча *1* и *2*, параллельные падающему и поляризованные в перпендикулярных плоскостях. Оптическая разность хода между ними представляется выражением

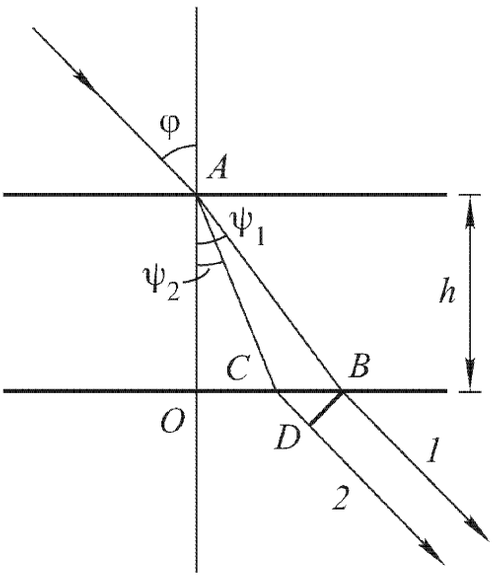


Рис. 277

$$\Delta = (ACD) - (AB) = n_2 \cdot AC + CD - n_1 \cdot AB,$$

где n_1 и n_2 — показатели преломления рассматриваемых волн, а D — основание перпендикуляра, опущенного из точки B на луч *2*. Если h — толщина пластинки, то

$$AC = h/\cos\psi_2, \quad AB = h/\cos\psi_1, \quad CD = h(\operatorname{tg}\psi_1 - \operatorname{tg}\psi_2)\sin\varphi.$$

По закону преломления $\sin\varphi = n_1 \sin\psi_1 = n_2 \sin\psi_2$. Используя эти выражения, получим

$$\Delta = h(n_2\cos\psi_2 - n_1\cos\psi_1). \tag{79.1}$$

Разность хода Δ между лучами *1* и *2* обусловлена двумя обстоятельствами: 1) различием показателей преломления n_1 и n_2; 2) различием углов преломления ψ_1 и ψ_2. Второе обстоятельство играет малую роль. В большинстве случаев им можно пренебречь и пользоваться приближенным выражением

$$\Delta = h\cos\psi(n_2 - n_1), \tag{79.2}$$

где угол ψ имеет любое промежуточное значение между углами преломления ψ_1 и ψ_2.

5. Разность хода Δ можно использовать для получения интерференции лучей *1* и *2*. Если бы свет, падавший на кристаллическую пластинку, был естественный, то интерференция была бы невозможна, так как в этом случае лучи *1* и *2* не были бы коррелированы между собой. Для получения коррелированных лучей *1* и *2* падающий свет должен быть поляризован —

линейно или эллиптически. Но и в этом случае при наложении лучей *1* и *2* интерференция все же не возникнет, так как лучи *1* и *2* поляризованы во взаимно перпендикулярных плоскостях. Нужно свести колебания в этих лучах к одному направлению, т.е. предварительно пропустить их через николь. В случае плоскопараллельных пластинок лучи сводятся вместе в фокальной плоскости линзы — получаются полосы равного наклона. В случае тонких пластинок переменной толщины наблюдаются полосы равной толщины, локализованные на самих пластинках.

Схема для получения так называемой *интерференции в параллельных лучах* приведена на рис. 278. Кристаллическая пластинка K вводится между поляризатором P и анализатором A. Линейно поляризованный свет, выйдя из поляризатора P, после

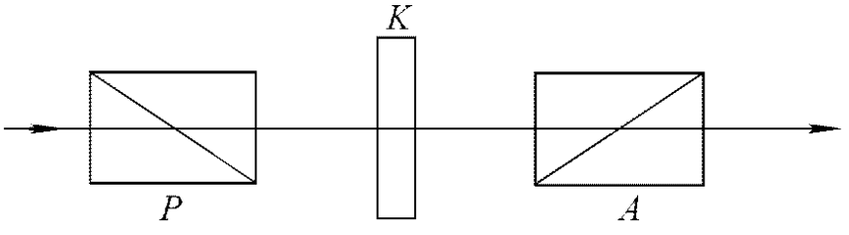


Рис. 278

прохождения через кристаллическую пластинку K превращается в поляризованный эллиптически, а затем проходит через анализатор A, превращаясь снова в линейно поляризованный. Интенсивность проходящего света зависит от ориентации анализатора. При вращении последнего интенсивность I будет меняться.

Разность фаз, возникающая при прохождении света через кристаллическую пластинку, равна $2\pi\Delta/\lambda$, т.е. зависит от длины волны λ. При неподвижном анализаторе она различна для разных длин волн. По этой причине интерференционная картина получается окрашенной. При вращении анализатора окраска меняется. При повороте анализатора на 90° окраска меняется на дополнительную — светлые места становятся темными и наоборот. Если главные сечения поляризатора и пластинки параллельны или перпендикулярны между собой, то через пластинку проходит только один луч. Ему не с чем интерферировать, и окрашивание пропадает. Все эти явления получили физически не совсем удачное название *хроматической поляризации*. Она была открыта Араго в 1911 г.

Обычно главные сечения поляризатора и анализатора устанавливаются параллельно (*николи параллельны*) или перпендикулярно (*николи скрещены*) друг к другу. В последнем случае установка весьма чувствительна к обнаружению анизотропии. Через два скрещенных николя свет не проходит (поле зрения темное). Если же между ними ввести какое-либо анизотропное

тело, то даже при наличии слабой анизотропии система начинает пропускать свет (поле зрения просветляется).

Для количественного расчета примем главные направления кристаллической пластинки за координатные оси X и Y (рис. 279). Пусть поляризатор пропускает колебания только в направлении P, а анализатор — только в направлении A. Углы между этими направлениями и осью X обозначим соответственно через α и β. Амплитуду света, вышедшего из поляризатора, примем за единицу. Тогда по выходе из пластинки амплитуда световых колебаний вдоль оси X будет $\cos\alpha$, а вдоль оси Y — $\sin\alpha$. Из этих колебаний анализатор пропустит колебания с амплитудами $a_1 = \cos\alpha\cos\beta$ и $a_2 = \sin\alpha\sin\beta$ соответственно, причем в результате прохождения через пластинку между обоими колебаниями возникнет разность фаз δ. От сложения обоих колебаний получится колебание, амплитуда a которого определяется соотношением

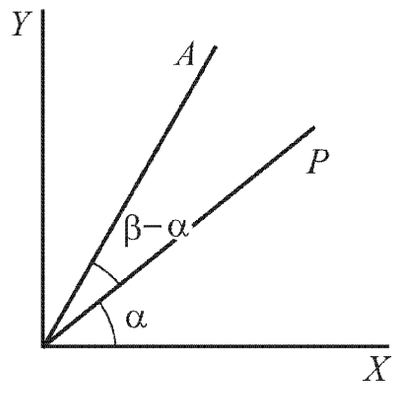


Рис. 279

$$a^2 = a_1^2 + a_2^2 + 2a_1a_2\cos\delta = (a_1 + a_2)^2 - 4a_1a_2\sin^2\frac{\delta}{2}.$$

После подстановки значений a_1 и a_2, приведенных выше, для интенсивности света, проходящего через анализатор, получим

$$I = \cos^2(\beta - \alpha) - \sin 2\alpha\sin 2\beta\sin^2\frac{\delta}{2}. \qquad (79.3)$$

В этой формуле в сочетании с формулой (79.1) содержится объяснение всех особенностей интерференции поляризованного света в параллельных лучах. Пока свет монохроматичен, а толщина пластинки всюду одинакова, все величины в (79.3) постоянны, так что получается *равномерная освещенность* пластинки. При повороте поляризатора или анализатора на 90° выражение (79.3) переходит в

$$I' = \sin^2(\beta - \alpha) + \sin 2\alpha\sin 2\beta\sin^2\frac{\delta}{2}.$$

Следовательно, $I + I' = 1$. Это очевидно и из общих соображений, так как при указанном повороте весь свет, пропускавшийся ранее, будет задержан, а задержавшийся начнет проходить. В частном случае, когда николи параллельны ($\alpha = \beta$),

$$I = 1 - \sin^2 2\alpha\sin^2\frac{\delta}{2}, \qquad (79.4)$$

а когда скрещены —

$$I' = \sin^2 2\alpha\sin^2\frac{\delta}{2}. \qquad (79.5)$$

В белом свете, если его разложить на спектральные составляющие, первое слагаемое в (79.3) остается постоянным, а второе зависит от длины волны и приводит к окрашиванию поля зрения. Если направление P совпадает с одним из главных направлений пластинки ($\alpha = 0$ или $\alpha = \pi/2$), то $I = \cos^2(\beta - \alpha)$, т.е. интенсивность I постоянна и не зависит от длины волны. В этом случае, как было уже отмечено выше, интерференции нет, и в белом свете получается равномерно освещенное неокрашенное поле. Интенсивность его можно менять поворотом николя-анализатора A, но оно все время остается белым.

6. Более сложные явления наблюдаются при *интерференции сходящихся поляризованных лучей*. Обычно применяемая схема для наблюдения таких явлений представлена на рис. 280. Парал-

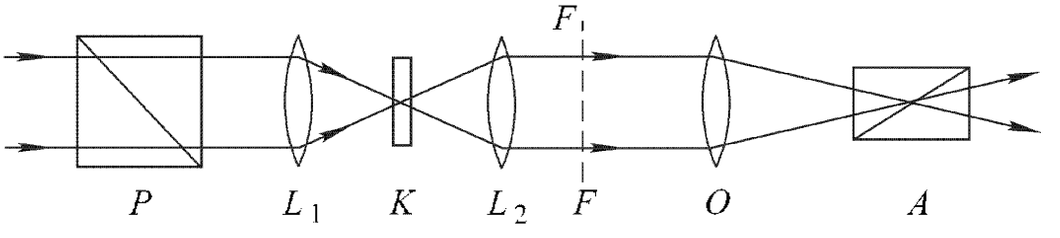


Рис. 280

лельный пучок линейно поляризованного света линзой L_1 превращается в сходящийся и концентрируется на кристаллической пластинке K. После прохождения через пластинку K каждый луч разделяется на два параллельных луча, между которыми возникает разность хода. При наложении таких лучей свет становится поляризованным эллиптически, причем разность хода между лучами, а с ней и характер эллиптической поляризации зависят от угла наклона падающего луча. Линза L_2, одинаковая с L_1, делает лучи снова параллельными. Распределение светового поля в фокальной плоскости FF линзы проецируется объективом O и николем-анализатором A на экран, где и наблюдается интерференционная картина.

Чтобы составить представление о форме получающихся интерференционных полос, введем понятие *изохроматической поверхности* (поверхности одинакового цвета). Возьмем в неограниченной кристаллической среде произвольную точку O и представим, что через нее во всевозможных направлениях и притом в одинаковых фазах проходят плоские волны. В каждом направлении волновой нормали распространяются две волны, линейно поляризованные во взаимно перпендикулярных плоскостях. На расстоянии r от точки O между ними возникнет разность хода $\Delta = (n_2 - n_1)r$, где n_1 и n_2 — показатели преломления этих волн. Фиксировав Δ, будем откладывать на волновых нормалях отрезки с длиной $r = \Delta/(n_2 - n_1)$. Геометрическое место концов

таких отрезков есть поверхность равной разности фаз между волнами, которые могут распространяться вдоль одной и той же волновой нормали. Она и называется *изохроматической поверхностью*. Придавая Δ всевозможные значения, получим *семейство изохроматических поверхностей* с общим центром O. Линии, получающиеся от сечения такого семейства плоскостями, называются *изохроматическими линиями*, или *изохроматами*. Очевидно, изохроматы есть *линии постоянной разности фаз* между соответствующими волнами, могущими распространяться вдоль одной и той же волновой нормали.

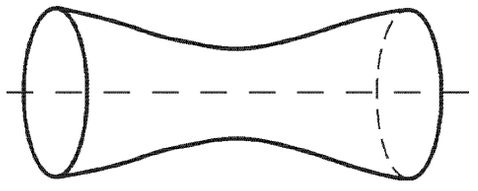


Рис. 281

Для одноосного кристалла изохроматическая поверхность есть поверхность вращения вокруг оптической оси (рис. 281). В направлении оптической оси она уходит в бесконечность, так как для этого направления оба показателя преломления n_1 и n_2 совпадают. В двуосных кристаллах таких направлений два (рис. 282). Это есть оптические оси, точнее — оси нормалей (см. § 80). Мы не будем заниматься выводом уравнения изохроматической поверхности, а ограничимся качественными соображениями.

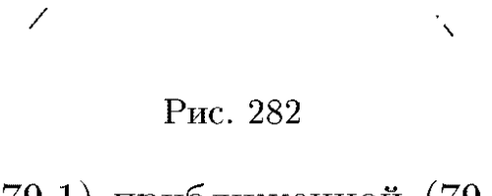


Рис. 282

Напомним, что интерферирующие волны возникают при двойном преломлении одной и той же падающей волны. Волновые нормали получившихся двух волн внутри кристаллической пластинки несколько отличаются друг от друга по направлению. Однако таким различием мы пренебрежем, как это уже делалось при замене точной формулы (79.1) приближенной (79.2). В этом приближении, как показывается ниже, интерференционные полосы можно отождествить с определенным семейством изохромат. В самом деле, представим, что на передней плоскости кристаллической пластинки K (см. рис. 280) помещена маленькая диафрагма. (В такой диафрагме нет надобности, если задний фокус линзы L_1 находится на передней поверхности пластинки.) Примем центр этой диафрагмы за центр O семейства изохроматических поверхностей. Задняя плоскость пластинки K пересечет это семейство вдоль изохроматических линий. Последние в указанном выше приближении будут совпадать с интерференционными линиями на зад-

ней поверхности пластинки K. Если оптическая система дает изображение этой поверхности на экране, то при этом получится и подобное изображение рассматриваемой системы интерференционных линий. При иной фокусировке, из-за преломления волновых нормалей при выходе из пластинки K, полного подобия между изохроматами и интерференционными линиями на экране не будет. Однако это обстоятельство не сказывается существенно на виде интерференционных полос. Изохроматы практически всегда качественно верно передают форму интерференционных полос.

Форма изохромат и интерференционных полос определяется симметрией кристалла и ориентацией его оптических осей. На рис. 283 показаны интерференционные полосы в случае пластинки одноосного кристалла, вырезанной перпендикулярно к оптической оси. В соответствии с осевой симметрией они имеют форму концентрических кругов. Картина получена в скрещенных николях. Кольца пересечены темным крестом. Происхождение его весьма простое. На пластинку K падает линейно поляризованный свет. Рассмотрим плоскость главного сечения и

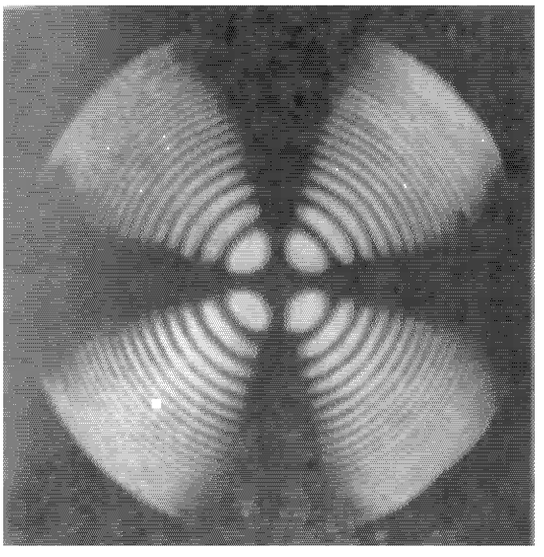


Рис. 283

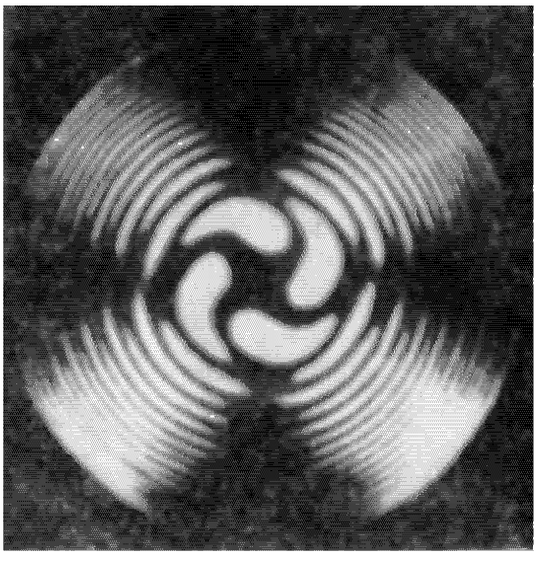


Рис. 284

плоскость, перпендикулярную к ней. Пусть обе плоскости пересекаются вдоль оси вращения системы. Лучи в этих плоскостях не испытывают двойного преломления, интерференции в таких лучах не возникает. После прохождения через пластинку обе системы лучей остаются линейно поляризованными с сохранением прежнего направления колебаний. Лучи с такими колебаниями гасятся николем-анализатором. Вокруг прямых, по которым рассматриваемые плоскости пересекаются с плоскостью экрана, и формируется темный крест. Если николи перевести в параллельное положение, то вся картина сменяется дополнительной, в частности, черный крест заменяется белым.

Если в предыдущем опыте пластинку K заменить двумя сложенными вместе пластинками из правого и левого кварца, вырезанными перпендикулярно к оптической оси, то получаются спиралеобразные интерференционные картины, называемые *спиралями Эйри* (рис. 284). Их происхождение связано с тем, что кварц *вращает плоскость поляризации*. Для правого кварца это вращение происходит в одну сторону, а для левого — в противоположную.

На рис. 285 показан вид изохромат для пластинки одноосного кристалла, вырезанной параллельно оптической оси, а на рис. 286 — для пластинки двуосного кристалла, вырезанной перпендикулярно к биссектрисе угла между оптическими осями.

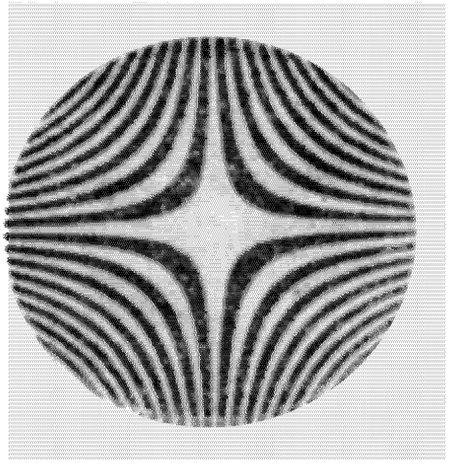


Рис. 285

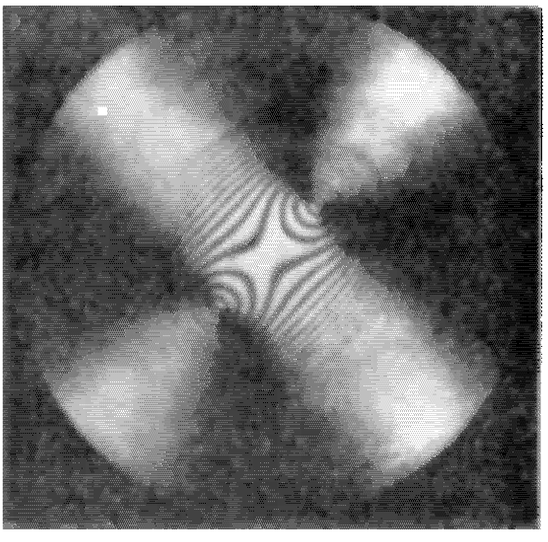


Рис. 286

7. Интерференция поляризованных лучей дает удобный и чувствительный метод обнаружения и исследования оптической анизотропии тел во всех случаях, когда она мала. Для этого обычно применяется установка в параллельных лучах с скрещенными николями, схематически изображенная на рис. 278. Вместо кристаллической пластинки ставится исследуемый образец. В отсутствие образца свет через систему не проходит. При введении образца, обладающего анизотропией, поле зрения просветляется и в случае белого света становится окрашенным. Таким методом исследуются различные случаи *искусственной анизотропии*.

Брюстером в 1815 г. было открыто *двойное преломление при механических деформациях* изотропных тел. Среда может состоять из анизотропных молекул. В отсутствие деформаций они ориентированы хаотически, вследствие чего среда макроскопически изотропна. При деформациях может возникнуть преимущественная ориентация и изменение расположения молекул в пространстве. Это и ведет к возникновению двойного преломле-

ния. В простейшем случае одностороннего растяжения или сжатия тело ведет себя как одноосный кристалл с оптической осью вдоль растяжения или сжатия. Как показали опыты, разность $n_o - n_e$ между обыкновенным и необыкновенным показателями преломления пропорциональна приложенному напряжению, а коэффициент пропорциональности зависит от длины волны. Эффект сохраняется и при остаточных деформациях. Он применяется для исследования упругих напряжений и остаточных деформаций в телах. Двойное преломление возникает при течении жидкостей с анизотропными молекулами (эффект Максвелла), а также при внесении тел в электрическое или магнитное поля (см. § 90 и 91).

ЗАДАЧИ

1. Почему никогда не может получиться интерференция обыкновенного и необыкновенного лучей, вышедших из пластинки двоякопреломляющего кристалла, настолько тонкой, что она не дает заметного разделения лучей?

2. Почему демонстрационные опыты по интерференции поляризованных лучей удобнее делать с тонкими, а не с толстыми пластинками? Почему даже с тонкими пластинками из исландского шпата трудно получить интерференционную картину в белом свете?

О т в е т. У исландского шпата очень велика разность обыкновенного и необыкновенного показателей преломления, так что даже в тонких пластинках получаются большие разности хода, при которых интерференция в белом свете невозможна.

3. Две толстые пластинки одноосного кристалла, одинаково ориентированные и весьма мало отличающиеся по толщине, в скрещенных николях дают порознь белый свет. Почему в тех же условиях может получиться окрашивание, если повернуть одну пластинку относительно другой на 90°?

Р е ш е н и е. Когда пластинки одинаково ориентированы, вносимая ими разность хода велика, т.е. соответствует высокому порядку интерференции. В таких условиях интерференция в белом свете наблюдаться не может. Если одну из пластинок повернуть относительно другой на 90°, то вносимая ими разность хода будет такая же, какая вносится пластинкой с толщиной, равной разности толщин рассматриваемых пластинок. При малой разности толщин разность хода может сделаться настолько малой, что станет возможна интерференция в белом свете.

4. Кварцевая пластинка толщиной в 1 мм вырезана перпендикулярно к оптической оси. Как определить, из право- или левовращающего кварца сделана пластинка, имея в своем распоряжении два николя и источник: 1) монохроматического света; 2) белого света?

О т в е т. 1) Если поместить пластинку кварца, вырезанную перпендикулярно к оптической оси, между скрещенными николями и осветить систему монохроматическим светом, то она будет пропускать свет; повернув анализатор на угол, меньший 90°, можно снова погасить свет. Если при этом наблюдатель должен вращать анализатор по направлению часовой стрелки, то кварц будет правовращающий, если же против часовой стрелки, то левовращающий.

2) Если осветить систему белым светом, то пластинка будет казаться окрашенной. Вращательная способность увеличивается с уменьшением

длины волны. Поэтому если вращать анализатор по часовой стрелке, то для правовращающего кварца окраска будет меняться в сторону коротких длин волн спектра. Для левовращающего кварца порядок изменения окраски будет обратный.

§ 80. Нормальные скорости и поляризация волн в двуосных кристаллах

1. Перейдем теперь к исследованию распространения волн в *оптически двуосных кристаллах*. В общем случае вектор $\mathbf{D}$ может зависеть не только от вектора $\mathbf{E}$, но и от его пространственных производных. Это явление называется *пространственной дисперсией* (см. § 96). В слабых полях такая зависимость, конечно, может считаться линейной. Для плоских монохроматических волн дифференцирование $\mathbf{E}$ по координатам x, y, z сводится к умножению его проекций на $-ik_x$, $-ik_y$, $-ik_z$. В этом случае зависимость от пространственных производных можно учесть прежней формулой (75.2), если диэлектрический тензор ε_{kj} считать комплексным. Формально так можно поступать и в случае неплоских волн. Однако волны должны предполагаться монохроматическими.

Для непоглощающих сред диэлектрический тензор должен быть *эрмитовым*, т.е. $\varepsilon_{kj} = \varepsilon^*_{jk}$. Действительно, для производной плотности электромагнитной энергии u по времени электродинамика дает

$$\frac{\partial u}{\partial t} = \frac{1}{4\pi}(\mathbf{E}\dot{\mathbf{D}} + \mathbf{H}\dot{\mathbf{B}})$$

(см. т. III, § 84). В случае монохроматического поля в непоглощающей среде среднее значение этой производной, согласно закону сохранения энергии, должно равняться нулю. Если пользоваться комплексной формой монохроматического поля, то это условие запишется в виде $(\mathbf{E}\dot{\mathbf{D}}^* + \mathbf{H}\dot{\mathbf{B}}^*) + \text{компл. сопр.} = 0$. А так как мы пренебрегаем различием между $\mathbf{B}$ и $\mathbf{H}$, то $\mathbf{E}\dot{\mathbf{D}}^* + \text{компл. сопр.} = 0$. Отсюда с учетом соотношений $\dot{\mathbf{D}} = i\omega\mathbf{D}$, $\dot{\mathbf{D}}^* = -i\omega\mathbf{D}^*$ получаем: $\mathbf{E}\mathbf{D}^* - \mathbf{E}^*\mathbf{D} = 0$, или

$$\sum E_\alpha \varepsilon^*_{\alpha\beta} E^*_\beta - \sum E^*_\alpha \varepsilon_{\alpha\beta} E_\beta = 0.$$

Заменим в первой сумме немой индекс α на β и наоборот. Тогда

$$\sum (\varepsilon_{\alpha\beta} - \varepsilon^*_{\beta\alpha}) E^*_\alpha E_\beta = 0.$$

Это соотношение должно выполняться для любого поля $\mathbf{E}$, что возможно тогда и только тогда, когда $\varepsilon_{\alpha\beta} = \varepsilon^*_{\beta\alpha}$. Действительно, пусть все компоненты вектора $\mathbf{E}$, за исключением одной E_α,

равны нулю. Тогда предыдущее соотношение переходит в $(\varepsilon_{\alpha\alpha} - -\varepsilon^*_{\alpha\alpha})E_\alpha E^*_\alpha = 0$, откуда $\varepsilon_{\alpha\alpha} = \varepsilon^*_{\alpha\alpha}$. Пусть теперь отличны от нуля две компоненты E_α и E_β, а третья компонента равна нулю. Тогда

$$(\varepsilon_{\alpha\beta} - \varepsilon^*_{\beta\alpha})E^*_\alpha E_\beta + (\varepsilon_{\beta\alpha} - \varepsilon^*_{\alpha\beta})E_\alpha E^*_\beta = 0.$$

Полагая здесь $E_\alpha = E_\beta$, получим

$$(\varepsilon_{\alpha\beta} - \varepsilon^*_{\beta\alpha}) + (\varepsilon_{\beta\alpha} - \varepsilon^*_{\alpha\beta}) = 0.$$

Полагая же $E_\beta = iE_\alpha$, найдем

$$(\varepsilon_{\alpha\beta} - \varepsilon^*_{\beta\alpha}) - (\varepsilon_{\beta\alpha} - \varepsilon^*_{\alpha\beta}) = 0.$$

Из этого и предыдущего соотношений следует: $\varepsilon_{\alpha\beta} = \varepsilon^*_{\beta\alpha}$. Таким образом, соотношение $\varepsilon_{\alpha\beta} = \varepsilon^*_{\beta\alpha}$ справедливо как для одинаковых, так и для разных индексов α и β, т.е. для непоглощающих кристаллов тензор $\varepsilon_{\alpha\beta}$ эрмитов. Для поглощающих кристаллов он не эрмитов.

Допустим теперь, что среда не обладает пространственной дисперсией или этим явлением, ввиду его малости, можно пренебречь. Тогда величины $\varepsilon_{\alpha\beta}$ вещественны, а потому $\varepsilon_{\alpha\beta} = \varepsilon_{\beta\alpha}$, т.е. тензор $\varepsilon_{\alpha\beta}$ будет *симметричен*. В дальнейшем мы ограничимся этим случаем. Кроме того, будем предполагать, что все диагональные элементы $\varepsilon_{\alpha\alpha}$ *положительны*. Только тогда среда будет *прозрачной*, т.е. плоские волны в ней будут распространяться без затухания. В противном случае возникнет затухание без поглощения, как это имеет место, например, в плазме (см. § 87).

Всякий симметричный тензор можно привести к так называемому *диагональному виду*, т.е. найти такую систему прямоугольных координат, в которой недиагональные компоненты тензора обращаются в нуль. Диагональные компоненты тензора в этой системе координат условимся обозначать через ε_x, ε_y, ε_z, т.е. характеризовать их не двойными, а только единичными индексами x, y, z. В рассматриваемой системе материальные уравнения имеют вид

$$D_\alpha = \varepsilon_\alpha E_\alpha. \tag{80.1}$$

Координатные оси, относительно которых тензор $\varepsilon_{\alpha\beta}$ диагонален, называются *главными осями тензора* или *диэлектрическими осями кристалла*, а величины ε_x, ε_y, ε_z — *главными диэлектрическими проницаемостями*. Эти оси мы и примем за координатные оси, причем названия осей X, Y, Z установим так, чтобы соблюдались неравенства

$$\varepsilon_x \leqslant \varepsilon_y \leqslant \varepsilon_z. \tag{80.2}$$

Так как компоненты тензора $\varepsilon_{\alpha\beta}$ могут зависеть от длины волны λ, могут зависеть от λ и направления диэлектри-

ческих осей. Это явление, называемое *дисперсией диэлектрических осей*, действительно встречается в триклинных и моноклинных кристаллах, характеризующихся наиболее низкой симметрией.

Отметим, наконец, что угол α между векторами $\mathbf{E}$ и $\mathbf{D}$ всегда острый. Это вытекает из того, что скалярное произведение $(\mathbf{ED}) = ED\cos\alpha$ пропорционально плотности электрической энергии, а она существенно положительна.

2. Обратимся теперь к исследованию плоских волн (75.3) в прозрачных кристаллах в общем виде. Фиксируем направление волновой нормали $\mathbf{N}$ и определим, какие плоские волны могут распространяться в этом направлении. Используя материальные уравнения (80.1), перепишем соотношение (75.7) в следующем виде:

$$(v^2 - a_\alpha^2)D_\alpha = -c^2(\mathbf{NE})N_\alpha, \tag{80.3}$$

где введено обозначение

$$a_\alpha = c/\sqrt{\varepsilon_\alpha}. \tag{80.4}$$

Разделив на $v^2 - a_\alpha^2$, получим

$$D_\alpha = -\frac{c^2}{v^2 - a_\alpha^2}(\mathbf{NE})N_\alpha. \tag{80.5}$$

Умножим обе части этого соотношения на N_α и просуммируем по α. Тогда

$$\sum_\alpha D_\alpha N_\alpha \equiv (\mathbf{DN}) = -c^2(\mathbf{NE})\sum_\alpha \frac{N_\alpha^2}{v^2 - a_\alpha^2} = 0,$$

так как $(\mathbf{DN}) = 0$. Скалярное произведение $(\mathbf{NE})$, вообще говоря, отлично от нуля. Поэтому

$$\sum \frac{N_\alpha^2}{v^2 - a_\alpha^2} = 0. \tag{80.6}$$

В развернутом виде

$$\frac{N_x^2}{v^2 - a_x^2} + \frac{N_y^2}{v^2 - a_y^2} + \frac{N_z^2}{v^2 - a_z^2} = 0, \tag{80.7}$$

а после освобождения от знаменателей

$$F(v^2) \equiv N_x^2(v^2 - a_y^2)(v^2 - a_z^2) + N_y^2(v^2 - a_z^2)(v^2 - a_x^2) + \\ + N_z^2(v^2 - a_x^2)(v^2 - a_y^2) = 0. \tag{80.8}$$

Хотя уравнение (80.8) мы и получили преобразованием уравнения (80.7), в действительности оно обладает большей общностью. Как видно из вывода, при получении (80.7) надо было вводить предположение, что $v^2 - a_\alpha^2 \neq 0$ и $(\mathbf{NE}) \neq 0$. Уравнение (80.7) теряет смысл, когда по крайней мере одна из разностей

$v^2 - a_\alpha^2$ обращается в нуль Уравнение же (80.8) остается справедливым и в этом случае, как показывает несложное математическое исследование, которое мы опускаем.

3. Уравнение (80.6), или (80.8), называется *законом Френеля для нормальной скорости распространения световых волн в кристалле*. Если задать направление $\mathbf{N}$, то из этих уравнений можно определить нормальную скорость v. Уравнение (80.8) второй степени относительно v^2. Докажем, что оно имеет вещественные и притом положительные корни. Для прозрачных кристаллов главные диэлектрические проницаемости, а с ними и величины a_α^2, существенно положительны. При этом ввиду условия (80.2)

$$a_x \geqslant a_y \geqslant a_z. \tag{80.9}$$

Придавая в функции $F(v^2)$ аргументу v^2 значения a_x^2, a_y^2, a_z^2, придем к неравенствам:

$$\begin{aligned} F(a_x^2) &= N_x^2(a_x^2 - a_y^2)(a_x^2 - a_z^2) \geqslant 0, \\ F(a_y^2) &= N_y^2(a_y^2 - a_z^2)(a_y^2 - a_x^2) \leqslant 0, \\ F(a_z^2) &= N_z^2(a_z^2 - a_x^2)(a_z^2 - a_y^2) \geqslant 0. \end{aligned}$$

Из них видно, что функция $F(v^2)$ дважды меняет знак: один раз между a_x^2 и a_y^2, другой — между a_y^2 и a_z^2. Следовательно, уравнение $F(v^2) = 0$ имеет два вещественных положительных корня: v_1^2 и v_2^2, причем

$$a_x \geqslant v_1 \geqslant a_y \geqslant v_2 \geqslant a_z. \tag{80.10}$$

Отсюда следует, что в направлении $\mathbf{N}$ могут распространяться две волны: одна с нормальной скоростью v_1, а другая с v_2. В частных случаях скорости v_1 и v_2 могут совпадать.

Если скорости v_1 и v_2 различны, то каждая из волн будет поляризована линейно. Это следует из соотношения

$$D_x : D_y : D_z = \frac{N_x}{v^2 - a_x^2} : \frac{N_y}{v^2 - a_y^2} : \frac{N_z}{v^2 - a_z^2}, \tag{80.11}$$

которое получается из (80.5) и в котором под v следует понимать либо v_1, либо v_2. Все величины, стоящие в правой части (80.11), вещественны. Значит, между компонентами D_x, D_y, D_z нет сдвигов фаз, отличающихся от 0 или π, а потому волна поляризована линейно.

Докажем, что если скорости v_1 и v_2 различны, то векторы $\mathbf{D}$ обеих волн, которые могут распространяться в направлении $\mathbf{N}$, взаимно перпендикулярны. Отмечая величины, относящиеся к одной из волн, индексом 1, а к другой — индексом 2, из (75.7)

получим

$$v_1^2\mathbf{D}_1 - c^2\mathbf{E}_2 = -c^2(\mathbf{N}\mathbf{E}_1)\mathbf{N},$$
$$v_2^2\mathbf{D}_2 - c^2\mathbf{E}_2 = -c^2(\mathbf{N}\mathbf{E}_2)\mathbf{N}.$$

Умножим первое уравнение скалярно на $\mathbf{D}_2$, второе на $\mathbf{D}_1$ и вычтем одно уравнение из другого. Так как $(\mathbf{D}_1\mathbf{N}) = (\mathbf{D}_2\mathbf{N}) = 0$, то в результате получим

$$(v_1^2 - v_2^2)\mathbf{D}_1\mathbf{D}_2 = c^2(\mathbf{E}_1\mathbf{D}_2 - \mathbf{E}_2\mathbf{D}_1).$$

Но, очевидно, $\mathbf{E}_1\mathbf{D}_2 = \mathbf{E}_2\mathbf{D}_1$, так как каждое из этих скалярных произведений равно $\sum \varepsilon_\alpha E_{1\alpha}E_{2\alpha}$. Следовательно,

$$(v_1^2 - v_2^2)\mathbf{D}_1\mathbf{D}_2 = 0.$$

Отсюда при $v_1 \neq v_2$ следует $(\mathbf{D}_1\mathbf{D}_2) = 0$, что и требовалось доказать. Аналогично докажем, что $(\mathbf{B}_1\mathbf{B}_2) = 0$.

Итак, в *каждом направлении в кристалле могут распространяться две линейно поляризованные волны, скорости которых, вообще говоря, различны. Обе волны поперечны относительно векторов* $\mathbf{D}$ *и* $\mathbf{B}$. *Векторы* $\mathbf{D}$ (*а также* $\mathbf{B}$) *в этих волнах взаимно перпендикулярны. Относительно вектора* $\mathbf{E}$ *обе волны в кристалле не поперечны, за исключением тех случаев, когда вектор* $\mathbf{E}$ *параллелен одной из диэлектрических осей кристалла.* Однако деление волн на обыкновенную и необыкновенную возможно *только для одноосных кристаллов.* В общем случае такое деление смысла не имеет — обе волны в кристалле ведут себя как «необыкновенные».

Чтобы выяснить физический смысл постоянных a_α, направим вектор $\mathbf{E}$ вдоль диэлектрической оси α. Тогда $\mathbf{D} = \varepsilon_\alpha\mathbf{E}$, и уравнение (75.8) перейдет в

$$v^2 = c^2\,\frac{\varepsilon_\alpha E^2}{\varepsilon_\alpha^2 E^2} = \frac{c^2}{\varepsilon_\alpha} = a_\alpha^2,$$

откуда $v = a_\alpha$. Таким образом, *величина* a_α *есть нормальная скорость распространения волны, у которой электрический вектор параллелен диэлектрической оси* α. Это утверждение становится очевидным, если заметить, что в частном случае, когда электрическое поле параллельно диэлектрической оси, уравнения распространения волн в кристалле не отличаются от уравнений в изотропных средах.

Величины a_α называются *главными скоростями распространения света в кристалле.* Наряду с главными скоростями, для характеристики оптических свойств кристаллов пользуются также *главными показателями преломления*, которые определяются выражениями

$$n_\alpha = c/a_\alpha = \sqrt{\varepsilon_\alpha}. \qquad (80.12)$$

Для волны произвольного направления показатель преломления кристалла определяется выражением

$$n = c/v. \tag{80.13}$$

Его значение, как видно из (75.8), однозначно определяется направлением вектора **D** или **E**. Каждому направлению нормали **N** соответствуют *два* значения показателя преломления в соответствии с двумя возможными поляризациями волны.

4. Для исследования уравнения Френеля применим геометрический метод. Из какой-то точки O в различных направлениях будем проводить прямые и на них откладывать отрезки, длины которых равны значениям нормальных скоростей в этих направлениях. Геометрическое место концов таких отрезков называется *поверхностью нормалей*. В кристалле каждому направлению нормали соответствуют два значения скорости. Поэтому поверхность нормалей в кристалле будет *двойной поверхностью*, т.е. состоит из двух слоев. Она представляет собой поверхность *шестого порядка* и имеет очень сложный вид. Чтобы составить представление о поверхности нормалей электромагнитных волн в кристалле, рассмотрим сечения ее координатными плоскостями XY, YZ, ZX.

Сечение плоскостью XY. Волновая нормаль лежит в плоскости XY, т.е. $N_z = 0$. Уравнение Френеля (80.8) принимает вид

$$(v^2 - a_z^2)[N_x^2(v^2 - a_y^2) + N_y^2(v^2 - a_x^2)] = 0.$$

Из него получаем два значения нормальных скоростей:

$$v_2 = a_z, \quad v_1^2 = N_x^2 a_y^2 + N_y^2 a_x^2. \tag{80.14}$$

Скорость v_2 не зависит от направления **N**. Ей соответствует круговое сечение поверхности нормалей (рис. 287). Скорость v_1 изменяется с изменением направления **N**. Ей соответствует сечение поверхности нормалей, имеющее *форму овала*. Из уравнений (80.14) следует: $v_1 \geqslant v_2$, так что круг находится целиком внутри овала. Вектор **D** должен быть перпендикулярен к **N**. Из соображений симметрии ясно, что вектор **D** одной волны параллелен оси Z, а вектор **D** другой волны параллелен плоскости XY. Первому направлению вектора **D** соответствует *круговое сечение* поверхности нормалей, второму — *овальное*.

Сечение плоскостью YZ. Волновая нормаль **N** лежит в плоскости YZ, т.е. $N_x = 0$. Уравнение Френеля принимает вид

$$(v^2 - a_x^2)[N_y^2(v^2 - a_z^2) + N_z^2(v^2 - a_y^2)] = 0.$$

Оно дает два значения нормальных скоростей:

$$v_1 = a_x, \quad v_2^2 = N_y^2 a_z^2 + N_z^2 a_y^2. \tag{80.15}$$

Скорость v_1 не зависит от направления $\mathbf{N}$. Ей соответствует *круговое сечение* поверхности нормалей и вектор $\mathbf{D}$, параллельный оси X. Скорость v_2 изменяется с изменением направления $\mathbf{N}$. Ей соответствует *овальное сечение* поверхности нормалей и вектор $\mathbf{D}$, параллельный плоскости YZ. Оно целиком помещается внутри круга, так как $v_2 \leqslant v_1$, как это следует из уравнений (80.15).

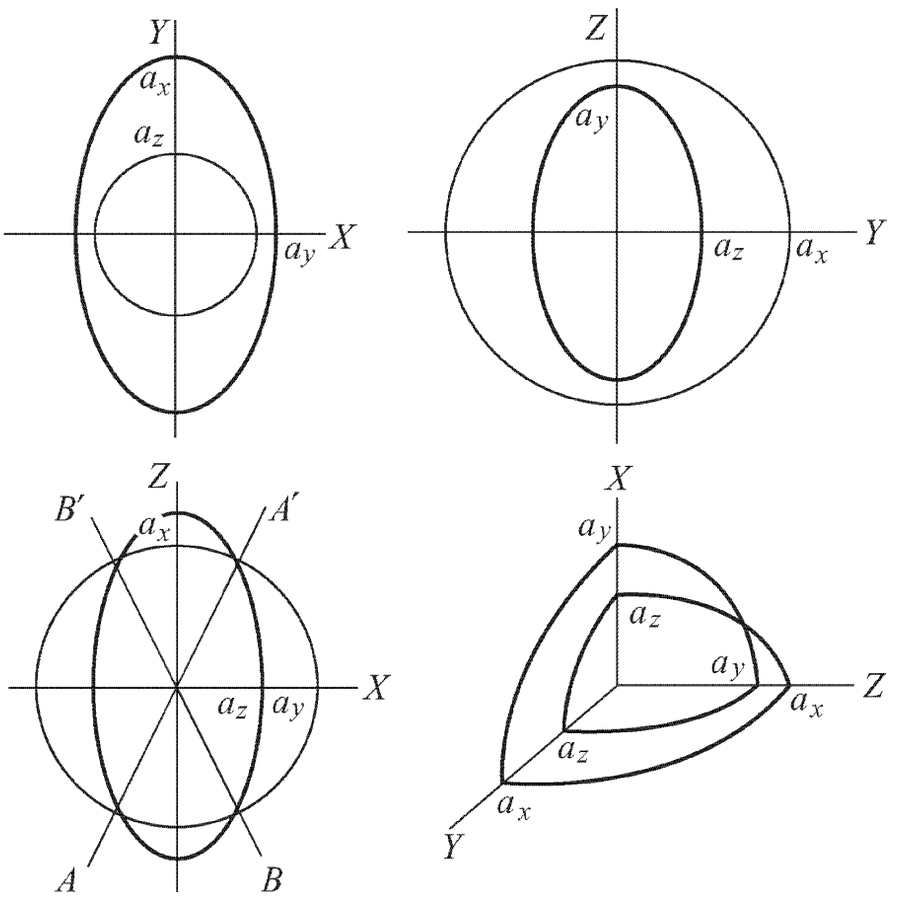


Рис. 287

С е ч е н и е п л о с к о с т ь ю ZX. Волновая нормаль $\mathbf{N}$ лежит в плоскости ZX, т.е. $N_y = 0$. Уравнение Френеля принимает вид

$$(v^2 - a_y^2)[N_z^2(v^2 - a_x^2) + N_x^2(v^2 - a_z^2)] = 0$$

и дает два значения нормальных скоростей:

$$v_1 = a_y, \quad v_2^2 = N_z^2 a_x^2 + N_x^2 a_z^2. \tag{80.16}$$

Скорость v_1 не зависит от направления $\mathbf{N}$. Ей соответствует *круговое сечение* поверхности нормалей и вектор $\mathbf{D}$, параллельный оси Y. Скорость v_2 изменяется с изменением направления $\mathbf{N}$. Ей соответствует *овальное сечение* поверхности нормалей и вектор $\mathbf{D}$, параллельный плоскости ZX.

5. Третий из рассмотренных случаев существенно отличается от первых двух. В первых двух случаях овал и круг не пересекаются. В третьем случае они пересекаются в четырех

точках (рис. 287). Это означает, что в плоскости ZX имеются *два направления* AA' и BB', симметричные относительно оси Z, вдоль которых обе волны распространяются с одной и той же нормальной скоростью. *Направления, вдоль которых совпадают нормальные скорости волн, называются оптическими осями второго рода, осями нормалей или бинормалями.*

Если в кристалле все три главные скорости a_x, a_y, a_z *различны, то в нем существуют две и только две оптические оси второго рода.* Действительно, если вектор $\mathbf{N}$ направлен вдоль оптической оси второго рода, то должно быть $v_1 = v_2$. Ввиду соотношений (80.10), это возможно только тогда, когда $v_1 = v_2 = a_y$. Но тогда уравнение (80.8) дает $N_y^2(a_y^2 - a_z^2)(a_y^2 - a_x^2) = 0$. Так как по предположению $a_x \neq a_y \neq a_z$, то отсюда следует, что $N_y = 0$. Это значит, что оптические оси лежат в плоскости ZX. Но в этой плоскости, как показано выше, имеются две и только две оптические оси второго рода. Они симметрично расположены относительно оси Z и наклонены к ней под некоторым углом β. Для нахождения β в уравнение $v_1 = v_2$ подставим значения v_1 и v_2 из формул (80.16). Получим

$$a_y^2 = N_z^2 a_x^2 + N_x^2 a_z^2, \qquad \text{или} \qquad a_y^2(N_x^2 + N_z^2) = N_z^2 a_x^2 + N_x^2 a_z^2,$$

откуда

$$\operatorname{tg}\beta = \frac{N_x}{N_z} = \sqrt{\frac{a_x^2 - a_y^2}{a_y^2 - a_z^2}}. \tag{80.17}$$

Теперь оправдан термин «оптически двуосный кристалл», которым мы уже пользовались.

Если две из трех главных скоростей совпадают между собой ($a_x = a_y$ или $a_y = a_z$), то оптические оси сливаются в одну ось, параллельную оси Z (когда $a_x = a_y$) или оси X (когда $a_y = a_z$). Кристалл становится *оптически одноосным.* Наконец, если все три главные скорости одинаковы, то любое направление в кристалле обладает свойством оптической оси. В таких кристаллах плоские волны, независимо от их поляризации и направления, распространяются с одной и той же скоростью — кристаллы в оптическом отношении ведут себя как *изотропные среды.* К ним относятся кристаллы кубической системы [1]).

[1]) Необходимо, однако, отметить, что при наличии *пространственной дисперсии* кристаллы кубической системы могут быть оптически анизотропными. На эту возможность указывал еще Г.А. Лоренц. Только в 1960 г. Е.Ф. Гросс и А.А. Каплянский, исследуя спектры поглощения на монокристаллических образцах Cu_2O, экспериментально обнаружили необычное для кубических кристаллов явление анизотропного поглощения света.

ЗАДАЧИ

1. Как надо ориентировать пластинку из двуосного кристалла, чтобы получить на кристалл-рефрактометре три главных показателя преломления?

О т в е т. Перпендикулярно к любой из диэлектрических осей кристалла.

2. Исходя из соображений симметрии, показать, что все кристаллы три-, тетра- и гексагональной систем оптически одноосны.

§ 81. Лучи, волновые нормали и связь между ними

1. Распространение света в кристаллах, как и любых волн в анизотропных средах, характеризуется замечательной *двойственностью*, или *взаимностью*. Она обусловлена тем, что в анизотропных средах каждой волновой нормали соответствует луч, т.е. прямая, вдоль которой происходит распространение энергии волны. Поскольку энергия распространяется с групповой скоростью, для исследования свойств лучей и обоснования самого понятия луча надо вычислить групповую скорость в анизотропной среде. В этом случае такую скорость называют также *лучевой скоростью*. Для ее вычисления воспользуемся формулой (8.16), подставив в нее $\omega = kv(k)$. Дифференцируя по k_i и учитывая, что $\partial k/\partial k_i = k_i/k$, получим

$$\frac{\partial\omega}{\partial k_i} = v\frac{k_i}{k} + k\frac{\partial v}{\partial k_i}.$$

Отсюда для вектора групповой скорости находим

$$\mathbf{u} = v\mathbf{N} + k\frac{\partial v}{\partial \mathbf{k}}, \tag{81.1}$$

где $\mathbf{N} = \mathbf{k}/k$ — единичный вектор волновой нормали, а v — *нормальная скорость*, т.е. скорость распространения фазы в направлении волновой нормали.

Групповая скорость $\mathbf{u}$ в анизотропной среде отличается от нормальной скорости $\mathbf{v}$ добавочным слагаемым $k\,\partial v/\partial\mathbf{k}$. Это слагаемое в свою очередь содержит составляющую вдоль нормали $\mathbf{N}$. Чтобы определить ее, заметим, что $\mathbf{k} = k\mathbf{N}$, а потому указанная составляющая равна $\mathbf{k}\,\partial v/\partial\mathbf{k}$. Поэтому для самой групповой скорости $\mathbf{u}_N$ в направлении волновой нормали $\mathbf{N}$ можно написать

$$\mathbf{u}_N = \left(v + k\frac{\partial v}{\partial k}\right)\mathbf{N} = \left(v - \lambda\frac{\partial v}{\partial\lambda}\right)\mathbf{N}. \tag{81.2}$$

Этот результат совпадает с формулой Рэлея (8.6) для групповой скорости в изотропной среде. Этого и следовало ожидать, так как он относится не ко всему вектору групповой скорости,

а только к его проекции на направление волновой нормали. А вычисление такой проекции можно провести в точности так же, как и вычисление групповой скорости в изотропной среде.

Отличие u_N от v обусловлено *дисперсией волн*, т.е. зависимостью нормальной скорости v от частоты ω. Дисперсия в равной мере свойственна и изотропным, и анизотропным средам. Специфика распространения световых волн в кристаллах обусловлена не столько дисперсией, сколько *отличием направлений волновых нормалей и лучей*. Чтобы не вводить излишних усложнений, пренебрежем совсем дисперсией, т.е. будем считать кристаллы *недиспергирующими*. Тогда $\partial v/\partial\lambda = 0$, а потому $\mathbf{u}_N = v\mathbf{N}$, или $u_N = v$. Но $u_N = (\mathbf{uN})$, так что

$$(\mathbf{uN}) = v. \tag{81.3}$$

2. Наряду с поверхностью нормалей, введенной в предыдущем параграфе, введем еще *лучевую поверхность*, называемую иногда также *волновой поверхностью*. Для этого из произвольной точки O во всевозможных направлениях будем проводить лучи и откладывать на них величины лучевой скорости в этих направлениях. Геометрическое место концов отложенных отрезков есть замкнутая поверхность, которая и называется *лучевой поверхностью*.

Если лучевую поверхность и поверхность нормалей строить из общего центра O, то между этими двумя поверхностями существует простая и важная связь. Для установления этой связи умножим формулу (81.3) на k и придадим ей вид

$$(\mathbf{uk}) = \omega. \tag{81.4}$$

Отсюда следует, что бесконечно малые изменения величин $\mathbf{u}$, $\mathbf{k}$, ω связаны соотношением $(\mathbf{u}\,\delta\mathbf{k}) + (\mathbf{k}\,\delta\mathbf{u}) = \delta\omega$. По определению групповой скорости $(\mathbf{u}\,\delta\mathbf{k}) = \delta\omega$. Следовательно, $(\mathbf{k}\,\delta\mathbf{u}) = 0$, или

$$(\mathbf{N}\,\delta\mathbf{u}) = 0. \tag{81.5}$$

Но $\mathbf{u}$ есть радиус-вектор лучевой поверхности, а потому всякий бесконечно малый вектор $\delta\mathbf{u}$ лежит в плоскости, касательной к этой поверхности в соответствующей точке касания. Поэтому из формул (81.5) и (81.3) следует, что *касательная плоскость к лучевой поверхности перпендикулярна к соответствующей волновой нормали и отсекает на ней отрезок, равный нормальной скорости волны*. Отсюда в свою очередь следует, что *лучевая поверхность есть огибающая плоских волн, распространившихся из ее центра за единицу времени в различных направлениях*. Этими теоремами и устанавливается искомая геометрическая связь между лучевой поверхностью и поверхностью нормалей.

Можно также сказать, что *касательная плоскость к лучевой поверхности есть фронт волны, соответствующий лучу, проведенному в точку касания.* В таком виде теорема допускает простую интерпретацию. Действительно, лучевая поверхность есть поверхность равных фаз, до которой световое возмущение от точечного источника доходит в течение одной секунды. Малый участок такой поверхности может рассматриваться как плоский. Если размеры участка очень велики по сравнению с длиной волны, то его распространение в течение ближайшего времени будет с достаточной точностью подчиняться законам геометрической оптики. Согласно этим законам, участок должен распространяться как безграничная плоская волна в направлении луча, причем лучевая и нормальная скорости будут связаны соотношением (81.3). Отсюда непосредственно следует, что волновой фронт есть касательная плоскость к лучевой поверхности.

Эта простая интерпретация не может, однако, заменить строгое доказательство. В ее основе лежит утверждение, что расходящийся пучок, исходящий из точечного источника, ведет себя совершенно так же, как система не зависящих друг от друга плоских волн, распространение которых чисто геометрически представляется с помощью лучевой поверхности. Впервые (1852 г.) Ламе (1795–1870) указал, что здесь необходимо решить сложную математическую задачу: точно представить волновой комплекс, исходящий в анизотропной среде из одного точечного центра (аналог шаровой волны в изотропной среде). Ламе решил эту задачу для *упругой анизотропной среды.* При этом он действительно (при исключении продольных волн) пришел к френелевой форме лучевой поверхности. В электромагнитной теории аналогичный вопрос сводится к решению задачи о поле *точечного диполя Герца, помещенного в однородную анизотропную среду.*

3. Все изложенное выше справедливо для *любых волн в анизотропных средах.* Специфичность электромагнитных волн в кристаллах состоит в том, что для них направление луча совпадает с направлением *вектора Пойнтинга.* Докажем это утверждение для рассматриваемого нами случая недиспергирующих кристаллов. В этом случае лучевая скорость вдоль волновой нормали равна $\mathbf{u}_N = v\mathbf{N} = v^2\mathbf{N}/v$, или на основании формулы (75.8)

$$\mathbf{u}_N = \frac{c^2}{D^2 v}(\mathbf{DE})\mathbf{N}. \tag{81.6}$$

Найдем теперь составляющую лучевой скорости $\mathbf{u}_\perp = k\,\partial v/\partial \mathbf{k}$, перпендикулярную к волновой нормали. Умножая формулу

(80.6) на k^2, представим ее в виде

$$\sum_\alpha \frac{k_\alpha^2}{v^2 - a_\alpha^2} = 0.$$

Дифференцируя это соотношение по k_i, получим

$$\sum_\alpha \frac{k_\alpha \partial k_\alpha/\partial k_i}{v^2 - a_\alpha^2} - \sum_\alpha \frac{k_\alpha^2}{(v^2 - a_\alpha^2)^2} v \frac{\partial v}{\partial k_i} = 0,$$

или, с учетом соотношений $\partial k_\alpha/\partial k_i = \delta_{\alpha i}$ и $k_i = kN_i$,

$$(u_\perp)_i v \sum_\alpha \frac{N_\alpha^2}{(v^2 - a_\alpha^2)^2} = \frac{N_i}{v^2 - a_i^2}.$$

Из формулы (80.5) находим

$$\frac{1}{v^2 - a_\alpha^2} = -\frac{D_\alpha}{c^2(\mathbf{NE})N_\alpha}, \quad \sum_\alpha \frac{N_\alpha^2}{(v^2 - a_\alpha^2)^2} = \frac{1}{c^4}\frac{D^2}{(\mathbf{NE})^2}.$$

После подстановки этих значений в предыдущее соотношение и перехода к векторной форме получим

$$\mathbf{u}_\perp = -\frac{c^2(\mathbf{NE})}{vD^2}\mathbf{D}. \tag{81.7}$$

Следовательно,

$$\mathbf{u} = \mathbf{u}_N + \mathbf{u}_\perp = \frac{c^2}{vD^2}\{(\mathbf{DE})\mathbf{N} - (\mathbf{NE})\mathbf{D}\} = \frac{c^2}{vD^2}[\mathbf{E}[\mathbf{ND}]].$$

Согласно первой формуле (75.5),

$$[\mathbf{ND}] = -\frac{c}{v}[\mathbf{N}[\mathbf{NH}]] = \frac{c}{v}\mathbf{H},$$

а по формуле (75.8) $v^2D^2 = c^2(\mathbf{DE}) = c^2H^2$, так что

$$vD = cH. \tag{81.8}$$

В результате получим

$$\mathbf{u} = \frac{c}{H^2}[\mathbf{EH}].$$

Таким образом, лучевая скорость $\mathbf{u}$, а с ней и самый луч действительно направлены вдоль вектора Пойнтинга. Так как векторы $\mathbf{E}$ и $\mathbf{H}$ взаимно перпендикулярны, то $[\mathbf{EH}] = EH\mathbf{s}$, где $\mathbf{s}$ — единичный вектор в направлении луча. С учетом этого

$$\mathbf{u} \equiv u\mathbf{s} = \frac{cE}{H}\mathbf{s}. \tag{81.9}$$

При доказательстве предполагалось, что скалярное произведение $(\mathbf{NE})$ не равно нулю. Если $(\mathbf{NE}) = 0$, то вектор $\mathbf{E}$ параллелен одной из диэлектрических осей кристалла. В этом случае теорема очевидна.

4. Обоснование понятия луча и определение его направления были даны выше через групповую скорость в анизотропной среде. Мы не воспользовались сразу теоремой Пойнтинга, чтобы показать, что это понятие и его общие свойства не специфичны для электромагнитных волн, а относятся к волнам *любой физической природы.* Кроме того, теорема Умова–Пойнтинга строго доказана для потоков энергии только через *замкнутые поверхности.* Локализация потока энергии в пространстве требует дополнительных соображений. Такая локализация не вызывает затруднений в рамках применимости геометрической оптики, т.е. в той же области, к которой относится понятие луча. Тогда теорема Умова–Пойнтинга открывает наиболее простой и общий метод для решения всех вопросов, связанных с понятием луча.

Пусть, например, AB — участок плоского волнового фронта, вырезаемый диафрагмой, поставленной на пути распространения электромагнитной волны (рис. 288). Если размеры участка велики по сравнению с длиной волны, то справедлива геометрическая оптика. Надо только определить направление световых лучей. С этой целью построим на основании AB цилиндр $ABA'B'$, образующие которого совпадают с направлениями вектора Пойнтинга. Тогда поток электромагнитной энергии через боковую поверхность цилиндра будет тождественно равен нулю. Останутся только потоки через основания AB и $A'B'$. Энергия, втекающая через AB, целиком выйдет через $A'B'$. Но так же ведут себя и световые лучи геометрической оптики. Поэтому направления лучей и вектора Пойнтинга должны совпадать и притом *не только в случае недиспергирующих сред* (как предполагалось в доказательстве, приведенном выше), *но и в случае сред, обладающих дисперсией.*

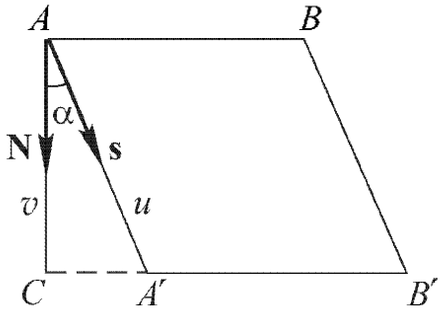


Рис. 288

5. Двойственность, характерная для кристаллооптики, о которой говорилось в начале этого параграфа, в электромагнитной теории выражается общим положением, называемом *теоремой обращения.* Эта важная теорема помогает ориентироваться в обилии сложных формул кристаллооптики, давая руководящую нить при установлении внутренних связей между ними. Выражаясь упрощенно, можно сказать, что теорема обращения сокращает вдвое число формул и теорем кристаллооптики, подлежащих запоминанию. Для вывода этой теоремы умножим первое уравнение (75.5) векторно на $\mathbf{s}$. Получим

$$[\mathbf{sD}] = -\frac{c}{v}[\mathbf{s}[\mathbf{NH}]] = \frac{c}{v}\{(\mathbf{Ns})\mathbf{H} - (\mathbf{sH})\mathbf{N}\} = \frac{c}{v}(\mathbf{Ns})\mathbf{H}.$$

Из (81.3) следует $u(\mathbf{Ns}) = v$. Исключая $(\mathbf{Ns})$, найдем $\mathbf{H} = \frac{u}{c}[\mathbf{sD}]$. Таким же путем можно получить $\mathbf{E} = -\frac{u}{c}[\mathbf{sH}]$. Таким образом, получаются два ряда формул:

$$\begin{aligned} \mathbf{D} &= -\frac{c}{v}[\mathbf{NH}], & \mathbf{E} &= -\frac{u}{c}[\mathbf{sH}], \\ \mathbf{H} &= \frac{c}{v}[\mathbf{NE}], & \mathbf{H} &= \frac{u}{c}[\mathbf{sD}], \\ \mathbf{N} &= \frac{1}{DH}[\mathbf{DH}], & \mathbf{s} &= \frac{1}{EH}[\mathbf{EH}], \\ D_\alpha &= \varepsilon_\alpha E_\alpha, & E_\alpha &= \frac{1}{\varepsilon_\alpha} D_\alpha. \end{aligned} \tag{81.10}$$

Все результаты, относящиеся к распространению плоских волн в однородных кристаллах, были получены нами как следствия первого ряда формул. Но формулы второго ряда можно формально получить из формул первого ряда заменой всех величин по следующей схеме:

$$\begin{array}{cccccc} \mathbf{D} & \mathbf{H} & \mathbf{N} & c & v & \varepsilon_\alpha \\ \mathbf{E} & \mathbf{H} & \mathbf{s} & 1/c & 1/u & 1/\varepsilon_\alpha \end{array} \tag{81.11}$$

Следовательно, *любое соотношение между величинами, характеризующими распространение плоских волн в однородных кристаллах, останется справедливым, если все входящие в него величины заменить на соответствующие согласно схеме* (81.11). Соответствующими считаются величины, стоящие друг под другом в рядах (81.11). Этот результат и называется *теоремой обращения.*

6. Применяя теорему обращения, легко найти лучевую скорость в кристалле, если известно направление луча $\mathbf{s}$. Для этого в уравнении (80.6) или (80.7) заменяем $\mathbf{N}$ на $\mathbf{s}$, v на $1/u$, a_α на $1/a_\alpha$ и получаем

$$\sum_\alpha \frac{s_\alpha^2}{(1/u)^2 - (1/a_\alpha)^2} = 0, \tag{81.12}$$

или

$$\sum_\alpha \frac{a_\alpha^2 s_\alpha^2}{u^2 - a_\alpha^2} = 0. \tag{81.13}$$

Это уравнение называется *законом Френеля для лучевой скорости в кристалле.* Оно вполне аналогично закону Френеля для нормальной скорости и может быть исследовано теми же способами. Но в этом нет необходимости, так как все результаты получаются непосредственно из теоремы обращения. Достаточно перечислить их.

Уравнение (81.13) *второй степени относительно* u^2. *Для каждого направления* $\mathbf{s}$ *оно имеет два положительных вещественных корня* u_1^2 *и* u_2^2, *причем*

$$a_x \geqslant u_1 \geqslant a_y \geqslant u_2 \geqslant a_z. \tag{81.14}$$

В каждом направлении в кристалле могут распространяться два линейно поляризованных луча, вообще говоря, с различными лучевыми скоростями u_1 *и* u_2. *Электрические векторы в этих двух лучах взаимно перпендикулярны.*

Лучевая поверхность, как и поверхность нормалей, состоит из *двух слоев*. Это есть поверхность *четвертого порядка*. Рассмотрим ее сечения координатными плоскостями XY, YZ и ZX. При этом можно воспользоваться прежним рис. 287, так как качественно сечения лучевой поверхности координатными плоскостями не отличаются от соответствующих сечений поверхности нормалей. Отличия, трудно передаваемые чертежом, лучше выразить словами или математическими формулами. При сечении поверхности нормалей получаются *круги* и *овалы*. Сечениями лучевой поверхности будут *круги* и *эллипсы*.

Сечение плоскостью XY. Луч $\mathbf{s}$ лежит в плоскости XY, т.е. $s_z = 0$. Лучевая скорость может иметь два значения:

$$u_2 = a_z, \quad \frac{1}{u_1^2} = \frac{s_x^2}{a_y^2} + \frac{s_y^2}{a_x^2}. \tag{81.15}$$

Скорость u_2 не зависит от направления луча. Ей соответствует *круговое сечение* лучевой поверхности. Скорость u_1 меняется с изменением направления луча. Соответствующее сечение имеет *форму эллипса*. Действительно, уравнение рассматриваемого сечения в векторной форме имеет вид $\mathbf{r} = u_1\mathbf{s}$, откуда $s_x = x/u_1$, $s_y = y/u_1$. Подставляя эти значения в (81.15), получаем

$$\frac{x^2}{a_y^2} + \frac{y^2}{a_x^2} = 1, \tag{81.16}$$

т.е. уравнение эллипса с полуосями a_y и a_x. Ввиду соотношений (81.14) $u_1 \geqslant u_2$, так что круг находится целиком внутри эллипса. Вектор $\mathbf{E}$ должен быть перпендикулярен к $\mathbf{s}$. Из соображений симметрии ясно, что вектор $\mathbf{E}$ одной волны параллелен оси Z, а вектор $\mathbf{E}$ другой волны параллелен плоскости XY. Первому направлению вектора $\mathbf{E}$ соответствует круговое, второму — эллиптическое сечение лучевой поверхности.

Сечение плоскостью YZ. В этой плоскости лучевые скорости могут иметь два значения:

$$u_1 = a_x, \quad \frac{1}{u_2^2} = \frac{s_y^2}{a_z^2} + \frac{s_z^2}{a_y^2}. \tag{81.17}$$

Скорость u_1 не зависит от направления луча. Ей соответствует круговое сечение лучевой поверхности и вектор $\mathbf{E}$, параллель-

ный оси X. Скорость u_2 меняется с изменением направления луча. Соответствующее сечение есть эллипс

$$\frac{y^2}{a_z^2}+\frac{z^2}{a_y^2}=1, \tag{81.18}$$

а вектор $\mathbf{E}$ лежит в плоскости ZX. Согласно (81.14) $u_1 \geqslant u_2$, так что эллипс целиком помещается внутри круга.

Сечение плоскостью ZX. Скорость луча определяется двумя выражениями:

$$u_1=a_y, \quad \frac{1}{u_2^2}=\frac{s_z^2}{a_x^2}+\frac{s_x^2}{a_z^2}. \tag{81.19}$$

Скорости u_1 соответствует круговое сечение лучевой поверхности и вектор $\mathbf{E}$, параллельный оси Y. Для u_2 получается эллипс

$$\frac{z^2}{a_x^2}+\frac{x^2}{a_z^2}=1, \tag{81.20}$$

а вектор $\mathbf{E}$ лежит в плоскости ZX. Эллипс и круг пересекаются друг с другом в четырех точках (рис. 287). В соответствии с этим в плоскости ZX имеются два направления AA' и BB', симметричные относительно оси Z, вдоль которых оба луча распространяются с одинаковыми лучевыми скоростями. Такие направления называются *оптическими осями первого рода*, *лучевыми осями* или *бирадиалями*.

Если в кристалле все три главные скорости a_x, a_y, a_z различны, то в нем существуют две и только две оптические оси первого рода. Они лежат в плоскости ZX и симметрично расположены относительно оси Z. Угол γ, образуемый одной из оптических осей первого рода с осью Z, определяется формулой

$$\operatorname{tg}\gamma=\frac{s_x}{s_z}=\sqrt{\frac{(1/a_x^2)-(1/a_y^2)}{(1/a_y^2)-(1/a_z^2)}}=\frac{a_z}{a_x}\sqrt{\frac{a_x^2-a_y^2}{a_y^2-a_z^2}}. \tag{81.21}$$

Сравнение этой формулы с формулой (80.17) приводит к соотношению

$$\operatorname{tg}\gamma=\frac{a_z}{a_x}\operatorname{tg}\beta, \tag{81.22}$$

из которого следует: $\gamma<\beta$, т.е. *оптические оси первого рода расположены ближе к оси Z, чем оптические оси второго рода*. Обычно a_x и a_z не очень сильно отличаются друг от друга. Поэтому угол между оптическими осями первого и второго рода, как правило, мал и при рассмотрении многих явлений может не приниматься во внимание. Для слюды он составляет около $40'$.

Если две из трех главных скоростей равны между собой, то оптические оси второго рода сливаются в одну ось, направленную либо параллельно оси Z (когда $a_x=a_y$), либо параллельно

оси X (когда $a_y = a_z$). В этом случае оптическая ось первого рода совпадает с оптической осью второго рода. Наконец, когда все три главные скорости равны между собой, любое направление в кристалле обладает свойствами оптической оси.

По числу оптических осей первого рода кристаллы разделяются на: 1) *двуосные*, 2) *одноосные* и 3) *оптически изотропные*. Эта классификация совпадает с классификацией, основанной на числе оптических осей второго рода.

7. Отметим еще одно следствие теоремы обращения, которое понадобится нам в следующем параграфе. Умножая скалярно уравнение (75.7) на $\mathbf{s}$ и принимая во внимание, что $(\mathbf{Es}) = 0$, получим

$$(\mathbf{Ds}) = -\frac{c^2}{v^2}(\mathbf{NE})(\mathbf{Ns}) = -\frac{c^2}{uv}(\mathbf{NE}). \tag{81.23}$$

Преобразуем (80.3) с помощью теоремы обращения:

$$\left(\frac{1}{u^2} - \frac{1}{a_\alpha^2}\right) E_\alpha = -\frac{1}{c^2}(\mathbf{sD}) s_\alpha,$$

или на основании (81.23)

$$\frac{a_\alpha^2 - u^2}{a_\alpha^2} E_\alpha = \frac{u}{v}(\mathbf{NE}) s_\alpha.$$

Умножим и разделим левую часть этого соотношения на ε_α и учтем, что $D_\alpha = \varepsilon_\alpha E_\alpha$, $a_\alpha^2 \varepsilon_\alpha = c^2$. Тогда

$$(a_\alpha^2 - u^2) D_\alpha = \frac{c^2 u}{v}(\mathbf{NE}) s_\alpha. \tag{81.24}$$

Сравнение этого соотношения с (80.3) дает

$$u(a_\alpha^2 - v^2) s_\alpha = v(a_\alpha^2 - u^2) N_\alpha, \tag{81.25}$$

откуда

$$\frac{u s_\alpha}{a_\alpha^2 - u^2} = \frac{v N_\alpha}{a_\alpha^2 - v^2}. \tag{81.26}$$

Умножая это соотношение на N_α, суммируя по α и принимая во внимание (80.6), получим

$$\sum_\alpha \frac{N_\alpha s_\alpha}{a_\alpha^2 - u^2} = 0, \tag{81.27}$$

или

$$N_x s_x (a_y^2 - u^2)(a_z^2 - u^2) + N_y s_y (a_z^2 - u^2)(a_x^2 - u^2) + \\ + N_z s_z (a_x^2 - u^2)(a_y^2 - u^2) = 0$$

— соотношение более общее, чем (81.27), так как справедливость его, как легко показать, не связана с предположением, что $a_\alpha^2 - u^2 \neq 0$.

8. Теорема о связи между лучевой поверхностью и поверхностью нормалей (см. п. 2) позволяет геометрически построить одну из этих поверхностей, если известна другая. Пусть, например, ACB — участок лучевой поверхности с центром O (рис. 289). В каждой точке этой лучевой поверхности проведем касательную плоскость и опустим на нее перпендикуляр из центра O. Геометрическое место оснований таких перпендикуляров будет поверхностью нормалей. Наоборот, чтобы по заданной поверхности нормалей построить лучевую поверхность, надо из центра O провести во всевозможных направлениях радиусы-векторы и в точках пересечения их с поверхностью нормалей построить плоскости, перпендикулярные к ним. Огибающая таких плоскостей и будет лучевой поверхностью.

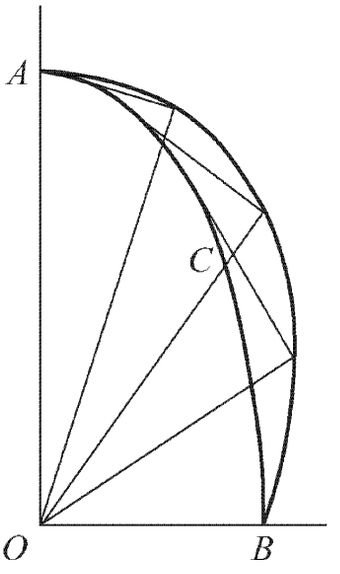


Рис. 289

Допустим, что в некоторый момент времени t в кристалле известно положение плоского волнового фронта. Для того чтобы построить волновой фронт в более поздний момент времени t', можно на основании доказанной теоремы поступить следующим образом. Из каждой точки исходного волнового фронта опишем элементарную волну, радиусы-векторы которой получаются умножением на $(t' - t)$ соответствующих радиусов-векторов лучевой поверхности. Плоскость, касательная ко всем элементарным волнам, и даст положение волнового фронта в момент времени t'. Из двух возможных касательных плоскостей следует выбрать ту, которой соответствует волна требуемой поляризации. Направление луча найдется соединением центра элементарной волны с соответствующей точкой касания.

Это построение является обобщением *построения Гюйгенса* для изотропных сред. Оно было впервые введено Гюйгенсом для объяснения двойного преломления. Гюйгенс постулировал, что элементарная волна в кристаллах состоит из двух волн: *сферической* и *эллипсоидальной*. Сферические волны порождают *обыкновенную*, а эллипсоидальные — *необыкновенную волны*. Это предположение Гюйгенса оправдалось, но оно верно только для *оптически одноосных кристаллов*.

§ 82. Коническая рефракция

1. Когда волновая нормаль **N** параллельна одной из оптических осей второго рода, нормальные скорости обеих волн v_1 и v_2 совпадают между собой, а направления векторов **D** становятся неопределенными. Значит, в

направлении оптической оси второго рода может распространяться плоская волна любой поляризации, причем скорость распространения не зависит от характера поляризации. В этом отношении рассматриваемый случай аналогичен распространению волн в изотропной среде. Однако если кристалл двуосный, между ними имеется существенное различие.

В изотропной среде направления векторов $\mathbf{D}$ и $\mathbf{E}$, а также $\mathbf{N}$ и $\mathbf{S}$ всегда совпадают. Для волны, распространяющейся вдоль оптической оси двуосного кристалла, положение меняется. В этом случае вектор $\mathbf{D}$ может принимать любое направление, перпендикулярное к $\mathbf{N}$. Так как $\mathbf{N}$ лежит в плоскости ZX, то одним из возможных направлений вектора $\mathbf{D}$ будет ось Y. Тогда и только тогда вектор $\mathbf{D}$ будет совпадать по направлению с вектором $\mathbf{E}$, а нормаль $\mathbf{N}$ — с лучом $\mathbf{s}$. Во всех остальных случаях направления луча и волновой нормали отличаются друг от друга. Придавая $\mathbf{D}$ всевозможные направления, перпендикулярные к оптической оси, получим бесчисленное множество направлений луча $\mathbf{s}$. Направление вектора $\mathbf{s}$ становится *неопределенным*. Докажем, что в рассматриваемом случае все лучи лежат на поверхности конуса.

Для доказательства воспользуемся соотношением (81.27), которое справедливо при любом направлении волновой нормали. Так как волновая нормаль $\mathbf{N}$ направлена вдоль оптической оси, то $N_y = 0$, и соотношение (81.27) принимает вид

$$\frac{N_x s_x}{a_x^2 - u^2} + \frac{N_z s_z}{a_z^2 - u^2} = 0.$$

Отсюда

$$a_z^2 N_x s_x + a_x^2 N_z s_z = u^2(\mathbf{Ns}).$$

Ввиду соотношения (81.3), $u(\mathbf{Ns}) = v = a_y$. Следовательно,

$$(a_z^2 N_x s_x + a_y^2 N_z s_z)(N_x s_x + N_z s_z) = a_y^2.$$

Рассмотрим произвольную точку на луче $\mathbf{s}$ с радиусом-вектором $\mathbf{r}(x, y, z)$. Очевидно, $x_\alpha = r s_\alpha$, и предыдущее соотношение переходит в

$$(a_z^2 N_x x + a_y^2 N_z z)(N_x x + N_z z) = a_y^2(x^2 + y^2 + z^2). \tag{82.1}$$

Это однородное уравнение второго порядка представляет *конус*. Образующими конуса являются лучи, соответствующие волновой нормали $\mathbf{N}$, параллельной одной из двух оптических осей второго рода. Конус (82.1) называется *конусом внутренней конической рефракции*. Волновая нормаль есть одна из образующих конуса (82.1). Это следует из того, что направления $\mathbf{s}$ и $\mathbf{N}$ совпадают, когда вектор $\mathbf{D}$ параллелен диэлектрической оси Y.

2. Конус внутренней конической рефракции пересекается фронтом волны

$$\mathbf{Nr} \equiv N_x x + N_z z = a_y \tag{82.2}$$

по кругу. В самом деле, линия пересечения определяется системой уравнений (82.1) и (82.2), равносильной системе

$$\begin{gathered} a_z^2 N_x x + a_y^2 N_z z = a_y(x^2 + y^2 + z^2), \\ \mathbf{Nr} = a_y. \end{gathered} \tag{82.3}$$

Первое уравнение есть уравнение сферы, второе — уравнение плоскости. Их пересечение есть круг, и наше утверждение доказано.

Определим угол раствора конуса внутренней конической рефракции, точнее — угол χ, получающийся от пересечения этого конуса плоскостью ZX, проходящей через оптическую ось кристалла (рис. 290). Когда вектор $\mathbf{D}$ направлен вдоль диэлектрической оси Y, векторы $\mathbf{s}$ и $\mathbf{N}$, а также оптическая ось второго рода совпадают по направлению. Если же вектор $\mathbf{D}$ лежит в плоскости ZX, то в той же плоскости будет лежать и луч $\mathbf{s}$, так как четыре вектора $\mathbf{E}$, $\mathbf{D}$, $\mathbf{s}$, $\mathbf{N}$ всегда должны лежать в одной плоскости (см. § 75). Искомый угол χ будет равен углу между векторами $\mathbf{D}$ и $\mathbf{E}$, поскольку стороны этих углов взаимно перпендикулярны. Его легко определить из формулы (75.8), так как в рассматриваемом случае нормальная скорость v равна a_y. Формула (75.8) дает

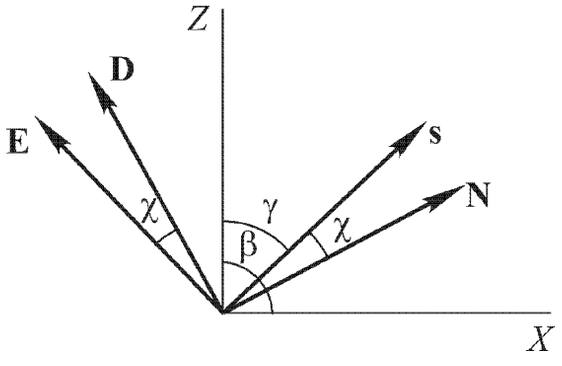


Рис. 290

$$\cos\chi = \frac{D}{E}\left(\frac{a_y}{c}\right)^2.$$

Как видно из рис. 290, $D_x = -D\cos\beta$, $D_z = D\sin\beta$, где β — угол между оптической осью второго рода и осью Z. Поэтому

$$E_x = \frac{D_x}{\varepsilon_x} = -\frac{D\cos\beta}{c^2}a_x^2, \quad E_z = \frac{D_z}{\varepsilon_z} = \frac{D\sin\beta}{c^2}a_z^2.$$

Подставляя эти значения в выражение для $\cos\chi$, получим

$$\cos\chi = \frac{a_y^2}{\sqrt{a_x^4\cos^2\beta + a_z^4\sin^2\beta}},$$

откуда

$$\operatorname{tg}\chi = \frac{1}{a_y^2}\sqrt{a_x^4\cos^2\beta + a_z^4\sin^2\beta - a_y^4}.$$

Использовав формулу (80.17), после несложных преобразований найдем

$$\operatorname{tg}\chi = \frac{1}{a_y^2}\sqrt{(a_x^2 - a_y^2)(a_y^2 - a_z^2)}. \tag{82.4}$$

Конус внутренней конической рефракции пересекает лучевую поверхность по кругу, вдоль которого ее касается фронт волны. Это непосредственно следует из теоремы, доказанной в § 81 (п. 2).

3. Теорема обращения распространяет полученные результаты на лучи. Если луч в двуосном кристалле направлен вдоль одной из оптических осей первого рода, то ему соответствует бесконечное множество волновых нормалей, образующих конус. Этот конус называется *конусом внешней конической рефракции.* Луч есть одна из образующих этого конуса. Сечение конуса внешней конической рефракции плоскостью, перпендикулярной лучу, есть круг. Угол раствора конуса определяется уравнением

$$\operatorname{tg}\psi = a_y^2\sqrt{\left(\frac{1}{a_x^2} - \frac{1}{a_y^2}\right)\left(\frac{1}{a_y^2} - \frac{1}{a_z^2}\right)} = \frac{1}{a_x a_z}\sqrt{(a_x^2 - a_y^2)(a_y^2 - a_z^2)}. \tag{82.5}$$

Проведем касательную плоскость к лучевой поверхности в точке S пересечения ее с лучевой осью. Такая плоскость будет перпендикулярна к волновой нормали. А так как волновых нормалей, соответствующих лучу, направленному вдоль лучевой оси, бесконечно много, то в точке S можно провести бесконечное множество касательных плоскостей к лучевой поверхности. Это означает, что в окрестности такой точки лучевая поверхность имеет воронкообразную форму.

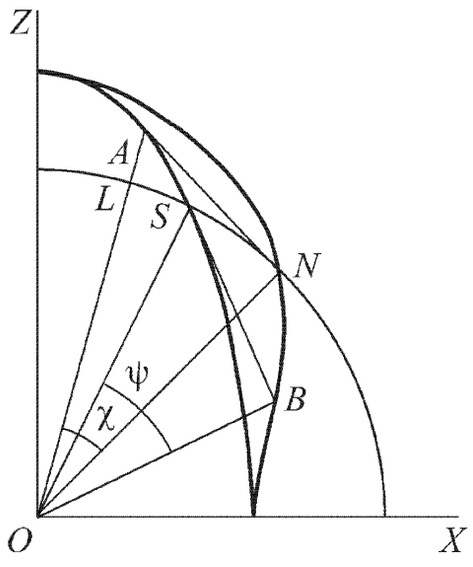


Рис. 291

На рис. 291 представлено сечение поверхности нормалей и лучевой поверхности плоскостью ZX. Точка N есть двойная точка поверхности нормалей, ON — оптическая ось второго рода. Перпендикуляр NA к этой оси дает сечение фронта волны плоскостью рисунка. Прямая NA касается лучевой поверхности в точке A, угол $\chi = \angle NOA$ есть угол раствора конуса внутренней конической рефракции, S — двойная точка лучевой поверхности, OS — лучевая ось. Касательная к лучевой поверхности в точке S пересекает поверхность нормалей в точке B; прямая OB будет одной из волновых нормалей, принадлежащих лучу OS. Сам луч OS является нормалью плоской волны, которая касается кругового сечения лучевой поверхности $u = a_y$ в точке S. Угол $\psi = \angle SOB$ есть угол раствора конуса внешней конической рефракции.

Каждому лучу, принадлежащему конусу внутренней конической рефракции, например лучу OA (рис. 291), соответствует вполне определенная линейная поляризация. В самом деле, в направлении OA могут распространяться два луча, электрические векторы которых взаимно перпендикулярны. Однако только один из них соответствует волне, распространяющейся вдоль волновой нормали ON. Другому лучу соответствует лучевая скорость OL и, следовательно, иное направление волновой нормали. Аналогично, каждой волновой нормали, принадлежащей конусу внешней конической рефракции, также соответствует вполне определенная линейная поляризация.

4. Волновой фронт AN, распространяющийся в направлении оптической оси второго рода ON, как было показано, касается лучевой поверхности по кругу, вдоль которого эта поверхность пересекается конусом внутренней конической рефракции. Такой волновой фронт не может пересекать лучевую поверхность. В самом деле, пересечем лучевую поверхность плоскостью ANO, проходящей через оптическую ось ON (рис. 292). В сечении получится кривая CBA. Если бы фронт AN пересекал лучевую поверхность, то, ввиду конечности лучевой скорости, на кривой CBA нашлись бы такие точки B, C, что проекции радиусов-векторов OB, OC на направление ON были бы минимальны или максимальны. Если плоскость ANO проводить во всевозможных направлениях, проходящих через оптическую ось ON, то точки

Рис. 292

B, C опишут замкнутые кривые. На этих кривых в свою очередь найдутся точки, проекции радиусов-векторов которых на направление бинормали ON будут максимальны или минимальны. Пусть B — одна из таких точек. Тогда касательная плоскость BN' к лучевой поверхности в точке B будет параллельна волновому фронту AN, т.е. перпендикулярна к бинормали ON. Значит, плоскость BN' сама является волновым фронтом, распространяющимся в направлении бинормали ON.

Таким образом, если бы волновой фронт NA мог пересекать лучевую поверхность, то вдоль бинормали могли бы распространяться две волны с различными нормальными скоростями $v = ON$ и $v' = ON'$, что противоречит определению бинормали. Поскольку волновой фронт касается лучевой поверхности по кругу (в точках A), из доказанного следует, что лучевая поверхность целиком лежит с той стороны волнового фронта AN, с которой находится ее центр O (рис. 291). Значит, в точке S лучевая поверхность имеет не просто воронкообразную форму, как было отмечено выше, но *воронкообразное углубление*.

5. После выяснения этих геометрических соотношений обратимся к рассмотрению внутренней конической рефракции, теоретически предсказанной Гамильтоном (1805–1865) в 1832 г. Примерный ход рассуждений Гамильтона был следующий. Пусть плоскопараллельная пластинка из двуосного кристалла прикрыта с одной стороны непрозрачным экраном с малым отверстием O (рис. 293). Осветим пластинку параллельным пучком неполяризованных лучей таким образом, чтобы после преломления на передней поверхности пластинки волновая нормаль оказалась направленной вдоль одной из оптических осей второго рода OA. Волновой нормали OA соответствует конус лучей. Энергия распространяется вдоль лучей, поэтому при достаточно малых размерах отверстия O световой пучок внутри пластинки развернется в конус OAB. После преломления на задней поверхности пластинки волновая нормаль примет свое исходное направление. А так как в изотропных средах направления лучей и волновых нормалей совпадают, то все лучи выйдут из пластинки параллельным пучком и расположатся по поверхности цилиндра. Если на их пути поместить экран, то на нем должно получиться светлое кольцо. По предложению Гамильтона отысканием этого явления занялся Ллойд, который и обнаружил его в 1833 г. на кристалле арагонита: на экране наблюдалось светлое эллиптическое кольцо. (Угол χ для арагонита равен $1°52'$.)

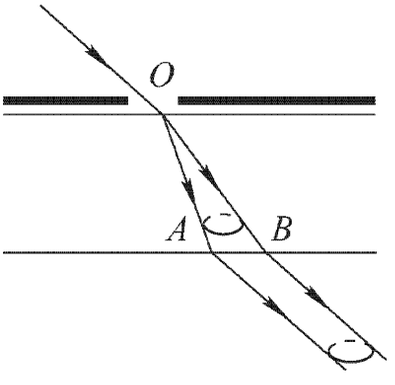


Рис. 293

Хотя Гамильтон и предсказал коническую рефракцию, его объяснение неправильно. При более детальном изучении оказалось, что явление выглядит иначе, чем предсказывал Гамильтон. Применяя более узкие отверстия в экране, Поггендорф (1796–1877) нашел, что кольцо в действительности *двойное*. Объяснение было дано Фохтом (1850–1919). Гамильтон рассматривал строго плоскую волну, распространяющуюся в кристалле точно в направлении оптической оси. Физически это реализовать невозможно. Если бы даже можно было осветить отверстие O строго плоской волной, то после прохождения через него волна перестала бы быть плоской из-за дифракции. Такая волна распадается на бесконечное множество плоских волн, направления распространения которых близки к направлению оптической оси. Нельзя ограничиться рассмотрением поведения только одной волны,

распространяющейся строго в направлении оптической оси. Это ясно уже из того, что на ее долю приходится исчезающе малая энергия, и физически ничего не изменится, если эту волну даже совсем удалить из волнового комплекса. Необходимо рассмотреть бесконечное множество плоских волн, волновые нормали которых группируются вблизи оптической оси. Это и было сделано Фохтом.

Если строго плоская волна распространяется в направлении оптической оси второго рода, то, как было показано, волновой фронт касается лучевой поверхности по кругу. Примем плоскость такого круга за плоскость рис. 294.

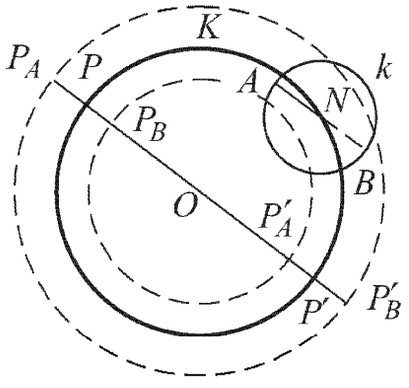


Рис. 294

Пусть N — точка пересечения плоскости рисунка с оптической осью. Допустим, что имеется бесконечная совокупность плоских волн, волновые нормали которых лежат в пределах небольшого конуса, ось которого совпадает с оптической осью. Этот конус пересекается плоскостью рисунка по кругу k. Каждой волновой нормали соответствует точка внутри или на границе круга k. Найдем, где расположатся соответствующие ей лучи. Волновой нормали N соответствует конус лучей, пересекающих плоскость рисунка по окружности K. Лучевая поверхность касается плоскости рисунка вдоль этой окружности, а поэтому кривизна лучевой поверхности в направлении окружности K равна нулю. Кроме того, лучевая поверхность должна лежать по одну сторону от плоскости рисунка. Ради определенности примем, что все волны распространяются к читателю. Тогда лучевая поверхность будет лежать за плоскостью рисунка, и следовательно, ее окрестность вблизи окружности K будет обращена к читателю своей выпуклостью. В такой окрестности лучевая поверхность имеет баранкообразную форму.

Пересечем лучевую поверхность двумя параллельными плоскостями P_AP_B и AB, перпендикулярными к плоскости рисунка и проходящими через центры окружностей K и k. Бесконечно малые отрезки P_AP_B и $P'_AP'_B$ перпендикулярны к окружности K, поэтому в направлениях этих отрезков кривизна лучевой поверхности будет максимальна, а в перпендикулярных направлениях равна нулю. Следовательно, перпендикуляры к этим бесконечно малым отрезкам должны лежать в плоскости $P_AP'_B$, т.е. они будут параллельны волновым нормалям, лежащим в плоскости AB. Значит, касательная плоскость к лучевой поверхности в какой-либо точке отрезка P_AP или отрезка P'_AP' будет перпендикулярна к соответствующей волновой нормали, проходящей через отрезок AN. Касательная же плоскость к лучевой поверхности в какой-либо точке отрезков PP_B в $P'P'_B$ будет перпендикулярна к волновой нормали, проходящей через отрезок NB. Это означает, что каждой волновой нормали, проходящей через отрезок AN, соответствуют два луча, из которых один проходит через отрезок P_AP, а другой — через отрезок P'_AP'. Каждой же волновой нормали, проходящей через отрезок NB, соответствуют два луча, проходящие через отрезки PP_B и $P'P'_B$. Таким образом, каждой волновой нормали, наклоненной под малым углом к оптической оси второго рода, соответствуют два луча, один из которых проходит внутри конуса внутренней конической рефракции, а другой вне этого конуса.

Теперь ясно происхождение двойного светлого кольца Поггендорфа. Волновым нормалям, пересекающим плоскость рисунка внутри малого кру-

га радиуса dr с центром в N, соответствует малая доля энергии, которая должна распределиться по сравнительно большой площади dS кольца по обе стороны от окружности K. Если же взять волновые нормали, пересекающие плоскость рисунка внутри кольца со средним радиусом r и той же толщиной dr, то таким волновым нормалям будет соответствовать значительно бóльшая энергия, поскольку она пропорциональна площади кольца $2\pi r\,dr$. Эта энергия должна распределиться по площади двух колец, одно из которых лежит внутри, а другое вне окружности K. Площади обоих колец с точностью до бесконечно малых высшего порядка по-прежнему равны dS. Поэтому освещенность обоих колец будет много больше освещенности центрального кольца в окрестности окружности K. Освещенность должна равняться нулю вдоль окружности K и непрерывно возрастать по мере удаления от этой окружности как в наружную, так и во внутреннюю стороны. Таким образом, там, где по Гамильтону должна была бы получаться максимальная освещенность, в действительности наблюдается темнота.

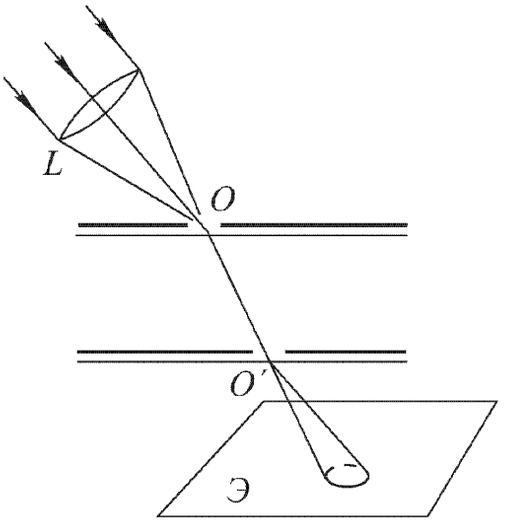


Рис. 295

6. Гамильтоном была предсказана еще *внешняя коническая рефракция*, экспериментально обнаруженная также Ллойдом в 1833 г. Она связана с тем, что световому лучу, идущему вдоль лучевой оси двуосного кристалла, соответствует бесконечная совокупность волновых нормалей, лежащих на конической поверхности. Нет необходимости входить в теорию этого явления. Достаточно сослаться на теорему обращения и описать явление, как оно наблюдалось в установке Ллойда. Обе поверхности плоскопараллельной арагонитовой пластинки были покрыты экранами с малыми отверстиями O и O' (рис. 295), центры которых лежали на оптической оси первого рода. Линза L концентрировала на O сходящийся пучок лучей. Диафрагмы O и O' выделяли только те лучи, которые шли вдоль оптической оси OO'. По выходе из пластинки лучи развертывались в конус, и на экране $Э$ наблюдалось светлое кольцо. Разумеется, при достаточно малых размерах отверстий O и O' кольцо будет *двойное*.

§ 83. Замечания об отражении и преломлении света на границе кристаллов

1. Как и в оптике изотропных сред, задача об отражении и преломлении в кристаллооптике может быть полностью решена на основе *граничных условий*, которым должны удовлетворять векторы электромагнитного поля: по обе стороны границы раздела тангенциальные компоненты векторов $\mathbf{E}$ и $\mathbf{H}$ должны быть равны. Задача распадается на две части: *геометрическую*, в которой определяются направления распространения, поляри-

зация и скорости отраженных и преломленных волн, а в случае их неоднородности также и затухание волн в пространстве, и *физическую*, в которой определяются амплитуды волн (вообще говоря, комплексные). Остановимся кратко только на геометрической стороне вопроса. Формулы для амплитуд, из-за их громоздкости, выводить не будем, хотя сам вывод и не встречает принципиальных затруднений.

Пусть на плоскую границу раздела падает плоская монохроматическая волна с волновым вектором $\mathbf{k}_1$. В случае изотропных сред получается только одна отраженная и только одна преломленная волна. Для анизотропных сред это, вообще говоря, не так. Однако, каково бы ни было число отраженных и преломленных волн, из линейности и однородности граничных условий непосредственно следует, что тангенциальные компоненты волновых векторов падающей, отраженных и преломленных волн должны быть одинаковы (см. § 69). Следовательно, *нормали падающей, отраженных и преломленных волн, а также нормаль к границе раздела все лежат в одной плоскости.* Кроме того, *преломление волновых нормалей подчиняется закону преломления Снеллиуса: отношение синуса угла падения к синусу угла преломления равно отношению соответствующих нормальных скоростей волн.* Практически от этого закона мало пользы. Его недостаточно для нахождения направления волновой нормали преломленной волны, так как нормальная скорость преломленной волны *неизвестна.* Она зависит от направления падающей волны и *сама подлежит определению.* Проще всего это можно сделать графическим методом.

2. Рассмотрим одну из преломленных волн. Пусть $\mathbf{k} = \overrightarrow{OA}$ или $\mathbf{k} = \overrightarrow{OB}$ означает ее волновой вектор. Ввиду граничных условий, тангенциальную составляющую этого вектора $\mathbf{k}_t$ можно считать известной: $\mathbf{k}_t = \mathbf{k}_{1t}$. Она лежит в плоскости падения. Остается найти нормальную составляющую $\mathbf{k}_n$. Прежде всего, она должна быть направлена вниз, в сторону второй среды, так как преломленная волна должна *уходить от границы раздела.* Для решения задачи можно применить следующее геометрическое построение. Будем проводить из точки O (рис. 296) во вторую среду прямые по всевозможным направлениям и на них откладывать длины волнового вектора $\mathbf{k}$. Поскольку направления $\mathbf{k}$ и $\mathbf{N}$ совпадают, а каждому направлению $\mathbf{N}$ соответствуют два значе-

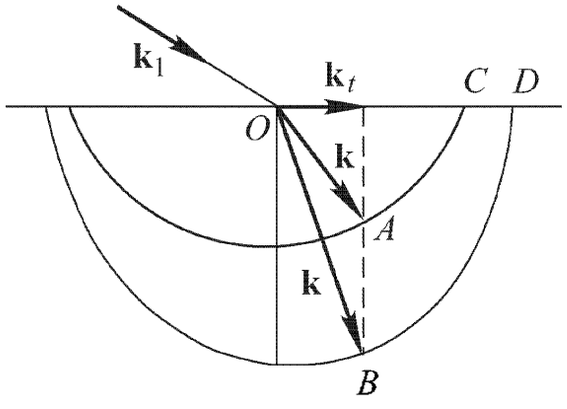


Рис. 296

ния нормальной скорости, можно сказать также, что каждому направлению нормали $\mathbf{N}$ соответствуют два значения волнового числа k. Поэтому в результате построения получится сложная поверхность, состоящая из двух слоев. Будем называть ее поверхностью *волновых векторов* [1]). Она просто связана с поверхностью нормалей. Поскольку $k = \omega/v$, поверхность волновых векторов по отношению к поверхности нормалей является *обратной*: длины ее радиусов-векторов обратно пропорциональны длинам параллельных им радиусов-векторов поверхности нормалей.

Пересечем поверхность волновых векторов плоскостью падения — в сечении получатся две кривые (рис. 296). Отложим от точки O вдоль границы раздела тангенциальную составляющую $\mathbf{k}_t = \mathbf{k}_{1t}$ волнового вектора $\mathbf{k}$ и из конца этой составляющей восстановим перпендикуляр к границе раздела. Он пересечет указанные две кривые, вообще говоря, в двух точках A и B. Соединив эти точки с точкой O, получим два вектора $\overrightarrow{OA}$ и $\overrightarrow{OB}$, каждый из которых может быть волновым вектором преломленной волны.

Тем самым определятся направления волновых нормалей $\mathbf{N}$ и соответствующие им значения нормальных скоростей. Поверхность волновых векторов позволяет построить поверхность нормалей, а затем лучевую поверхность и найти направления соответствующих лучей и лучевые скорости. Векторы $\mathbf{N}$ и $\mathbf{s}$ определят направление магнитного поля $\mathbf{H}$, поскольку оно перпендикулярно к плоскости $(\mathbf{N}, \mathbf{s})$. Определятся и направления векторов $\mathbf{D}$ и $\mathbf{E}$, т.е. поляризация обеих волн в кристалле. В общем случае направления лучей и волновых нормалей не совпадают. В оптически двуосных кристаллах оба луча, как правило, выходят из плоскости падения.

Если волна падает на кристалл из изотропной среды, то она может быть поляризована как угодно. В кристалл она вступит в виде двух линейно поляризованных волн. Если же первая среда также кристаллическая, то в направлении $\mathbf{k}_1$ могут распространяться две волны с различными поляризациями и с различными нормальными скоростями. Их надо рассматривать как *независимые волны*. Каждая из них в кристалле расщепится на *две линейно поляризованные волны*. Во второй среде получится всего *четыре волны*.

[1]) Если вместо величины $k = \omega/v$ откладывать величину c/v, т.е. показатель преломления n, то получится поверхность, подобная поверхности волновых векторов. В кристаллооптике ее называют *индикатрисой* или *поверхностью показателей преломления*. Радиус-вектор этой поверхности дает значение показателя преломления для плоской волны, нормаль которой совпадает по направлению с этим радиусом-вектором.

Может случиться, что длина вектора $\mathbf{k}_{1t}$ окажется больше длины одного или обоих отрезков OC и OD, отсекаемых на оси X поверхностью волновых векторов. Тогда построение, выполненное на рис. 296, станет невозможным. Возникнет либо одна преломленная волна, либо ни одной. Одна из волн или обе волны испытают *полное отражение.*

3. Таким же способом может быть решен вопрос о направлении и поляризации отраженного света. Надо только построить поверхность волновых векторов не во второй, а в первой среде. Если первая среда изотропна, то поверхность волновых векторов будет сферой. Возникнет только одна отраженная волна, причем угол падения будет равен углу отражения. Если же первая среда анизотропна, то при отражении волна расщепится на две линейно поляризованные волны, идущие в разных направлениях.

ГЛАВА VIII

МОЛЕКУЛЯРНАЯ ОПТИКА

§ 84. Классическая теория дисперсии света

1. Все среды, за исключением абсолютного вакуума, обладают *дисперсией*, т.е. зависимостью показателя преломления электромагнитных волн от их частоты. Наилучшим приближением к вакууму является межпланетное и межзвездное пространство. По астрофизическим данным средняя плотность вещества в межпланетном пространстве составляет около одного атома (иона) на 1 см^3. В межзвездном пространстве нашей Галактики средняя концентрация вещества — около 10^{-2} атомов на 1 см^3, а в межгалактическом пространстве еще меньше. В лучших же вакуумных приборах она не меньше 10^4 атомов на 1 см^3, а обычно гораздо больше.

Отсутствие дисперсии у межзвездного пространства доказывается астрономическими наблюдениями над затмениями *двойных звезд*. Допустим, например, что красные лучи в межзвездном пространстве распространяются быстрее синих. Тогда при начале затмения должно было бы наблюдаться изменение цвета звезды от нормального к синему, а при окончании — от красного к нормальному. При колоссальных расстояниях до звезд этот эффект не мог бы ускользнуть от наблюдения, даже если бы разница в скоростях красных и синих лучей оказалась ничтожно малой. В действительности в видимой области спектра он обнаружен не был. Еще Араго на основании своих наблюдений над двойной звездой Альголь пришел к заключению, что разность между скоростями красных и синих лучей во всяком случае не может превышать одной стотысячной скорости света.

Конечно, отсутствие дисперсии видимого света в межзвездном пространстве экспериментально доказано лишь с той точностью, которую может обеспечить эксперимент. Наблюдения над пульсарами с помощью современных радиоприемных устройств

позволили установить наличие дисперсии межзвездного пространства в области радиодиапазона (см. § 87).

2. Феноменологическая теория Максвелла, не содержащая никаких атомно-молекулярных констант, не давала объяснения дисперсии света. Для этого необходимы атомистические представления. Дисперсия света возникает в результате *вынужденных колебаний* заряженных частиц — электронов и ионов — под действием переменного поля электромагнитной волны.

В классической теории Г.А. Лорентца строение колеблющихся систем — атомов и молекул — и их колебания описываются на основе классических представлений о движении и законов Ньютона. В нашем курсе мы можем в основном ограничиться только такой классической теорией. Теоретическому рассмотрению проще всего поддается *дисперсия в газах*, так как в этом случае в первом приближении можно не учитывать сложное взаимодействие атомов и молекул среды. Для не очень плотных газов основные предположения теории выполняются с меньшей натяжкой, чем в случае конденсированных сред. Поэтому экспериментальную проверку этих предположений лучше всего производить именно на газах, для которых и теория разработана лучше. В дальнейшем мы в основном ограничимся этим простейшим случаем.

Все электроны, входящие в атом, можно разделить на *периферийные*, или *оптические*, и *электроны внутренних оболочек*. На излучение и поглощение света в оптической области спектра оказывают влияние практически *одни только оптические электроны*. Собственные частоты электронов внутренних оболочек слишком велики, так что их колебания в поле световой волны практически не возбуждаются. Поэтому в теории дисперсии можно ограничиться рассмотрением одних только оптических электронов.

Для простоты предположим сначала, что в атоме есть *всего один оптический электрон*. В классической теории дисперсии оптический электрон рассматривается как *затухающий гармонический осциллятор*, колебания которого в поле световой волны описываются уравнением

$$m\ddot{\mathbf{r}} = -k\mathbf{r} - g\dot{\mathbf{r}} + e\mathbf{E}',$$

где m — масса, e — заряд электрона, $-k\mathbf{r}$ — *квазиупругая возвращающая сила*, стремящаяся вернуть электрон в положение равновесия, $-g\dot{\mathbf{r}}$ — сила, аналогичная *силе трения* и введенная для учета *поглощения света*, $\mathbf{E}'$ — напряженность электрического поля, *действующего* на электрон. Поле $\mathbf{E}'$, вообще говоря, отличается от среднего макроскопического поля $\mathbf{E}$, входящего в уравнения Максвелла (см. т. III, § 35), но в случае неплотных газов этим различием можно пренебречь. Разделив на m,

приведем предыдущее уравнение к виду

$$\ddot{\mathbf{r}} + 2\gamma\dot{\mathbf{r}} + \omega_0^2\mathbf{r} = \frac{e}{m}\mathbf{E}', \tag{84.1}$$

где $\omega_0^2 = k/m$, $2\gamma = g/m$. Магнитной силой $\frac{e}{c}[\mathbf{vH}]$, действующей на электрон, мы пренебрегли, так как скорость электрона v пренебрежимо мала по сравнению со скоростью света c. Эта сила проявляется лишь в *световом давлении*, которое в обычных условиях мало и в разбираемом нами вопросе не играет существенной роли (см. задачу 2 к этому параграфу).

Все силы, действующие внутри атомов и молекул, имеют *электрическую природу*. Такими силами определяется и внутренняя структура этих частиц. Однако объяснить этими силами существование и структуру атомов и молекул на основе классических представлений оказалось невозможным. Это было сделано только в рамках *квантовой физики*. Никаких квазиупругих сил и сил трения, пропорциональных скорости заряженных частиц, в атомах и молекулах нет. Правильная теория дисперсии должна принимать во внимание только *реально существующие силы* и основываться на *квантовых законах*. Такую теорию дает квантовая механика. Однако она приводит к поразительному результату, что *в отношении дисперсии и поглощения света атомы и молекулы ведут себя так, как если бы среда представляла собой набор осцилляторов с различными собственными частотами и коэффициентами затухания, подчиняющихся классическим уравнениям движения Ньютона.*

Собственные частоты и коэффициенты затухания не могут быть вычислены на основе классической модели. В классической теории на них надо смотреть как на *формально введенные постоянные*. Вычисление этих постоянных и раскрытие их истинного физического смысла возможно только в *рамках квантовой теории*. На классическую же теорию надо смотреть как на теорию, дающую *модель диспергирующей среды*, которая приводит к правильным окончательным результатам, если к ней применять законы классической механики.

После этого замечания можно перейти к изложению классической теории дисперсии, не опасаясь, что смысл этой теории будет понят неправильно.

3. Предположим, что поле $\mathbf{E}$ представляется плоской волной $\mathbf{E} = \mathbf{A}e^{i(\omega t - \mathbf{kr})} = \mathbf{E}_0(\mathbf{r})e^{i\omega t}$. Амплитуда поля $\mathbf{E}_0 = \mathbf{A}e^{-i\mathbf{kr}}$ меняется от точки к точке. Значит, в различных точках траектории электрон подвергается действию поля различной амплитуды. (Указанное обстоятельство приводит к явлению так называемой *пространственной дисперсии*.) Однако мы пренебрежем этим обстоятельством, предполагая, что амплитуда колебаний электрона $\mathbf{r}_0$ мала по сравнению с длиной волны λ. Для теории

дисперсии имеет значение не общее, а только *частное решение* уравнения (84.1), представляющее *вынужденные колебания* осциллятора $\mathbf{r} = \mathbf{r}_0 e^{i\omega t}$. Амплитуда $\mathbf{r}_0$ найдется подстановкой этого выражения в (84.1). Если пренебречь различием между $\mathbf{E}$ и $\mathbf{E}'$, то получится

$$\mathbf{r} = \frac{e/m}{\omega_0^2 - \omega^2 + 2i\omega\gamma}\mathbf{E}. \tag{84.2}$$

Атом в электрическом поле приобретает дипольный момент $\mathbf{p} = e\mathbf{r} = \beta\mathbf{E}$, где β — поляризуемость атома, определяемая формулой

$$\beta = \frac{e^2/m}{\omega_0^2 - \omega^2 + 2i\omega\gamma}. \tag{84.3}$$

Если N — число атомов в единице объема, то вектор поляризации среды будет $\mathbf{P} = N\mathbf{p} = N\beta\mathbf{E}$, а индукция

$$\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P} = \varepsilon\mathbf{E}, \tag{84.4}$$

где

$$\varepsilon = 1 + \frac{4\pi Ne^2/m}{\omega_0^2 - \omega^2 + 2i\omega\gamma}. \tag{84.5}$$

Материальное уравнение $\mathbf{D} = \varepsilon\mathbf{E}$ принимает такой же вид, как и в формальной теории Максвелла, с тем существенным отличием, что диэлектрическая проницаемость ε теперь *зависит от частоты* ω. Поэтому сохраняют силу все ранее полученные результаты, если в них постоянную ε заменить функцией $\varepsilon(\omega)$. Функция $\varepsilon(\omega)$ получилась *комплексной*. Этого и следовало ожидать, так как в нашей модели учтено *поглощение света*. Введем комплексный показатель преломления по формуле

$$\sqrt{\varepsilon} = n - i\varkappa, \tag{84.6}$$

где n — *вещественный показатель преломления*, а $\varkappa$ — *показатель затухания среды*. Обе функции $n(\omega)$ и $\varkappa(\omega)$ зависят от частоты. Таким образом, получают принципиальное, хотя и формальное объяснение не только дисперсия, но и абсорбция света.

Возведем (84.6) в квадрат и сравним вещественные и мнимые части полученного соотношения и соотношения (84.5). Тогда получим

$$n^2 - \varkappa^2 = 1 + 4\pi\frac{Ne^2/m}{(\omega_0^2 - \omega^2)^2 + 4\omega^2\gamma^2}(\omega_0^2 - \omega^2), \tag{84.7}$$

$$n\varkappa = 4\pi\frac{Ne^2/m}{(\omega_0^2 - \omega^2)^2 + 4\omega^2\gamma^2}\omega\gamma. \tag{84.8}$$

С помощью этих формул показатели преломления и затухания n и $\varkappa$ можно выразить через частоту ω и молекулярные параметры N и e^2/m.

Вдали от собственной частоты ω_0, где $2\omega\gamma \ll |\omega_0^2 - \omega^2|$, мнимой частью в (84.5) можно пренебречь и получить приближенную формулу:

$$\varepsilon = n^2 = 1 + \frac{4\pi N e^2/m}{\omega_0^2 - \omega^2}. \tag{84.9}$$

Формула для n такого вида впервые была получена Зельмейером в 1871 г. [1]) в механической теории эфира. Конечно, в формуле самого Зельмейера нет никакого намека на связь показателя преломления с диэлектрической проницаемостью ε и вообще со всякими величинами, характеризующими электрические свойства среды.

На рис. 297 сплошной линией представлен график функции $\varepsilon = \varepsilon(\omega)$. При переходе через точку $\omega = \omega_0$ эта функция претерпевает скачок от $+\infty$ до $-\infty$. Показатель преломления n представлен штриховой, а показатель затухания $\varkappa$ — штрихпунктирной линией. Величина n обращается в нуль на участке, где $\varepsilon(\omega) < 0$. Показатель затухания $\varkappa$, наоборот, на этом участке отличен от нуля, обращаясь в нуль при всех остальных частотах. Величина $\varkappa$ описывает именно *затухание, а не поглощение волн*, так как кривые рис. 297 относятся к модели *непоглощающей среды* (диэлектрическая проницаемость ε всюду вещественна).

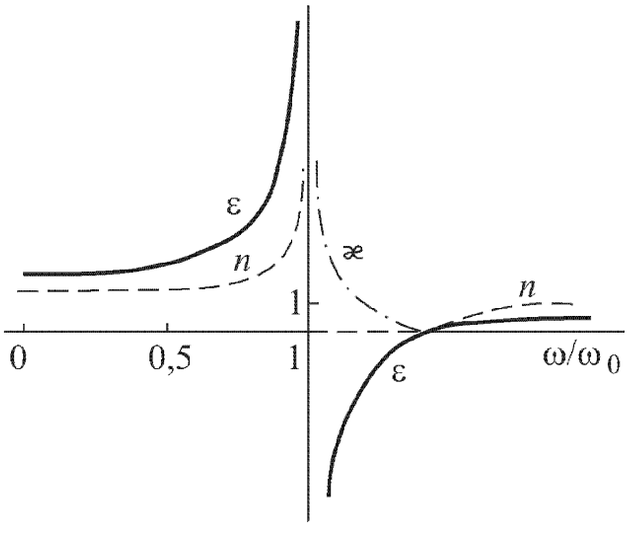


Рис. 297

Впрочем, эти выводы имеют формальный характер, так как вблизи собственной частоты ω_0 соотношение $2\omega\gamma \ll |\omega_0^2 - \omega^2|$ не выполняется, а потому формула (84.9) неприменима. В действительности при переходе через точку $\omega = \omega_0$ величины ε, n, $\varkappa$ должны меняться *непрерывно*, нигде не обращаясь в бесконечность. В частности, поднимающаяся ветвь кривой $\varepsilon = \varepsilon(\omega)$ левее точки $\omega = \omega_0$ должна переходить в опускающуюся, чтобы сомкнуться с другой поднимающейся ветвью, расположенной правее той же точки.

[1]) Как отметил Рэлей, формула (84.9) была получена двумя годами раньше Максвеллом, который еще в 1869 г. изложил основы механической теории дисперсии в виде экзаменационной задачи. Формулы Максвелла содержали также члены с затуханием.

Это замечание позволяет без особого труда представить ход кривых n и $\varkappa$ в окрестности полосы поглощения. Типичный вид кривых $n = n(\omega)$ и $\varkappa = \varkappa(\omega)$ приведен на рис. 298. Там, где поглощение невелико, показатель преломления $n(\omega)$ возрастает

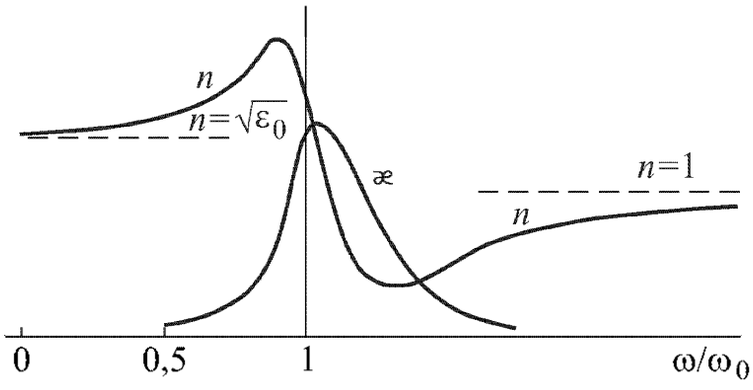


Рис. 298

с частотой. В этом случае говорят, что *дисперсия нормальная*. В области сильного поглощения $n(\omega)$ уменьшается с частотой. Такую дисперсию называют *аномальной*. Ее трудно наблюдать из-за сильного поглощения.

4. Теоретические соображения, изложенные выше, справедливы не только для электронов, *но и для ионов*, и притом в последнем случае классические представления более обоснованы ввиду относительно больших масс ионов. Во всех телах наблюдается не одна, а несколько полос поглощения. Чтобы это учесть, в классической модельной теории принимается, что вещество построено из частиц *различного типа* — электронов и ионов, которые ведут себя как затухающие гармонические осцилляторы с *различными собственными частотами*. В газах можно пренебречь их взаимодействием. Тогда формула (84.5) принимает вид

$$\varepsilon = 1 + \sum_k \frac{4\pi N_k e_k^2/m_k}{\omega_k^2 - \omega^2 + 2i\omega\gamma_k}, \tag{84.10}$$

где N_k, m_k, e_k, ω_k, γ_k — концентрация, масса, заряд, собственная частота и коэффициент затухания осциллятора k-го типа, а суммирование ведется по всем k. Каждой собственной частоте соответствует своя линия поглощения, вблизи которой показатель преломления меняется аномально. Общий ход показателя преломления в зависимости от частоты представлен схематически на рис. 299.

Сравнивая формулу (84.10) с опытом, можно найти коэффициенты $C_k \equiv N_k e_k^2/m_k$. По этим данным можно оценить удельные заряды e_k/m_k и вывести заключения относительно природы осцилляторов. Действительно, заряды e_k по порядку величины

не могут существенно отличаться от элементарного заряда $e = 4{,}8 \cdot 10^{-10}$ СГСЭ, а N_k — от числа атомов в единице объема, т.е. от $N_{\text{Ав}}\rho/A$, где $N_{\text{Ав}}$ — число Авогадро, ρ — плотность вещества, A — атомная масса (по старой терминологии — атомный вес). Таким образом,

$$C_k = \frac{N_k e_k^2}{m_k} \approx \rho \frac{N_{\text{Ав}} e}{A} \frac{e_k}{m_k} = \rho \frac{F}{A} \frac{e_k}{m_k},$$

где $F \equiv N_{\text{Ав}} e = 2{,}9 \cdot 10^{14}$ СГС·г·экв$^{-1}$ — число Фарадея. С помощью этой формулы и можно оценить величины e_k/m_k. Таким путем, в согласии с теоретическими представлениями, было найдено, что все осцилляторы отчетливо подразделяются на две

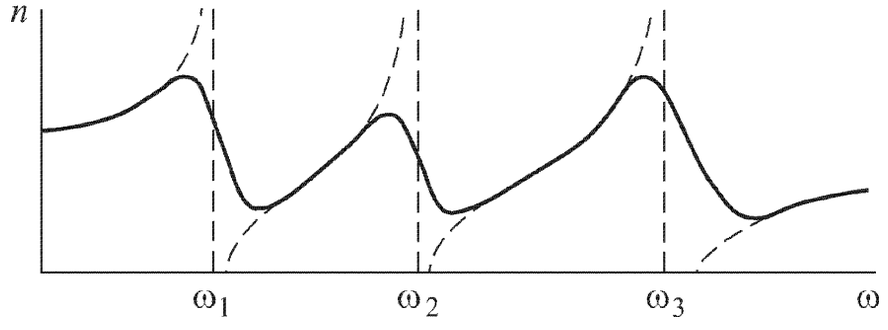


Рис. 299

группы: у одной удельные заряды по порядку величины такие же, как у электронов, у другой — как у ионов. Первым соответствуют полосы поглощения, лежащие в *ультрафиолетовой* (реже — *в видимой*), вторым — в *инфракрасной области спектра.*

5. Формула (84.10) и соответствующие ей формулы для n и $\varkappa$ применимы и в случае *металлов.* В инфракрасной области спектра основную роль играют *свободные электроны*, для которых следует положить $\omega_k = 0$. Влияют также ионы вблизи их собственных частот. В видимой и примыкающей к ней ультрафиолетовой областях спектра формула (84.10) удовлетворительно согласуется с опытом, если наряду со свободными учесть также электроны, *связанные в ионах металла* (колебания самих ионов в этих областях спектра не играют роли).

6. Учтем теперь, хотя далеко не безупречным способом, взаимодействие молекул, поляризованных электрическим полем световой волны. Для этого воспользуемся методом *действующего поля.* В кристаллах кубической системы, построенных из точечных и изотропных атомов, поле $\mathbf{E}'$, действующее на молекулу, связано со средним максвелловским полем $\mathbf{E}$ соотношением

$$\mathbf{E}' = \mathbf{E} + \frac{4\pi}{3}\mathbf{P} \qquad (84.11)$$

(см. т. III, § 35). То же соотношение приближенно справедливо для изотропных газов и жидкостей, построенных из изотропных молекул, если под $\mathbf{E}'$ понимать действующее поле, усредненное по положениям центров молекул в физически бесконечно малом объеме среды. Формула (84.11) выведена в электростатике, но ею можно приближенно пользоваться и в случае переменных полей, когда длина волны велика по сравнению с межмолекулярными расстояниями.

Будем предполагать, что все молекулы одинаковы и каждая из них содержит только один оптический электрон. Кроме того, пренебрежем затуханием, предполагая, что частота ω далека от полосы поглощения. В этих предположениях все, что нужно сделать, — это заменить в уравнении (84.1) $\mathbf{E}'$ выражением (84.11). Последующие вычисления крайне элементарны, но мы опустим их, так как они уже были выполнены в электростатике (см. т. III, § 35). Там по методу, в точности совпадающему с излагаемым здесь, была получена формула Мосотти–Клаузиуса:

$$\frac{\varepsilon-1}{\varepsilon+2}=\frac{4\pi}{3}N\beta. \tag{84.12}$$

Подставив сюда $\varepsilon=n^2$ и выражение для β из формулы (84.3), придем к соотношению

$$\frac{n^2-1}{n^2+2}=\frac{4\pi}{3}\frac{Ne^2/m}{\omega_0^2-\omega^2}. \tag{84.13}$$

Эта формула была одновременно (1880 г.) получена в электронной теории голландским физиком Г.А. Лорентцом и датским физиком Л. Лоренцом (1829–1891) еще из домаксвелловских представлений о свете. Она называется *формулой Лорентца–Лоренца*. Из нее следует, что для данного вещества при неизменной частоте ω должно выполняться соотношение

$$r\equiv\frac{1}{\rho}\frac{n^2-1}{n^2+2}=\text{const}, \tag{84.14}$$

поскольку концентрация N молекул пропорциональна плотности вещества ρ. Величина, стоящая в левой части этого соотношения, называется *удельной рефракцией*. Таким образом, согласно формуле Лорентца–Лоренца *удельная рефракция вещества не должна изменяться при изменении плотности.*

В ряде случаев формула (84.14) хорошо выполняется. Так, для воздуха при изменении давления от 1 до 200 атм постоянство удельной рефракции выполняется с ошибкой, не превышающей 0,1 %. Нередко удельная рефракция остается приблизительно постоянной даже при переходе вещества из парообразного состояния в жидкое. Так, при ожижении газообразного кислорода (изменение плотности в 800 раз) или конденсации водяного пара (изменение плотности в 1200 раз) удельная рефракция остает-

ся постоянной с точностью до 2–3 %. Эти факты удивительны, так как в приведенных примерах предположения, в которых выведена формула (84.11), несомненно, не выполняются. Имеется, однако, много случаев отступлений от формулы (84.14).

7. Еще более удивительно эмпирическое правило, по которому можно приближенно вычислять рефракцию смеси или химического соединения. Если вещество состоит из одинаковых частиц (атомов или молекул), то $\delta = Nm$, где m — масса одной частицы. Поэтому из формулы (84.12) с учетом соотношения $\varepsilon = n^2$ получаем

$$\beta = \frac{3}{4\pi} mr. \tag{84.15}$$

Поляризуемость частицы β, а с ней и величина mr есть свойство *только самой частицы*. Правило, которое было упомянуто выше, состоит в том, что *поляризуемость частицы аддитивна, т.е. поляризуемость сложной частицы равна сумме поляризуемостей частиц, из которых она состоит*. Следовательно,

$$mr = \sum m_i r_i, \tag{84.16}$$

где m и m_i — массы составной и составляющих частиц, а r и r_i — удельные рефракции соответствующих им химических веществ. Удивительно, что правило это относится не только к механическим смесям, но и *к химическим соединениям*. Возьмем в качестве составляющих частиц атомы, а составных — молекулы. Тогда $m_i = A_i m_{\text{H}}$, $m = M m_{\text{H}}$, где m_{H} — масса атома водорода, а M и A_i — соответствующие молекулярная и атомные массы. Поэтому формула (84.16) переходит в

$$Mr = \sum A_i r_i. \tag{84.17}$$

Произведение удельной рефракции на атомную массу называется *атомной*, а на молекулярную массу — *молекулярной рефракцией*. Таким образом, *молекулярная рефракция химического соединения равна сумме атомных рефракций*.

В книге Борна [1]) приведены обширные таблицы, подтверждающие формулу Лорентца–Лоренца и аддитивность атомной рефракции. Там же приведены примеры и значительных отступлений от этих правил.

8. Помимо дисперсии, связанной с вынужденными колебаниями электронов и ионов, существует другой вид дисперсии, на который обратил внимание Дебай. Она проявляется в сре-

[1]) *Макс Борн.* Оптика.— Харьков; Киев: Гос. научно-техн. изд-во Украины. 1937.

дах с *полярными молекулами*, главным образом в жидкостях. Полярные — это такие молекулы, которые обладают «твердыми» дипольными моментами, обусловленными несимметричным распределением зарядов внутри молекулы. Полярные молекулы могут обладать и «индуцированными» дипольными моментами, возбуждаемыми внешним электрическим полем. Как правило, индуцированные моменты полярных молекул малы по сравнению с твердыми.

В отсутствие электрического поля дипольные моменты полярных молекул изотропной среды ориентируются в пространстве по всем направлениям и притом хаотически. В этом случае среда никакой электрической поляризацией не обладает. При наложении электрического поля оно стремится ориентировать дипольные моменты молекул вдоль поля, причем эта ориентация постоянно нарушается из-за теплового движения. В результате в среде возникает электрическая поляризация (см. т. III, § 36). Если электрическое поле меняется во времени с частотой ω, то с той же частотой будет меняться эта «ориентационная» поляризация и обусловленная ею диэлектрическая проницаемость среды. Амплитуды вынужденных вращений полярных молекул зависят от частоты ω приложенного электрического поля. С этим связана дисперсия *вращательной части* диэлектрической проницаемости и показателя преломления.

Не входя в подробное количественное рассмотрение, ограничимся качественными замечаниями.

В полях низких частот успевает установиться такая же мгновенная поляризация среды, как и в статических полях. В такой области частот величины ε и n практически постоянны, т.е. не зависят от ω. Если дипольные моменты полярных молекул велики, как, например, у воды и спиртов, то в этой области частот значения ε и n велики. Так, в случае воды $\varepsilon = 81$, $n = \sqrt{\varepsilon} = 9$.

В другом предельном случае очень высоких частот дипольные молекулы не успевают заметно поворачиваться за времена порядка периода световых колебаний. В этой области частот практически нет вынужденных вращений молекул, т.е. вращательные части ε и n выпадают. Поэтому должна существовать промежуточная область частот, в которой величины ε и n аномально сильно уменьшаются до тех значений, которые они принимают в оптической области спектра. Такая область частот у воды и спиртов лежит в диапазоне сантиметровых радиоволн. За пределами этой области (со стороны высоких частот) молекулы начинают вести себя так, как если бы они не были полярными. Этим объясняется резкое расхождение между показателем преломления воды n в оптической области спектра и значением $\sqrt{\varepsilon}$ в электростатике и области низкочастотных электромагнитных волн.

Этим же объясняется, почему в оптике магнитную проницаемость μ почти всегда можно считать равной единице для всех веществ. Действительно, механизм намагничивания — такой же, как и поляризации полярных диэлектриков. Он состоит в ориентации магнитных моментов атомов и молекул во внешнем магнитном поле. Но этот эффект не имеет места в высокочастотных полях, видимой и примыкающих к ней областях спектра.

9. Подводя итоги, дадим схематический обзор дисперсии во всем диапазоне частот электромагнитных волн. В области радиоволн длиннее примерно 1 см существенна только одна собственная частота $\omega_0 = 0$, которой обладают свободные электроны или ионы. В указанном радиодиапазопе дисперсией практически могут обладать *только ионизованные газы* (см. § 86). Если свободных электронов или ионов нет, то в этом диапазоне нет и дисперсии.

В полярных диэлектриках дисперсия появляется примерно с сантиметрового диапазона радиоволн. Показатель преломления убывает с частотой, т.е. дисперсия *аномальна*, а поглощение велико.

По мере приближения к собственным частотам ионов дисперсия становится *нормальной*, а внутри самой полосы поглощения, обусловленной колебаниями ионов, — *аномальной*. После прохождения через инфракрасную область поглощения вдали от нее вынужденные колебания ионов практически прекращаются. Здесь, а также в видимой и ультрафиолетовой областях спектра дисперсия вызывается колебаниями только *связанных оптических электронов*. В прозрачной области (точнее, области слабого поглощения) она имеет *нормальный*, а в области сильного поглощения — *аномальный* характер. Полоса поглощения лежит обычно в области ультрафиолета, а иногда в видимой области.

При переходе через полосу поглощения, т.е. область аномальной дисперсии, сбусловленную колебаниями электронов, дисперсия снова становится *нормальной*. Если частота ω велика по сравнению со всеми собственными частотами электронов, то в формуле (84.10) величинами ω_k^2 можно пренебречь по сравнению с ω^2. Мы вступили в область *мягкого рентгеновского излучения*. Если не учитывать поглощение и заметить, что в колебаниях принимают участие одни только электроны, то дисперсионной формуле (84.10) в указанной области спектра можно придать вид

$$n^2 = 1 - \frac{Ne^2}{m\omega^2}. \qquad (84.18)$$

Показатель преломления здесь *меньше единицы*, хотя и очень мало отличается от нее. Его удалось измерить, наблюдая отклонение рентгеновских лучей в призме из различных материалов, хотя этот метод и не может дать большой точности.

Более падежные результаты дают измерения предельного угла полного отражения, испытываемого рентгеновскими лучами при переходе из воздуха в твердое тело ($n < 1$). Оказалось, что показатель преломления рентгеновских лучей меньше единицы на величину порядка $(10^{-5} - 10^{-6})\rho$, где ρ — плотность вещества. Так, в случае стекла (крон с плотностью 2,52) для $\lambda = 0{,}1279$ нм он оказался равным $n = 0{,}999995 = 1 - 5 \cdot 10^{-6}$.

При увеличении жесткости рентгеновского излучения резонансные явления могут проявляться на электронах *внутренних оболочек атомов*, связанных с ними более сильно, чем оптические электроны. Действительно, удалось наблюдать аномальную дисперсию рентгеновских лучей вблизи частот их истинного фотоэлектрического поглощения.

ЗАДАЧИ

1. Найти максимальную скорость вынужденных колебаний свободного электрона в поле солнечного излучения вблизи земной поверхности (см. задачу к § 5). Определить также отношение максимальной силы F_B, действующей на такой электрон со стороны магнитного поля, к максимальной силе F_E, действующей со стороны электрического поля. Поле солнечного излучения заменить монохроматическим $\mathbf{E} = \mathbf{E}_0 \cos \omega t$ с длиной волны $\lambda = 550$ нм.

Р е ш е н и е. В уравнении движения свободного электрона $m\dot{\mathbf{v}} = e\left(\mathbf{E} + \frac{1}{c}[\mathbf{vB}]\right)$ пренебрежем действием магнитного поля. Тогда

$$m\dot{\mathbf{v}} = e\mathbf{E} = e\mathbf{E}_0 \cos \omega t, \quad \mathbf{v} = \frac{e}{m\omega}\mathbf{E}_0 \sin \omega t, \quad v_{\text{макс}} = \frac{eE_0}{m\omega} = \frac{eE_0\lambda}{2\pi mc}.$$

Так как $\overline{E^2} = \frac{1}{2}E_0^2 = 5{,}85 \cdot 10^{-4}$, то $E_0 = \sqrt{2\overline{E^2}} = 0{,}034$ СГСЭ. Подставляя это значение, получим

$$v_{\text{макс}} = 5{,}3 \text{ см/с}, \quad \frac{F_B}{F_E} = \frac{v_{\text{макс}}}{2c} = 0{,}9 \cdot 10^{-10}.$$

2. Определить среднюю силу светового давления на гармонический затухающий осциллятор, колеблющийся в поле монохроматической однородной световой волны. Рассчитать также среднюю энергию ε, поглощаемую осциллятором в одну секунду, и выразить через нее среднюю силу светового давления. Показать, что если бы поглощения света не было, то средняя сила светового давления равнялась бы нулю.

Р е ш е н и е. В месте нахождения осциллятора электромагнитное поле имеет вид $\mathbf{E} = A\mathbf{e}e^{i\omega t}$, $\mathbf{B} = A\mathbf{b}e^{i\omega t}$, где $\mathbf{e}$ и $\mathbf{b}$ — взаимно перпендикулярные единичные векторы. Пренебрегая сначала действием магнитного поля, дифференцированием (84.2) находим

$$\mathbf{v} = \frac{i\omega e/m}{(\omega_0^2 - \omega^2) + 2i\omega\gamma}\mathbf{E}.$$

Сила, действующая на электрон со стороны магнитного поля, $\mathbf{F} = \frac{e}{c}[\mathbf{vB}]$. Поскольку, однако, вычисления ведутся в комплексной форме, следует заменить это выражение на

$$\mathbf{F} = \frac{e}{4c}[(\mathbf{v}+\mathbf{v}^*)(\mathbf{B}+\mathbf{B}^*)].$$

Для средней силы получаем

$$\overline{\mathbf{F}} = \frac{e}{4c}\{[\mathbf{v}^*\mathbf{B}] + [\mathbf{v}\mathbf{B}^*]\},$$

или после простых преобразований

$$\overline{\mathbf{F}} = \frac{m\gamma}{c}\mathbf{v}\mathbf{v}^*[\mathbf{eb}].$$

Отсюда видно, что средняя сила светового давления направлена в сторону векторного произведения $[\mathbf{eb}]$, т.е. в сторону распространения волны.

Энергия, поглощаемая в единицу времени:

$$\int_0^1 2m\gamma\left(\frac{\mathbf{v}+\mathbf{v}^*}{2}\right)^2 dt,$$

а ее среднее значение

$$\varepsilon = \int_0^1 m\gamma\mathbf{v}\mathbf{v}^*\,dt = m\gamma\mathbf{v}\mathbf{v}^*.$$

Таким образом,

$$\overline{F} = \varepsilon/c. \tag{84.19}$$

Эта формула может быть получена из простых соображений. Излучение, поглощаемое атомом вещества, передает ему не только энергию ε, но и связанный с ней импульс $g = \varepsilon/c$, который проявляется в силах светового давления.

Если свет распространяется через поглощающий газ, то, согласно изложенному, он должен производить на него давление. Впервые такое предположение высказал Кеплер, изучая формы кометных хвостов. Экспериментальное доказательство существования светового давления на газы было дано П.Н. Лебедевым (1910 г.).

§ 85. Понятие о дисперсионной формуле квантовой механики

1. Классическая дисперсионная формула (84.10) для газов с большой точностью описывает фактически наблюдаемый ход показателя преломления вблизи отдельных линий поглощения ω_k, но лишь при том условии, если коэффициенты N_k, а также собственные частоты ω_k и коэффициенты затухания γ_k рассматриваются как *эмпирические постоянные, определяемые из самой кривой дисперсии и фактического положения спектральных линий в спектре излучения или поглощения вещества.* В

частности, для согласования с опытом оказалось необходимым в этой формуле величины N_k заменить на $f_k N_k$, где f_k — постоянные коэффициенты, меньшие единицы, называемые *силами осцилляторов.*

Теорию дисперсии в квантовой механике можно строить по той же схеме, что и в классической физике. Задача сводится к вычислению *поляризуемости атомов и молекул* в электрическом поле световой волны. Но при решении этой задачи надо пользоваться не классическими, а *квантомеханическими* уравнениями движения, например, в форме *волнового уравнения Шредингера* (1887–1961). Поскольку в данной книге квантовая механика не предполагается известной, систематическое изложение квантовой теории дисперсии в ней невозможно. Можно дать только общую характеристику и некоторые результаты этой теории.

2. Квантовая теория дисперсии (и в этом одно из преимуществ ее перед классической) не пользуется моделью квазиупруго связанного электрона и не вводит никаких в действительности не существующих сил. Квантовая механика объясняет строение и устойчивость атомных систем с помощью *одних только электрических сил.* Для простоты будем предполагать, что в атоме есть всего только один *валентный электрон*, определяющий его химические и оптические свойства. Остальные электроны прочно удерживаются на внутренних оболочках, обладая значительно бóльшими энергиями связи. Состояния таких электронов практически не возмущаются слабым электрическим полем световой волны. Их роль сводится только к изменению электрического поля атомного ядра, в котором движется валентный электрон. Именно по этой причине атом может рассматриваться как *одноэлектронный.* Такая модель атома весьма близко соответствует действительности.

В отсутствие внешних полей атом может находиться только во вполне определенных *стационарных состояниях*, в которых его энергия может принимать также вполне определенные дискретные значения $W_1, W_2, W_3, \ldots$ (*уровни энергии*). Их принято нумеровать в порядке возрастания энергии. При переходе атома с высшего энергетического уровня k на более низкий уровень n происходит излучение света с испусканием кванта энергии

$$\hbar\omega_{kn} \equiv h\nu_{kn} = W_k - W_n. \tag{85.1}$$

При обратном переходе атома с низшего энергетического уровня n на более высокий уровень k происходит возбуждение атома с поглощением такого же кванта. Таким образом, в отличие от классического гармонического осциллятора, атом, даже если он одноэлектронный, излучает не одну частоту ω_0, а целый спектр

частот ω_{kn}, которые в квантовой теории дисперсии и играют роль собственных частот атома. Если переход происходит с более низкого уровня k на более высокий уровень n (поглощение), то для сохранения без изменения соотношения (84.10) удобно ввести *отрицательные частоты* ω_{kn}. Если нет внешних возмущений (отсутствие силовых полей, невысокие температуры), то в результате процессов излучения все атомы перейдут на *низший* или *основной* энергетический уровень, т.е. в так называемое *основное* или *нормальное состояние.* На основном уровне изолированный атом будет находиться неограниченно долго, пока в результате внешнего воздействия он не перейдет на другой уровень.

При включении внешнего поля атом выйдет из исходного основного стационарного состояния. Возникшее состояние, согласно квантовой механике, можно характеризовать как суперпозицию стационарных состояний, каждому из которых приписывается определенная вероятность. Если электрическое поле слабое и гармонически меняется во времени с частотой ω, то эти вероятности также будут меняться во времени с той же частотой. Появится переменный дипольный момент атома, индуцированный электрическим полем световой волны. Его можно представить в виде $\mathbf{p} = \beta\mathbf{E}$. В общем случае поляризуемость атома β есть *тензор*, но мы ограничимся простейшим случаем, когда он *скаляр*. Если не учитывать поглощения (что допустимо вдали от собственных частот ω_{nk}), то квантовая механика приводит к результату

$$\beta = \frac{e^2}{m}\sum_k \frac{f_{nk}}{\omega_{nk}^2 - \omega^2}, \tag{85.2}$$

где f_{nk} — постоянные. Формула предполагает, что до включения электрического поля атом находился в стационарном состоянии n. Суммирование ведется по всем возможным значениям k, как бóльшим, так и меньшим n.

Для выяснения физического смысла постоянных f_{nk} заметим, что существует определенная вероятность, что за определенное время (например, равное единице) атом в отсутствие внешних полей *самопроизвольно* перейдет из высшего стационарного состояния k в более низкое стационарное состояние n. Постоянная f_{nk} пропорциональна этой вероятности. Самопроизвольный переход из низшего состояния n в высшее состояние k, конечно, невозможен. В этом случае f_{kn} определяется соотношением $f_{kn} = -f_{nk}$. Следовательно, при $n = k$ величины f_{nk} равны нулю, так что суммирование в (85.2) производится только по $k \neq n$. Величина f_{nk} и называется *силой осциллятора с частотой* ω_{nk}.

Используя выражения (85.2), приходим к следующей дисперсионной формуле для атомов с одним валентным электроном:

$$n^2 = 1 + 4\pi N \frac{e^2}{m} \sum_{k \neq n} \frac{f_{nk}}{\omega_{nk}^2 - \omega^2}. \tag{85.3}$$

Сравнение этой формулы с классической формулой (84.10) (без учета поглощения) приводит к следующей наглядной аналогии. Все атомы, находящиеся в одном и том же исходном состоянии n, как бы разбиваются на группы, нумеруемые индексом k. Атомы одной и той же группы ведут себя как *классические гармонические осцилляторы* с собственной частотой ω_{nk}. Если $k > n$, то частота ω_{nk} считается *положительной*, а при $k < n$ — *отрицательной*. Число атомов k-й группы в единице объема определяется выражением $N_k = f_{nk} N$. Так как $\sum N_k = N$, то должно быть

$$\sum_k f_{nk} = 1. \tag{85.4}$$

Если в атоме Z валентных электронов, то это соотношение переходит в соотношение

$$\sum_k f_{nk} = Z. \tag{85.5}$$

Полученные формулы выражают так называемое *правило сумм*, найденное независимо друг от друга Томасом и Рейхе, с одной стороны, и Куном, с другой (1925 г.). Приведенные наводящие рассуждения не могут служить доказательством формул (85.4) и (85.5). Это видно уже из того, что при $k < n$ силы осцилляторов f_{nk}, а потому и числа атомов N_k отрицательны. Однако сами формулы (85.4) и (85.5) могут быть строго доказаны методами квантовой механики.

3. Если до включения электрического поля атомы газа не находились в одном и том же стационарном состоянии, а были распределены по таким состояниям, то формулу (85.3) следует заменить на формулу

$$n^2 = 1 + 4\pi N \frac{e^2}{m} \sum_n \sum_k \frac{f_{nk}}{\omega_{nk}^2 - \omega^2}. \tag{85.6}$$

Таким же путем должны быть обобщены и правила сумм (85.4) и (85.5) — в них суммирование надо производить не только по всем значениям k, но и по всем значениям n.

Принципиально новое явление, предсказанное квантовой механикой, состоит в том, что в формулу (85.6) могут входить слагаемые не только с положительными f_{nk} (когда $k > n$), но

и с отрицательными (когда $k < n$). Так как $\omega_{nk}^2 = \omega_{kn}^2$, то в дисперсионную формулу войдут только суммы $(f_{nk} + f_{kn})$, а не каждое слагаемое в отдельности. Если на нижнем энергетическом уровне находится больше атомов, чем на верхнем, то $(f_{nk} + f_{kn}) > 0$, в противоположном случае $(f_{nk} + f_{kn}) < 0$. В первом случае собственная частота $|\omega_{nk}|$ внесет в дисперсионную формулу (85.6) слагаемое того же знака, что и в классическую формулу, во втором — противоположного. В соответствии с этим различают *положительную* и *отрицательную дисперсию*. Ход показателя преломления $n(\omega)$ вблизи собственной частоты $\omega_0 = |\omega_{nk}|$ в обоих случаях представлен схематически на рис. 300 (n_+ для положительной, n_- для отрицательной дисперсии). Что касается коэффициента затухания γ, то в случае отрицательной дисперсии он также *отрицателен*, т.е. при распространении света имеет место его *усиление*, а не ослабление. Это происходит, конечно, за счет переходов атомов с более высокого энергетического уровня на более низкий с испусканием квантов света (индуцированное излучение, см. § 119).

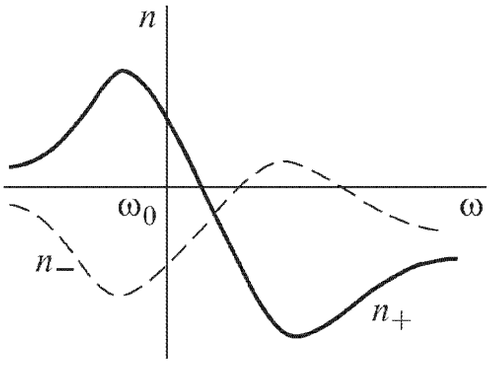


Рис. 300

В обычных условиях до включения внешнего поля атомы среды находятся на самом низком — основном — энергетическом уровне. В этом случае световая волна будет переводить атомы только на более высокие уровни, так что все f_{nk} окажутся положительными. То же самое будет происходить и при тепловом возбуждении в случае термодинамического равновесия, так как, в соответствии с формулой Больцмана, число атомов в исходном состоянии будет убывать с возрастанием номера энергетического уровня. Однако, применяя нетепловые методы возбуждения, например, пропуская через вещество сильные электрические токи, можно создавать *термодинамически неравновесные* метастабильные состояния вещества с *инверсной заселенностью энергетических уровней*, как это в действительности и делается в лазерах. Под инверсной заселенностью двух различных энергетических уровней понимают такое состояние, когда на верхнем уровне находится больше атомов, чем на нижнем. В этом случае можно получать и действительно получают среду с отрицательной дисперсией. Влияние отрицательных слагаемых в дисперсионной формуле впервые (1930 г.) наблюдал Ладенбург в газе при прохождении через него сильного электрического разряда, хотя дисперсия в целом в его опытах и оставалась положительной.

§ 86. Методы экспериментального исследования аномальной дисперсии

1. Аномальная дисперсия впервые наблюдалась Леру (1832–1907) в 1861 г. Он обнаружил, что в призме, наполненной парами йода, синие лучи преломляются меньше красных (промежуточные лучи парами йода сильно поглощаются и ускользают от наблюдения). Отсюда следует, что при переходе от красных к синим лучам показатель преломления паров йода должен убывать, т.е. в указанной области спектра дисперсия паров йода *аномальна*.

Открытие Леру не обратило на себя должного внимания, пока Христиансен (1848–1917) в 1870 г. не обнаружил и детально изучил аномальную дисперсию в растворе фуксина в спирте. Здесь показатель преломления наименьший ($n = 1{,}285$) для фиолетовых и наибольший ($n = 1{,}561$) для желтых лучей, зеленые лучи поглощаются.

Систематические экспериментальные исследования аномальной дисперсии были выполнены Кундтом (1839–1894). Он установил важный экспериментальный результат, что *аномальный ход дисперсии всегда сопровождается поглощением*. После экспериментальных работ Кундта и разработки теории дисперсии стало ясно, что аномальная дисперсия не есть какое-то редкое явление природы, а должна наблюдаться у всех веществ в тех областях спектра, где имеется сильное поглощение.

При изучении аномальной дисперсии Кундт пользовался *методом скрещенных призм*, который применялся еще Ньютоном в его исследованиях по дисперсии света. Идея метода состоит в следующем. Берутся две призмы, из которых первая изготовлена из вещества с *нормальной дисперсией* (обычно из стекла), вторая — из исследуемого вещества. Источником света, как во всяком спектроскопе, служит освещаемая щель, помещаемая в фокальной плоскости коллиматорной линзы. Первая призма, ребро которой устанавливается *вертикально*, разлагает падающий свет в цветную горизонтальную полоску (спектр). Вторая призма, преломляющее ребро которой *горизонтально*, помещается за первой призмой. Она смещает каждую точку полоски вверх или вниз, в зависимости от того, куда обращена эта призма своим основанием: вверх или вниз. Смещение зависит от длины волны. Вследствие этого полоска искривляется и становится наклонной. Если дисперсия второй призмы нормальная, то полоска монотонно поднимается или опускается (рис. 301 *а*). Если же она аномальная, то в результате поглощения лучей с аномальной преломляемостью полоска разрывается на две части, края которых, примыкающие к полосе поглощения, загибаются в противоположные стороны (рис. 301 *б*). Разумеется, для по-

лучения описанной картины должна применяться система линз, дающая на экране изображение освещаемой щели в различных цветах спектра.

В изящной лекционной демонстрации Кундта, широко применяющейся до настоящего времени, роль второй призмы играло конусообразное пламя горелки, в которое вводился металлический натрий. В результате прохождения света через пламя горелки в спектре наблюдалась не только темная полоска, соот-

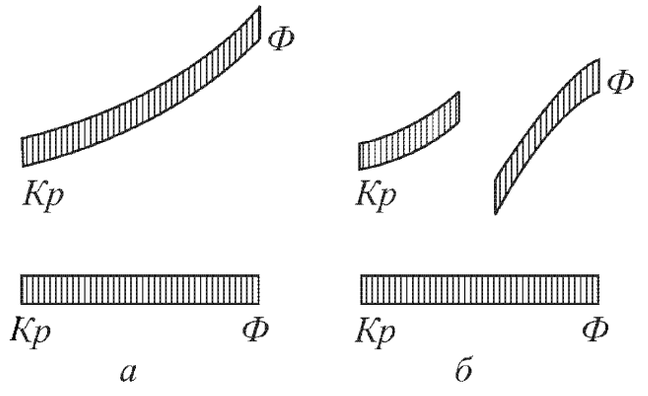


Рис. 301

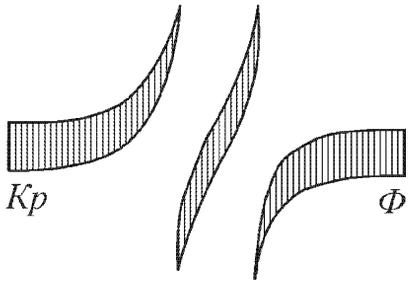


Рис. 302

ветствующая желтой линии поглощения паров натрия, но и загиб в противоположные стороны разорванных частей спектра, примыкающих к области поглощения (рис. 301 *б*). D-линия поглощения паров натрия в желтой части спектра состоит из двух близко расположенных тонких линий с длинами волн $\lambda_{D_1} = 589{,}0$ и $\lambda_{D_2} = 589{,}6$ нм. В описанной демонстрации плотность паров натрия была настолько велика, что обе линии сливались в одну полоску, и детали явления не разрешались. Их можно разрешить, уменьшая плотность паров натрия и улучшая условия опыта. Тогда можно наблюдать две области аномальной дисперсии, соответствующие линиям поглощения с длинами волн λ_{D_1} и λ_{D_2} (рис. 302).

2. Как уже указывалось в п. 2 § 84, проверку теории лучше всего производить на газах малой плотности, характеризующихся тонкими и резкими линиями поглощения. Однако при малой плотности, необходимой для работы вблизи линии поглощения, показатели преломления газов малы. Для их измерения наилучшим является интерференционный метод, предложенный Пуччианти (1901 г.) и усовершенствованный Д.С. Рождественским в его классических работах по аномальной дисперсии в парах натрия (1912 г.).

В этом методе интерферометр (например, Жамена) скрещивается со спектрографом (дифракционной решеткой или призмой с большой дисперсией). Интерферометр устанавливают так, что он дает в белом свете горизонтальные интерференцион-

ные полосы на щели спектрографа, установленной вертикально. Цветную интерференционную картину, полученную на щели, спектрограф развертывает в горизонтальном направлении. Спектр оказывается пересеченным в продольном направлении интерференционными полосами, каждая из которых характеризуется одним и тем же порядком интерференции. Цвет полосы меняется вдоль ее длины от красного к фиолетовому, а расстояния между полосами при этом уменьшаются из-за уменьшения длины волны. Интерференционным минимумам соответствуют темные линии. Таким образом, спектр будет пересечен в продольном направлении темными линиями, сужающимися от красного конца спектра к фиолетовому.

Перед измерением прибор регулируется так, чтобы нулевая интерференционная полоса была прямолинейной и горизонтальной. Примем ее за ось X с направлением от красного конца к фиолетовому, а ось Y направим параллельно щели спектрографа. Ширина интерференционной полосы пропорциональна λ и может быть представлена в виде $\Delta y = \alpha\lambda(x)$, где α — постоянная прибора, практически не зависящая от номера полосы (порядка интерференции) k. Поэтому ордината k-й полосы будет $y_k = k\,\Delta y = \alpha k\lambda(x)$. Это есть уравнение кривой, определяющей форму рассматриваемой полосы.

Введем теперь в одно из плеч интерферометра слой исследуемого вещества толщины l с показателем преломления $n(\lambda)$. Тогда добавится разность хода $(n-1)l$, в результате чего полосы интерференции сместятся вверх или вниз на $(n-1)l/\lambda$ полос, т.е. на расстояние $(n-1)l/\lambda(\Delta y)$. Уравнение, определяющее форму k-й полосы, примет вид

$$y_k = \alpha[k\lambda(x) \pm (n-1)l] \tag{86.1}$$

(знак определяется тем, в какое из плеч интерферометра введен слой исследуемого вещества). В частности, при $k = 0$ получаем $y_0 = \alpha(n-1)l$. Отсюда видно, что нулевая интерференционная полоса вычерчивает в определенном масштабе график зависимости показателя преломления n от абсциссы x, а следовательно, и от длины волны λ, т.е. определяет дисперсию исследуемого вещества. Полосы ненулевого порядка имеют дополнительный наклон, изменяющийся с изменением k (см. рис. 303, заимствованный из работ Рождественского).

3. Небольшое изменение метода позволило Рождественскому значительно повысить точность измерений в окрестности полосы поглощения. Измененный метод получил название *метода крюков*. Допустим, что в одно из плеч интерферометра введено исследуемое вещество (газ или пар), а в другое — стеклянная пластинка толщины $l_{\text{ст}}$ с показателем преломления $n_{\text{ст}}$. Пластинка вносит между интерферирующими пучками разность хо-

да $(n_{\text{ст}}-1)l_{\text{ст}}$, смещая интерференционную картину вверх или вниз на расстояние $(n_{\text{ст}}-1)l_{\text{ст}}\alpha$. Теперь ордината k-й полосы будет определяться выражением

$$y_k = \alpha[k\lambda(x) \pm (n-1)l \mp (n_{\text{ст}}-1)l_{\text{ст}}]. \qquad (86.2)$$

Интерференционная полоса нулевого порядка уйдет из поля зрения. Ее место в поле зрения займет другая интерференционная

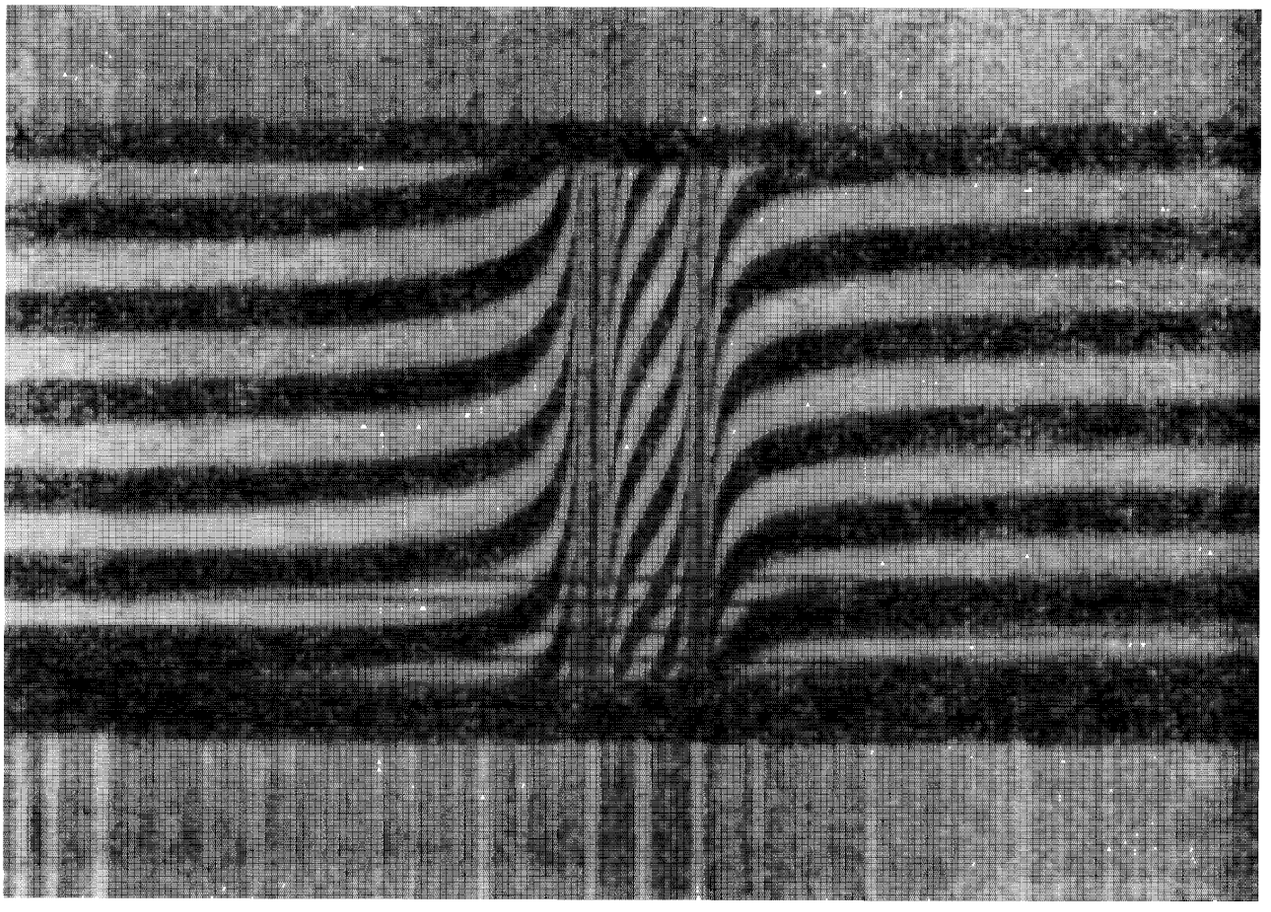


Рис. 303

полоса, порядок k которой найдется, если в (86.2) положить $y_k = 0$. Это дает

$$k = \pm\frac{n_{\text{ст}}-1}{\lambda}\,l_{\text{ст}} \mp \frac{n-1}{\lambda}\,l. \qquad (86.3)$$

Из-за близости к единице показателей преломления газов последнее слагаемое мало и при вычислении k может не учитываться. Таким путем для пластинки с $l_{\text{ст}} = 1$ мм, $n_{\text{ст}} = 1{,}5$ при длине волны $\lambda = 600$ нм получаем $k \sim 10^4$. В белом свете интерференция столь высокого порядка невозможна. Интерференция в белом свете становится возможной благодаря наличию спектроскопа, играющего роль монохроматора, пространственно разлагающего белый свет на квазимонохроматические составляющие. Поэтому длина волны λ вдоль каждой интерференционной полосы изменяется, т.е. является функцией координаты x.

Дифференцируя y_k по x при постоянном k, находим наклон k-й интерференционной полосы:

$$\frac{dy_k}{dx} = \alpha \frac{d\lambda}{dx}\left(k \pm l\frac{dn}{d\lambda} \mp l_{\text{ст}}\frac{dn_{\text{ст}}}{d\lambda}\right). \tag{86.4}$$

Благодаря малой дисперсии стекла последнее слагаемое мало и может быть отброшено. Если в отсутствие стеклянной пластинки и исследуемого вещества нулевая интерференционная полоса горизонтальна, то внесение одной только пластинки делает интерференционные полосы наклонными. Наклон полосы определяется производной $\dfrac{dy_k}{dx} = k\alpha\dfrac{d\lambda}{dx}$, т.е. он тем больше, чем больше порядок интерференции k. Наклон, вызываемый стеклянной пластинкой, происходит в одну и ту же сторону во всей исследуемой области спектра. При перенесении пластинки в другое плечо интерферометра знак k, а с ним и наклон интерференционных полос меняются на противоположные. Поместим пластинку в то плечо интерферометра, чтобы было $k > 0$, т.е.

$$k \approx \frac{n_{\text{ст}} - 1}{\lambda}\, l_{\text{ст}}. \tag{86.5}$$

Тогда при нашем выборе положительного направления оси X ($d\lambda/dx < 0$) получится $dy_k/dx < 0$, т.е. полосы интерференции будут наклонены сверху (от положительного конца оси Y) вниз направо (в сторону положительного — фиолетового — конца оси X).

Допустим теперь, что во второе плечо интерферометра введен исследуемый газ. Показатель преломления газа вдали от линии поглощения очень близок к единице и на наклон полос практически не оказывает никакого влияния. Зато вблизи линии поглощения велика производная $dn/d\lambda$, вызывающая сильный загиб интерференционных полос. В отсутствие пластинки наклон будет определяться производной

$$\frac{dy_k}{dx} = -\alpha l \frac{d\lambda}{dx}\frac{dn}{d\lambda}.$$

Следовательно, в области нормальной дисперсии ($dn/d\lambda < 0$) получится $dy_k/dx > 0$, т.е. в этом случае интерференционные полосы пойдут сверху вниз налево.

Таким образом, в области нормальной дисперсии стеклянная пластинка и исследуемый газ вызывают наклон полос в противоположные стороны. При одновременном действии газа и стеклянной пластинки с надлежаще подобранной толщиной эти действия в определенной точке интерференционной полосы могут компенсировать друг друга. В такой точке касательная к интерференционной полосе станет горизонтальной, т.е. на полосе получится *крюк*, по выражению Рождественского. Вблизи линии

поглощения таких крюков получится два. Оба они расположены в области нормальной дисперсии. Слева (со стороны длинных волн) крюк будет обращен вершиною вниз, а справа (со стороны коротких волн) — вверх (рис. 304). В вершине крюка должно быть $dy_k/dx = 0$, или $dy_k/d\lambda = 0$, т.е.

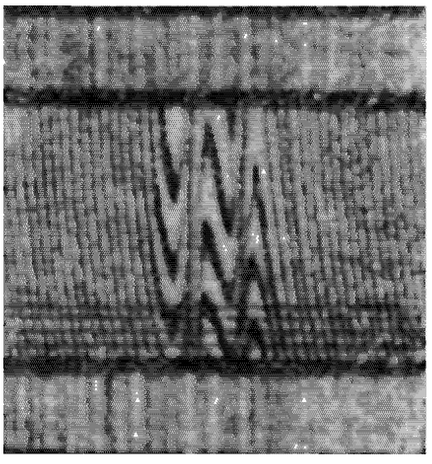


Рис. 304

$$k + l\frac{dn}{d\lambda} - l_{\text{ст}}\frac{dn_{\text{ст}}}{d\lambda} = 0,$$

или

$$l\frac{dn}{d\lambda} = -k + l_{\text{ст}}\frac{dn_{\text{ст}}}{d\lambda}. \qquad (86.6)$$

Последнее слагаемое, как уже указывалось, пренебрежимо мало, а первое может быть вычислено по формуле (86.5). Таким образом, по формуле (86.6) можно вычислить значения дисперсии газа $dn/d\lambda$ для тех значений λ, которые соответствуют вершинам крюков, т.е. точкам загиба интерференционных полос.

Уже в начале своих исследований Рождественский убедился, что вдали от линии поглощения формула Зельмейера правильно передает ход показателя преломления в зависимости от длины волны. Эту формулу следует писать в виде

$$n^2 = 1 + 4\pi\frac{Nfe^2/m}{\omega_0^2 - \omega^2} = 1 + \frac{Nf\lambda_0^2\lambda^2 e^2/m}{\pi c^2(\lambda^2 - \lambda_0^2)}, \qquad (86.7)$$

где f — сила осциллятора. Обозначим через $\Delta\lambda$ расстояние вершины одного из крюков от линии поглощения λ_0 (тогда расстояние между вершинами обоих крюков будет $2\Delta\lambda$). Вычислим по формуле (86.7) производную $dn/d\lambda$, учитывая при этом, что $|\Delta\lambda| \ll \lambda_0$ и $n - 1 \ll 1$. Тогда из формулы (86.6), пренебрегая последним слагаемым, найдем

$$f = \frac{4\pi c^2 k}{Nl\lambda_0^3 e^2/m} = \frac{4\pi c^2(n_{\text{ст}} - 1)l_{\text{ст}}}{Nl\lambda_0^4 l e^2/m}(\Delta\lambda)^2. \qquad (86.8)$$

Таким образом, силу осциллятора можно найти по расстоянию между вершинами крюков.

§ 87. Дисперсия плазмы

1. Плазма есть ионизованный газ, в котором электроны и ионы могут рассматриваться как свободные частицы с собственными частотами, равными нулю (см. т. III, § 121). Диэлектрическая проницаемость плазмы определяется в основном *свободными электронами*. Влиянием ионов можно пренебречь, так как

их массы практически бесконечно велики по сравнению с массами легких электронов. Полагая в формуле (84.5) $\omega_0 = 0$ и пренебрегая затуханием, получим для плазмы

$$\varepsilon = 1 - \left(\frac{\omega_p}{\omega}\right)^2, \tag{87.1}$$

где введено обозначение

$$\omega_p^2 = 4\pi N e^2/m, \tag{87.2}$$

а N означает концентрацию свободных электронов. Величина ω_p называется *плазменной* или *ленгмюровской частотой*. Частота ω_p играет для плазмы роль *собственной частоты*. Однако она характеризует не отдельные частицы, а *весь коллектив заряженных частиц*, из которых состоит плазма.

Чтобы понять, как может появиться такая «собственная частота» у коллектива частиц, каждая из которых в отдельности собственными частотами не обладает, рассмотрим следующий пример. Допустим, что нейтральная плазма занимает пространство между бесконечными плоскостями, перпендикулярными к оси X. Среднее электрическое поле в такой нейтральной плазме равно нулю. Сместим все электроны плазмы параллельно оси X на малое расстояние x (рис. 305), а ионы оставим несмещенными. Тогда на границах плазмы возникнут электрические заряды с поверхностной плотностью $\sigma = Nex$, и в плазме возникнет электрическое поле $E = 4\pi\sigma = 4\pi Nex$. На каждый электрон будет действовать квазиупругая сила $F = 4\pi Ne^2 x$. Если плазму предоставить самой себе, то возникнет свободное гармоническое колебание электронов с собственной частотой $\sqrt{4\pi Ne^2/m}$. Но это и есть плазменная частота.

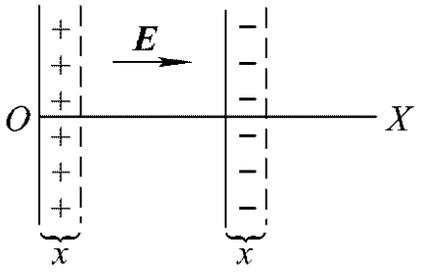


Рис. 305

2. Для плазменной частоты $\omega = \omega_p$ диэлектрическая проницаемость ε обращается в нуль. При $\omega > \omega_p$ величина ε (а с ней и показатель преломления $n = \sqrt{\varepsilon}$) положительна, но меньше единицы. При $\omega < \omega_p$ диэлектрическая проницаемость ε отрицательна, а показатель преломления чисто мнимый, т.е. $n = -i\varkappa$. Поэтому длинные электромагнитные волны (частота которых $\omega < \omega_p$) в плазме распространяться не могут. Они могут проникать только в тонкий поверхностный слой плазмы, испытывая от него полное отражение. Действительно, предположим, что падающая волна поляризована перпендикулярно к плоскости падения. (Случай другой поляризации разбирается так же.)

Тогда по формуле Френеля

$$\frac{R}{\mathcal{E}} = \frac{\cos\varphi - \sqrt{\varepsilon}\cos\psi}{\cos\varphi + \sqrt{\varepsilon}\cos\psi} = \frac{\cos\varphi + i\varkappa\cos\psi}{\cos\varphi - i\varkappa\cos\psi},$$

причем $\sin\varphi/\sin\psi = \sqrt{\varepsilon} = -i\varkappa$. Следовательно, $\cos\psi = \sqrt{1-\sin^2\psi} = \sqrt{1+\sin^2\varphi/\varkappa^2}$. Таким образом, $\cos\psi$ — величина вещественная, а потому $|R/\mathcal{E}| = 1$, что и доказывает наше утверждение.

Изложенное играет исключительно важную роль в осуществлении на Земле *дальней радиосвязи*. В земной атмосфере имеется ионизованная область, называемая *ионосферой*. Она начинается примерно с высоты 60 км и простирается, по-видимому, до высот $\sim 20\,000$ км. Основными источниками ионизации ионосферы являются ультрафиолетовое излучение Солнца и мягкое (от 0,8 до 30 нм) рентгеновское излучение солнечной короны. Другим источником служит корпускулярное излучение Солнца. Концентрация электронов N меняется с высотой неравномерно. Имеется несколько относительных максимумов ионизации, расположенных на различных высотах. Область ионосферы, содержащая один из таких максимумов, условно называется *ионосферным слоем*. Слои, расположенные в порядке возрастания высоты, обозначаются через D, E_1, E_2, F_1, F_2. Максимумы электронной концентрации в них меняются примерно в пределах $10^4 - 10^6$ электронов на 1 см^3. Концентрация электронов зависит от географической широты места и испытывает регулярные суточные и годичные изменения. Летом она больше, чем зимой, днем больше, чем ночью. Кроме того, наблюдаются *спорадические* изменения концентрации, вызванные вспышками на Солнце и пр.

Посмотрим теперь, как радиоволны, излученные какой-либо радиостанцией A, находящейся на земной поверхности, могут достигнуть приемника B, расположенного также на земной поверхности на расстоянии нескольких тысяч километров. Прямой путь через землю исключен, так как радиоволны в земле сильно поглощаются из-за ее высокой электрической проводимости. Показатель преломления неионизованного воздуха очень мало отличается от единицы, так что рефракция радиоволн практически не играет роли. Если бы не было ионосферы, то единственным способом достигнуть приемника B была бы дифракция. Но приемник B расположен глубоко в области геометрической тени на расстоянии в тысячи или десятки тысяч длин волн от ее границы. При таких условиях интенсивность дифрагированной волны в точке нахождения приемника B будет ничтожно мала, и никакой приемник практически не сможет обнаружить эту волну. Положение меняется при наличии ионосферы, так

как радиоволна может *отразиться от ионосферы* и таким путем достигнуть приемника. Только благодаря такому отражению возможна передача радиосигналов на земной поверхности на многие тысячи километров.

3. Найдем связь между фазовой v и групповой u скоростями электромагнитных волн в плазме при $\omega_p > \omega$. Используя выражение (87.1), для волнового числа k получаем

$$c^2k^2 = \omega^2\varepsilon = \omega^2 - \omega_p^2.$$

Дифференцирование этого соотношения дает: $c^2k\,dk = \omega\,d\omega$, т.е. $(\omega/k)(d\omega/dk) = c^2$, или

$$vu = c^2. \tag{87.3}$$

Фазовая скорость в плазме

$$v = \frac{c}{\sqrt{\varepsilon}} = \frac{c}{\sqrt{1 - \omega_p^2/\omega^2}} \tag{87.4}$$

всегда больше скорости света в вакууме. Для групповой скорости соотношение (87.3) дает

$$u = c^2/v = c\sqrt{1 - \omega_p^2/\omega^2}. \tag{87.5}$$

Она всегда меньше c, как это и должно быть.

Отметим интересное астрофизическое применение формулы (87.5). После открытия Хьюишем в 1967 г. *пульсаров* (*нейтронных звезд*) сразу же было обнаружено, что *длинноволновые сигналы доходят от пульсаров до Земли медленнее коротковолновых.* (В этом можно убедиться, принимая один и тот же сигнал с помощью двух радиоприемников, настроенных на разные частоты.) Это было объяснено влиянием *межзвездной плазмы*, через которую проходит сигнал. Квазимонохроматический сигнал распространяется в межзвездной плазме с групповой скоростью (87.5). Время распространения сигнала от пульсара до Земли определяется интегралом $t = \int dx/u$ по всему пути сигнала. Концентрация свободных электронов N, а с ней и плазменная частота ω_p имеют разные значения в разных точках пути. Однако всюду $\omega_p \ll \omega$, так что можно ограничиться первым членом разложения подынтегрального выражения по степеням отношения ω_p^2/ω^2. Это дает

$$t = \frac{1}{c}\int \left(1 - \omega_p^2/\omega^2\right)^{-1/2} dx = \frac{1}{c}\int \left(1 + \frac{1}{2}\frac{\omega_p^2}{\omega^2}\right) dx.$$

По сравнению с вакуумом время распространения сигнала увеличивается на

$$\Delta t = \frac{1}{2c\omega^2}\int \omega_p^2\,dx = \frac{\lambda^2 e^2}{2\pi c m c^2}\int N\,dx = \frac{\lambda^2 r}{2\pi c}\int N\,dx,$$

где $r = e^2/(mc^2) \approx 2{,}8 \cdot 10^{-13}$ см — *классический радиус электрона.*

1. Интеграл $\int N\,dx$ имеет смысл полного числа электронов в цилиндрическом канале, поперечное сечение которого равно 1 см^2, а длина — пути, пройденному сигналом от пульсара до Земли. Он является одной из интегральных характеристик межзвездной плазмы на пути распространения сигнала. Несмотря на ничтожную концентрацию такой плазмы, из-за колоссальности расстояний до пульсаров величина этого интеграла оказалась достаточной, чтобы обнаружить запаздывание (в области сантиметровых волн) длинноволновых сигналов относительно коротковолновых. Таким путем впервые были оценены расстояния до пульсаров. В предположении, что на пути от пульсаров к Земле около 10 % атомов водорода ионизовано, было найдено, что расстояния до большинства зарегистрированных пульсаров лежат между 200 и 7000 световых лет.

§ 88. Средняя плотность электромагнитной энергии в диспергирующих средах

1. Выражение для плотности электромагнитной энергии $w = (\varepsilon E^2 + \mu H^2)/(8\pi)$ получается в предположении, что ε и μ *постоянны*, т.е. не зависят от частоты ω (см. т. III, § 84). *В случае диспергирующих сред это выражение неприменимо.* Не разбирая этот вопрос в общем виде, выведем выражение для средней плотности электромагнитной энергии в непоглощающей диспергирующей среде на частном примере, принадлежащем М.Л. Левину.

Пусть вещество с диэлектрической проницаемостью $\varepsilon(\omega)$ и магнитной проницаемостью $\mu(\omega)$ заполняет плоский конденсатор с емкостью $C = \varepsilon(\omega)C_0$ и тонкий соленоид с индуктивностью $L = \mu(\omega)L_0$, соединенные в колебательный контур (рис. 306). Здесь C_0 и L_0 — значения емкости и индуктивности для того случая, когда в пространстве между обкладками конденсатора и внутри соленоида вакуум. При отсутствии сопротивления в контуре будут совершаться свободные гармонические колебания с циклической частотой $\omega = 1/\sqrt{L(\omega)C(\omega)}$. Если в некоторый момент времени ввести в контур малое сопротивление R, то, начиная с этого момента, колебания сделаются затухающими и первоначально запасенная электромагнитная энергия будет переходть в джоулево тепло, выделяющееся в сопротивлении R. Полное количество тепла, выделившееся в сопротивлении R за время, когда колебания прекратятся, будет равно электромагнитной энергии, запасенной в контуре до введения сопротивления. Поэтому задача сводится к вычислению джоулева тепла.

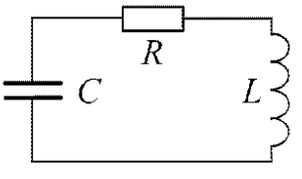


Рис. 306

Пусть при $t < 0$ в контуре совершаются свободные колебания:

$$I = I_0 e^{i\omega t}, \quad V = V_0 e^{i\omega t}.$$

где I — сила тока в контуре, а V — напряжение на обкладках конденсатора, связанные между собой соотношением $L\dot{I}+V=0$, или $i\omega LI+V=0$. Если в момент $t=0$ в контур ввести сопротивление R, то, начиная с этого момента, колебания будут описываться уравнением

$$L(\widetilde{\omega})\ddot{I}+R\dot{I}+\frac{I}{C(\widetilde{\omega})}=0,$$

откуда

$$I=I_0e^{i\widetilde{\omega}t},\quad t>0,$$

где $\widetilde{\omega}$ — комплексная частота, определяемая уравнением

$$\widetilde{\omega}L(\widetilde{\omega})-\frac{1}{\widetilde{\omega}C(\widetilde{\omega})}=iR.$$

Если R исчезающе мало, то $\widetilde{\omega}$ должна отличаться от ω также на исчезающе малую величину. Но ω удовлетворяет уравнению

$$\omega L(\omega)-\frac{1}{\omega C(\omega)}=0.$$

Вычитая его из предыдущего соотношения и заменяя все разности дифференциалами, получим

$$\left[\frac{d}{d\omega}(\omega L)+\frac{1}{\omega^2C^2}\frac{d}{d\omega}(\omega C)\right](\widetilde{\omega}-\omega)=iR,$$

откуда $\widetilde{\omega}=\omega+i\delta$, причем

$$\frac{R}{\delta}=\frac{d(\omega L)}{d\omega}+\frac{1}{\omega^2C^2}\frac{d(\omega C)}{d\omega}=\frac{d(\omega L)}{d\omega}+\frac{L}{C}\frac{d(\omega C)}{d\omega}.$$

Для определения джоулева тепла надо проинтегрировать выражение RI^2 по времени. Поскольку возведение в квадрат — нелинейная операция, необходимо перейти к вещественной форме, т.е. сделать замену

$$I\to\operatorname{Re}(I)=(I+I^*)/2.$$

Энергия, первоначально запасенная в колебательном контуре, равна

$$W=\int_0^\infty R\left(\frac{I+I^*}{2}\right)^2dt=\frac{R|I_0|^2}{4}\left\{\frac{\delta}{\omega^2+\delta^2}+\frac{1}{\delta}\right\}$$

или в пределе при $\delta\to0$

$$W=\frac{|I_0|^2}{4}\frac{R}{\delta}.$$

Подставляя сюда значение для R/δ и пользуясь соотношением $\omega L|I_0|=|V_0|$, получим

$$W=\frac{L_0|I_0|^2}{4}\frac{d(\omega\mu)}{d\omega}+\frac{C_0|V_0|^2}{4}\frac{d(\omega\varepsilon)}{d\omega}.$$

Если бы между обкладками конденсатора и внутри соленоида был вакуум, то для средних по времени значений магнитной и электрической энергий можно было бы написать

$$\frac{L_0|I_0|^2}{4}=\frac{1}{8\pi}\overline{\mathbf{H}^2}\tau_m,\quad\frac{C_0|V_0|^2}{4}=\frac{1}{8\pi}\overline{\mathbf{E}^2}\tau_e,$$

где τ_m и τ_e — объемы соленоида и конденсатора, а $\mathbf{E}$ и $\mathbf{H}$ — напряженности электрического и магнитного полей, когда амплитуды напряжения на конденсаторе и тока в соленоиде равны V_0 и I_0. Но при заданных V_0 и I_0 поля $\mathbf{E}$ и $\mathbf{H}$ не зависят от среды, заполняющей конденсатор и соленоид. Поэтому предыдущие соотношения остаются справедливыми и в том случае, когда конденсатор и соленоид заполнены веществом. Используя их, получаем следующие выражения для средних по времени значений плотностей электрической и магнитной энергий:

$$\begin{aligned} \overline{w}_e &= \frac{\overline{W}_e}{\tau_e} = \frac{1}{8\pi}\frac{d(\omega\varepsilon)}{d\omega}\overline{\mathbf{E}^2}, \\ \overline{w}_m &= \frac{\overline{W}_m}{\tau_m} = \frac{1}{8\pi}\frac{d(\omega\mu)}{d\omega}\overline{\mathbf{H}^2}. \end{aligned} \tag{88.1}$$

Принципиальный недостаток приведенного вывода состоит в том, что в нем дифференцирование функций ωL и ωC производится вдоль *мнимой оси* (так как разность частот $\widetilde{\omega} - \omega = iR$ — величина чисто мнимая), а в окончательном выражении (88.1) производится подмена дифференцированием по *вещественной переменной* ω. Так можно поступать, когда функции ωL и ωC *аналитичны*. Поэтому для полноты доказательства надо было бы доказать аналитичность этих функций, чего в выводе Левина нет. Это можно сделать в общей теории дисперсии, исследуя аналитические свойства функций $\varepsilon(\omega)$ и $\mu(\omega)$. Однако рассмотрение этого вопроса выходит за рамки нашей книги.

2. Смысл формулы (88.1) полезно уяснить на примере газа классических гармонических осцилляторов в монохроматическом электрическом поле с частотой ω. Вдали от собственной частоты ω_0 осциллятора можно пренебречь затуханием. Тогда смещение осциллятора из положения равновесия выразится формулой

$$\mathbf{r} = \frac{e\mathbf{E}}{m(\omega_0^2 - \omega^2)}.$$

В этом случае энергия слагается из *энергии самого электромагнитного поля* (т.е. поля в вакууме) и из *энергии частиц*, находящихся в поле. Последняя энергия в свою очередь состоит из *кинетической* и *потенциальной энергии* колеблющихся осцилляторов. В статических полях кинетической энергии нет. Это приводит к формуле $w = \varepsilon E^2/(8\pi)$. В переменных полях кинетическую энергию надо учитывать, что и делается ниже.

Плотность собственно энергии электрического поля равна

$$w_1 = \frac{1}{8\pi}\left(\frac{\mathbf{E} + \mathbf{E}^*}{2}\right)^2 = \frac{\mathbf{E}^2}{32\pi} + \frac{\mathbf{E}\mathbf{E}^*}{32\pi} + \text{компл. сопр.}$$

Плотность потенциальной энергии:

$$w_2 = \frac{Nm\omega_0^2}{2}\left(\frac{\mathbf{r} + \mathbf{r}^*}{2}\right)^2 = \frac{Nm\omega_0^2}{8}(\mathbf{r}^2 + \mathbf{r}\mathbf{r}^*) + \text{компл. сопр.}$$

Плотность кинетической энергии:

$$w_3 = \frac{Nm}{2}\left(\frac{\dot{\mathbf{r}} + \dot{\mathbf{r}}^*}{2}\right)^2 = -\frac{Nm\omega^2}{8}(\mathbf{r}^2 - \mathbf{r}\mathbf{r}^*) + \text{компл. сопр.}$$

Подставляя сюда выражение для $\mathbf{r}$ и замечая, что из формулы Зельмейера следует

$$\frac{d(\omega\varepsilon)}{d\omega} = 1 + \frac{(\varepsilon - 1)(\omega_0^2 + \omega^2)}{\omega_0^2 - \omega^2},$$

получим для плотности электрической энергии:

$$w_e = \frac{\varepsilon \mathbf{E}^2}{32\pi} + \frac{1}{32\pi}\frac{d(\omega\varepsilon)}{d\omega}(\mathbf{E}\mathbf{E}^*) + \text{компл. сопр.}$$

Усредняя по времени, получаем первую формулу (88.1). Для плотности магнитной энергии имеем обычное выражение, как в недиспергирующей среде.

ЗАДАЧИ

1. Рэлей предложил определять *среднюю скорость движения энергии* u в плоской бегущей волне как отношение средней плотности потока энергии к средней плотности самой энергии. Пользуясь выражением для вектора Пойнтинга, показать, что так определенная скорость в случае монохроматической электромагнитной волны совпадает с групповой скоростью.

Р е ш е н и е. Для средних плотностей энергии и ее потока нетрудно получить

$$\overline{w} = \frac{c}{8\pi}\sqrt{\frac{\varepsilon}{\mu}}\,\frac{dk}{d\omega}(\mathbf{E}\mathbf{E}^*), \quad \overline{S} = \frac{c}{8\pi}\sqrt{\frac{\varepsilon}{\mu}}\,(\mathbf{E}\mathbf{E}^*),$$

откуда и следует требуемый результат.

2. Показать, что если $\varepsilon(\omega)$ и $\mu(\omega)$ положительны, то фазовая и групповая скорости в электромагнитной волне направлены в одну сторону.

Р е ш е н и е. Средняя плотность электромагнитной энергии

$$\overline{w} = \frac{1}{8\pi}\frac{d(\omega\varepsilon)}{d\omega}(\mathbf{E}\mathbf{E}^*) + \frac{1}{8\pi}\frac{d(\omega\mu)}{8\pi}(\mathbf{H}\mathbf{H}^*)$$

— существенно положительная величина. В плоской волне $\varepsilon(\mathbf{E}\mathbf{E}^*) = \mu(\mathbf{H}\mathbf{H}^*)$. Поэтому

$$\frac{d(\omega\varepsilon)}{d\omega} + \frac{\mu}{\varepsilon}\frac{d(\omega\mu)}{d\omega} > 0.$$

Это неравенство должно соблюдаться для любых сред, у которых знаки ε и μ совпадают, поскольку оно выведено в предположении, что в среде может распространяться однородная монохроматическая волна, для которой $k^2 = \varepsilon\mu\omega^2/c^2 > 0$. В этом же предположении имеет смысл говорить о групповой скорости. Преобразовав предыдущее неравенство к виду

$$\mu\frac{\omega}{k}\frac{d\omega}{dk} = \mu v u > 0,$$

легко получить требуемый результат.

§ 89. Поглощение света и уширение спектральных линий

1. В классической теории дисперсии поглощение (затухание) излучения учитывается формально с помощью тормозящей силы $-g\dot{\mathbf{r}} = -2m\gamma\mathbf{v}$ в уравнении (84.1). Благодаря этому амплитуда колебаний убывает во времени экспоненциально по закону $\exp(-\gamma t)$, а энергия колебаний — по закону

$$\mathcal{E} = \mathcal{E}_0 e^{-2\gamma t} = \mathcal{E}_0 e^{-t/\tau}. \tag{89.1}$$

Физическая природа тормозящей силы $-g\mathbf{v}$ при этом остается нераскрытой.

Планк развил теорию дисперсии и поглощения света в предположении, что колеблющийся осциллятор (электрон), двигаясь ускоренно, непрерывно теряет энергию на излучение. Эта убыль энергии на излучение определяется формулой

$$-\frac{d\mathcal{E}}{dt} = \frac{2e^2}{3c^3}\dot{\mathbf{v}}^2 \tag{89.2}$$

(см. т. III, § 141). Применим эту формулу к осциллятору. Если потеря энергии осциллятора за период колебаний относительно мала, то колебания будут отличаться от гармонических мало. Тогда можно считать, что средние за период колебаний значения кинетической и потенциальной энергий осциллятора одинаковы, а потому среднее значение его полной энергии равно удвоенному среднему значению кинетической энергии. Но полная энергия $\mathcal{E}$ слабо затухающего осциллятора в течение периода колебаний остается почти постоянной, так что $\overline{\mathcal{E}} \approx \mathcal{E}$. В случае слабо затухающих колебаний $\mathbf{v} \approx \mathbf{v}_0 \sin(\omega_0 t + \delta)$, $\dot{\mathbf{v}} = \omega_0 \mathbf{v}_0 \cos(\omega t + \delta)$. Возведя последнее соотношение в квадрат и усреднив по периоду колебаний, получим

$$\langle \dot{\mathbf{v}}^2 \rangle = \omega_0^2 \langle \mathbf{v}^2 \rangle = \omega_0^2 \mathcal{E}/m.$$

Усреднив теперь по времени формулу (89.2), найдем

$$-\frac{d\mathcal{E}}{dt} = \frac{2e^2\omega_0^2}{3mc^3}\,\mathcal{E}.$$

Отсюда снова получается формула вида (89.1). Таким образом, излучение проявляется так, как если бы на колеблющийся заряд действовала тормозящая сила $-2m\gamma\mathbf{v}$, причем

$$\gamma = \frac{1}{3}\frac{e^2\omega_0^2}{mc^3}. \tag{89.3}$$

Затухание, обусловленное излучением, называется *естественным затуханием*. Время естественного затухания $\tau_{\text{изл}}$, в течение которого энергия убывает в e раз, равно

$$\tau_{\text{изл}} = \frac{1}{2\gamma} = \frac{3mc^3}{2e^2\omega_0^2} = \frac{3mc}{8\pi^2 e^2}\lambda^2 = \frac{3}{8\pi^2}\left(\frac{\lambda}{r}\right)\left(\frac{\lambda}{c}\right), \tag{89.4}$$

где $r = e^2/(mc^2) = 2{,}8 \cdot 10^{-13}$ см — *классический радиус электрона*. Для длины волны $\lambda = 600$ нм эта формула дает $\tau_{\text{изл}} = 1{,}63 \cdot 10^{-8}$ с.

2. Затухание делает колебания немонохроматическими. Это приводит к *уширению спектральных линий*. Ширина спектральной линии, обусловленная затуханием вследствие излучения, называется *естественной шириной спектральной линии*. Для ее

вычисления поле излучения $E = E_0 \exp(-\gamma t) \sin \omega_0 t$ надо разложить в интеграл Фурье. Квадраты коэффициентов этого разложения определяют *относительные спектральные плотности излучения* $I(\omega)$, т.е. распределение энергии в спектральной линии по спектру частот. Результат выражается ранее выведенной формулой (29.10), которую здесь мы запишем в виде

$$I(\omega) = \frac{1}{(\omega - \omega_0)^2 + 1/(2\tau_{\text{изл}})^2}, \qquad (89.5)$$

заменив прежнее время τ на $2\tau_{\text{изл}}$. (Прежнее время τ определяло *затухание по амплитуде*, а здесь через $\tau_{\text{изл}}$ обозначено *время затухания по энергии.*) Форма спектральной линии приведена на рис. 307 для значения параметра $\omega_0 \tau_{\text{изл}} = 5$, которое, конечно, слишком мало. (В приведенном выше численном примере $\omega_0 \tau_{\text{изл}} = 2\pi c \tau_{\text{изл}}/\lambda \approx 5 \cdot 10^5$) Время затухания для спектральных линий, излучаемых газами, в действительности много больше, а сами спектральные линии много у́же, чем на кривой рис. 307.

Ширину спектральной линии принято характеризовать расстоянием между точками A и A', расположенными по разные стороны от центра линии, в которых интенсивность света равна половине интенсивности в центре линии. Это расстояние называется *полушириной спектральной линии.* Естественная полуширина спектральной линии, как это следует из формулы (89.5), равна

$$\Delta\omega = 1/\tau_{\text{изл}} = 2\gamma. \qquad (89.6)$$

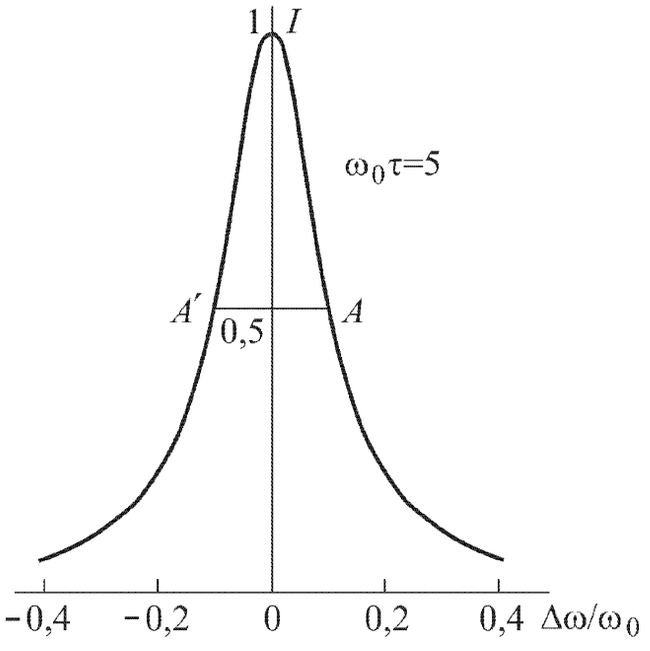


Рис. 307

В квантовой физике, где излучение связывают с переходом атома с более высокого энергетического уровня на более низкий (из возбужденного состояния в нормальное), величина $\tau_{\text{изл}}$ имеет смысл *времени жизни возбужденного состояния.* Здесь, как и в классической физике, формула (89.6) является следствием общего соотношения (29.8), согласно которому $\Delta\omega \cdot \tau_{\text{изл}} \sim 1$.

Естественное затухание свечения изолированных атомов и соответствующее ему уширение спектральных линий на опыте исследовалось В. Вином в 1919–1927 гг. В его опытах каналовые лучи, состоявшие из светящихся атомов, проходили через узкое отверстие ($0{,}1 \times 3$ мм^2) в пространство, где с помощью мощных

насосов поддерживался высокий вакуум ($< 0{,}001$ мм рт. ст.). В этом пространстве атомы двигались без столкновений, но их свечение постепенно затухало по мере удаления от входного отверстия. По затуханию свечения можно было оценить время естественного затухания. Для этого надо было знать среднюю скорость движения атомов каналовых лучей. Она измерялась по доплеровскому смещению спектральных линий при наблюдении вдоль направления каналовых лучей и оказалась порядка $5 \cdot 10^7$ см/с (для атомов водорода). Из своих опытов Вин нашел для времени затухания τ около 10^{-8} с. Эта величина несколько менялась от одного вещества к другому и от одной спектральной линии к другой. Полученные результаты совпадали с предсказаниями теории, но только по порядку величины, а простая зависимость (89.4) времени затухания от длины волны ($\tau \sim \lambda^2$) не подтвердилась. Впрочем, полного количественного согласия и нельзя было ожидать от простой классической теории, основанной на модели гармонического осциллятора.

3. Все изложенное относится к излучению *изолированного атома*. В случае *среды*, состоящей из атомов, надо принять во внимание, что атом не только теряет энергию на излучение, но и получает энергию, излучаемую другими атомами. Если бы среда была *оптически однородна*, то, как показал Л.И. Мандельштам, эти два процесса в точности компенсировали бы друг друга. В отсутствие других причин затухания колебания атома были бы незатухающими. Плоская бегущая волна распространялась бы в среде без ослабления.

Дело меняется при нарушении оптической однородности среды. Однородность может нарушиться либо за счет *включения посторонних частиц* (например, пыли или капелек тумана в воздухе), либо за счет *тепловых флуктуаций*, возникающих в среде. В этом случае потеря энергии атомами на излучение при распространении волны будет приводить к ее ослаблению. Однако это не есть истинное поглощение света с переходом электромагнитной энергии волны в другие формы энергии (тепло), а *рассеяние света*, где такого превращения нет. *Истинное поглощение*, когда энергия волны превращается в другие формы (тепло), вызывается другими причинами. Кроме того, ослабление волны, обусловленное излучением, слишком мало по сравнению с тем ослаблением, которое во многих случаях наблюдается в действительности.

4. Одна из причин поглощения света состоит в том, что атомы, внутри которых происходят колебания, совершая тепловое движение, претерпевают *столкновения друг с другом*. При каждом столкновении резко и неправильно меняются амплитуды и фазы гармонических колебаний, происходит переход в тепло

энергии регулярных колебаний, т.е. *поглощение света.* Исходя из этих представлений, Г.А. Лорентц развил теорию уширения спектральных линий, обусловленного столкновениями между атомами (молекулами) газа. Такое уширение называется *ударным уширением.* Лорентц показал, что в газах столкновения между молекулами при тепловом движении статистически приводят также к экспоненциальному закону затухания интенсивности волны и к форме спектральных линий такого же вида, что и при естественном затухании.

Мы не будем проводить детальные вычисления, как это сделал сам Лорентц. При таких вычислениях надо было бы вводить специальные предположения относительно характера нарушения регулярности амплитуд и фаз колебаний при столкновениях, так что результат оказался бы верным все равно с точностью до числового коэффициента. А в пределах такой точности его можно получить из общих соображений без всяких вычислений, если воспользоваться общей формулой (29.8).

Обозначим через $\tau_{\text{уд}}$ среднее время свободного пробега атома от одного столкновения до следующего. Можно считать, что между двумя последовательными столкновениями колебания внутри атома происходят гармонически, а при каждом столкновении обрываются. Затем они опять возобновляются, но уже с новыми значениями амплитуд и фаз, никак не связанными с амплитудами и фазами до столкновения. Но такие обрывающиеся во времени колебания при разложении в спектр будут заполнять интервал частот $\Delta\omega$, подчиняющийся соотношению $\Delta\omega \cdot \tau_{\text{уд}} \approx 1$. Отсюда следует, что ударное уширение спектральных линий определяется формулой

$$\Delta\omega \sim 1/\tau_{\text{уд}}. \tag{89.7}$$

Число столкновений, претерпеваемых атомом в единицу времени, определяется формулой кинетической теории газов $z = N\sigma v$, где N — число атомов в единице объема, σ — газокинетическое поперечное сечение атома, $v \approx \sqrt{kT/m}$ — средняя скорость теплового движения атома (см. т. II, § 86). Следовательно,

$$\tau_{\text{уд}} \approx \frac{1}{N\sigma v}. \tag{89.8}$$

Ударные затухание и уширение спектральных линий особенно существенны в *плотных газах* и при *высоких температурах.* Для уменьшения влияния столкновений надо уменьшать плотность газа. Вот почему в опытах Вина при изучении естественного затухания свечения атомов каналовые лучи направлялись в высокий вакуум. В обычных условиях столкновения значительно сильнее влияют на затухание волны, чем излучение. Однако формула (89.8) дает для времени затухания все же бóльшие, а

следовательно, для уширения спектральных линий — меньшие значения, чем наблюдаются на опыте. Следовательно, должны существовать другие причины поглощения света и уширения спектральных линий.

Такими причинами могут быть воздействия на излучающий атом различных *силовых полей*, в первую очередь электрических полей, создаваемых окружающими атомами. В постоянных электрических полях спектральные линии претерпевают расщепление (эффект Штарка). В переменных полях изменение спектральных линий носит более сложный характер. Это — одна из основных причин уширения спектральных линий.

5. Другой причиной уширения спектральных линий является изменение частоты излучаемого света из-за *эффекта Доплера*. Пусть ω_0 — частота волны, излучаемой неподвижной молекулой (атомом). Если молекула равномерно движется, то частота волны, излучаемой ею в направлении оси X, в лабораторной системе отсчета изменяется и в нерелятивистском приближении определяется формулой

$$\omega = \omega_0(1 + v_x/c), \tag{89.9}$$

где v_x — скорость молекулы в направлении оси X.

Согласно закону распределения скоростей Максвелла, вероятность того, что x-составляющая скорости молекулы газа лежит в интервале $(v_x, v_x + dv_x)$, равна

$$dw = A_1 \exp\left(-\frac{mv_x^2}{2kT}\right) dv_x,$$

где m — масса молекулы, а A_1 — постоянная нормировки (см. т. II, § 72). Заменив в этом выражении переменную v_x на ω по формуле (89.9), получим

$$dw = A_1 \frac{c}{\omega_0} \exp\left[-\frac{mc^2}{2kT}\left(\frac{\omega - \omega_0}{\omega_0}\right)^2\right] d\omega.$$

Так как излучения различных молекул газа *некогерентны*, то эта величина определяет вероятность излучения в спектральный интервал $(\omega, \omega + d\omega)$. Интенсивность света $I(\omega)\,d\omega$, приходящаяся на этот интервал, будет пропорциональна dw. В результате получится

$$I(\omega) = I_0 \exp\left[-\frac{mc^2}{2kT}\left(\frac{\omega - \omega_0}{\omega_0}\right)^2\right], \tag{89.10}$$

где I_0 — новая постоянная. Очевидно, $I_0 = I(\omega_0)$, т.е. I_0 есть спектральная интенсивность в середине линии. По обе стороны от середины линии спектральная интенсивность убывает экспоненциально по закону (89.10). Полуширина спектральной линии,

как нетрудно вычислить, определяется формулой

$$\Delta\omega = 2\sqrt{\frac{2kT}{mc^2}\ln 2}\,\omega_0 = 2\sqrt{\ln 2}\,\frac{v_m}{c}\,\omega_0 = 1{,}67\,\frac{v_m}{c}\,\omega_0, \qquad (89.11)$$

где $v_m = \sqrt{2kT/m}$ — наиболее вероятная скорость молекулы.

Эффект Доплера является основной причиной уширения спектральных линий в разреженных газах при высоких температурах.

6. В основе теории дисперсии и абсорбции света, как она изложена в этой главе, лежат *линейные уравнения*. По этой причине показатели преломления n и затухания $\varkappa$ в этой теории *постоянны*, т.е. *не зависят от интенсивности света*. Амплитуда монохроматической волны при ее распространении убывает по закону $A = A_0 \exp(-2\pi\varkappa x/\lambda)$, а интенсивность — по закону

$$I = I_0 e^{-\alpha x}, \qquad (89.12)$$

где $\alpha = 4\pi\varkappa/\lambda$ — *коэффициент затухания* (*поглощения*), зависящий, вообще говоря, от длины волны. Формула (89.12) выражает так называемый *закон Бугера* (1698–1758), постулированный им в 1729 г. Сущность этого закона состоит в том, что для монохроматического света коэффициент затухания (поглощения) α не зависит от интенсивности света. С.И. Вавилов, исследуя на опыте поглощение света в некоторых растворах, установил, что постоянство коэффициента поглощения α соблюдается с точностью не менее 5 % в крайне широком интервале изменения интенсивности света — от 20 до $2{,}5\cdot10^{-19}$ Вт/см^2, т.е. от интенсивности прямых солнечных лучей, сконцентрированных в фокусе линзы, до ничтожных интенсивностей, соответствующих порогу зрительного восприятия глаза.

Интерпретируем эти результаты с точки зрения *квантовых представлений* о поглощении света. При поглощении световая волна переводит молекулу (атом) с основного энергетического уровня W_0 на возбужденный уровень W_n, увеличивая тем самым запас поглощенной энергии. Способность молекулы поглощать свет в возбужденном состоянии иная, чем в основном. Если в опытах Вавилова коэффициент поглощения оставался одним и тем же при всех интенсивностях, то отсюда следует, что в этих опытах число возбужденных молекул было пренебрежимо мало по сравнению с числом невозбужденных, т.е. было относительно мало среднее время жизни молекулы в возбужденном состоянии. И действительно, для веществ, с которыми имел дело Вавилов в своих опытах, это время не превышало 10^{-8} с. Можно было бы ожидать, что при больших временах жизни и больших интенсивностях света число молекул в основном состоянии (которые практически только и поглощают свет) уменьшится и

закон Бугера перестанет выполняться. Выбрав специально вещества с большими временами жизни в возбужденном состоянии, Вавилов наблюдал уменьшение коэффициента поглощения при увеличении интенсивности света. Это было первым примером, где наблюдалось *нарушение принципа суперпозиции в оптике.*

После открытия лазеров, сделавших возможным получение световых волн с напряженностями электрических полей, сравнимыми с внутримолекулярными и внутриатомными, класс подобных «нелинейных» оптических явлений быстро расширился. Их изучением занимается *нелинейная оптика* — самостоятельная обширная и быстро развивающаяся наука. Краткое представление об этих явлениях будет дано в последней (XI) главе этого тома.

§ 90. Двойное преломление в электрическом и магнитном полях

1. Электромагнитная природа света и вещества заставляет ожидать, что оптические свойства вещества должны изменяться при его внесении в электрическое и магнитное поле. Действительно, такие изменения были обнаружены на опыте. Сначала Фарадей в 1846 г. открыл магнитооптическое явление — *вращение плоскости поляризации света в магнитном поле.* Затем Керр (1824–1907) в 1875 г. открыл «электрооптическое явление», названное его именем. В 1877 г. тот же Керр нашел, что намагниченное железное зеркало иначе отражает поляризованный свет, чем ненамагниченное. В дальнейшем были открыты и другие *электрооптические* и *магнитооптические явления.*

Электрооптический эффект Керра состоит в том, что многие изотропные тела при введении в постоянное электрическое поле становятся оптически анизотропными. Они начинают вести себя подобно одноосным двупреломляющим кристаллам, оптическая ось которых параллельна приложенному электрическому полю.

Сам Керр наблюдал это явление сначала на стекле, вводимом в электрическое поле. В этом случае, а также в случае всех других твердых тел оптическая анизотропия может быть вторичным явлением, а именно может возникать в результате механических деформаций, производимых электрическим полем. Однако последующие исследования самого Керра и других ученых показали, что явление наблюдается также в жидкостях и даже газах, где механическая анизотропия исключена. Кроме того, «чистый эффект Керра» можно отделить от эффекта, вызванного механическими деформациями, исследуя явление в переменных полях: при внезапном включении или выключении

электрического поля механические деформации возникают и исчезают не сразу, а с некоторой задержкой, тогда как явление Керра в чистом виде (наблюдаемое в жидкостях и газах) происходит *практически безынерционно* (см. п. 5).

Как и всякое двойное преломление, эффект Керра можно обнаружить, помещая между скрещенными николями N_1 и N_2 плоский конденсатор, заполненный исследуемым веществом (рис. 308, ср. с рис. 278). В отсутствие внешнего электрического поля свет через систему не проходит. При наложении электрического поля возникает *двойное преломление*, вследствие чего свет по выходе из конденсатора становится *поляризованным эллиптически* и частично проходит через николь N_2, — поле зрения просветляется. Вводя компенсатор K, можно измерить разность фаз между обоими лучами, возникшими при двойном преломлении, и таким образом найти разность $n_e - n_o$ между необыкновенным n_e и обыкновенным n_o показателями преломления исследуемого вещества в электрическом поле. Ввиду малых значений указанной разности ее измерение возможно только интерференционными методами, аналогичными изложенному.

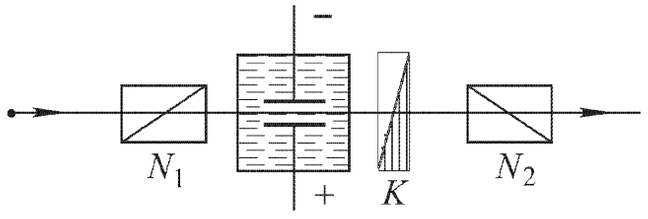


Рис. 308

Небольшим изменением описанного опыта можно для жидкостей найти обе разности $n_e - n$ и $n_o - n$. Для этого по предложению Л.И. Мандельштама можно воспользоваться, например, интерферометром Жамена. Исследуемая жидкость наливается в кювету, а конденсатор погружается в нее целиком (рис. 309). Один из интерферирующих лучей проходит внутри, а другой — снаружи конденсатора. Николь-поляризатор N устанавливается так, чтобы свет был поляризован либо в плоскости рисунка, либо перпендикулярно к ней. При наложении электрического поля показатель преломления n наружного луча не изменяется, тогда как луч, проходящий между пластинами конденсатора, становится либо необыкновенным, либо обыкновенным (в зависимости от установки николя). По смещению интерференционных полос можно вычислить $n_e - n$ и $n_o - n$.

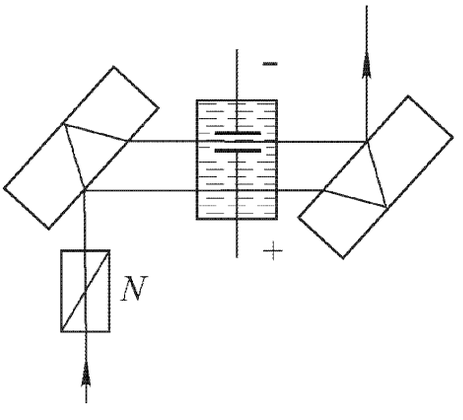


Рис. 309

2. В дальнейшем предполагается, что внешнее электрическое поле *однородно*, а луч распространяется перпендикулярно к нему. Внешнее поле обозначается через $\mathbf{E}_0$, в отличие от поля $\mathbf{E}$ самой волны. Очевидно, разность $n_e - n_o$ является функцией компонент поля E_{0x}, E_{0y}, E_{0z}. Если (свободная от поля) среда *изотропна*, то эта функция не может зависеть от направления вектора $\mathbf{E}_0$. В частности, она не должна изменяться при изменении знаков своих аргументов, т.е. быть *четной* по отношению к каждому из них. Поэтому ее разложение в степенной ряд должно содержать только *четные степени*. В слабых электрических полях, какими являются все поля, применяемые на опыте, можно сохранить только первые — *квадратичные* — члены этого разложения. В силу уже отмеченной изотропии среды они войдут только в виде квадратов E_{0x}^2, E_{0y}^2, E_{0z}^2 и притом с одним и тем же коэффициентом. Поэтому их можно объединить, написав

$$n_e - n_o = qE_0^2, \tag{90.1}$$

где коэффициент q зависит только от рода вещества и его состояния, а также от длины световой волны λ. Опыт действительно подтвердил формулу (90.1). Разность фаз между необыкновенным и обыкновенным лучами после прохождения через конденсатор будет

$$\varphi = \frac{2\pi}{\lambda}(n_e - n_o)l = 2\pi BlE^2, \tag{90.2}$$

где l — толщина проходимого слоя вещества, а $B \equiv q/\lambda$ — так называемая *постоянная Керра*. Она увеличивается при уменьшении длины волны (дисперсия) и сильно уменьшается с повышением температуры.

Таблица 9

Вещество	Температура, °C	Постоянная Керра B, СГСЭ-ед.	Вещество	Температура, °C	Постоянная Керра B, СГСЭ-ед.
Жидкости			*Газы*		
Нитробензол	20	$2{,}2 \cdot 10^{-5}$	Этиловый спирт	18	$9{,}2 \cdot 10^{-10}$
Нитротолуол	20	$1{,}2 \cdot 10^{-5}$	Ацетон	83	$5{,}4 \cdot 10^{-10}$
Хлорбензол	20	$1 \cdot 10^{-6}$	Сероуглерод	57	$3{,}6 \cdot 10^{-10}$
Вода	20	$4{,}7 \cdot 10^{-7}$	Этиловый эфир	63	$-0{,}66 \cdot 10^{-10}$
Сероуглерод	20	$3{,}2 \cdot 10^{-7}$	Этиленовый		
Бензол	20	$6 \cdot 10^{-8}$	спирт	20	$-1{,}7 \cdot 10^{-10}$
Хлороформ	20	$-3{,}5 \cdot 10^{-8}$			

В табл. 9 приведены значения постоянной Керра для некоторых жидкостей и газов ($\lambda = 589$ нм, давление в случае газов 760 мм рт. ст.). Для большинства веществ $n_e > n_o$, т.е. постоянная Керра положительна. Это соответствует анизотропии *положительного кристалла*. Значительно реже встречаются случаи, когда $B < 0$ (этиловый эфир, многие масла и спирты). Максимальным значением B из всех исследованных веществ обладает нитробензол.

Чтобы составить представление о порядке величины эффекта Керра, приведем следующий пример. Пусть расстояние между пластинами «конденсатора Керра» равно 1 мм, а напряжение между ними 1500 В, так что $E_0 = 15000$ В/см $= 50$ СГСЭ-ед. Если конденсатор заполнен жидким нитробензолом, то при $l = 5$ см возникшая разность фаз будет $\varphi = \pi/2$. Такой конденсатор может служить пластинкой $\lambda/4$. Нетрудно обнаружить и значительно меньшие разности фаз. Поэтому эффект Керра в нитробензоле находит широкие технические применения.

3. Эффект Керра — явление более простое и теоретически лучше изучено, чем искусственная анизотропия при механических деформациях. Это и понятно. В последнем случае проявляется воздействие сложных молекулярных полей на молекулы тела, подвергшегося деформации. Оно плохо изучено и значительно сложнее соответствующего сравнительно простого воздействия постоянного однородного электрического поля. Кроме того, явление Керра удалось наблюдать в газах, для которых теория развита наиболее глубоко и подробно. Изучение явления Керра в газах совместно с рефракцией и деполяризацией рассеянного света позволяет определить тензор поляризуемости молекул.

Явление Керра объясняется анизотропией самих молекул. Количественная теория для газов была развита Ланжевеном (1872–1946) в 1910 г. В ней анизотропия молекулы характеризовалась *только тензором поляризуемости.* В отсутствие электрического поля анизотропные молекулы ориентированы в пространстве хаотически, так что среда в целом макроскопически изотропна. При наложении внешнего электрического поля молекулы преимущественно ориентируются осями наибольшей поляризуемости вдоль поля, вследствие чего среда становится анизотропной. В общем случае произвольного тензора поляризуемости вычисления очень громоздки. Однако сущность теории и ее основные результаты, по крайней мере качественно, можно передать, предполагая, что молекулы *полностью анизотропны*, что очень сильно упрощает вычисления. Это и делается ниже. *Полностью анизотропной* называется такая молекула, внутри которой электрические заряды могут смещаться только в определенном направлении, называемом *осью молекулы.* Моделью

такой молекулы может служить палочка, вдоль которой и может происходить смещение зарядов [1]).

При проведении расчета (подчеркнем это еще раз) надо соблюдать осторожность, чтобы не спутать внешнее поле $\mathbf{E}_0$ с полем $\mathbf{E}$ электромагнитной волны, которое предполагается монохроматическим. Роль постоянного поля $\mathbf{E}_0$ сводится к созданию в среде определенного распределения молекул по их направлениям в пространстве. Переменное же поле $\mathbf{E}$ создает в ней переменную электрическую поляризацию $\mathbf{P}$, определяющую показатели преломления электромагнитной волны.

Переменный дипольный момент $\mathbf{p}$ полностью анизотропной молекулы создается только составляющей поля $\mathbf{E}$, параллельной ее оси. Он равен $\mathbf{p} = \beta E_s \mathbf{s} = \beta(\mathbf{E}\mathbf{s})\mathbf{s}$, где $\mathbf{s}$ — единичный вектор в направлении оси молекулы, а β — ее поляризуемость в том же направлении. Очевидно, $\mathbf{p}$ можно представить в виде $\mathbf{p} = \beta E s_E \mathbf{s}$. Направление внешнего поля $\mathbf{E}_0$ примем за ось X, а направление распространения волны — за ось Z. Допустим сначала, что электрический вектор $\mathbf{E}$ волны направлен по оси X. Тогда $s_E = s_x$, и следовательно,

$$p_x = \beta E s_x^2, \quad p_y = \beta E s_x s_y, \quad p_z = \beta E s_x s_z.$$

Все направления, перпендикулярные к оси X, для распределения молекул равновероятны. Усредняя по этим направлениям, найдем: $\overline{s_x s_y} = \overline{s_x s_z} = 0$. Следовательно, для средних значений проекций дипольного момента получится

$$\overline{p}_x = \beta E \overline{s_x^2}, \quad \overline{p}_y = \overline{p}_z = 0,$$

а для вектора поляризации

$$P_x = N\beta E \overline{s_x^2}, \quad P_y = P_z = 0.$$

Вектор поляризации среды, таким образом, направлен по $\mathbf{E}$ и равен

$$\mathbf{P} = N\beta \mathbf{E} \overline{s_x^2}.$$

Соответствующая поляризуемость среды

$$\alpha_x = N\beta \overline{s_x^2}.$$

Показатель преломления будет *необыкновенным*. Для него и диэлектрической поляризуемости получаем

$$\varepsilon_x = n_e^2 = 1 + 4\pi N\beta \overline{s_x^2}.$$

[1]) Полностью анизотропных молекул, конечно, не существует. Поляризуемость реальной молекулы можно характеризовать *главными значениями* β_x, β_y, β_z тензора поляризуемости. Если одно из этих главных значений очень велико по сравнению с остальными двумя, то такая молекула может служить хорошим приближением к полной анизотропии.

Так как показатель преломления газа близок к единице, то

$$n_e \approx 1 + 2\pi N\beta\overline{s_x^2} = 1 + 2\pi N\beta\overline{\cos^2\vartheta}, \tag{90.3}$$

где ϑ — угол между осью молекулы и внешним полем $\mathbf{E}_0$ (т.е. осью X). Если распределение молекул изотропно (т.е. внешнего поля нет), то $\overline{s_x^2} = \overline{s_y^2} = \overline{s_z^2}$. Но $s_x^2 + s_y^2 + s_z^2 = 1$, а потому $\overline{s_x^2} = 1/3$. Значит, среда будет оптически изотропна, причем

$$n = 1 + \frac{2\pi}{3}N\beta. \tag{90.4}$$

Если поле $\mathbf{E}$ направлено вдоль оси Y, то $s_E = s_y$. В этом случае показатель преломления будет *обыкновенным*. Для него расчет, аналогичный приведенному выше, дает

$$n_o = 1 + 2\pi N\beta\overline{s_y^2}.$$

Чтобы вычислить $\overline{s_y^2}$, представим проекции единичного вектора $\mathbf{s}$ в полярной системе координат с полярной осью вдоль поля $\mathbf{E}$:

$$s_x = \cos\vartheta, \quad s_y = \sin\vartheta\cos\varphi, \quad s_z = \sin\vartheta\sin\varphi,$$

где φ — соответствующий азимут. Он с равной вероятностью может принимать все значения от 0 до 2π. Поэтому

$$\overline{s_y^2} = \overline{\sin^2\vartheta\cos^2\varphi} = \overline{\sin^2\vartheta}\,\overline{\cos^2\varphi} = \frac{1}{2}\overline{\sin^2\vartheta}.$$

Следовательно,

$$n_o = 1 + \pi N\beta\overline{\sin^2\vartheta}. \tag{90.5}$$

Прежде чем проводить дальнейшие вычисления, отметим одно следствие теории. Вычитая соотношение (90.4) из соотношений (90.3) и (90.5), получим

$$n_e - n = 2\left(\overline{\cos^2\vartheta} - \frac{1}{3}\right)\pi N\beta,$$

$$n_o - n = \left(\overline{\sin^2\vartheta} - \frac{2}{3}\right)\pi N\beta = \left(\frac{1}{3} - \overline{\cos^2\vartheta}\right)\pi N\beta.$$

Отсюда

$$\frac{n_e - n}{n_o - n} = -2. \tag{90.6}$$

Это соотношение, как показали тщательные измерения, хорошо выполняется для большинства веществ.

Найдем теперь постоянную Керра B. Из (90.3) и (90.5) получаем

$$n_e - n_o = 2\pi N\beta\left(\overline{\cos^2\vartheta} - \frac{1}{2}\overline{\sin^2\vartheta}\right) = 3\pi N\beta\left(\overline{\cos^2\vartheta} - \frac{1}{3}\right),$$

или на основании (90.4)

$$n_e - n_o = 9\frac{n-1}{2}\left(\overline{\cos^2\vartheta} - \frac{1}{3}\right). \tag{90.7}$$

Для дальнейших вычислений воспользуемся формулой Больцмана

$$dw = \text{const} \cdot e^{-U/kT}\, d\Omega, \tag{90.8}$$

которая определяет вероятность dw направления оси молекулы в пределах телесного угла $d\Omega$. Потенциальная энергия молекулы в электрическом поле $\mathbf{E}_0$ равна $U = p^2/(2\beta) - pE_0 \cos\vartheta$. Первый член есть энергия, затраченная на создание диполя с дипольным моментом $p = \beta E_0 \cos\vartheta$, второй — потенциальная энергия этого диполя в электрическом поле $\mathbf{E}_0$. Сумма обоих членов $U = -\frac{1}{2}\beta E_0^2 \cos^2\vartheta$. Поскольку она зависит только от угла ϑ, в качестве телесного угла $d\Omega$ удобно взять $2\pi \sin\vartheta\, d\vartheta$ и переписать формулу Больцмана в виде

$$dw = Ce^{-U/kT} \sin\vartheta\, d\vartheta, \tag{90.9}$$

где C — постоянная нормировки. Предполагая, что выполнено условие $|U/kT| \ll 1$, разложим экспоненциальную функцию в ряд и оборвем его на линейном члене:

$$e^{-U/kT} = 1 - \frac{U}{kT} = 1 + \frac{1}{2kT}\beta E_0^2 \cos^2\vartheta.$$

Постоянную C определим из условия нормировки:

$$C\int\limits_0^\pi \left(1 + \frac{1}{2kT}\beta E_0^2 \cos^2\vartheta\right)\sin\vartheta\, d\vartheta = 2C\left(1 + \frac{1}{3}\frac{\beta E_0^2}{2kT}\right) = 1.$$

Следовательно,

$$dw = \frac{1}{2}\frac{1 + \beta E_0^2 \cos^2\vartheta/(2kT)}{1 + \beta E_0^2/(6kT)}\sin\vartheta\, d\vartheta = \left[\frac{1}{2} + \frac{\beta E_0^2}{4kT}\left(\cos^2\vartheta - \frac{1}{3}\right)\right]\sin\vartheta\, d\vartheta,$$

$$\overline{\cos^2\vartheta} - \frac{1}{3} = \int\limits_{\vartheta=0}^{\vartheta=\pi}\left(\cos^2\vartheta - 1\right)dw = \frac{2}{45}\frac{\beta E_0^2}{kT}.$$

После подстановки в формулу (90.7) получим

$$n_e - n_o = \frac{n-1}{5}\frac{\beta E_0^2}{kT}, \tag{90.10}$$

а для постоянной Керра

$$B = \frac{n-1}{5\lambda kT}\beta. \tag{90.11}$$

4. Таким образом, разность $n_e - n_o$ пропорциональна *квадрату внешнего поля* E_0, как этого и следовало ожидать из соображений симметрии. Но в теории Ланжевена постоянная Керра всегда положительна и притом не только для полностью анизотропных молекул, но и для молекул с произвольным тензором

поляризуемости. Борн (1882–1971) в 1916 г. устранил этот недостаток теории, распространив ее на *полярные молекулы* со значительными постоянными дипольными моментами $\mathbf{p}_0$, направления которых могут не совпадать с направлениями наибольшей поляризуемости молекул. Если $\mathbf{p}_0$ велико по сравнению с дипольным моментом молекулы, индуцированным внешним полем $\mathbf{E}_0$, то ориентация молекул в таком поле будет определяться преимущественно постоянными моментами. Постоянные моменты $\mathbf{p}_0$ будут стремиться ориентироваться вдоль поля $\mathbf{E}_0$, определяющего в среде направление оптической оси. Направление же наибольшей поляризуемости среды может составлять с ним заметный угол. Если эти два направления совпадают, то постоянная Керра B будет положительна. Если они взаимно перпендикулярны, то $B < 0$. В промежуточных случаях может получиться и тот и другой знак.

Конечно, наличие постоянного дипольного момента заметно усложняет вычисления, даже если молекулы полностью анизотропны. Ограничимся простейшим случаем, когда направления вектора $\mathbf{p}_0$ и наибольшей поляризуемости полностью анизотропной молекулы совпадают между собой. Тогда единственное отличие от случая, разобранного выше, состоит в том, что теперь потенциальная энергия молекулы в электрическом поле $\mathbf{E}_0$ будет $U = -(\mathbf{p}_0\mathbf{E}_0) = -p_0E_0\cos\vartheta$, и в разложении соответствующего экспоненциального множителя появится линейный член по E_0:

$$e^{-U/kT} = 1 + \frac{p_0E_0\cos\vartheta}{kT} + \frac{p_0^2E_0^2\cos^2\vartheta}{2k^2T^2}.$$

Но этот член не сказывается на вычислении $\overline{\cos^2\vartheta}$, так как интегралы от $\cos\vartheta\sin\vartheta$ и $\cos^3\vartheta\sin\vartheta$ обращаются в нуль. Таким образом, чтобы получить окончательный результат, достаточно в формуле (90.11) поляризуемость β заменить на p_0^2/kT. Это дает

$$B = \frac{n-1}{5\lambda}\left(\frac{p_0}{kT}\right)^2. \qquad (90.12)$$

Формула (90.12) показывает, что и в случае полярных молекул разность $n_e - n_o$ пропорциональна квадрату поля E_0.

5. Поскольку процессы поляризации молекул и их ориентация в электрическом поле происходят за очень короткие времена, следует ожидать, что и время возникновения и уничтожения эффекта Керра при наложении и снятии электрического поля должно быть такого же порядка. Действительно, опыты Абрагама и Лемуана, начатые еще в 1899 г. и неоднократно продолженные другими исследователями, показали, что это время меньше 10^{-8} с, а в некоторых случаях даже меньше 10^{-9} с.

Точность всех этих опытов была недостаточна, чтобы высказать более определенные суждения. Это удалось только в результате развития лазерной техники, позволившей получать короткие импульсы (с длительностью порядка 10^{-12} с) мощного лазерного излучения (со средней напряженностью электрического поля в несколько десятков кВ/см).

Такие световые поля, хотя они и не являются статическими, способны вызывать двойное преломление в ячейках Керра. Принципиальная схема опыта приведена на рис. 310. Справа на-

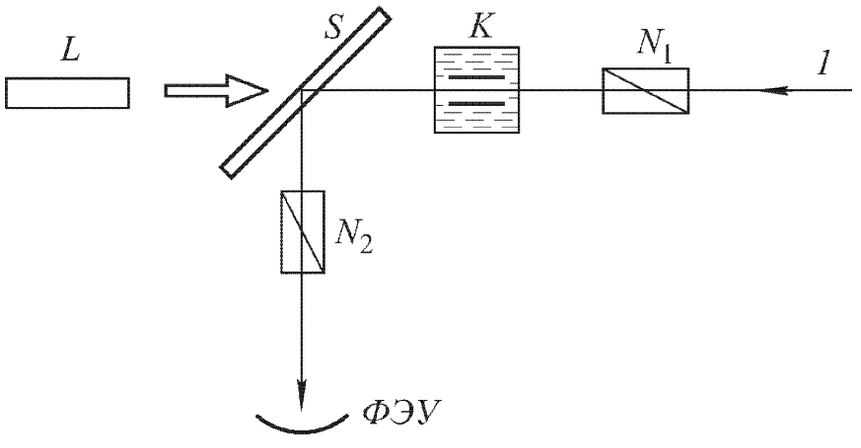


Рис. 310

лево через ячейку Керра K проходит луч *1* голубого света, претерпевающий затем отражение от полупрозрачной пластинки S и попадающий в фотоумножитель ФЭУ. Если николи N_1 и N_2 скрещены, то свет в фотоумножитель не попадет. Однако если через ячейку Керра слева направо пропустить мощный импульс света от лазера L, то при прохождении через ячейку он вызовет в исследуемой жидкости двойное преломление луча *1*, и свет в фотоумножитель начнет поступать. Если на ту же ячейку наложить статическое электрическое поле и подобрать его напряженность так, чтобы возникла такая же разность фаз между необыкновенным и обыкновенным лучами, что и в предыдущем случае, то можно сравнить постоянные Керра в статическом и переменном полях лазерного излучения. Оказалось, что в недипольных жидкостях постоянные Керра в обоих случаях практически совпадают. В дипольных жидкостях при переходе от статического поля к переменному полю лазерного излучения постоянная Керра уменьшается. Так, в случае нитробензола она уменьшается приблизительно в 100 раз.

Уже описанный опыт позволяет оценить длительность существования эффекта Керра. Более точные результаты дают измерения по схеме, приведенной на рис. 311. Мощный инфракрасный импульс от лазера с длительностью порядка 10^{-12} с и длиной волны $\lambda = 1060$ нм проходит через кристалл дигид-

рофосфата калия (KDP). Из-за нелинейного эффекта удвоения частоты (см. § 124) незначительная часть света превращается в зеленый свет с длиной волны $\lambda = 530$ нм. Зеркало S_1 пропускает инфракрасный луч и отражает зеленый. Зеркало S_2, наоборот, пропускает зеленый луч, но отражает инфракрасный. Оба луча проходят через ячейку Керра K с исследуемой жидкостью, расположенную между скрещенными николями N_1 и N_2. За николем N_2 поставлен светофильтр F, задерживающий инфракрасное излучение и пропускающий на фотоумножитель ФЭУ зеленое. Если после зеркала S_2 импульс зеленого излучения идет впереди мощного инфракрасного импульса, не перекрываясь с

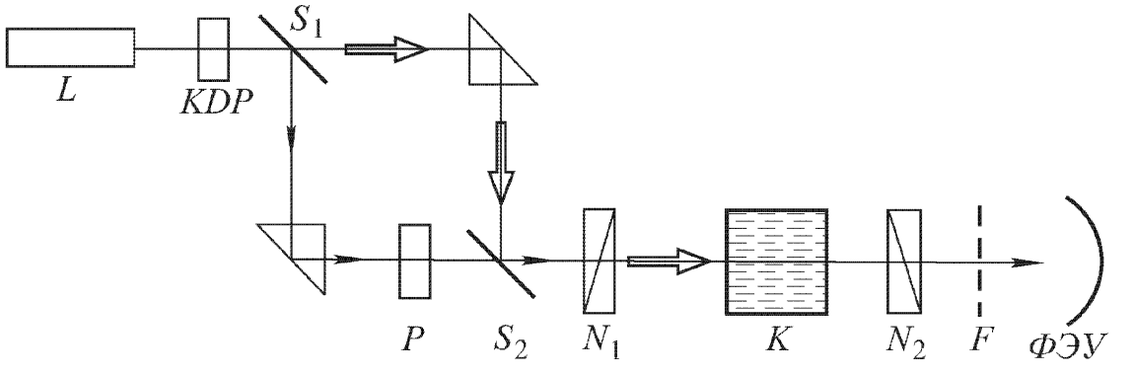


Рис. 311

ним, то, разумеется, зеленый свет не сможет достигнуть фотоумножителя. Помещая на пути зеленого луча пластинку P, можно задержать зеленый импульс относительно инфракрасного и по толщине пластинки измерить время задержки τ. Располагая набором пластинок разной толщины, можно менять время τ. Опыты показали, что в простейшем случае интенсивность проходящего света, регистрируемая фотоумножителем, меняется экспоненциально по закону $I = I_0 \exp(-\tau/\tau_0)$. Постоянная τ_0 имеет смысл времени релаксации анизотропии жидкости, вызванной инфракрасным импульсом. Она и является мерой длительности эффекта Керра. Измерения дали для сероуглерода $\tau_0 = 2 \cdot 10^{-12}$ с, а для нитробензола $\tau_0 = 5 \cdot 10^{-11}$ с.

6. Рассмотрим теперь механизм возникновения двойного преломления в переменных полях. Если молекулы полярные, то в одну половину периода их постоянные дипольные моменты стремятся ориентироваться в направлении внешнего электрического поля. В следующую половину периода, когда внешнее поле изменит направление на противоположное, возникнет такая же, но противоположно направленная ориентация. Эффекты ориентации, вызванные такими полями, вычитаются. В низкочастотных полях (в радиодиапазоне) в каждый момент времени устанавливается ориентация, соответствующая мгновенному значе-

нию электрического поля. В этом случае явление протекает так же, как в статических полях.

Однако в высокочастотных полях (с длинами волн короче примерно 1 см) ориентация постоянных дипольных моментов практически прекращается. В таких полях постоянные дипольные моменты $\mathbf{p}_0$ не могут играть роли в возникновении анизотропии, а с ней и двойного преломления среды. Ориентация обусловлена *только индуцированными дипольными моментами.* В одну половину периода, когда электрическое поле направлено в определенную сторону, индуцированные дипольные моменты создают моменты сил, стремящиеся приблизить оси наибольшей поляризуемости молекул к направлению электрического поля. В следующую половину периода направления всех моментов меняются на противоположные. Однако они также будут приближать оси наибольшей поляризуемости молекул к тому же направлению внешнего поля. Действительно, в отличие от вектора, у оси наибольшей поляризуемости нет одностороннего направления, *оба направления ее совершенно эквивалентны.* Не имеет никакого значения, каким концом эта ось приближается к направлению вектора $\mathbf{E}_0$. Таким образом, эффекты ориентации осей наибольшей поляризуемости в соседние половины периода, несмотря на противоположные направления векторов $\mathbf{E}_0$ в эти полупериоды, будут складываться, а не вычитаться. Хотя за каждый полупериод ориентация и ничтожна, благодаря такому сложению за время релаксации анизотропии τ_0 возникает достаточная ориентация, чтобы вызвать двойное преломление.

Таким образом, механизм возникновения двойного преломления в переменных полях высокой частоты — такой же, какой принимается в теории Ланжевена.

7. Благодаря чрезвычайной быстроте установления и исчезновения явления Керра оно нашло широкие научные и технические применения в качестве *быстродействующих затворов и модуляторов света.* Керровский модулятор света представляет собой ячейку Керра, конденсатор которой питается электрическим полем высокой частоты. Он позволяет осуществить громадное число (до 10^9) прерываний в секунду, недостижимое другими (например, механическими) средствами. Ячейка Керра, на которую подается кратковременный импульс электрического поля, может служить фотографическим затвором, время действия которого определяется длительностью этого импульса. Если в качестве электрического импульса взять мощный световой импульс от лазера, то время экспозиции можно довести до 10^{-12} с. Керровские затворы и модуляторы света применяются в лазерной технике для управления режимом работы лазеров.

8. Аналогом эффекта Керра является *эффект Коттона–Мутона*, открытый в 1910 г. Если молекулы среды анизотропны

и обладают постоянными магнитными моментами, то они могут преимущественно ориентироваться постоянным магнитным полем. В достаточно сильных магнитных полях возникает анизотропия и связанное с ней двойное лучепреломление среды. Это и есть эффект Коттона–Мутона. Среда ведет себя подобно одноосному кристаллу, оптическая ось которого параллельна магнитному полю **B**.

Двойное преломление можно обнаружить и измерить так же, как это делается при изучении явления Керра. И законы обоих явлений совершенно аналогичны. Разность необыкновенного и обыкновенного показателей преломления определяется выражением $n_e - n_o = DB^2$, а соответствующая разность фаз

$$\varphi = \frac{2\pi}{\lambda}(n_e - n_o)l = 2\pi ClB^2, \tag{90.13}$$

где D и $C = D/\lambda$ — постоянные, зависящие от свойств среды и ее физического состояния. Для «постоянной Коттона–Мутона» в нитробензоле измерения дают $C = 2{,}23 \cdot 10^{-12}$ СГСМ. В магнитном поле $B = 20\,000$ Гс на пути $l = 10$ см возникает разность фаз $\varphi = 0{,}056$ рад, т.е. всего 3,2 градуса.

§ 91. Линейный электрооптический эффект Поккельса

1. Свободные колебания гармонического осциллятора описываются уравнением $\ddot{r} + 2\gamma\dot{r} + \omega_0^2 r = 0$. Они происходят около положения равновесия $r = 0$. Допустим теперь, что осциллятор находится в постоянном электрическом поле E_0. Тогда в отсутствие других внешних сил будет

$$\ddot{r} + 2\gamma\dot{r} + \omega_0^2 r = -\frac{e}{m}E_0.$$

Теперь положение равновесия сместится от начала координат на расстояние $r_0 = -eE_0/(m\omega_0^2)$. Обозначим через q расстояние колеблющейся частицы от нового положения равновесия. Тогда $r = r_0 + q$, и после подстановки в предыдущее уравнение имеем

$$\ddot{q} + 2\gamma\dot{q} + \omega_0^2 q = 0.$$

Отсюда видно, что в постоянном внешнем электрическом поле колебания осциллятора останутся гармоническими с *прежней частотой* ω_0, но они будут происходить около нового положения равновесия. Таким образом, постоянное электрическое поле не изменяет собственную частоту гармонического осциллятора, а только смещает положение равновесия, около которого совершаются свободные колебания.

В случае колебаний с большой амплитудой модель гармонического осциллятора может оказаться непригодной. В простей-

шем случае к квазиупругой силе $m\omega_0^2 r$ надо добавить член, пропорциональный квадрату смещения частицы из положения равновесия (начала координат). Свободные колебания такого *ангармонического осциллятора* описываются уравнением $\ddot{r} + 2\gamma\dot{r} + \omega_0^2 r + \beta r^2 = 0$, где β — постоянная. При наличии внешнего постоянного электрического поля E_0 уравнение колебаний переходит в

$$\ddot{r} + 2\gamma\dot{r} + \omega_0^2 r + \beta r^2 = -\frac{e}{m}E_0.$$

Теперь положение равновесия $r = r_0$ определится из уравнения

$$\omega_0^2 r_0 + \beta r_0^2 = -eE_0/m.$$

Из двух корней этого квадратного уравнения надо взять тот, который мало отличается от ранее найденного значения $r_0 = -eE_0/(m\omega_0^2)$ без учета ангармоничности (так как последняя предполагается малой). Пусть по-прежнему q означает отклонение колеблющейся частицы от нового положения равновесия, так что $r = r_0 + q$. Предполагая колебания малыми, пренебрежем квадратами q. Тогда

$$\ddot{q} + 2\gamma\dot{q} + (\omega_0^2 + 2\beta r_0)q = 0.$$

Отсюда видно, что во внешнем постоянном электрическом поле малые колебания ангармонического осциллятора в рассматриваемом приближении опять будут гармоническими. Однако *при наличии ангармоничности внешнее поле E_0 не только смещает положение равновесия, но и изменяет собственную частоту осциллятора.* Изменение квадрата собственной частоты осциллятора приближенно равно $\Delta\omega_0^2 = 2\beta r_0$, или в том же приближении

$$\Delta\omega_0^2 = -\frac{2e\beta}{m\omega_0^2}E_0. \tag{91.1}$$

2. Смещение собственных частот меняет кривую дисперсии, т.е. показатель преломления n среды. В простейшем случае, когда собственная частота ω_0 одна, величина n вдали от линии поглощения зависит только от разности $\omega^2 - \omega_0^2$, как это видно из формулы (84.9). Тогда изменение n в статическом электрическом поле E_0 определяется выражением

$$\Delta n = \frac{\partial n}{\partial \omega_0^2}\Delta\omega_0^2 = -\frac{\partial n}{\partial \omega_0^2}\frac{2e\beta}{m\omega_0^2}E_0.$$

Это выражение можно преобразовать, заметив, что $\partial f/\partial\omega_0^2 = -\partial f/\partial\omega^2$. Тогда

$$\Delta n = \frac{\partial n}{\partial \omega^2}\frac{2e\beta}{m\omega_0^2}E_0 = \frac{\partial n}{\partial \omega}\frac{e\beta}{m\omega\omega_0^2}E_0. \tag{91.2}$$

При фиксированном направлении внешнего поля $\mathbf{E}_0$ величина Δn зависит от направления распространения света. Это сказывается на двойном преломлении среды. *Изменение двойного преломления вещества из-за смещения собственной частоты во внешнем электрическом поле называется электрооптическим эффектом Поккельса.* В этом эффекте изменения показателей преломления пропорциональны *первой степени* внешнего поля E_0, в отличие от эффекта Керра, где они пропорциональны квадрату поля.

Эффект Поккельса может наблюдаться только в кристаллах, не обладающих центром симметрии. Дело в том, что он *линеен* относительно внешнего поля E_0. Поэтому при изменении направления поля $\mathbf{E}_0$ на противоположное должен меняться на противоположный и знак изменения Δn показателя преломления. Но в кристаллах с центром симметрии это невозможно, так как оба взаимно противоположных направления внешнего поля физически совершенно эквивалентны.

Из механизма явления ясно, что эффект Поккельса по крайней мере столь же *безынерционен*, что и эффект Керра. Поэтому он, наряду с эффектом Керра, нашел применение (например, в технике лазеров) в качестве *оптических затворов* и *высокочастотных модуляторов света.* Соответствующее устройство называется *ячейкой Поккельса.* Она представляет собой кристалл, помещаемый между двумя скрещенными николями. Такое устройство действует так же, как и ячейка Керра. Николи не пропускают свет, когда нет внешнего электрического поля, но при наложении такого поля пропускание появляется. Необходимо, чтобы кристалл до наложения внешнего электрического поля не давал двойного преломления. Этого можно достигнуть, если взять оптически одноосный кристалл, вырезанный перпендикулярно к оптической оси, а свет направить вдоль этой оси. Внешнее поле $\mathbf{E}_0$ может быть направлено либо перпендикулярно (*поперечный модулятор света*), либо параллельно распространению света (*продольный модулятор*).

§ 92. Эффект Зеемана

1. В 1896 г. Зееман (1865–1943) обнаружил, что спектральные линии определенным образом расщепляются, если источник света поместить в магнитное поле. В опыте Зеемана исследовалась очень узкая зелено-голубая линия кадмия и применялись магнитные поля с напряженностью 10 000–15 000 Гс. Г.А. Лоренц, развивавший в то время электронную теорию, сразу же объяснил *явление Зеемана* и тем самым придал дальнейшим исследованиям планомерный и целенаправленный характер.

Применяемая схема для наблюдения и исследования явления приведена на рис. 312. Источник света с линейчатым спектром (например, газоразрядная трубка или вакуумная дуга) помещается между полюсами электромагнита, создающего достаточно однородное магнитное поле. Исследуемый свет попадает на щель

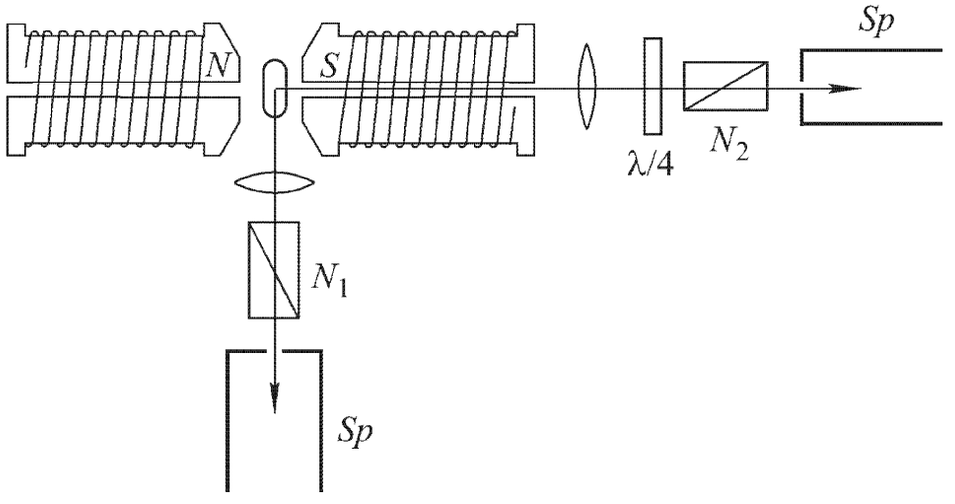


Рис. 312

спектроскопа или спектрографа Sp с разрешающей силой около 100 000 или выше (дифракционную решетку или интерференционный спектральный аппарат). Николи N_1, N_2 и пластинка $\lambda/4$ служат для исследования поляризации излучаемого света. При фотографировании наблюдаемой картины применяются иногда многочасовые экспозиции. В течение всего этого времени должно быть обеспечено с достаточной точностью постоянство магнитного поля и температуры источника, чтобы картина оставалась неизменной во времени и можно было использовать спектральный аппарат высокой разрешающей силы.

В первых опытах Зееман обнаружил, что *при наблюдении поперек поля спектральная линия расщепляется на три линейно поляризованные компоненты.* Средняя компонента не смещена, крайние смещены в противоположные стороны на одинаковые расстояния (в шкале частот). Смещение пропорционально напряженности внешнего магнитного поля $\mathbf{B}$. В средней компоненте электрический вектор направлен *параллельно магнитному полю* (такие компоненты называются *π-компонентами*), в крайних — *перпендикулярно к нему* (такие компоненты называются *σ-компонентами*). Интенсивность π-компоненты вдвое, а каждой из σ-компонент в четыре раза меньше интенсивности исходной линии.

При наблюдении *вдоль магнитного поля* получается *такое же смещение* (при одинаковой напряженности магнитного поля), что и в предыдущем случае, но *несмещенная компонента отсутствует.* Интенсивность каждой компоненты вдвое мень-

ше интенсивности исходной спектральной линии. Обе компоненты *поляризованы по кругу* в противоположных направлениях (их принято называть также σ-компонентами). Если свет распространяется в направлении магнитного поля, то σ-компонента с меньшей частотой поляризована *по правому*, а с большей — *по левому* кругу. При изменении направления магнитного поля на противоположное меняется на противоположную и круговая поляризация обеих компонент.

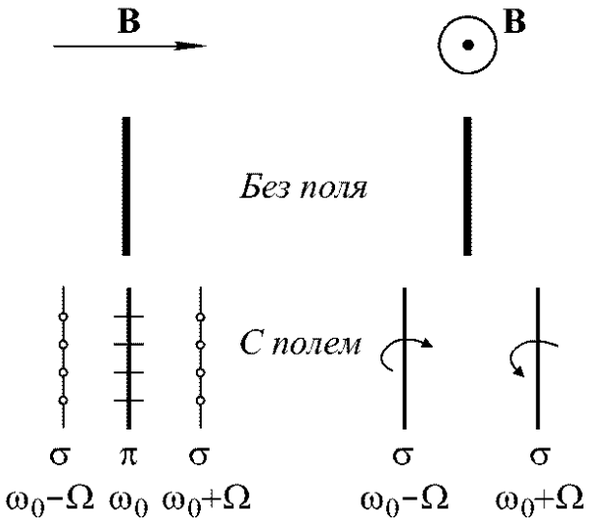


Рис. 313

Картина, наблюдаемая поперек и вдоль магнитного поля, представлена схематически на рис. 313. Предполагается, что в случае продольного эффекта свет распространяется вдоль магнитного поля, направленного к читателю. Относительные интенсивности линий показаны их толщиной, поляризация π-компоненты — штрихами, параллельными магнитному полю, а σ-компонент — кружочками.

2. Описанная картина расщепления спектральных линий объясняется *классической теорией Лорентца*. Как и классическая теория дисперсии, это есть *модельная теория*, в простейшей форме которой излучающими центрами являются *гармонические осцилляторы* в виде *квазиупруго связанных электронов*. В отсутствие внешнего магнитного поля уравнение движения такого электрона имеет вид $\ddot{\mathbf{r}} + \omega_0^2 \mathbf{r} = 0$, где ω_0 — собственная частота электрона. При наличии постоянного магнитного поля на электрон действует еще сила Лорентца $-\frac{e}{c}[\dot{\mathbf{r}}\mathbf{B}]$ (заряд электрона обозначен через $-e$). Уравнение движения электрона принимает вид

$$\ddot{\mathbf{r}} + \omega_0^2 \mathbf{r} = -\frac{e}{mc}[\dot{\mathbf{r}}\mathbf{B}],$$

где m — масса электрона. Введя ларморовскую частоту

$$\boldsymbol{\Omega} = \frac{e}{2mc}\mathbf{B}, \tag{92.1}$$

приведем его к виду

$$\ddot{\mathbf{r}} + 2[\dot{\mathbf{r}}\boldsymbol{\Omega}] + \omega_0^2 \mathbf{r} = 0 \tag{92.2}$$

(см. т. III, § 86). Классическая теория сводится к решению этого уравнения. Для решения уравнения (92.2) перейдем к координатной форме. Направим ось Z прямоугольной системы координат вдоль магнитного поля $\mathbf{B}$. Тогда предыдущее уравнение

сведется к системе трех скалярных уравнений

$$\begin{aligned} \ddot{x} + 2\Omega\dot{y} + \omega_0^2 x &= 0, \\ \ddot{y} - 2\Omega\dot{x} + \omega_0^2 y &= 0, \\ \ddot{z} + \omega_0^2 z &= 0. \end{aligned} \tag{92.3}$$

Из последнего уравнения видно, что магнитное поле не влияет на движение электрона вдоль магнитного поля. Это и понятно, так как при таком движении не возникает силы, действующей со стороны магнитного поля. Интегрирование первых двух уравнений (92.3) удобно провести в *комплексной форме*. Объединим x и y в комплексную координату $\zeta = x + iy$. Она определяет положение электрона в координатной плоскости (X, Y) совершенно так же, как это делается с помощью двумерного вектора $\boldsymbol{\zeta}$ с составляющими x и y. Заметив, что $-i\dot{\zeta} = \dot{y} - i\dot{x}$, умножим второе уравнение (92.3) на i и сложим с первым. Тогда

$$\ddot{\zeta} - i \cdot 2\Omega\dot{\zeta} + \omega_0^2 \zeta = 0.$$

Ищем решение этого уравнения в виде $\zeta = e^{i\omega t}$. Постоянная ω найдется из квадратного уравнения

$$-\omega^2 + 2\Omega\omega + \omega_0^2 = 0,$$

которое дает

$$\omega = \Omega \pm \sqrt{\omega_0^2 + \Omega^2}.$$

Даже в очень сильных магнитных полях квадратом ларморовской частоты можно пренебречь по сравнению с ω_0^2. Например, если $B = 10^4$ Гс, то формула (92.1) дает $\Omega \approx 10^{11}$ с$^{-1}$, тогда как для видимого света ($\lambda = 500$ нм) $\omega \sim 4 \cdot 10^{15}$ с$^{-1}$, а потому $(\Omega/\omega)^2 \sim 10^{-9}$. Максимальное магнитное поле, в котором измерялось зеемановское расщепление спектральных линий, получено в 1938 г. П.Л. Капицей (1894–1984). Оно было $3{,}2 \cdot 10^5$ Гс. Даже в этом случае $\Omega/\omega_0 \sim 1{,}4 \cdot 10^{-3}$, $(\Omega/\omega_0)^2 \sim 2 \cdot 10^{-6}$. Таким образом, с большой точностью $\omega = \pm\omega_0 + \Omega$. Чтобы не пользоваться отрицательными частотами, введем переобозначение, положив $\omega_1 = \omega_0 + \Omega$, $\omega_2 = \omega_0 - \Omega$. Тогда полученные два решения запишутся в виде

$$\zeta_1 = e^{i\omega_1 t}, \quad \zeta_2 = e^{-i\omega_2 t}.$$

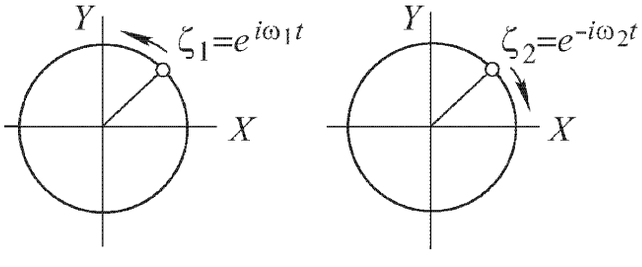


Рис. 314

Первое решение представляет *круговое движение*, в котором электрон вращается *против часовой стрелки* с угловой частотой ω_1, второе — также *круговое движение*, но *по часовой стрел-*

же и с частотой ω_2 (рис. 314). Общее решение соответствует наложению таких двух вращений и представляется в виде $\zeta = C_1\zeta_1 + C_2\zeta_2$, где C_1 и C_2 — произвольные постоянные.

3. Чтобы нагляднее уяснить полученные результаты, разложим первоначальное движение электрона (т.е. движение в отсутствие магнитного поля) на два движения: на гармоническое колебание в направлении оси Z и на движение в плоскости XY. Второе движение в свою очередь разложим на два круговых вращения с одной и той же угловой частотой ω_0, но совершающиеся в противоположных направлениях. Тогда в постоянном магнитном поле колебание вдоль оси Z остается неизменным. Частоты же обоих круговых вращений изменяются на одну и ту же величину Ω: если вращение совершается против часовой стрелки, то частота увеличивается, а если по часовой стрелке, то уменьшается.

Для изменения частоты при вращении по кругу можно привести простое объяснение. Центростремительная сила, действующая на вращающийся электрон в отсутствие магнитного поля, равна $m\omega_0^2 r$. В магнитном поле к ней добавляется сила $\pm\frac{e}{c}vB = \pm\frac{e}{c}\omega rB$, так что новая центростремительная сила становится равной

$$\left(m\omega_0^2 \pm \frac{eB}{c}\omega\right) r = m(\omega_0^2 \pm 2\Omega\omega)r.$$

Выбор знака зависит от направления вращения. Приравнивая это выражение $m\omega^2 r$, приходим к уравнению $\omega^2 = \omega_0^2 \pm 2\Omega\omega$, из которого для положительных корней находим $\omega \approx \omega_0 \pm \Omega$. Это совпадает с результатами, полученными выше.

При включении магнитного поля кинетическая энергия вращения электрона изменяется. Возникает вопрос, как это может происходить, если сила, действующая со стороны магнитного поля, перпендикулярна к скорости электрона $\mathbf{v}$ и, следовательно, работы не совершает? Ответ состоит в том, что последнее утверждение относится к *постоянным магнитным полям*, которые только и учитываются уравнением (92.2). Но при включении магнитного поля оно *нарастает во времени* от нуля до максимального значения, а в дальнейшем вплоть до выключения остается постоянным. Во время же нарастания магнитного поля, согласно закону индукции Фарадея, возбуждается *вихревое электрическое поле*, которое и совершает работу над электроном, меняя его кинетическую энергию. Когда магнитное поле становится постоянным, электрическое поле пропадает и дальнейшее изменение кинетической энергии вращения электрона прекращается, пока не будет выключено магнитное поле. К этим установившимся вращениям и относятся движения, найденные выше. Подробное рассмотрение механизма изменения кинетиче-

ской энергии вращения электрона было приведено в учении об электричестве (см. т. III, § 88).

4. Перейдем теперь к объяснению расщепления спектральных линий в магнитном поле. Колеблющийся электрон излучает электромагнитные волны. Излучение максимально в направлении, перпендикулярном к ускорению электрона, а в направлении ускорения отсутствует. Согласно классической теории, частота излучаемого света совпадает с частотой колебания электрона. Но последняя меняется при включении магнитного поля. Поэтому должна измениться и частота излучаемого света. При наблюдении вдоль магнитного поля колебание в том же направлении излучения не дает. Излучение создается *только круговыми вращениями электрона.* В результате наблюдаются две σ-компоненты с *круговой поляризацией* и частотами $\omega_0 + \Omega$ и $\omega_0 - \Omega$. Если свет идет в направлении вектора $\mathbf{B}$, то поляризация первой линии будет *левой*, а второй — *правой.* При изменении направления магнитного поля на противоположное меняется на противоположную и круговая поляризация каждой линии. При наблюдении поперек магнитного поля $\mathbf{B}$ колебания электрона, параллельные $\mathbf{B}$, дают максимум излучения. Им соответствует *несмещенная π-компонента*, в которой электрический вектор *параллелен* $\mathbf{B}$. Оба круговых движения совершаются в плоскости, перпендикулярной к $\mathbf{B}$. Разложим каждое из них на гармоническое колебание вдоль линии наблюдения и перпендикулярное к нему. Только колебания, перпендикулярные к линии наблюдения, сопровождаются излучением и дают две *σ-компоненты* с частотами $\omega + \Omega$ и $\omega - \Omega$, в которых электрические векторы *перпендикулярны к* $\mathbf{B}$.

Таково объяснение расщепления спектральных линий, наблюдавшееся в первых опытах Зеемана. Если учесть, что в отсутствие магнитного поля все направления движения электрона равновероятны, то нетрудно объяснить и относительные интенсивности спектральных линий в этих опытах.

Как видно из численного примера, приведенного выше ($B = 10^4$ Гс), $\Omega/\omega_0 \approx 2 \cdot 10^{-5}$. Для разрешения такого расщепления требуются спектральные приборы с разрешающей силой ω_0/Ω не менее $5 \cdot 10^4$, т.е. дифракционные решетки или интерференционные спектроскопы. Но в опытах П.Л. Капицы ($B = 3{,}2 \cdot 10^5$ Гс) были уже достаточны призменные спектроскопы.

Исследуя характер круговой поляризации линий в продольном эффекте Зеемана, можно определить *знак зарядов*, вызывающих этот эффект. Он оказался *отрицательным.* Измеряя же величину расщепления, можно определить удельный заряд e/m. Он оказался таким же, как и при измерениях по отклонениям катодных лучей в электрических и магнитных полях

($e/m = 1{,}759 \cdot 10^7$ СГСМ). Это не оставляет сомнений в том, что заряженные частицы, определяющие оптическое поведение атомов, действительно являются *электронами.*

5. Дальнейшие опыты показали, что явление Зеемана в том виде, в каком оно наблюдалось сначала и нашло объяснение в теории Лорентца — лоренцовский триплет, состоящий из одной π-компоненты и двух σ-компонент, а также дублет из двух σ-компонент, поляризованных по кругу, — наблюдается крайне редко. Такое расщепление называется *нормальным* или *простым эффектом Зеемана.* Простой эффект дают так называемые *синглетные*, т.е. одиночные, практически монохроматические спектральные линии. Подавляющее большинство спектральных линий являются *мультиплетами* (*дублетами*, *триплетами*, *квартетами* и т.д.), т.е. состоят из нескольких тесно расположенных спектральных линий.

Простейшим примером мультиплета (дублета) может служить двойная D-линия натрия. Она состоит из двух близко расположенных линий с длинами волн $\lambda_{D_1} = 589{,}5930$ нм и $\lambda_{D_2} = 588{,}96963$ нм, причем интенсивность линии D_2 вдвое больше интенсивности линии D_1.

Мультиплеты в магнитных полях дают значительно более сложную картину расщепления, чем расщепление в простом эффекте Зеемана. Так, линия D_1 натрия расщепляется на *четыре линии*: средние из них являются π-, а крайние — σ-компонентами. Линия же D_2 расщепляется на *шесть компонент*: две средние являются π-, а четыре крайние — σ-компонентами. Таким образом, весь дублет расщепляется на 10 линий. Наблюдаются и значительно более сложные картины расщепления мультиплетов. Такие расщепления называются *аномальным* или *сложным эффектом Зеемана.* Предпочтителен термин «сложный эффект», так как именно *сложный*, а не простой эффект *является правилом, а не исключением.*

Объяснение сложного эффекта Зеемана дала *квантовая теория*, да и то после того, как был открыт *спин* (т.е. собственный момент количества движения) и связанный с ним *магнитный момент электрона.* В случае синглетных спектральных линий квантовая теория приводит к тем же результатам, что и простая теория Лорентца. Мы вернемся ко всем этим вопросам в пятом томе нашего курса.

§ 93. Понятие об эффекте Штарка

Как было показано в § 91, однородное электрическое поле не меняет собственную частоту ω_0 гармонического осциллятора. Однако при учете ангармоничности такое изменение должно происходить. Если ограничиться линейными членами по внеш-

нему полю E_0, то оно будет выражаться формулой (91.1), из которой следует $\Delta\omega_0 = -\beta e E_0/(m\omega_0^3)$. В этом приближении внешнее поле E_0 не вызывает расщепления спектральных линий, а только производит *небольшое смещение их в сторону*, пропорциональное напряженности поля E_0. При учете членов высших степеней можно ожидать расщепления спектральных линий с расстояниями между компонентами, пропорциональными *квадрату поля* E_0, а именно

$$\delta\omega_0 \sim [e/(m\omega_0^3)]E_0^2.$$

На возможность такого расщепления указывал Фохт (1850–1919). Однако из-за экспериментальных трудностей ему не удалось наблюдать ожидаемое явление. Трудность состояла в том, что для наблюдения эффекта требуются электрические поля E_0 порядка 100 000 В/см. В обычных же газоразрядных трубках, излучающих узкие спектральные линии (например, водородной серии Бальмера), такие сильные поля создавать и поддерживать невозможно из-за *сильной ионизации газа в трубках.*

Штарк (1874–1957) в 1913 г. преодолел эту трудность, создавая сильное электрическое поле за катодом в узком зазоре (~ 1 мм) между F и K (рис. 315), в котором непрерывной откачкой поддерживался высокий вакуум, так что в этом зазоре практически не происходило ионизации частиц из-за столкновений между ними. Возбуждение свечения атомов газа происходило в другой части трубки, как в обычной газоразрядной трубке. Светящиеся атомы (каналовые лучи) проникали через отверстия (каналы) в катоде и попадали в пространство между F и K, где и подвергались воздействию внешнего электрического поля.

Направление наблюдения

Рис. 315

С помощью такой установки Штарку удалось наблюдать расщепление спектральных линий в электрическом поле. Это явление было названо его именем. Оно оказалось совсем *не похожим* на то, что ожидал Фохт. Штарк исследовал расщепление спектральных линий водородной серии Бальмера. В видимой области спектра она содержит четыре линии: H_α ($\lambda = 656{,}285$ нм), H_β ($\lambda = 486{,}132$ нм), H_γ ($\lambda = 434{,}046$ нм), H_δ ($\lambda = 410{,}173$ нм). Расщепление *в водороде*, а также в случае всех *одноэлектронных ионов* оказалось пропорциональным *не второй степени элек-*

трического поля, а первой, т.е. значительно более сильным. По этой причине такое расщепление стали называть *линейным эффектом Штарка.*

Картина расщепления оказалась довольно сложной и различной для различных спектральных линий. Каждая спектральная линия водорода расщеплялась на несколько компонент. Для линии H_α было найдено расщепление на 15 компонент, для H_β на 20, для H_γ на 27 и для H_δ на 32 компоненты. В электрическом поле $E_0 = 104\,000$ В/см для расстояния между крайними компонентами Штарк нашел 2,3; 3,88; 5,88 и 7,5 нм для линий H_α, H_β, H_γ, H_δ соответственно. Эти расстояния довольно велики по сравнению с соответствующими расстояниями между крайними компонентами простого триплета Зеемана. (В поле $B = 104\,000$ Гс для указанных линий зеемановские расщепления составляли бы соответственно 0,42; 0,23; 0,18 и 0,16 нм.)

При наблюдении поперек поля $\mathbf{E}_0$ компоненты *поляризованы линейно*, причем часть из них является π-компонентами (у которых электрическое поле параллельно внешнему полю $\mathbf{E}_0$), а часть σ-компонентами (у которых электрическое поле перпендикулярно к внешнему полю $\mathbf{E}_0$). При наблюдении вдоль поля $\mathbf{E}_0$ π-компоненты не появляются, а σ-компоненты *не поляризованы.* Более сильные π-компоненты в общем случае лежат снаружи, а более сильные σ-компоненты — внутри. Расщепление и поляризация у водорода по обе стороны от первоначальной линии симметричны; у других атомов часто встречается несимметричное расположение. Расстояния (в шкале частот) компонент от первоначальной линии оказались кратными наименьшему возможному расстоянию между компонентами, причем опыты показали, что последнее одно и то же для всех линий бальмеровской серии водорода. В сильных полях (превосходящих примерно 100 000 В/см) были обнаружены отступления от линейной зависимости между расщеплением и напряженностью внешнего поля E_0. Линии атомов или ионов, содержащих *более одного электрона*, дают не линейный, а *квадратичный эффект Штарка*, в котором расщепление пропорционально *квадрату электрического поля* E_0.

Классическая теория не дает объяснения эффекта Штарка, как и всякого явления, связанного с излучением и поглощением спектральных линий. Его теория — существенно квантовая и не может быть изложена в этом томе.

§ 94. Вращение плоскости поляризации

1. Если линейно поляризованный свет проходит через плоскопараллельный слой вещества, то в некоторых случаях плос-

кость поляризации света оказывается повернутой относительно своего исходного положения. Это явление называется *вращением плоскости поляризации* или *оптической активностью*. Если вещество не находится во внешнем магнитном поле, то оптическая активность и вращение плоскости поляризации называются *естественными*. В противоположном случае говорят о *магнитном вращении плоскости поляризации* или *эффекте Фарадея*.

Естественная активность была открыта в 1811 г. Араго на пластинках кварца, вырезанных перпендикулярно к оптической оси. В 1815 г. Био подробно исследовал это явление, а также обнаруженное им вращение плоскости поляризации в растворах сахара. Затем естественное вращение плоскости поляризации было найдено у многих других тел. К концу прошлого века число известных естественно-активных веществ превышало 700. Теперь их известно гораздо больше, хотя для большинства веществ, где обнаружено это явление, оно выражено очень слабо.

Для наблюдения явления можно установить на оптической скамье два скрещенных николя. Такая система не пропускает свет. Однако, если между николями ввести пластинку кварца, вырезанную перпендикулярно к оптической оси, или слой какого-либо другого оптически активного вещества, то свет через систему будет проходить. Но его можно погасить вращением одного из николей. Отсюда следует, что после прохождения через активное вещество свет остается линейно поляризованным, но его плоскость поляризации оказывается повернутой. Для успеха опыта падающий свет, если он белый, необходимо монохроматизировать, пропустив его через светофильтр, так как *угол поворота плоскости поляризации зависит от длины волны*. Кварц — одноосный кристалл. В описанном опыте свет распространяется вдоль оптической оси, когда кварц ведет себя как изотропное тело, не давая обычного (линейного) двойного лучепреломления.

В зависимости от взятого вещества естественное вращение плоскости поляризации может происходить вправо или влево, причем эти два направления условились относить к наблюдателю, к которому свет приближается. В соответствии с этим различают *право-* и *левовращающие вещества*. Вращение вправо считается *положительным*, а влево — *отрицательным*.

2. Явление вращения плоскости поляризации указывает на определенную *дисимметрию*, свойственную оптически активным средам. Она выражается в том, что в таких средах направления вращения по и против часовой стрелки физически не эквивалентны. Поэтому в среде не может быть плоскости симметрии, проходящей через направление нормали к фронту волны. Иначе, как это следует из общих соображений симметрии, плоскость поляризации света не могла бы вращаться, если бы

она совпадала с любой из плоскостей симметрии. В то же время естественно-активные среды, если они жидкие, *полностью изотропны*, т.е. все направления в них совершенно эквивалентны. Это проявляется, в частности, в том, что естественно-активная жидкость вращает плоскость поляризации *в одну и ту же сторону, независимо от направления распространения света.* Поэтому естественно-активную жидкость можно охарактеризовать как *дисимметрично-изотропную среду.* В кристаллах нет изотропии, но в одноосных кристаллах всякие два взаимно противоположные направления оптической оси также эквивалентны, по крайней мере в оптическом отношении.

Отмеченная дисимметрия напоминает дисимметрию винтовой спирали. Будем смотреть на один из торцов спирали. Пусть по спирали движется точка, вращаясь по часовой стрелке. Если при этом точка удаляется от нас, то спираль называется правой (в противоположном случае она называется левой). Если посмотреть на спираль с противоположного торца, то вращение той же точки будет происходить против часовой стрелки, но в этом случае точка будет приближаться к нам. Чтобы она удалялась, надо направление вращения изменить на противоположное. Таким образом, свойство спирали быть правой или левой не зависит от того, с какого торца на нее смотреть. Так и *свойство естественно-активной среды быть право- или левовращающей не зависит от того, в каком из двух прямо противоположных направлений распространяется свет.*

Таким образом, *если плоскость поляризации в естественно-активной среде вращается, например, вправо, то при изменении направления распространения света на противоположное она по-прежнему будет вращаться вправо. Однако направления вращения «вправо» и «влево» относятся к разным наблюдателям: к каждому из них свет должен приближаться. Объективно, независимо от выбора наблюдателя, вращения происходят в противоположные стороны, если лучи распространяются навстречу друг другу.* Если свет заставить пройти туда и обратно через естественно-активное вещество, отразив его от зеркала, то плоскость поляризации возвратится к своему исходному направлению.

3. Кварц встречается в природе в виде двух модификаций: *правовращающего* и *левовращающего* (короче — *правого* и *левого*). Это явление называется *энантиоморфизмом* и встречается у кристаллов, не содержащих центров и плоскостей симметрии. Обе энантиоморфные модификации кристалла отличаются друг от друга внешней формой и внутренней кристаллической структурой. По своей симметрии они отличаются друг от друга примерно так же, как правая спираль отличается от левой или правая рука от левой. Таким образом, обе модификации *не кон-*

груэнтны, т.е. правая не может быть наложена на левую и наоборот. Но зеркальное изображение одной из этих модификаций может быть совмещено с другой. По-видимому, все естественно-активные кристаллы существуют в двух энантиоморфных модификациях, хотя не во всех случаях известны обе модификации. Некоторые жидкости, например винная кислота, могут встречаться также в виде двух модификаций, вращающих плоскость поляризации в противоположных направлениях.

4. Био установил на опыте, что угол поворота χ плоскости поляризации пропорционален толщине l оптически активного вещества: $\chi = \alpha l$, где коэффициент α называется *вращением на единицу длины*. Он зависит от длины волны, природы вещества и температуры. Для кварца при температуре 20 °C и желтого света натрия ($\lambda = 589{,}3$ нм) $\alpha = \pm 21{,}728$, для хлорноватистокислого натрия ($NaClO_3$) $\alpha = 3{,}170$ угловых градуса на миллиметр. Для некоторых жидких кристаллов α может достигать 40 000 градусов на миллиметр. Для правых и левых модификаций кварца и всех остальных кристаллов значения вращения *α одинаковы по величине, но противоположны по знаку*. Вращение α увеличивается с уменьшением длины волны. Био нашел, что величина α обратно пропорциональна квадрату длины волны λ^2. Но такая зависимость грубо приближенна. В области прозрачности и малого поглощения хорошо согласуется с опытом формула Друде

$$\alpha = \sum_i \frac{B_i}{\lambda^2 - \lambda_i^2},$$

где B_i — постоянные, а λ_i — длины волн, соответствующие собственным частотам рассматриваемого вещества.

Оптическую активность жидкостей принято характеризовать *удельным вращением* $[\alpha]$. В случае химически чистых жидкостей это есть отношение вращения α к плотности жидкости ρ, т.е. $[\alpha] = \alpha/\rho$. В случае растворов речь идет обычно о растворах *активного вещества* в *неактивном растворителе*. В этом случае при определении удельного вращения $[\alpha]$ под ρ понимают *плотность активного вещества*, т.е. массу его в единице объема раствора. *Молекулярным вращением* жидкости или раствора называют произведение удельного вращения $[\alpha]$ на молекулярную массу активного вещества $[M] = M[\alpha]$. Зависимость вращения плоскости поляризации от концентрации активного вещества используется в *сахарометрии* для определения содержания сахара в растворе.

Вращение плоскости поляризации в одноосных кристаллах имеет место и в тех случаях, когда свет распространяется *под углом к оптической оси*. Но в этих случаях явление осложняется

обычным двойным лучепреломлением. Еще больше усложнение в оптически двуосных кристаллах, так как вращение может быть разным вдоль каждой из оптических осей. Наконец, встречаются естественно-активные кристаллы кубической системы, не обнаруживающие обычного двойного лучепреломления (хлорноватистокислый натрий $NaClO_3$, бромноватистокислый натрий $NaBrO_3$). Для них вращение α не зависит от ориентации кристалла.

5. Согласно Френелю, вращение плоскости поляризации, какова бы ни была его физическая природа, есть проявление особого, так называемого *кругового двойного лучепреломления.* Однако две волны, которые могут распространяться в оптически активной среде с определенными, но различными скоростями, поляризованы не линейно, как при обычном (линейном) двойном лучепреломлении, а *по кругу*: *одна по левому, другая по правому.*

Для обоснования достаточности такого предположения допустим, что плоская волна с вращающейся плоскостью поляризации распространяется в положительном направлении оси Z. Так как волна поперечна, ее электрический вектор $\mathbf{E}$ будет содержать только две компоненты E_x и E_y. Если A — амплитуда колебаний, то $E_x = A\cos\chi\cos(\omega t - kz)$, $E_y = A\sin\chi\cos(\omega t - kz)$, где χ — угол между направлением электрического вектора $\mathbf{E}$ и осью X. Этот угол может зависеть от координаты z, так что можно положить $\chi = -\alpha z$, где α — введенный ранее коэффициент, определяющий вращение плоскости поляризации. Когда коэффициент α положителен, то при распространении волны угол χ уменьшается, т.е. вращение происходит *по часовой стрелке* (вправо). При отрицательном α плоскость поляризации вращается *против часовой стрелки* (влево). Подставив значение $\chi = -\alpha z$ в выражения для E_x и E_y и воспользовавшись известными формулами тригонометрии, представим их в виде

$$E_x = \frac{A}{2}\cos(\omega t - kz + \alpha z) + \frac{A}{2}\cos(\omega t - kz - \alpha z),$$

$$E_y = \frac{A}{2}\cos\left(\omega t - kz + \alpha z + \frac{\pi}{2}\right) + \frac{A}{2}\cos\left(\omega t - kz - \alpha z - \frac{\pi}{2}\right).$$

Разложим полученную волну на две: $\mathbf{E} = \mathbf{E}^{\text{п}} + \mathbf{E}^{\text{л}}$, где компоненты векторов $\mathbf{E}^{\text{п}}$ и $\mathbf{E}^{\text{л}}$ определяются выражениями

(94.1)

(94.2)

и введены обозначения

$$k^{\text{п}} = k - \alpha, \quad k^{\text{л}} = k + \alpha. \tag{94.3}$$

В волне (94.1) компонента E_y опережает по фазе компоненту E_x на $\pi/2$, а в волне (94.2) отстает на такую же величину. Обе волны поляризованы по кругу: в волне (94.1) электрический вектор вращается от положительного конца оси Y к положительному концу оси X, в волне (94.2) вращение происходит в противоположном направлении (см. рис. 314). Иначе говоря, волна (94.1) поляризована по правому, а волна (94.2) — по левому кругу. Таким образом, *кинематически волну с вращающейся плоскостью поляризации можно рассматривать как наложение двух волн, поляризованных по кругу: правому и левому.* Скорости этих волн определяются выражениями

$$v^{\text{п}} = \frac{\omega}{k-\alpha}, \quad v^{\text{л}} = \frac{\omega}{k+\alpha}, \tag{94.4}$$

а показатели преломления — выражениями

$$n^{\text{п}} = \frac{c}{v^{\text{п}}} = \frac{c}{\omega}(k-\alpha), \quad n^{\text{л}} = \frac{c}{v^{\text{л}}} = \frac{c}{\omega}(k+\alpha). \tag{94.5}$$

Отсюда

$$n^{\text{л}} - n^{\text{п}} = \frac{2c}{\omega}\,\alpha, \quad \alpha = \frac{\omega}{2c}(n^{\text{л}} - n^{\text{п}}). \tag{94.6}$$

Если $n^{\text{л}} - n^{\text{п}} > 0$, то $\alpha > 0$, т.е. вращение плоскости поляризации происходит *вправо*. Если же $n^{\text{л}} - n^{\text{п}} < 0$, то плоскость поляризации вращается *влево*. Иначе говоря, *плоскость поляризации вращается в ту же сторону, что и электрический вектор поляризованной по кругу волны с меньшим показателем преломления, т.е. с большей фазовой скоростью.*

6. Приведенное рассуждение отнюдь не доказывает, что каждая из поляризованных по кругу волн (94.1) и (94.2) может в отдельности существовать в среде. Мы исходили из опытного факта, что в оптически активной среде может реально существовать волна с вращающейся плоскостью поляризации. Такая волна, конечно, должна быть *решением системы фундаментальных уравнений Максвелла*, дополненной материальными уравнениями в оптически активной среде. Должна удовлетворять этой системе уравнений и суперпозиция поляризованных по кругу волн (94.1) и (94.2), так как мы доказали, что такая суперпозиция дает волну с вращающейся плоскостью поляризации. Но любое решение всякой системы уравнений можно представить (и притом бесконечным числом способов) в виде суммы нескольких слагаемых, которые *вовсе не обязательно должны быть решениями той же системы.* Поэтому из того факта, что в оптически активной среде возможна волна с вращающейся плоскостью поляризации, еще не следует, что в ней возможны и одиночные волны с круговой поляризацией. Однако *Френель предположил, что поляризованные по кругу волны со скоростями* (94.4) *действительно могут распространяться в оптически активной среде.*

Их можно назвать *нормальными волнами*, т.е. такими волнами, которые распространяются в среде с сохранением своей формы и характера поляризации. Это предположение есть *гипотеза*, которую теоретически Френель доказать не мог, так как для этого необходимо было бы располагать полной системой уравнений волновой теории света в оптически активных средах. Но Френель подтвердил свою гипотезу *экспериментально*.

Френель изготовил сложную призму, состоящую из трех кварцевых призм: двух призм ABM и DCM из правовращающего кварца и одной AMD из левовращающего с тупым углом $M = 152°$ (рис. 316). Оптические оси всех трех призм были параллельны основанию AD. Падающий луч, параллельный оптической оси, на грани AB не испытывал преломления, но разлагался внутри призмы ABM на два луча, поляризованные по кругу. Луч с правой поляризацией, как показывают формулы (94.5), имел меньший, а луч с левой поляризацией — больший показатель преломления ($n^{\text{п}} < n^{\text{л}}$). В левовращающей призме AMD соотношение между этими показателями преломления было обратным ($n^{\text{п}} > n^{\text{л}}$). Поэтому на грани AM лучи испытывали разное преломление: правый луч приближался к основанию AD, а левый удалялся. Угол расхождения между лучами еще больше увеличивался при преломлении на гранях DM и CD. В результате из призмы выходили два луча: один отклонялся вниз, а другой вверх, угол расхождения между которыми составлял около $4'$ (см. задачу к этому параграфу). Исследование с помощью параллелепипеда Френеля или пластинки $\lambda/4$ показало, что луч, отклонившийся вниз, был поляризован по правому, а вверх — по левому кругу.

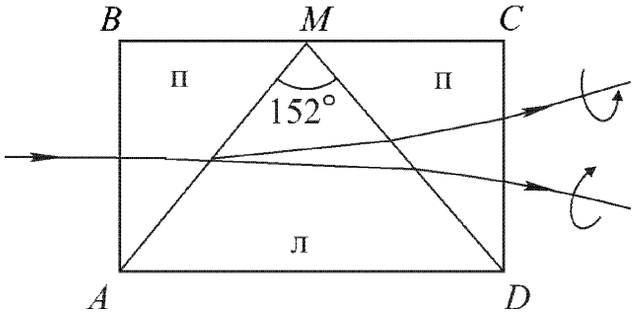


Рис. 316

Френель предлагал проверить свою гипотезу и для жидкостей путем комбинации большого ряда призм, попеременно наполненных жидкостями, вращающими плоскость поляризации вправо и влево. Позднее такой опыт был осуществлен и показал, что и в оптически активных жидкостях также могут существовать две волны с различной круговой поляризацией: правой и левой.

Существование двойного кругового лучепреломления в кварце можно обнаружить и с помощью одной призмы, вырезанной так, что ее оптическая ось перпендикулярна к плоскости, делящей пополам двугранный преломляющий угол. При наимень-

шем отклонении луч внутри призмы идет параллельно оптической оси, не испытывая обычного двойного преломления. Если падающий свет естественный или поляризован линейно и содержит только одну спектральную линию, то в минимуме отклонения при выходе из призмы спектральная линия расщепляется на две линии, из которых одна поляризована по правому, а другая по левому кругу.

Таким образом, Френель доказал экспериментально, что *при вступлении в оптически активную среду луч света испытывает двойное круговое лучепреломление: лучи, поляризованные по правому и левому кругу, идут внутри оптически активной среды с различными фазовыми скоростями. Если падающий свет был поляризован линейно, то при выходе из такой среды эти волны складываются снова в линейно поляризованную волну, но с повернутой плоскостью поляризации.* Тем самым задача объяснения вращения плоскости поляризации была сведена к задаче объяснения *кругового двойного лучепреломления.*

Поляризованные по кругу в противоположных направлениях волны в окрестности полос или линий поглощения могут отличаться не только скоростями распространения, но и коэффициентами поглощения. Тогда они выйдут с различными амплитудами. Если падающий свет был поляризован линейно, то выходящий будет поляризован *эллиптически.* Это явление называется *круговым дихроизмом.* Оно впервые (1847 г.) было замечено Гайдингером (1795–1871) на аметисте и подробно изучено (1895 г.) Коттоном (1869–1951) на растворах солей винной кислоты.

ЗАДАЧА

Пользуясь данными, приведенными в тексте, вычислить угол расхождения $\delta\varphi$ между лучами, поляризованными по правому и левому кругу, вышедшими из призмы $ABCD$ в опыте Френеля (рис. 316). Зная обыкновенный показатель преломления кварца ($n_o = 1{,}544$ для $\lambda = 589{,}3$ нм), найти угловое расхождение между теми же лучами $\delta_1\varphi$ после преломления на плоскости AM и $\delta_2\varphi$ после последующего преломления на плоскости MD.

О т в е т. $\delta\varphi = \dfrac{4\lambda\alpha}{\pi}\operatorname{tg}\dfrac{M}{2} = 3{,}92'$; $\delta_1\varphi = \dfrac{\delta\varphi}{2n_o} = 1{,}27'$; $\delta_2\varphi = \dfrac{\delta\varphi}{n_o} = 2{,}54'$.

§ 95. Магнитное вращение плоскости поляризации

1. В 1846 г. Фарадей обнаружил, что естественно-неактивное вещество, помещенное в постоянное сильное магнитное поле, вращает плоскость поляризации, когда свет распространяется вдоль магнитного поля. Для наблюдения явления можно воспользоваться электромагнитом, вдоль оси которого просвер-

лен цилиндрический канал, как это показано схематически на рис. 312. Электромагнит помещается между скрещенными николями, исследуемое вещество вводится между полюсами электромагнита. Если наблюдатель смотрит в направлении магнитного поля, то вращение вправо условились считать *положительным*, а влево — *отрицательным*. В первом случае вращение происходит в ту же сторону, куда течет ток по виткам электромагнита, во втором — в противоположную сторону. Фарадей наблюдал только положительное вращение. Таким вращением обладает большинство веществ. Из диамагнитных веществ единственным исключением является хлористый титан. Среди парамагнитных веществ встречается значительное число отрицательно вращающих, например соли железа и редких земель, $K_2Cr_2O_7$, $KCrO_4$ и др.

Опыты самого Фарадея, а затем более точные опыты Верде (1824–1866) показали, что *угол поворота плоскости поляризации пропорционален длине пути l света в веществе и напряженности внешнего магнитного поля, т.е. магнитной индукции B в веществе*:

$$\chi = RlB, \tag{95.1}$$

где коэффициент R называется *постоянной Верде* или *магнитной вращательной способностью*. Этот коэффициент зависит от рода вещества, его физического состояния и длины волны. В табл. 10 приведены характеристики магнитного вращения для некоторых веществ.

Таблица 10

Вещество	χ			
	$\lambda = 656$ нм	$\lambda = 589$ нм	$\lambda = 486$ нм	
Вода (25 °C)	0,0102′	0,0130′	0,0197′	На 1 см длины луча в поле 1 Гс
Сероуглерод (25 °C)	0,0319′	0,0415′	0,0667′	
Кварц	0,0136′	0,0166′	0,0250′	
Кислород	0,484′	0,559′	0,721′	При давл. 1 ат на 1 см длины луча в поле 10^5 Гс
Водород	0,430′	0,537′	0,805′	
Углекислота	0,691′	0,862′	0,286′	
Железо	217°	195°	145°	На 10^{-3} см в поле 15 000 Гс
Никель	92°	75°	64°	

В ферромагнитных металлах угол поворота χ не пропорционален B, но в сильных магнитных полях с возрастанием B возрастает очень медленно, приближаясь асимптотически к предельному значению. Согласно Кундту, угол χ пропорционален

намагниченности магнетика $I = (\mu - 1)B/(4\pi\mu)$, где μ — магнитная проницаемость. Этот угол определяется выражением

$$\chi = KlI, \tag{95.2}$$

где K — так называемая *постоянная Кундта*. При $B = 15\,000$ Гс ферромагнетик можно считать намагниченным до насыщения.

2. Физическая природа магнитного вращения плоскости поляризации была выяснена после открытия явления Зеемана и объяснения его, данного Лоренцом. Допустим для простоты, что в отсутствие магнитного поля вещество испускает *только одну спектральную линию*, т.е. обладает *одной собственной (резонансной) частотой* ω_0. При внесении в магнитное поле $\mathbf{B}$ у осцилляторов вещества появляются две новые резонансные частоты $\omega_0 + \Omega$ и $\omega_0 - \Omega$ (где Ω — ларморовская частота), соответствующие круговым вращениям электрона (см. § 92). Эти собственные частоты проявляются не только в *испускании* (*прямой эффект Зеемана*), но и в *поглощении света* (*обратный эффект Зеемана*). Прежняя резонансная частота ω_0 проявляется только при распространении света поперек магнитного поля и под углом к нему. Нормальные волны, которые могут распространяться вдоль магнитного поля, поляризованы по кругу. Когда направления распространения света и магнитного поля совпадают, большей собственной частоте $\omega_+ = \omega_0 + \Omega$ соответствует вращение по, а меньшей $\omega_- = \omega_0 - \Omega$ — против часовой стрелки, если смотреть в направлении магнитного поля (см. рис. 313).

Поскольку собственные частоты ω_+ и ω_- различны, скорости распространения обеих поляризованных по кругу волн также различны. Этим принципиально дано объяснение магнитного вращения плоскости поляризации, так как, согласно Френелю, при сложении таких поляризованных по кругу волн получается волна с вращающейся плоскостью поляризации. Угол поворота плоскости поляризации найдется из формулы (94.6). Она дает

$$\chi = \frac{\omega l}{2c}(n_- - n_+) = \frac{\pi l}{\lambda}(n_- - n_+), \tag{95.3}$$

где через n_+ и n_- обозначены показатели преломления соответственно для волн с собственными частотами ω_+ и ω_-. При выборе знака учтено, что положительные вращения плоскости поляризации в случае естественного и магнитного вращения в рассмотренном нами случае принято выбирать противоположно.

Формулу (95.3) удобно исследовать графически. На рис. 317 штриховыми линиями представлены графики для показателей преломления n_- и n_+ как функций частоты света ω. Их разность представлена сплошной линией. Как видно из рисунка, вне интервала (ω_-, ω_+) разность $n_- - n_+$, а с ней и вращение плоскости поляризации *положительны*, а внутри этого интервала — *отрицательны*.

Во всех случаях направления ларморовского вращения Ω, а следовательно, и вращения плоскости поляризации определяются *только направлением магнитного поля* $\mathbf{B}$ *и не зависят от направления распространения света.* При изменении на противоположное направления распространения света направление вращения плоскости поляризации *не изменяется*, в противо-

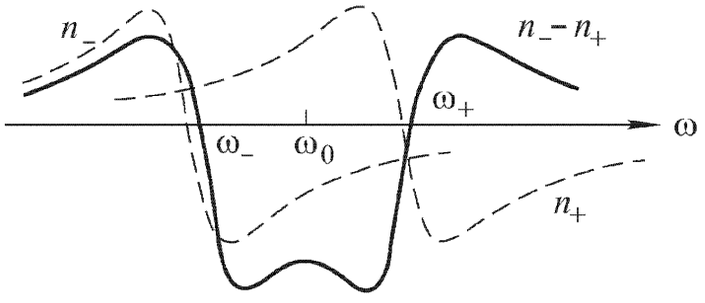


Рис. 317

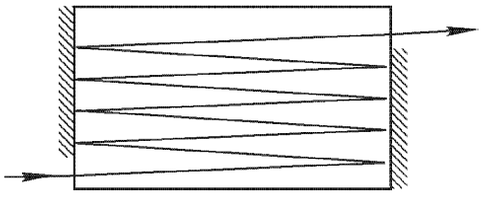


Рис. 318

положность тому, что происходит при естественном вращении. Этим воспользовался Фарадей для усиления магнитного вращения плоскости поляризации. Он увеличил расстояние l, проходимое светом в веществе, заставив луч многократно отразиться от концов исследуемого образца (рис. 318). Для этого концы образца серебрились (за исключением мест входа и выхода светового луча).

3. Вне интервала (ω_-, ω_+) вдали от резонансных частот ω_- и ω_+ поглощением можно пренебречь и воспользоваться для показателей преломления n_- и n_+ формулой Зельмейера (84.9), заменив в ней ω_0 на ω_- и ω_+ для показателей преломления n_- и n_+ соответственно. Согласно этой формуле, n_- и n_+ являются *одной и той же функцией* аргументов $(\omega^2 - \omega_-^2)$ и $(\omega^2 - \omega_+^2)$. Такой же функцией, но от аргумента $(\omega^2 - \omega_0^2)$ выражается и показатель преломления n вещества в отсутствие магнитного поля. Следовательно,

$$n_- - n_+ = n(\omega^2 - \omega_-^2) - n(\omega^2 - \omega_+^2) = \\ = (\omega_+^2 - \omega_-^2)\frac{dn}{d\omega^2} = \frac{(\omega_+ + \omega_-)(\omega_+ - \omega_-)}{2\omega}\frac{dn}{d\omega}.$$

Вращение плоскости поляризации имеет заметную величину только не очень далеко от полосы поглощения, а потому можно принять, что $\omega_+ + \omega_- = 2\omega$. Кроме того, $\omega_+ - \omega_- = 2\Omega = eB/(mc)$. Следовательно,

$$n_- - n_+ = \frac{eB}{mc}\frac{dn}{d\omega} = -\frac{eB}{2\pi mc^2}\lambda^2\frac{dn}{d\lambda}.$$

Подстановка этого значения в (95.3) приводит к формуле (95.1),

причем для постоянной Верде получается выражение

$$R = -\frac{e}{2mc^2}\lambda\frac{dn}{d\lambda}. \tag{95.4}$$

Нелишне напомнить, что здесь e означает заряд электрона *по абсолютной величине*. Формулу (95.4) обычно называют *формулой Беккереля* (1852–1908).

4. Перейдем теперь к исследованию магнитного вращения плоскости поляризации *в области полосы поглощения вещества*. Здесь показатель преломления, а также коэффициент поглощения меняются с частотой *быстро* и *аномально*. Поэтому магнитное вращение плоскости поляризации велико и носит также *аномальный характер*. Кроме того, внутри самой полосы поглощения между резонансными частотами ω_- и ω_+ знак вращения *отрицателен*, как это видно из рис. 317. Такое аномальное поведение плоскости поляризации называется *эффектом Макалюзо и Корбино*. Эти итальянские физики открыли его экспериментально в 1898 г., а затем Фохт дал теорию явления.

Макалюзо и Корбино показали, что пары натрия в магнитном поле вызывают очень сильное вращение плоскости поляризации в непосредственной близости к желтым линиям D_1 и D_2. Знак вращения одинаков по обе стороны этой двойной линии. Между полюсами электромагнита помещалось небольшое натровое пламя, через которое проходил пучок линейно поляризованного белого света в направлении магнитного поля. На пути лучей ставился николь-анализатор, который при выключенном магнитном поле полностью тушил свет. При включении тока в обмотке электромагнита поле зрения становилось освещенным ярким желтым светом: в спектроскопе были видны две желтые линии, симметрично расположенные по обе стороны линии D. Дело в том, что свет с такими длинами волн испытывал в магнитном поле поворот плоскости поляризации на 90° и потому проходил через скрещенные поляризатор и анализатор. Поворачивая николь-анализатор сначала в одном направлении, а затем в другом, можно было убедиться, что по обе стороны полосы поглощения знак вращения был один и тот же.

Вуд усовершенствовал установку Макалюзо и Корбино, произведя опыты с парами натрия, заключенными в стальной трубе. В его опытах плотность паров натрия можно было варьировать в значительных пределах. При сравнительно малой плотности наблюдалась такая же картина, как и в опытах Макалюзо и Корбино. Разница состояла в том, что в опытах Макалюзо и Корбино двойная D-линия натрия не разрешалась, а в опытах Вуда разрешалась. Желтые линии появлялись по обе стороны как компоненты D_1, так и компоненты D_2. При увеличении плотности паров натрия расстояние между наружными желты-

ми линиями увеличивалось и между ними появлялись новые линии, соответствующие вращению на 270°, 450°, 630° и т.д. Центр между линиями D_1 и D_2 последовательно становился то темным, то светлым. В некоторых случаях Вуд наблюдал до восьми таких перемен, что соответствует вращению на 1440°.

5. Изложенная теория основана на *классической теории*, т.е. *на простом эффекте Зеемана*. Полная теория должна учитывать *сложный эффект Зеемана*, т.е. быть *квантовой*. Кроме того, для объяснения отрицательного вращения плоскости поляризации необходимо учитывать *парамагнетизм атомов*. На этих вопросах мы останавливаться не можем.

ЗАДАЧИ

1. Между двумя николями N_1 и N_2 (рис. 319), главные сечения которых повернуты относительно друг друга на 45°, в продольном магнитном поле помещен слой вещества, вращающий плоскость поляризации. На какой угол

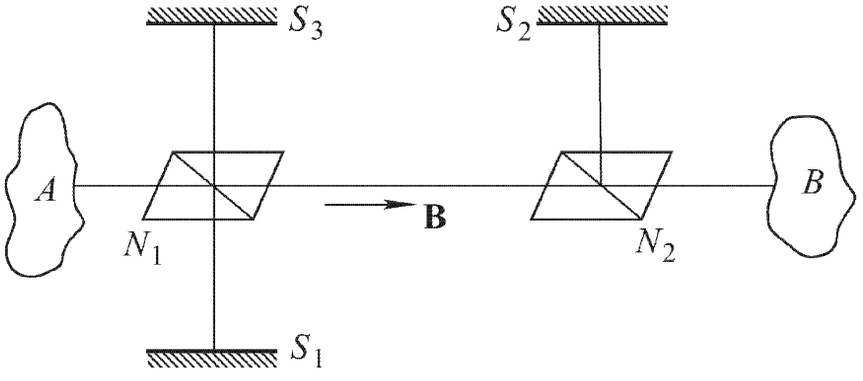


Рис. 319

этот слой должен вращать плоскость поляризации, чтобы система была *оптическим вентилем*, т.е. свет проходил бы через нее только в одном направлении (например, от тела A к телу B), а в обратном направлении не проходил?

О т в е т. На угол $45° + 180°m$ (m — целое число).

2. Как будет вести себя оптический вентиль, описанный в предыдущей задаче, если направление магнитного поля изменить на противоположное, оставляя без изменения взаимное расположение николей и величину напряженности магнитного поля?

О т в е т. Направление, в котором пропускается свет, изменится на противоположное.

3. Может ли установка, описанная в задаче 1, быть оптическим вентилем, если между николями поместить слой *естественно-активного вещества без магнитного поля*?

О т в е т. Нет.

4. Разъяснить парадокс Вина, заключающийся в следующем. В адиабатической оболочке помещены два одинаково нагретых тела A и B (рис. 319), а между ними световой вентиль. Свет, излученный телом A, падает на николь N_1. Половина падающей энергии I, т.е. $I/2$, проходит через николь,

а другая половина, испытав полное отражение в николе, возвращается зеркалом S_1 к телу A (применяя зеркало, можно не учитывать нагревание николя). Другая половина энергии $I/2$, пройдя через николь N_1, полностью пройдет через николь N_2 и поглотится телом B. Свет же от тела B не пройдет через вентиль. Половина излученной этим телом энергии I, равная $I/2$, испытает полное отражение в николе N_2 и после отражения от зеркала S_2 вернется к телу B. Другая половина энергии, пройдя через николь N_1 и вращающую среду, испытает поворот плоскости поляризации на 45° и поэтому не сможет пройти через николь N_1, а претерпит в нем полное отражение. Добавочным зеркалом S_3 эта часть энергии будет отражена обратно и после полного отражения в николе N_1 вернется к телу B. Таким образом, тело A получит энергию $I/2$, а тело B — энергию $3I/2$ и будет нагреваться в противоречии со вторым началом термодинамики.

Р е ш е н и е. Свет, излученный телом B (см. рис. 319) и возвращающийся к нему обратно после отражения от николя N_1, испытает при прохождении через вращающее вещество дополнительный поворот плоскости поляризации на 45° и поэтому не пройдет через николь N_2. Испытав в николе N_2 полное отражение, этот свет вернется (если поставить еще одно зеркало напротив S_2) к телу A. Приведенное решение парадокса Вина было дано Рэлеем.

§ 96. Временная и пространственная дисперсия. Теория естественной оптической активности

1. Многие оптические явления находят удовлетворительное объяснение в предположении, что связь между векторами $\mathbf{D}$ и $\mathbf{E}$ (а также между $\mathbf{B}$ и $\mathbf{H}$) *локальна во времени и пространстве*. Это значит, что вектор $\mathbf{D}$ в любой точке пространства $\mathbf{r}$ и в любой момент времени t определяется значением вектора $\mathbf{E}$ *в той же точке и в тот же момент времени*. (То же относится к векторам $\mathbf{B}$ и $\mathbf{H}$. В целях сокращения подобные замечания в дальнейшем подразумеваются, а все изложение ведется для векторов $\mathbf{D}$ и $\mathbf{E}$.) Однако для истолкования некоторых явлений предположения о локальной связи *недостаточно*. Пространственно-временную нелокальность можно разбить на *чисто пространственную* и *чисто временную*. Отвлечемся сначала от чисто пространственной нелокальности и учтем нелокальность временную. Среда во всем дальнейшем предполагается *однородной*.

2. Временная нелокальность проявляется в *оптической дисперсии среды*. Действительно, для истолкования дисперсии необходимо учитывать инерционные свойства электронов, атомных ядер и ионов вещества. А такая инерционность и приводит к нелокальной связи по времени между $\mathbf{D}$ и $\mathbf{E}$. Пренебрегая пространственной нелокальностью, будем рассуждать так, как если бы во всем пространстве электромагнитное поле было *однородно* и менялось *только во времени*. Действием магнитно-

го поля будем пренебрегать. Рассмотрим сначала *импульсные воздействия* на среду. Допустим, что от момента $t = 0$ в течение времени dt среда подверглась воздействию электрического поля $\mathbf{E}$, а по истечении этого промежутка поле в среде снова обратилось в нуль. Такое воздействие может рассматриваться как *электрический толчок*, возбуждающий регулярные колебания электронов, атомных ядер и ионов среды. Эти колебания не прекращаются после прекращения действия поля, но продолжаются дальше в силу *конечности масс* колеблющихся частиц. В этом и проявляется временная нелокальность связи между отклонением частиц из положений равновесия и электрическим полем, а следовательно, между поляризацией $\mathbf{P}$ или индукцией $\mathbf{D}$ среды и полем $\mathbf{E}$.

В рамках *линейной электродинамики* поляризация среды $d\mathbf{P}$, вызванная электрическим толчком, пропорциональна $\mathbf{E}\,dt$. Предполагая, что среда *однородна* и *изотропна*, для вектора $d\mathbf{P}$ в момент времени t можно написать

$$d\mathbf{P}(t) = f(t)\mathbf{E}(0)\,dt, \tag{96.1}$$

где функция $f(t)$ зависит только от свойств среды и от времени t, которое прошло с момента действия толчка до момента наблюдения. Функция $f(t)$ должна обращаться в нуль при $t = 0$, так как из-за своей инерционности электроны, атомные ядра и ионы не могут мгновенно получить конечные смещения. Она должна обращаться в нуль и при $t = \infty$, так как все реальные среды *диссипативны*, так что всякое свободное колебание в них должно в конце концов затухнуть.

Если поле $\mathbf{E}$ действует в течение длительного промежутка времени, то этот промежуток можно разбить на бесконечно малые промежутки и таким путем свести воздействие электрического поля на среду к действию последовательных толчков. Вклад в поляризацию среды в момент времени t, внесенный более ранним электрическим толчком $\mathbf{E}(t')\,dt'$, будет $d\mathbf{P}(t) = f(t - t')\mathbf{E}(t')\,dt'$. В линейной электродинамике справедлив *принцип суперпозиции*, а потому полный вектор поляризации в момент времени t будет

$$\mathbf{P}(t) = \int\limits_{-\infty}^{t} f(t - t')\mathbf{E}(t')\,dt', \tag{96.2}$$

или, вводя новую переменную интегрирования $\vartheta = t - t'$,

$$\mathbf{P}(t) = \int\limits_{0}^{\infty} f(\vartheta)\mathbf{E}(t - \vartheta)\,d\vartheta. \tag{96.3}$$

Следовательно,

$$\mathbf{D}(t) = \mathbf{E}(t) + 4\pi \int_0^\infty f(\vartheta)\mathbf{E}(t-\vartheta)\,d\vartheta. \tag{96.4}$$

Интегрирование производится по времени, *предшествующему* рассматриваемому моменту t. Этого требует *принцип причинности*. В своей нерелятивистской форме он означает, что *каждое событие определяется только прошедшими событиями, но не может зависеть от будущих.*

Если $\mathbf{E} = \mathbf{E}_0 \exp(i\omega t)$, то после подстановки этого выражения в (96.4) получится

$$\mathbf{D}(t) = \varepsilon(\omega)\mathbf{E}(t), \tag{96.5}$$

где

$$\varepsilon(\omega) = 1 + 4\pi \int_0^\infty f(\vartheta)e^{-i\omega\vartheta}\,d\vartheta. \tag{96.6}$$

Таким образом, для монохроматического поля связь между $\mathbf{D}$ и $\mathbf{E}$ формально может быть записана в локальной форме с помощью диэлектрической проницаемости как функции частоты ω. Оптическая дисперсия, т.е. зависимость ε от ω, эквивалентна *нелокальной связи по времени между* $\mathbf{D}(t)$ и $\mathbf{E}(t)$. Поэтому такую дисперсию называют *временной* или *частотной дисперсией*, в отличие от *пространственной дисперсии*, о которой говорится ниже.

3. Учтем теперь *пространственную нелокальность*. Если электрическое поле *неоднородно*, то для нахождения индуцированного дипольного момента молекулы недостаточно знать вектор $\mathbf{E}$ в одной точке пространства, а требуется знание функции $\mathbf{E}(\mathbf{r})$ *во всем объеме, занимаемом молекулой*. Это эквивалентно знанию поля $\mathbf{E}$ и его пространственных производных всех порядков в какой-то одной точке внутри молекулы (которую условно можно назвать *центром молекулы*), так как тогда функция $\mathbf{E}(\mathbf{r})$ представится рядом Тэйлора по координатам x_1, x_2, x_3 вектора $\mathbf{r}$.

Легко оценить порядок последовательных членов этого ряда в монохроматическом поле световой волны. Производная $\partial E_l/\partial x_m$ будет порядка E/λ, а координата x_m — порядка размеров молекулы a, так что член $x_m \cdot \partial E_l/\partial x_m$ будет порядка $(a/\lambda)E$. Так же оцениваются и порядки членов, содержащих высшие производные. Если нулевой член ряда принять за единицу, то члены, содержащие первые, вторые и последующие производные, будут порядка a/λ, $(a/\lambda)^2$, $(a/\lambda)^3$ и т.д.

Как видно, в неоднородном поле связь между индуцированным дипольным моментом молекулы и электрическим полем $\mathbf{E}$

пространственно нелокальна. Это ведет к *пространственной нелокальности связи между* $\mathbf{P}$ *и* $\mathbf{E}$, *а также между* $\mathbf{D}$ *и* $\mathbf{E}$. При этом в слабых электрических полях, с которыми имеют дело линейная электродинамика и оптика, можно пренебречь влиянием квадратов и высших степеней поля $\mathbf{E}$ и его пространственных производных. В этом «линейном приближении» в монохроматическом поле световой волны можно написать

$$D_j = \varepsilon_{jl} E_l + \gamma_{jlm} \frac{\partial E_l}{\partial x_m} + \alpha_{jlmn} \frac{\partial^2 E_l}{\partial x_m \, \partial x_n} + \ldots \tag{96.7}$$

Для общности предполагается, что среда *анизотропна.* Ее оптические свойства характеризуются тензорами ε_{jl}, γ_{jlm}, ..., являющимися функциями частоты ω. В соответствии с общепринятой тензорной символикой по дважды встречающимся координатным индексам подразумевается суммирование. Как выяснено выше, если нулевой член ряда (96.7) принять за единицу, то последующие члены будут порядка a/λ, $(a/\lambda)^2$ и т.д.

В случае плоской монохроматической волны

$$\mathbf{E} = \mathbf{E}_0 \exp\left[i(\omega t - \mathbf{kr})\right]$$

дифференцирование по координате x_m сводится к умножению на $-ik_m$. Поэтому соотношение (96.7) можно записать в виде

$$D_j(\mathbf{r}, t) = \varepsilon_{jl}(\omega, \mathbf{k}) E_l(\mathbf{r}, t), \tag{96.8}$$

где введено обозначение

$$\varepsilon_{jl}(\omega, \mathbf{k}) = \varepsilon_{jl}(\omega) - ik_m \gamma_{jlm} + (-ik_m)(-ik_n)\alpha_{jlmn} + \ldots \tag{96.9}$$

Таким образом, *в поле плоской монохроматической волны связь между* $\mathbf{D}$ *и* $\mathbf{E}$ *опять принимает формально локальный характер.* Однако тензор диэлектрической проницаемости $\varepsilon_{jl}(\omega, \mathbf{k})$ теперь зависит не только от ω, но и от $\mathbf{k}$. Зависимость этого тензора от волнового вектора $\mathbf{k}$ называется *пространственной дисперсией.*

4. Ввиду малости параметра a/λ эффекты пространственной дисперсии в оптике малы и трудно наблюдаемы. Долгое время единственно известным из таких эффектов было *естественное вращение плоскости поляризации.* Оно наблюдается в таких средах, у которых тензор γ_{jlm} отличен от нуля.

Найдем вид тензора γ_{jlm} для дисимметрично изотропных сред — жидкостей и кристаллов кубической системы, у которых тензор $\varepsilon_{jl}(\omega)$ вырождается в скаляр. Отбросив в выражении (96.7) все члены, содержащие тензоры четвертого и высших порядков, напишем

$$D_j(\mathbf{r}, t) = \varepsilon E_j(\mathbf{r}, t) + \gamma_{jlm} \frac{\partial E_l}{\partial x_m}. \tag{96.10}$$

В развернутом виде для x-составляющей вектора $\mathbf{D}$ это соотношение гласит

$$D_x = \varepsilon E_x + \gamma_{xxx}\frac{\partial E_x}{\partial x} + \gamma_{xxy}\frac{\partial E_x}{\partial y} + \gamma_{xxz}\frac{\partial E_x}{\partial z} + \gamma_{xyx}\frac{\partial E_y}{\partial x} + \\ + \gamma_{xyy}\frac{\partial E_y}{\partial y} + \gamma_{xyz}\frac{\partial E_y}{\partial z} + \gamma_{xzx}\frac{\partial E_z}{\partial x} + \gamma_{xzy}\frac{\partial E_z}{\partial y} + \gamma_{xzz}\frac{\partial E_z}{\partial z}. \quad (96.11)$$

Повернем теперь координатную систему вокруг оси X на угол 90° (рис. 320). Ввиду изотропии среды, все коэффициенты в предыдущем соотношении останутся неизменными. Однако y и E_y перейдут в $-z$ и $-E_z$, а z и E_z — в y и E_y. Остальные координаты и компоненты векторов $\mathbf{E}$ и $\mathbf{D}$ останутся неизменными. Произведя в (96.11) соответствующую замену, получим

$$D_x = \varepsilon E_x + \gamma_{xxx}\frac{\partial E_x}{\partial x} - \gamma_{xxy}\frac{\partial E_x}{\partial z} + \gamma_{xxz}\frac{\partial E_x}{\partial y} - \gamma_{xyx}\frac{\partial E_z}{\partial x} + \\ + \gamma_{xyy}\frac{\partial E_z}{\partial z} - \gamma_{xyz}\frac{\partial E_z}{\partial y} + \gamma_{xzx}\frac{\partial E_y}{\partial x} - \gamma_{xzy}\frac{\partial E_y}{\partial z} + \gamma_{xzz}\frac{\partial E_y}{\partial y}. \quad (96.11a)$$

Но в силу определения тензора γ_{jlm} в повернутой системе координат соотношение между D_x и компонентами вектора $\mathbf{E}$ можно записать в прежней форме (96.11). Сравнение обоих соотношений дает $\gamma_{xxy} = \gamma_{xxz}$, $\gamma_{xxz} = -\gamma_{xxy}$, откуда $\gamma_{xxy} = \gamma_{xxz} = 0$ и т.д. Кроме того, $\gamma_{xyz} = -\gamma_{xzy}$ и т.д.

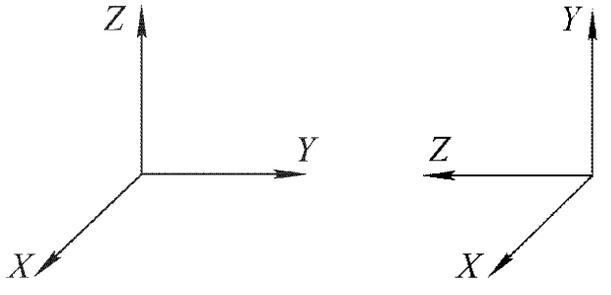


Рис. 320

Что касается коэффициента γ_{xxx}, то он равен нулю. Действительно, в силу симметрии $\gamma_{xxx} = \gamma_{yyy} = \gamma_{zzz}$. Повернем исходную систему координат вокруг оси Z на угол 90°, чтобы ось Y приняла отрицательное направление прежней оси X. Тогда в (96.11), оставляя коэффициенты неизменными, следует сделать замену $x \to -y$, $D_x \to -D_y$, $E_x \to -E_y$, что дает

$$-D_y = -\varepsilon E_y + \gamma_{xxx}\frac{\partial E_y}{\partial y} + \ldots$$

С другой стороны, на основании определения тензора γ_{jlm} в повернутой системе можно сразу написать

$$D_y = \varepsilon E_y + \gamma_{yyy}\frac{\partial E_y}{\partial y} + \ldots$$

Отсюда, ввиду равенства $\gamma_{xxx} = \gamma_{yyy}$, получаем $\gamma_{xxx} = 0$.

Таким образом, все компоненты тензора γ_{jlm} обращаются в нуль, если какие-либо два из индексов j, l, m одинаковы, независимо от значения третьего индекса. Отличны от нуля только

компоненты, у которых все три индекса различны. При этом при перестановке любых соседних индексов составляющая тензора γ_{jlm} меняет знак. Следовательно, можно написать

$$\gamma_{xyz} = -\gamma_{yxz} = \gamma_{yzx} = -\gamma_{zyx} = \gamma_{zxy} = -\gamma_{xzy} = -g. \qquad (96.12)$$

В результате (96.11) перейдет в

$$D_x = \varepsilon E_x + g\left(\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z}\right),$$

или в векторной форме

$$\mathbf{D} = \varepsilon \mathbf{E} + g\,\mathrm{rot}\,\mathbf{E}. \qquad (96.13)$$

Если каждая точка среды является *центром симметрии*, то при отражении в этом центре среда переходит сама в себя, а потому тензор γ_{jlm} при таком отражении должен оставаться неизменным. Но при этом правая система координат переходит в левую, а знаки координат x, y, z и компонент полярных векторов $\mathbf{E}$ и $\mathbf{D}$ меняются на противоположные, так что (96.10) переходит в

$$-D_j = -\varepsilon E_j + \gamma_{jlm}\frac{\partial E_l}{\partial x_m}.$$

Следовательно, $\gamma_{jlm} = 0$, т.е. вращение плоскости поляризации невозможно. Для возможности вращения необходимо, чтобы молекулы жидкости или кристаллов кубической системы *не имели центров симметрии.*

Заметим еще, что в случае изотропной естественно-активной среды величина g есть *псевдоскаляр*, а не истинный скаляр (см. т. I, § 7). При переходе от правой системы координат к левой или наоборот знак этой величины меняется на противоположный. Это непосредственно видно из соотношения (96.12), которое показывает, что γ_{jlm} есть *полностью антисимметричный псевдотензор.*

5. По аналогии с формулой (96.13) можно написать

$$\mathbf{B} = \mathbf{H} + g'\,\mathrm{rot}\,\mathbf{H}, \qquad (96.14)$$

где g' — новый *псевдоскаляр*. Магнитную проницаемость μ мы при этом приняли равной единице. Введение добавочного члена $g'\,\mathrm{rot}\,\mathbf{H}$ необходимо для выполнения закона сохранения энергии. Действительно, используя уравнения Максвелла, приведем соотношения (96.13) и (96.14) к виду

$$\mathbf{D} = \varepsilon\mathbf{E} - \frac{g}{c}\frac{\partial \mathbf{B}}{\partial t}, \quad \mathbf{B} = \mathbf{H} + \frac{g'}{c}\frac{\partial \mathbf{D}}{\partial t}.$$

Эти уравнения можно упростить. Для этого подставим второе выражение в первое, а первое во второе и отбросим при этом

произведение gg', как величину *более высокого порядка малости*. Тогда получим

$$\mathbf{D} = \varepsilon\mathbf{E} - \frac{g}{c}\frac{\partial\mathbf{B}}{\partial t}, \quad \mathbf{B} = \mathbf{H} + \frac{\varepsilon g'}{c}\frac{\partial\mathbf{E}}{\partial t}. \tag{96.15}$$

Теперь воспользуемся результатом электродинамики, согласно которому величина $\mathbf{E}\dot{\mathbf{D}} + \mathbf{H}\dot{\mathbf{B}}$ равна производной по времени от (умноженной на 8π) плотности электромагнитной энергии (см. т. III, § 84). Используя (96.15), преобразуем это выражение к виду

$$\mathbf{E}\dot{\mathbf{D}} + \mathbf{H}\dot{\mathbf{B}} = (\varepsilon\mathbf{E}\dot{\mathbf{E}} + \mathbf{H}\dot{\mathbf{H}}) + \frac{1}{c}(\varepsilon g'\mathbf{E}\ddot{\mathbf{E}} - g\mathbf{H}\ddot{\mathbf{H}}).$$

Первый член справа есть производная от $\frac{1}{2}(\varepsilon\mathbf{E}^2 + \mathbf{H}^2)$. Следовательно, и второй член должен быть производной по времени от некоторой функции. Это будет действительно так, если выполняется соотношение $\varepsilon g' = g$, так как тогда

$$\varepsilon g'\mathbf{E}\ddot{\mathbf{E}} - g\mathbf{H}\ddot{\mathbf{H}} = g\frac{\partial}{\partial t}(\mathbf{E}\dot{\mathbf{E}} - \mathbf{H}\dot{\mathbf{H}}).$$

Тем самым доказана необходимость введения второго члена в формуле (96.14), а формулы (96.15) приводятся к окончательному виду

$$\mathbf{D} = \varepsilon\mathbf{E} - \frac{g}{c}\frac{\partial\mathbf{H}}{\partial t}, \quad \mathbf{B} = \mathbf{H} + \frac{g}{c}\frac{\partial\mathbf{E}}{\partial t}. \tag{96.16}$$

6. Теперь мы располагаем полной системой уравнений для монохроматических волн в однородной естественно-активной среде. Из уравнений $\operatorname{div}\mathbf{D} = 0$ и $\operatorname{div}\mathbf{B} = 0$ следует, что плоские волны в такой среде *поперечны относительно векторов* $\mathbf{D}$ *и* $\mathbf{B}$. *Они поперечны также относительно векторов* $\mathbf{E}$ *и* $\mathbf{H}$, так как из уравнений (96.13) и (96.14) следует, что $\operatorname{div}\mathbf{E} = \operatorname{div}\mathbf{H} = 0$. Подставив далее выражения (96.16) в уравнения Максвелла

$$\operatorname{rot}\mathbf{H} = \frac{1}{c}\frac{\partial\mathbf{D}}{\partial t}, \quad \operatorname{rot}\mathbf{E} = -\frac{1}{c}\frac{\partial\mathbf{B}}{\partial t},$$

получим

$$\begin{aligned} \operatorname{rot}\mathbf{H} &= \frac{\varepsilon}{c}\frac{\partial\mathbf{E}}{\partial t} - \frac{g}{c^2}\frac{\partial^2\mathbf{H}}{\partial t^2}, \\ \operatorname{rot}\mathbf{E} &= -\frac{1}{c}\frac{\partial\mathbf{H}}{\partial t} - \frac{g}{c^2}\frac{\partial^2\mathbf{E}}{\partial t^2}. \end{aligned} \tag{96.17}$$

Допустим, что волна плоская и распространяется в положительном направлении оси Z. Тогда отличными от нуля будут только компоненты E_x и E_y, H_x и H_y, причем эти величины зависят только от одной координаты z. С учетом этого запишем

уравнения (96.17) в координатной форме

$$\begin{aligned}\frac{\partial E_x}{\partial z} &= -\frac{1}{c}\frac{\partial H_y}{\partial t} - \frac{g}{c^2}\frac{\partial^2 E_y}{\partial t^2},\\ \frac{\partial E_y}{\partial z} &= \frac{1}{c}\frac{\partial H_x}{\partial t} + \frac{g}{c^2}\frac{\partial^2 E_x}{\partial t^2}\end{aligned} \tag{96.18}$$

и аналогично для производных $\partial H_x/\partial z$ и $\partial H_y/\partial z$. Здесь все величины вещественные. Для упрощения рассуждений удобно ввести комплексные комбинации

$$\begin{aligned}E_+ &= E_x + iE_y, \quad E_- = E_x - iE_y,\\ H_+ &= H_x + iH_y, \quad H_- = H_x - iH_y.\end{aligned} \tag{96.19}$$

Вещественная часть комплексного числа E_+ дает компоненту E_x, а коэффициент при мнимой части — компоненту E_y, и т.д. Однако при исследовании явлений круговой поляризации удобнее оперировать непосредственно с самими комплексными комбинациями, не переходя к вещественной форме. Например, если совершаются гармонические колебания $E_x = A\cos\omega t$, $E_y = A\sin\omega t$, то $E_+ = Ae^{i\omega t}$. Точка, изображающая комплексное число E_+, движется в комплексной плоскости по кругу в направлении от оси X к оси Y, т.е. представляет *волну, поляризованную по левому кругу.* Аналогично, комплексная комбинация E_- описывает *волну, поляризованную по правому кругу.*

Умножив второе уравнение (96.18) на i, почленным сложением и вычитанием этих уравнений найдем выражения для производных $\partial E_+/\partial z$ и $\partial E_-/\partial z$. Аналогично поступаем с магнитным полем. В результате получим

$$\begin{aligned}\frac{\partial E_+}{\partial z} &= \frac{i}{c}\frac{\partial H_+}{\partial t} + \frac{ig}{c^2}\frac{\partial^2 E_+}{\partial t^2},\\ \frac{\partial H_+}{\partial z} &= -\frac{i\varepsilon}{c}\frac{\partial E_+}{\partial t} + \frac{ig}{c^2}\frac{\partial^2 H_+}{\partial t^2}.\end{aligned} \tag{96.20}$$

$$\begin{aligned}\frac{\partial E_-}{\partial z} &= -\frac{i}{c}\frac{\partial H_-}{\partial t} - \frac{ig}{c^2}\frac{\partial^2 E_-}{\partial t^2},\\ \frac{\partial H_-}{\partial z} &= \frac{i\varepsilon}{c}\frac{\partial E_-}{\partial t} - \frac{ig}{c^2}\frac{\partial^2 H_-}{\partial t^2}.\end{aligned} \tag{96.21}$$

Уравнения разделились на две группы *независимых уравнений.* В одну группу входят комбинации E_+ и H_+, описывающие волны с левой круговой поляризацией, в другую — комбинации E_- и H_-, которым соответствует правая круговая поляризация. Ввиду однотипности обеих групп достаточно исследовать одну из них, например группу (96.20). Исключим из нее стандартным способом величину H_+. Отбрасывая члены, содержащие g^2, по-

лучим

$$\frac{\partial^2 E_+}{\partial z^2} = \frac{\varepsilon}{c^2}\frac{\partial^2 E_+}{\partial t^2} + \frac{2ig}{c^2}\frac{\partial^3 E_+}{\partial t^2\,\partial z}$$

и такое же уравнение для H_+. В монохроматической плоской волне $E_+ = C\exp[i(\omega t - kz)]$ дифференцирование по z эквивалентно умножению на $-ik$. Поэтому

$$\frac{\partial^2 E_+}{\partial z^2} = \frac{\varepsilon}{c^2}\frac{\partial^2 E_+}{\partial t^2} + \frac{2gk}{c^2}\frac{\partial^2 E_+}{\partial t^2}.$$

Здесь в последнем члене вместо k можно подставить волновое число в нулевом приближении, т.е. $k = \omega\sqrt{\varepsilon}/c$. Тогда получится волновое уравнение

$$\frac{\partial^2 E_+}{\partial z^2} = \frac{1}{v_+^2}\frac{\partial^2 E_+}{\partial t^2}, \tag{96.22}$$

в котором

$$v_+ = \frac{c}{\sqrt{\varepsilon}}\frac{1}{[1 + 2g\omega/c\sqrt{\varepsilon}]^{1/2}} = \frac{c}{\sqrt{\varepsilon}} - \frac{g\omega}{\varepsilon},$$

или

$$v_+ = \frac{c}{n} - \frac{g\omega}{n^2}, \tag{96.23}$$

где n — показатель преломления среды. Аналогично, для волны с левой круговой поляризацией

$$v_- = \frac{c}{n} + \frac{g\omega}{n^2}. \tag{96.24}$$

Скорости v_+ и v_- различны. Поэтому должно существовать *круговое двойное лучепреломление*, а следовательно, и *вращение плоскости поляризации*.

Убедимся еще, что в каждой поляризованной по кругу плоской волне электрический и магнитный векторы взаимно перпендикулярны. Для этого в уравнениях (96.20) достаточно произвести замену $\partial/\partial t \to i\omega$, $\partial/\partial z \to -ik$. Тогда первое уравнение перейдет в

$$\frac{\omega}{c}H_+ = i\left(k - \frac{g\omega^2}{c}\right)E_+.$$

Отсюда видно, что отношение E_+/H_+ *чисто мнимое*, а это эквивалентно утверждению, что векторы $\mathbf{E}$ и $\mathbf{H}$ в рассматриваемой волне взаимно перпендикулярны.

7. В средах, обладающих *центром симметрии*, разложение (96.9) тензора $\varepsilon_{jl}(\omega, \mathbf{k})$ не может содержать линейных членов по $\mathbf{k}$. В таких случаях пространственная дисперсия может быть обусловлена квадратичными членами. С ней связана слабая зависимость поглощения кубических кристаллов закиси меди Cu_2O от поляризации света (Гросс и Каплянский, 1960 г.), а также слабая анизотропия показателя преломления кубических кристаллов кремния (Пастернак и Ведам, 1971 г.). Эффекты

очень малы, так как они определяются квадратом $(a/\lambda)^2$, т.е. величиной порядка 10^{-6}–10^{-5} (a — размер молекулы или постоянная кристаллической решетки). Существуют и другие эффекты пространственной дисперсии, в обсуждение которых мы входить не будем.

§ 97. О тепловых флуктуациях

В этом вспомогательном параграфе сообщаются краткие сведения о тепловых флуктуациях, необходимые для изложения вопросов молекулярного рассеяния света.

1. Пусть f — любая флуктуирующая величина. *Флуктуацией* этой величины называется отклонение $\Delta f = f - \overline{f}$ мгновенного значения этой величины от ее среднего значения. Обычно пользуются *средним квадратом флуктуации*, т.е. величиной $\overline{(\Delta f)^2}$. Квадратный корень из этой величины $\sqrt{\overline{(\Delta f)^2}}$ называется *среднеквадратичной флуктуацией*, а ее отношение к среднему значению $\overline{f}$, т.е. $\sqrt{\overline{(\Delta f)^2}}/\overline{f}$, — *среднеквадратичной относительной флуктуацией*.

Усредняя выражение $(\Delta f)^2 = (f - \overline{f})^2 = f^2 - 2f\overline{f} + (\overline{f})^2$, получим

$$\overline{(\Delta f)^2} = \overline{f^2} + 2\overline{f}\,\overline{f} - (\overline{f})^2.$$

Но $\overline{f}$ — величина постоянная, а потому $\overline{f}\,\overline{f} = \overline{f}\,\overline{f} = (\overline{f})^2$. Следовательно,

$$\overline{(\Delta f)^2} = \overline{f^2} - (\overline{f})^2. \tag{97.1}$$

Усредним теперь произведение двух флуктуирующих величин:

$$fg = (\overline{f} + \Delta f)(\overline{g} + \Delta g) = \overline{f}\,\overline{g} + \overline{g}\,\Delta f + \overline{f}\,\Delta g + \Delta f\,\Delta g.$$

Так как $\overline{f}$ и $\overline{g}$ — величины постоянные, а $\overline{\Delta f} = \overline{\Delta g} = 0$, то

$$\overline{fg} = \overline{f}\,\overline{g} + \overline{\Delta f\,\Delta g}. \tag{97.2}$$

Формула (97.1) содержится здесь как частный случай, если взять $f = g$.

Величины f и g называются *статистически независимыми*, если $\overline{\Delta f\,\Delta g} = 0$. Для таких величин

$$\overline{fg} = \overline{f}\,\overline{g}. \tag{97.3}$$

2. Рассмотрим теперь любую физическую систему, состоящую из N независимых одинаковых частей. Примером такой системы может служить идеальный газ, а составных частей — отдельные молекулы. Пусть f_i — произвольная *аддитивная величина*, характеризующая i-ю подсистему, например в приведенном примере — кинетическая энергия i-й молекулы. Тогда в силу предполагаемой аддитивности соответствующая величина для всей системы будет $F = \sum f_i$. Выразим средний квадрат флуктуации величины F через аналогичный квадрат для величины f_i. Очевидно, $\overline{F} = \sum \overline{f}_i = N\overline{f}$, где опущен индекс i, так как предполагается, что все составные части си-

стемы тождественны. Далее,

$$F^2 = \Big(\sum f_i\Big)^2 = \sum f_i^2 + \sum\sum_{i\neq j} f_i f_j.$$

А так как эти части независимы, то $\overline{f_i f_j} = \overline{f}_i \overline{f}_j = (\overline{f})^2$. Следовательно,

$$\overline{F^2} = N\overline{f^2} + N(N-1)(\overline{f})^2.$$

Подставляя эти значения в формулу (97.1), получим

$$\overline{(\Delta F)^2} = \overline{F^2} - (\overline{F})^2 = N(\overline{f^2} - \overline{f}^2). \tag{97.4}$$

Отсюда на основании (97.1)

$$\frac{\sqrt{\overline{(\Delta F)^2}}}{\overline{F}} = \frac{\sqrt{N}\,\sqrt{\overline{(\Delta f)^2}}}{N\overline{f}} = \frac{1}{\sqrt{N}}\,\frac{\sqrt{\overline{(\Delta f)^2}}}{\overline{f}}. \tag{97.5}$$

Таким образом, относительная флуктуация величины F обратно пропорциональна квадратному корню из N. При больших N относительные флуктуации ничтожны. Этот вывод верен и для неаддитивных величин. С ним связана *достоверность термодинамических результатов для макроскопических систем.*

3. Применим формулу (97.4) к вычислению флуктуаций числа молекул в фиксированном объеме V идеального газа. Пусть в большом закрытом сосуде объема V в отсутствие силовых полей находятся N молекул идеального газа. Разделим объем V на $z = V/v$ одинаковых объемчиков v_i величиной v каждый. Если n_i — число молекул в объемчике v_i, а N — в большом объеме V, то $N = \sum n_i$. Среднее число молекул $\overline{n}_i$ в объемчике v_i будет $\overline{n}_i = \overline{n} = Nv/V$, т.е. одно и то же во всех объемчиках. Представим его в виде $\overline{n} = Np$, где $p = v/V$ — вероятность нахождения молекулы в объемчике v. Возьмем теперь в качестве f_i следующие величины: $f_i = 1$, если i-я молекула находится внутри объемчика v, и $f_i = 0$, если она находится в оставшемся объеме $V - v$. Тогда число молекул n в объемчике v можно представить в виде $n = \sum f_i$, предполагая, что суммирование ведется по всем N молекулам объема V. Ясно, что $f_i = f_i^2 = f_i^3 = \ldots$, а потому $\overline{f}_i = \overline{f_i^2} = \overline{f_i^3} = \ldots = p$. Следовательно, по формуле (97.1)

$$\overline{\Delta f_i^2} = \overline{f_i^2} - (\overline{f}_i)^2 = p - p^2 = p(1-p).$$

А так как в случае идеального газа величины f_1, f_2, f_3, ... статистически независимы, то по формуле (97.4)

$$\overline{\Delta n^2} = Np(1-p) = (1-p)\overline{n}. \tag{97.6}$$

Если объем V бесконечно велик, то $p \to 0$, а следовательно,

$$\overline{\Delta n^2} = \overline{n}. \tag{97.7}$$

Отсюда получаем для относительной флуктуации плотности газа

$$\sqrt{\overline{\Delta\rho^2}}/\overline{\rho} = 1/\sqrt{\overline{n}}. \tag{97.8}$$

4. Более общий метод вычисления флуктуаций плотности, применимый также к жидкостям и твердым телам, основан на *теореме о равномерном*

распределении кинетической энергии по степеням свободы. Рассмотрим малую часть жидкости или газа, окруженную такой же жидкой или газообразной средой, температура которой T поддерживается постоянной (термостатом). С целью упрощения и наглядности вычислений предположим, что эта малая часть жидкости или газа заключена в цилиндр с поршнем. Стенки цилиндра идеально проводят тепло, а поршень может ходить в нем без трения. Тогда наличие стенок цилиндра и поршня не будет препятствовать обмену энергией и выравниванию давлений между веществом в цилиндре и термостатом. Благодаря тепловому движению поршень будет совершать броуновское движение. К нему мы и применим теорему о равномерном распределении кинетической энергии по степеням свободы.

Поршень можно рассматривать как гармонический осциллятор, совершающий тепловые колебания. Среднее значение его потенциальной энергии при смещении на x из положения равновесия $x = 0$ равно $\frac{1}{2}\varkappa\overline{x^2} = \frac{1}{2}kT$, где $\varkappa$ — коэффициент упругости, соответствующий такому смещению. Если S — площадь поршня, а ΔV — изменение объема системы, то $\Delta V = Sx$. Таким образом, $\overline{(\Delta V)^2} = S^2\overline{x^2} = S^2kT/\varkappa$. Сила, возвращающая поршень в положение равновесия, будет $F = S\dfrac{\partial P}{\partial x}x$, где P — давление газа или жидкости. Поэтому $\varkappa = -S\,\partial P/\partial x = -S^2\partial P/\partial V$. В результате получим

$$(\overline{\Delta V^2})_T = -\frac{kT}{(\partial P/\partial V)_T} = -kT(\partial V/\partial P)_T. \tag{97.9}$$

Значок T указывает, что в выводе предполагалось *постоянство температуры* окружающей среды (термостата). Если бы вещество внутри объема V было адиабатически изолировано, то T следовало бы заменить на S (постоянство энтропии), т.е.

$$(\overline{\Delta V^2})_S = -kT(\partial V/\partial P)_S. \tag{97.10}$$

Формулы (97.9) и (97.10) выражают флуктуации объема *одной и той же массы вещества*, находящейся в термодинамическом равновесии с окружающей средой. Для идеального газа при постоянстве температуры $PV = \mathrm{const}$, так что $(\partial V/\partial P)_T = -V/P$. А так как $PV = nkT$, где n — число молекул в объеме V, то из формулы (97.9) получаем $\overline{(\Delta V)^2} = V^2/n$. Пусть теперь объем V фиксирован, а число молекул в нем n меняется из-за тепловых флуктуаций. Если бы объем V увеличился на ΔV, то число молекул в прежнем объеме V уменьшилось бы на $\Delta n = n\,\Delta V/V$. Отсюда

$$\overline{(\Delta n)^2} = (n/V)^2\overline{(\Delta V)^2} = n = \overline{n},$$

что совпадает с прежним результатом (97.7). В окончательном результате различием между n и $\overline{n}$ можно пренебречь, что мы и сделали.

5. Перейдем теперь к вычислению *флуктуаций энергии.* С целью лучшего уяснения метода начнем с вычисления флуктуаций кинетической энергии $\mathcal{E}$ молекулы одноатомного идеального газа. Согласно максвелловскому закону распределения скоростей,

$$\overline{\mathcal{E}} = \frac{1}{Z}\int \mathcal{E}e^{-\alpha\mathcal{E}}\,d\Gamma, \tag{97.11}$$

где $\alpha = 1/(kT)$, $d\Gamma$ — элемент объема пространства скоростей, а Z определяется условием нормировки:

$$Z = \int e^{-\alpha\mathcal{E}}\,d\Gamma. \tag{97.12}$$

Дифференцируя это соотношение по параметру α, получим

$$\frac{dZ}{d\alpha} = -\int \mathcal{E} e^{-\alpha\mathcal{E}}\, d\Gamma,$$

и формула (97.11) перейдет в

$$\overline{\mathcal{E}} = -\frac{1}{Z}\frac{dZ}{d\alpha}. \tag{97.13}$$

Отсюда

$$\frac{d\overline{\mathcal{E}}}{d\alpha} = -\frac{1}{Z}\frac{d^2Z}{d\alpha^2} + \frac{1}{Z^2}\left(\frac{dZ}{d\alpha}\right)^2 = -\frac{1}{Z}\frac{d^2Z}{d\alpha^2} + \overline{\mathcal{E}}.$$

Аналогично,

$$\overline{\mathcal{E}^2} = \frac{1}{Z}\int \mathcal{E}^2 e^{-\alpha\mathcal{E}}\, d\Gamma = \frac{1}{Z}\frac{d^2Z}{d\alpha^2}. \tag{97.14}$$

Сравнение этой формулы с предыдущей дает

$$\overline{(\Delta\mathcal{E})^2} = \overline{\mathcal{E}^2} - (\overline{\mathcal{E}})^2 = -\frac{d\overline{\mathcal{E}}}{d\alpha},$$

или после подстановки значения параметра $\alpha = 1/(kT)$

$$\overline{(\Delta\mathcal{E})^2} = kT^2\frac{d\overline{\mathcal{E}}}{dT}. \tag{97.15}$$

Так как для идеального одноатомного газа $\overline{\mathcal{E}} = {}^3\!/_2 kT$, то

$$\overline{(\Delta\mathcal{E})^2} = \frac{3}{2}(kT)^2. \tag{97.16}$$

6. Приведенный метод вычисления $\overline{(\Delta\mathcal{E})^2}$ может быть распространен практически без изменений на более общие и важные случаи. Выделим, например, малую часть (подсистему) изотропной среды (жидкости или газа), находящуюся в статистическом равновесии со всей средой, температура T которой поддерживается постоянной. По отношению к выделенной подсистеме окружающая среда играет роль *термостата*. Из-за обмена энергией между термостатом и подсистемой энергия последней будет непрерывно флуктуировать. Беспорядочные изменения энергии подсистемы подчиняются статистическому закону, вполне аналогичному максвелловскому закону распределения кинетической энергии между молекулами. Поэтому среднее значение энергии подсистемы будет выражаться прежней формулой (97.13), где α имеет прежнее значение, а интегрирование в выражении (97.12) производится по многомерному пространству координат и импульсов подсистемы. В этом единственное отличие рассматриваемого случая от предыдущего. Но оно совсем не отражается на последующих выкладках. Поэтому окончательный результат (97.15) применим к рассматриваемой подсистеме без всяких изменений.

Результат (97.15) остается верным и в квантовой статистике. Чтобы убедиться в этом, достаточно заменить интеграл (97.12) суммой

$$Z = \sum g_i \exp(-\alpha\mathcal{E}_i),$$

где $\mathcal{E}_i$ — энергии квантовых уровней подсистемы, а g_i — кратности этих уровней. Суммирование производится по всем квантовым состояниям подсистемы (см. т. II, § 85).

В случае макроскопической подсистемы, объем которой поддерживается постоянным, $\mathcal{E}$ имеет смысл *внутренней энергии подсистемы*, а $d\mathcal{E}/dT$ — ее *теплоемкости* C_V при постоянном объеме. Тогда из формулы (97.15) получается

$$(\overline{\Delta\mathcal{E}^2})_V = kT^2 C_V. \tag{97.17}$$

Значок V слева указывает на то, что величина $(\overline{\Delta\mathcal{E}^2})_V$ есть средний квадрат флуктуации энергии подсистемы при сохранении ее объема V постоянным.

7. Рассмотрим теперь *флуктуации энтальпии* I подсистемы. Для этого воспользуемся следующим искусственным приемом. Предположим, что подсистема заключена в оболочку с идеально проводящими подвижными стенками, так что объем подсистемы не сохраняется постоянным. Пусть оболочка снаружи подвергается действию постоянных внешних сил, поддерживающих внешнее давление P постоянным. Эти силы увеличивают потенциальную энергию подсистемы на величину PV. Если под $\mathcal{E}$ понимать ту же энергию, что и в предыдущем выводе, то с учетом дополнительной потенциальной энергии PV среднее значение полной энергии подсистемы будет $\overline{\mathcal{E}}+P\overline{V}$. Но это есть энтальпия подсистемы I. Все предыдущие рассуждения можно повторить без изменений, заменив $\mathcal{E}$ на $\mathcal{E}+PV$. В результате вместо формулы (97.15) получится

$$(\overline{\Delta I^2})_P = kT^2 \frac{dI}{dT}. \tag{97.18}$$

Но при $P = \text{const}$ производная dI/dT есть теплоемкость C_P подсистемы при постоянном давлении, а потому

$$(\overline{\Delta I^2})_P = kT^2 C_P. \tag{97.19}$$

8. Распространим теперь термодинамический метод вычисления флуктуаций, изложенный выше, на *любые величины*, характеризующие макроскопические свойства подсистем. Ограничимся при этом изотропными телами. Для них любая термодинамическая величина в состоянии термодинамического равновесия есть функция двух других термодинамических величин, которые могут быть приняты за *независимые переменные*. Термодинамические величины макроскопических подсистем хотя и испытывают флуктуации, но в случае малости таких подсистем их мгновенные состояния *практически равновесны*. Они также определяются двумя независимыми переменными. Поэтому задача сводится к вычислению тепловых флуктуаций таких двух независимых переменных. В окончательном результате, определяющем значение среднего квадрата той или иной флуктуации, необходимо указывать, какая из двух величин, выбранных для характеристики состояния подсистемы, *поддерживается постоянной*. Иначе самый результат будет неопределенным, а потому и бессмысленным.

Покажем на примерах, как применяется изложенный метод к вычислению флуктуаций различных физических величин.

Начнем с флуктуаций температуры, предполагая, что рассматриваемая макроскопическая подсистема находится в тепловом контакте с термостатом. Считая температуру подсистемы функцией независимых переменных V и $\mathcal{E}$, напишем

$$\Delta T = \left(\frac{\partial T}{\partial V}\right)_{\mathcal{E}} \Delta V + \left(\frac{\partial T}{\partial \mathcal{E}}\right)_V \Delta\mathcal{E}.$$

В силу независимости V и $\mathcal{E}$ имеем: $\overline{\Delta V \cdot \Delta\mathcal{E}} = 0$, а потому

$$\overline{(\Delta T)^2} = \left(\frac{\partial T}{\partial V}\right)^2_{\mathcal{E}} \overline{(\Delta V)^2} + \left(\frac{\partial T}{\partial \mathcal{E}}\right)^2_V \overline{(\Delta\mathcal{E})^2}.$$

При постоянном объеме производная $(\partial\mathcal{E}/\partial T)_V$ есть теплоемкость подсистемы C_V. Тогда из формулы (97.17) находим

$$(\overline{\Delta T^2})_V = \frac{1}{C_V^2}(\overline{\Delta\mathcal{E}^2})_V = \frac{kT^2}{C_V}. \tag{97.20}$$

Вычислим теперь флуктуации энтропии S подсистемы. В качестве независимых переменных выберем V и $\mathcal{E}$. Рассуждая как в предыдущем примере, напишем

$$\overline{\Delta V\cdot\Delta\mathcal{E}} = 0, \quad (\overline{\Delta S^2})_V = \left(\frac{\partial S}{\partial\mathcal{E}}\right)_V^2 (\overline{\Delta\mathcal{E}^2})_V.$$

Так как $\partial S/\partial\mathcal{E} = 1/T$, то с учетом (97.17)

$$(\overline{\Delta S^2})_V = kC_V. \tag{97.21}$$

Если бы за независимые переменные были приняты P и I, то получилось бы

$$(\overline{\Delta S^2})_P = kC_P. \tag{97.22}$$

Рассмотрим теперь флуктуации давления P. Примем за независимые переменные V и T. Тогда

$$\overline{\Delta V\cdot\Delta T} = 0, \quad (\overline{\Delta P^2})_T = \left(\frac{\partial P}{\partial V}\right)_T^2 (\overline{\Delta V^2})_T,$$

или на основании формулы (97.9)

$$(\overline{\Delta P^2})_T = -kT\left(\frac{\partial P}{\partial V}\right)_T. \tag{97.23}$$

Приняв за независимые переменные P и S, мы получили бы

$$(\overline{\Delta P^2})_S = -kT\left(\frac{\partial P}{\partial V}\right)_S = -\gamma kT\left(\frac{\partial P}{\partial V}\right)_T, \tag{97.24}$$

где $\gamma = C_P/C_V$. При этом было использовано термодинамическое соотношение $\left(\frac{\partial P}{\partial V}\right)_S \Big/ \left(\frac{\partial P}{\partial V}\right)_T = \gamma$ (см.т. II, § 47).

Вычислим, наконец, *флуктуации плотности* вещества ρ в объеме V. Задача сводится просто к преобразованию формулы (97.9). Прежде всего заметим, что величина $V\,\partial P/\partial V$ не зависит от величины объема V. Поэтому в таком выражении V можно заменить удельным объемом вещества v. Тогда

$$(\overline{\Delta V^2})_T = -\frac{VkT}{v(\partial P/\partial v)_T}.$$

Так как масса вещества $V\rho$ в объеме V остается постоянной, то $V\,\Delta\rho + \rho\,\Delta V = 0$. Следовательно, $(\Delta\rho)^2 = \left(\frac{\rho}{V}\right)^2 (\Delta V)^2$, а потому

$$V(\overline{\Delta\rho^2})_T = -\frac{\rho^2 kT}{v(\partial P/\partial v)_T}. \tag{97.25}$$

Чем меньше объем V, тем больше относительные флуктуации плотности в нем.

§ 98. Рассеяние света

1. В прозрачной однородной среде бегущая плоская волна распространяется только в прямом направлении, не испытывая рассеяния в стороны. (Мы отвлекаемся от дифракции, предполагая, что ширина фронта волны достаточно велика, а следовательно, угол дифракционной расходимости мал.) Допустим теперь, что оптическая однородность среды нарушена, например множеством мельчайших частиц постороннего вещества, беспорядочно распределенных по объему среды. Примерами могут служить пыльный воздух, туман, дым, эмульсии и суспензии с взвешенными в них посторонними частицами. Тогда показатель преломления будет меняться в пространстве весьма нерегулярно, но среднее значение его во всяком малом объеме, содержащем еще очень много макроскопических неоднородностей, будет оставаться одним и тем же во всей среде. Такую среду называют *оптически мутной*. В оптически мутных средах свет распространяется не только в прямом направлении, но и *рассеивается в стороны*. Рассеяние света в мутных средах на частицах постороннего вещества экспериментально впервые исследовал Тиндаль (1820–1893) в 1869 г. Поэтому это явление получило название *тиндалевского рассеяния* или *эффекта Тиндаля*. Его теория была дана Рэлеем.

2. В неоднородной неподвижной изотропной среде распространение света описывается уравнениями Максвелла

$$\begin{aligned} \operatorname{rot}\mathbf{H} &= \frac{\varepsilon}{c}\frac{\partial\mathbf{E}}{\partial t}, \qquad \operatorname{div}(\varepsilon\mathbf{E}) = 0, \\ \operatorname{rot}\mathbf{E} &= -\frac{1}{c}\frac{\partial\mathbf{H}}{\partial t}, \qquad \operatorname{div}\mathbf{H} = 0, \end{aligned} \tag{98.1}$$

где диэлектрическая проницаемость ε является функцией координат. Выделим из нее постоянную часть ε_0, полагая $\varepsilon = \varepsilon_0 + \delta\varepsilon$. В проблеме рассеяния света интерес представляет случай, когда $\delta\varepsilon$ мало по сравнению с ε_0, но пока мы не будем вводить этого ограничения. Более того, постоянное слагаемое ε_0 в принципе можно было бы выбрать произвольно. От этого, если вычисления производить точно, окончательный результат зависеть не может. Однако удобно и естественно понимать под ε_0 диэлектрическую проницаемость среды, из которой удалены частицы постороннего вещества.

Представим электромагнитное поле в виде $\mathbf{E} = \mathbf{E}_0 + \mathbf{E}'$, $\mathbf{H} = \mathbf{H} + \mathbf{H}'$, где $\mathbf{E}_0$, $\mathbf{H}_0$ удовлетворяют уравнениям Максвелла в однородной среде

$$\begin{aligned} \operatorname{rot}\mathbf{H}_0 &= \frac{\varepsilon_0}{c}\frac{\partial\mathbf{E}_0}{\partial t}, \qquad \operatorname{div}(\varepsilon_0\mathbf{E}_0) = 0, \\ \operatorname{rot}\mathbf{E}_0 &= -\frac{1}{c}\frac{\partial\mathbf{H}_0}{\partial t}, \qquad \operatorname{div}\mathbf{H}_0 = 0. \end{aligned}$$

В задаче о рассеянии света это есть *падающая волна*, которая распространялась бы в среде, если бы в ней не было оптических неоднородностей, а $\mathbf{E}'$, $\mathbf{H}'$ — поле *рассеянного света*. Вычитая предыдущие уравнения из (98.1), получим

$$\begin{gathered}\operatorname{rot}\mathbf{H}'-\frac{\varepsilon_0}{c}\frac{\partial\mathbf{E}'}{\partial t}=\frac{\delta\varepsilon}{c}\frac{\partial\mathbf{E}}{\partial t},\quad \operatorname{div}(\varepsilon_0\mathbf{E}')=-\operatorname{div}(\delta\varepsilon\mathbf{E}),\\ \operatorname{rot}\mathbf{E}'-\frac{1}{c}\frac{\partial\mathbf{H}'}{\partial t}=0,\qquad\qquad \operatorname{div}\mathbf{H}'=0.\end{gathered}\tag{98.2}$$

Таким образом, для поля $\mathbf{E}'$, $\mathbf{H}'$ получились такие же уравнения Максвелла, как в однородной среде с диэлектрической проницаемостью ε_0. Только первые два из этих уравнений содержат правые части, которые можно рассматривать как *дополнительные источники электромагнитных волн*. Если ввести обозначение

$$\delta\mathbf{P}=\frac{\delta\varepsilon}{4\pi}\,\mathbf{E},\tag{98.3}$$

то эти два уравнения перейдут в

$$\operatorname{rot}\mathbf{H}'-\frac{\varepsilon_0}{c}\frac{\partial\mathbf{E}'}{\partial t}=\frac{4\pi}{c}\frac{\partial}{\partial t}\,\delta\mathbf{P},\quad \operatorname{div}(\varepsilon_0\mathbf{E}')=-4\pi\operatorname{div}(\delta\mathbf{P}).\tag{98.4}$$

Из них видно, что в среде появляется *дополнительная поляризация* $\delta\mathbf{P}$, определяемая выражением (98.3), так что каждый малый элемент объема среды δV получает *дополнительный дипольный момент* $\delta V\cdot\delta\mathbf{P}$. Меняясь во времени, он излучает электромагнитные волны как колеблющийся *диполь Герца*. Это и есть свет, рассеянный элементом объема δV.

3. Допустим теперь, что оптическая неоднородность создается одинаковыми шариками радиуса a, беспорядочно распределенными по объему, занятому средой. Пусть среднее расстояние между шариками велико по сравнению с a, а сами шарики малы по сравнению с длиной волны. Тогда при вычислении электрического поля $\mathbf{E}$ внутри шарика можно считать внешнее поле $\mathbf{E}_0$ световой волны однородным. Как показано в электростатике (см. т. III, § 16), поле $\mathbf{E}$ также однородно и определяется выражением

$$\mathbf{E}=\frac{3}{\varepsilon/\varepsilon_0+2}\,\mathbf{E}_0=\frac{3\varepsilon_0}{\varepsilon+2\varepsilon_0}\,\mathbf{E}_0,\tag{98.5}$$

где ε — диэлектрическая проницаемость шарика, а ε_0 — окружающей среды. Дополнительная поляризация, согласно формуле (98.3), будет отлична от нуля только внутри шариков, где она равна

$$\delta\mathbf{P}=\frac{\varepsilon-\varepsilon_0}{4\pi}\,\mathbf{E}=\frac{\varepsilon-\varepsilon_0}{4\pi}\,\frac{3\varepsilon_0}{\varepsilon+2\varepsilon_0}\,\mathbf{E}_0,$$

а дополнительный дипольный момент шарика

$$\mathbf{p}=\frac{\varepsilon_0}{\varepsilon+2\varepsilon_0}\,(\varepsilon-\varepsilon_0)a^3\mathbf{E}_0.\tag{98.6}$$

Предположим сначала, что падающая волна *поляризована линейно*. Тогда векторы $\mathbf{p}$ и $\mathbf{E}$ все время будут параллельны одному и тому же неизменному направлению. Электрическое поле диполя $\mathbf{p}$ на больших расстояниях r от него (в волновой зоне) определяется выражением

$$E_1 = \frac{\sin\vartheta}{c^2 r}[\ddot{p}]_{t-r/v} = -\frac{\omega^2 \sin\vartheta}{c^2 r}[p]_{t-r/v}, \tag{98.7}$$

где $v = c/\sqrt{\varepsilon} = c/n$ — скорость света в рассматриваемой среде, а ϑ — угол между осью диполя $\mathbf{p}$ и направлением рассеянного излучения (см. т. III, § 141). *Рассеянный свет поляризован линейно*, причем электрический вектор лежит в плоскости, проходящей через ось диполя $\mathbf{p}$ и направление излучения. *Под интенсивностью света здесь и в дальнейшем будем понимать усредненное по времени численное значение вектора Пойнтинга.* Для интенсивности света, рассеянного одним шариком, электродинамика дает

$$I_1 = \frac{\sin^2\vartheta}{4\pi\varepsilon_0 v^3 r^2}\overline{\ddot{p}^2} = \frac{\omega^4 \sin^2\vartheta}{4\pi\varepsilon_0 v^3 r^2}\overline{p^2} \tag{98.8}$$

(см. т. III, § 141). Интенсивность прямой волны равна

$$I_0 = \frac{c}{4\pi}\overline{E_0 H_0} = \frac{v}{4\pi}\varepsilon_0\overline{E_0^2}. \tag{98.9}$$

Воспользовавшись выражением (98.6), получим

$$I_1 = 9\varepsilon_0^2\left(\frac{\varepsilon-\varepsilon_0}{\varepsilon+2\varepsilon_0}\right)^2\frac{\omega^4 a^6 \sin^2\vartheta}{c^4 r^2}I_0, \tag{98.10}$$

или

$$I_1 = 9\varepsilon_0^2\left(\frac{\varepsilon-\varepsilon_0}{\varepsilon+2\varepsilon_0}\right)^2\frac{\pi^2 V_1^2}{\lambda^4}\frac{\sin^2\vartheta}{r^2}I_0, \tag{98.11}$$

где λ — длина волны в вакууме, а $V_1 = {}^4\!/_3\pi a^3$ — объем шарика. Энергия $\mathcal{P}_1$, рассеиваемая шариком в единицу времени по всем направлениям, найдется интегрированием величины (98.11) по сфере радиуса r. Взяв в качестве элемента поверхности $2\pi r^2 \sin\vartheta\, d\vartheta$, получим

$$\mathcal{P}_1 = 24\pi^3\varepsilon_0^2\left(\frac{\varepsilon-\varepsilon_0}{\varepsilon+2\varepsilon_0}\right)^2\frac{V_1^2}{\lambda^4}I_0. \tag{98.12}$$

Допустим теперь, что падающий свет *естественный*. Направление его распространения примем за ось Z. Пусть рассеянный свет наблюдается в направлении OA под углом θ к оси Z. Угол θ называется *углом рассеяния* (рис. 321). Направим ось X перпендикулярно к OA и OZ. Так как $\mathbf{p}$ и $\mathbf{E}_0$ коллинеарны, то вектор $\mathbf{p}$ параллелен плоскости XY. Разложим его по осям X и Y. Интенсивности излучений дипольных моментов $\mathbf{p}_x$ и $\mathbf{p}_y$ найдутся по формуле (98.8), если в ней положить сначала $\vartheta = \pi/2$,

а затем $\vartheta = \frac{\pi}{2} - \theta$. Так как падающий свет естественный, то эти излучения *некогерентны*, так что для нахождения I_1 надо сложить их интенсивности. В результате формула (98.8) перейдет в

$$I_1 = \frac{\omega^4}{4\pi\varepsilon_0 v^3 r^2}(\overline{p_x^2} + \overline{p_y^2}\cos^2\theta) = \frac{\omega^4}{4\pi\varepsilon_0 v^3 r^2}\frac{1+\cos^2\theta}{2}\overline{p^2},$$

так как в случае естественного света $\overline{p_x^2} = \overline{p_y^2} = {}^1\!/_2\overline{p^2}$. Следовательно, вместо формулы (98.11) получится

$$I_1 = 9\varepsilon_0^2\left(\frac{\varepsilon-\varepsilon_0}{\varepsilon+2\varepsilon_0}\right)^2\frac{\pi^2 V_1^2}{\lambda^4}\frac{1+\cos^2\theta}{2r^2}I_0. \qquad (98.13)$$

Формула (98.12), очевидно, останется без изменения.

Теперь рассеянный свет будет поляризован частично. *Полная линейная поляризация будет наблюдаться только в тех случаях, когда линия наблюдения OA перпендикулярна к направлению распространения падающего света*, так как в этом случае дипольный момент $\mathbf{p}_y$ излучения не дает.

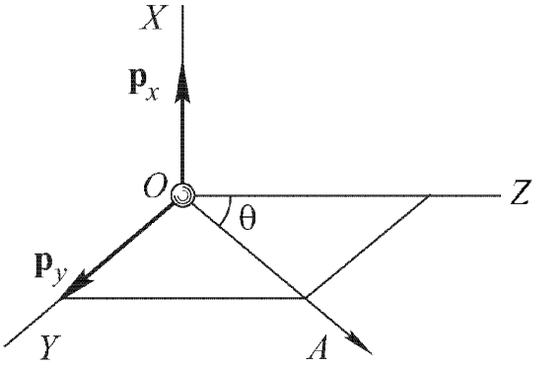


Рис. 321

Найдем теперь интенсивность I света, рассеиваемого объемом V, содержащим очень много шариков. Их среднее число в этом объеме равно NV, где N — среднее число шариков в единице объема. Так как расстояния между шариками велики по сравнению с a и они распределены по объему V беспорядочно, то для нахождения I надо сложить интенсивности, рассеиваемые отдельными шариками. Предположим, что расстояние от объема V до точки наблюдения велико по сравнению с линейными размерами самого объема V. Тогда в формуле (98.13) все расстояния r можно считать одинаковыми и написать

$$I = 9\varepsilon_0^2\left(\frac{\varepsilon-\varepsilon_0}{\varepsilon+2\varepsilon_0}\right)^2\pi^2\frac{V_1^2}{\lambda^4}\frac{1+\cos^2\theta}{2r^2}NVI_0. \qquad (98.14)$$

4. Рассчитаем убывание интенсивности I_0 падающего света из-за рассеяния. Выделим в среде произвольный цилиндр, площадь поперечного сечения которого равна единице, а образующие параллельны оси Z. Вырежем из него бесконечно короткий цилиндрик, ограниченный плоскостями $z = \text{const}$ и $z + dz = \text{const}$. Через первое основание такого цилиндрика ежесекундно вступает энергия $I_0(z)$, а через второе выходит $I_0(z+dz)$. Разность этих энергий $-dI_0$ есть рассеянная энергия $\mathcal{P}_1 N\,dz$.

Приравнивая оба выражения, получим

$$dI_0 = -\gamma I_0\, dz, \tag{98.15}$$

где в соответствии с формулой (98.12)

$$\gamma = 24\pi^3 \varepsilon_0^2 \left(\frac{\varepsilon - \varepsilon_0}{\varepsilon + 2\varepsilon_0}\right)^2 \frac{NV_1^2}{\lambda^4}. \tag{98.16}$$

Таким образом, из-за рассеяния интенсивность падающей волны убывает экспоненциально:

$$I_0 = \text{const} \cdot e^{-\gamma z}. \tag{98.17}$$

Величина γ называется *коэффициентом рассеяния.*

5. Согласно формуле (98.14), впервые полученной Рэлеем, интенсивность рассеянного света обратно пропорциональна *четвертой степени длины волны.* Этот результат называется *законом Рэлея.* Он справедлив для рассеивающих частиц, линейные размеры которых малы по сравнению с длиной волны.

Закон Рэлея качественно подтверждается уже ранними исследованиями Тиндаля, который наблюдал, что белый свет при рассеянии становится синеватым. На основе этих наблюдений Тиндаль высказал мысль, что синий цвет и поляризация неба определяются рассеянием солнечного света на мелких частицах пыли, всегда имеющихся в достаточном количестве в земной атмосфере.

Количественная теория была развита Рэлеем. Если бы рассеяния света не было, то небо было бы совершенно черным. На этом черном фоне звезды и другие небесные светила выделялись бы более ярко и контрастно. Именно таким видят небо космонавты из космических кораблей. При наличии же атмосферы значительная доля прямого солнечного излучения рассеивается в стороны. Она тем больше, чем короче длина волны. Поэтому рассеянный свет обогащен короткими волнами, чем и объясняется *синий цвет неба.* При восходе и заходе Солнца прямой солнечный свет проходит через большую толщу атмосферы, и при этом большая часть коротковолнового излучения теряется на рассеяние. Из прямого света до поверхности Земли доходят преимущественно красные лучи. Вот почему при восходе и заходе Солнце красное. Так же объясняется *красный цвет зари.*

По мере поднятия над земной поверхностью содержание пыли и других посторонних частиц в воздухе уменьшается. Казалось бы, что при этом насыщенность рассеянного света синими лучами должна также уменьшаться. Однако наблюдения в высокогорных обсерваториях показали, что дело обстоит как раз наоборот. Чем чище воздух, чем меньше в нем содержится посторонних частиц, тем насыщеннее излучение неба синими лучами и тем полнее его поляризация. На этом основании Рэлей

пришел к заключению, подтвержденному всеми последующими экспериментальными и теоретическими исследованиями, что здесь рассеяние вызывается не посторонними частицами, а *самими молекулами воздуха.* Такое рассеяние света называется *рэлеевским* или *молекулярным рассеянием.* Однако физическая природа молекулярного рассеяния была понята только в 1908 г. М. Смолуховским (1872–1917). *Молекулярное рассеяние вызывается тепловыми флуктуациями показателя преломления,* которые и делают среду оптически мутной. Теория рассеяния света в жидкостях и газах, построенная на этой основе, была создана в 1910 г. Эйнштейном. Она применима в тех случаях, когда длина световой волны настолько велика, что среду можно разбить на объемчики, малые по сравнению с кубом длины волны, каждый из которых содержит, однако, еще очень много молекул. К таким объемчикам еще можно применять макроскопические уравнения Максвелла, не учитывая явно молекулярную структуру вещества. Флуктуации показателя преломления в таких объемчиках и играют роль *макроскопических неоднородностей,* вызывающих рассеяние света, подобно шарикам в предыдущем рассмотрении.

Поле $\mathbf{E}'$, $\mathbf{H}'$ рассеянного излучения в общем случае можно рассчитать с помощью уравнений (98.2). В них теперь ε_0 означает среднее значение диэлектрической проницаемости среды, а $\delta\varepsilon$ — ее флуктуацию. Решение можно получить *методом последовательных приближений.* В нулевом приближении в уравнениях (98.2) пренебрегаем неоднородностями среды, т.е. правыми частями. Тогда рассеянного излучения $\mathbf{E}'$, $\mathbf{H}'$ не будет — останется только падающая волна $\mathbf{E}_0$, $\mathbf{H}_0$. Для нахождения первого приближения в правых частях (98.2) заменяем поле $\mathbf{E}$ его значением $\mathbf{E}_0$ в нулевом приближении. Решая полученные уравнения, находим $\mathbf{E}'$ и $\mathbf{H}'$, а затем $\mathbf{E}$ и $\mathbf{H}$ в первом приближении. Используя полученное решение, уточняем правые части уравнений (98.2) и находим $\mathbf{E}$ и $\mathbf{H}$ во втором приближении, и т.д.

Практически достаточно ограничиться первым приближением, полагая $\delta\mathbf{P} = \delta\varepsilon\mathbf{E}_0/(4\pi)$. Разделим среду на элементарные объемчики $\delta_i V$, малые по сравнению с кубом длины волны, но содержащие еще очень много молекул. Дополнительный дипольный момент объемчика $\delta_i V$, обусловленный флуктуациями диэлектрической проницаемости, будет

$$\mathbf{p} = \frac{\delta_i\varepsilon \cdot \delta_i V}{4\pi}\mathbf{E}_0.$$

Это выражение отличается от (98.6) только коэффициентом при $\mathbf{E}_0$. Поэтому, предполагая падающий свет естественным, для средней интенсивности света, рассеиваемого объемчиком $\delta_i V$,

можно сразу написать

$$I_i = \frac{\pi^2}{\lambda^4}\frac{1+\cos^2\theta}{2r^2} I_0 (\delta_i V)^2 \overline{(\delta_i \varepsilon)^2}. \qquad (98.18)$$

Соответствующее электрическое поле будет пропорционально $\delta_i V$ и $\delta_i \varepsilon$.

6. Рассмотрим сначала молекулярное рассеяние в *идеальных газах*. В этом случае

$$\varepsilon_i = 1 + 4\pi\beta \frac{N_i}{\delta_i V},$$

где β — поляризуемость молекулы, а N_i — число молекул в объемчике $\delta_i V$. Поскольку объемчик $\delta_i V$ фиксирован в пространстве, т.е. не флуктуирует, из последней формулы получаем $\delta_i V\, \delta\varepsilon_i = 4\pi\beta\delta N_i$, т.е. рассеяние света вызывается *флуктуациями числа молекул* в объемчиках $\delta_i V$. Согласно (97.7), $\overline{(\delta N_i)^2} = \overline{N_i} = N\delta_i V$, где N — число молекул в единице объема. Таким образом,

$$I_i = \frac{8\beta^2\pi^4}{\lambda^4}\frac{1+\cos^2\theta}{r^2} I_0 N \delta_i V.$$

Для газов $n^2 - 1 = 4\pi N\beta$, причем показатель преломления n близок к единице. Поэтому вместо этой формулы можно взять $n - 1 = 2\pi N\beta$. Исключив β, найдем

$$\overline{I}_i = \frac{2\pi^2 (n-1)^2}{N\lambda^4}\frac{1+\cos^2\theta}{r^2} I_0 \delta_i V.$$

Чтобы найти интенсивность I света, рассеянного конечным объемом V, заметим, что различные объемчики $\delta_i V$ и $\delta_j V$ рассеивают *некогерентно*. Действительно, возьмем все объемчики $\delta_i V$ одинаковыми по величине. Тогда электрические поля рассеянных ими волн представятся в виде $\mathbf{E}_i = \mathbf{A}\,\delta N_i$, где коэффициент $\mathbf{A}$ одинаков для всех объемчиков. Полное поле рассеянной волны будет $\sum \mathbf{E}_i$, а интенсивность

$$I \sim \overline{\Big(\sum \mathbf{E}_i\Big)^2} = A^2 \overline{\sum\sum_{i\neq j} \delta N_i\, \delta N_j} + A^2 \overline{\sum (\delta N_i)^2}.$$

Но в идеальных газах числа молекул в различных объемчиках флуктуируют *независимо*. Поэтому двойная сумма в предыдущей формуле обращается в нуль, что и доказывает наше утверждение. Значит, для нахождения I надо $\overline{I}_i$ умножить на число элементарных объемчиков в объеме V, т.е. на $V/\delta_i V$. Это дает

$$I = \frac{2\pi^2 (n-1)^2 V}{N\lambda^4}\frac{1+\cos^2\theta}{r^2} I_0. \qquad (98.19)$$

Если свет линейно поляризован, то множитель $(1+\cos^2\theta)/2$ следует заменить на $\sin^2\vartheta$.

Формула (98.19) была впервые получена Рэлеем в 1899 г., но недостаточно обоснована им. Рэлей вывел ее в предположении, что рассеяние происходит на *отдельных молекулах газа*, которые ведут себя совершенно аналогично независимым шарикам, о которых шла речь при выводе формулы (98.14). Результирующую интенсивность рассеянного света он вычислял, складывая интенсивности рассеянных волн от отдельных молекул, как если бы эти волны были некогерентны. Он полагал, что некогерентность возникает из-за теплового движения молекул, но не учитывал явно флуктуации числа частиц в рассеивающих объемчиках.

Л.И. Мандельштам в 1907 г. указал, что если бы все (одинаковые) малые пространственно неподвижные объемчики $\delta_i V$ содержали одинаковое число молекул, то излучаемые ими вторичные волны были бы *когерентны, независимо от того, движутся в них молекулы или нет.* При сложении напряженностей полей таких волн происходило бы *интерференционное гашение* их во всех направлениях, за исключением направления падающей волны.

Полного гашения на самом деле не будет, так как (на что в 1908 г. указал Смолуховский) числа молекул в элементарных объемчиках испытывают флуктуации. Из-за таких флуктуаций излучения объемчиков можно разложить на когерентную и некогерентную части. Когерентная часть влияет на падающую волну, изменяя ее фазовую скорость (см. § 68). Некогерентные излучения появляются из-за флуктуаций числа частиц. С ними и связано рассеяние света. Только благодаря случайному соотношению (97.7) окончательный результат (98.19) в случае идеальных газов получается таким же, как если бы молекулы рассеивали свет некогерентно.

7. Формула (98.19) хорошо согласуется с опытом во всех случаях, когда молекулы газа обладают сферической симметрией, т.е. *изотропны.* Таковы, например, молекулы инертных газов. Первое количественное подтверждение эта формула получила при исследовании рассеяния солнечного света в земной атмосфере (синевы неба). Измерив коэффициент рассеяния, можно с помощью формулы (98.19) вычислить число Лошмидта N, т.е. число молекул воздуха в единице объема. Полученное таким путем значение N прекрасно согласуется со значениями этого числа, определенными другими методами.

Если молекулы газа *анизотропны*, например CO_2, то наблюдаются отступления от изложенной теории. Прежде всего, если падающий свет поляризован линейно, то рассеянный свет поляризован только частично, а не полностью, как требует эта теория. Такая *деполяризация рассеянного света* вызывается именно анизотропией молекул. Пусть электрический вектор $\mathbf{E}_0$ па-

дающей волны параллелен оси X. Если бы молекула была изотропна, то ее индуцированный дипольный момент $\mathbf{p} = \beta\mathbf{E}_0$ имел бы то же направление. Свет, рассеянный молекулой, получился бы поляризованным линейно, с плоскостью колебаний, проходящей через ось диполя $\mathbf{p}$ и линию наблюдения. Но если молекула анизотропна, то параллельности между $\mathbf{p}$ и $\mathbf{E}$, вообще говоря, уже не будет. Появятся составляющие вектора $\mathbf{p}$ вдоль осей Y и Z. А так как при тепловом движении ориентация молекулы в пространстве непрерывно и беспорядочно меняется, то поляризуемости молекулы вдоль координатных осей X, Y, Z будут также флуктуировать. Составляющие $\mathbf{p}_y$ и $\mathbf{p}_z$ дают рассеянные волны, поляризация которых отлична от поляризации излучения, даваемого составляющей $\mathbf{p}_x$. Это и приводит к деполяризации рассеянного света.

Флуктуации в ориентации анизотропных молекул вызывают не только деполяризацию, но и влияют на интенсивность рассеянного света. Однако в такие детали теории мы входить не можем.

8. Перейдем теперь к молекулярному рассеянию в *неидеальных газах* и *жидкостях*, когда они находятся в состоянии, не очень близком к критической точке. Исходной будет формула (98.18). Диэлектрическая проницаемость ε является функцией плотности ρ и температуры T среды. От температуры она зависит слабо. Этой зависимостью можно пренебречь. Тогда

$$\delta_i\varepsilon = \frac{d\varepsilon}{d\rho}\,\delta_i\rho,$$

где $\delta_i\rho$ — *флуктуация плотности* в объеме $\delta_i V$. Полагая в формуле (97.25) $V = \delta_i V$, можем написать

$$\delta_i V\overline{(\delta_i\rho)^2} = -\rho^2\frac{kT}{v(\partial P/\partial v)_T}.$$

Следовательно, формула (98.18) переходит в

$$I_i = \frac{\pi^2}{\lambda^4}\frac{1+\cos^2\theta}{2r^2}I_0\delta_i V\left(\rho\frac{d\varepsilon}{\delta\rho}\right)^2\frac{kT}{(-v\,\partial P/\partial v)_T}.$$

Предположим теперь, что рассеивающий объем V не мал. В газах и жидкостях вдали от критической точки линейные размеры элементарных объемчиков $\delta_i V$ можно взять малыми по сравнению с λ, но большими по сравнению с радиусом действия молекулярных сил (порядка 10^{-7} см). В таком случае различные элементарные объемчики будут рассеивать свет *некогерентно*. Для нахождения результирующей интенсивности света, рассеиваемого всем объемом V, надо I_i умножить на число этих объемчиков, т.е. на $V/\delta_i V$. (Мы берем объемчики $\delta_i V$ одинаковой величины, что, очевидно, несущественно и не может влиять на

окончательный результат.) Таким путем получаем

$$I = \frac{\pi^2}{\lambda^4}\frac{1+\cos^2\theta}{2r^2}VI_0\left(\rho\frac{d\varepsilon}{d\rho}\right)^2\frac{kT}{(-v\,\partial P/\partial v)_T}. \qquad (98.20)$$

Эта формула была выведена Эйнштейном в 1910 г. Формула Рэлея (98.19) содержится в ней как частный случай.

Флуктуации плотности становятся очень большими при приближении к *критической точке* рассматриваемого вещества, так как при этом производная $\partial P/\partial v$ стремится к нулю. Этим объясняется так называемая *критическая опалесценция*, т.е. очень сильное рассеяние света в окрестности критической точки. Явление это было известно задолго до создания Смолуховским и Эйнштейном теории флуктуаций, но его причина оставалась неясной вплоть до появления работ этих ученых. В самой критической точке формула Эйнштейна (98.20) дает для интенсивности рассеянного света бесконечное значение. Отсюда следует, что в окрестности критической точки эта формула неприменима. Причина заключается в том, что флуктуации плотности в малых объемчиках $\delta_i V$ в окрестности критической точки уже *нельзя считать статистически независимыми*. Орнштейн и Цернике обобщили формулу Эйнштейна на случай рассеяния света вблизи критической точки. Так как здесь линейные размеры неоднородностей, на которых рассеивается свет, уже не малы по сравнению с λ, то интенсивность рассеянного света не пропорциональна четвертой степени частоты, а возрастает с частотой *более медленно*. Кроме того, рассеяние перестает быть симметричным относительно плоскости, перпендикулярной к падающим лучам: *вперед рассеивается больше света, чем назад*. На всех этих вопросах мы не можем останавливаться.

Формула Эйнштейна учитывает не все причины рассеяния света. В частности, она не учитывает *флуктуации анизотропии* в случае жидкостей, состоящих из анизотропных молекул. Учет этого обстоятельства увеличивает интенсивность рассеянного света и объясняет его деполяризацию. В случае жидкостей соответствующие теоретические расчеты, конечно, не столь надежны, а их согласие с опытом не такое хорошее, как в случае газов.

9. Молекулярное рассеяние света может вызываться не только флуктуациями плотности, но и другими причинами. Интересным примером может служить *рассеяние света в растворах*. В обычных условиях молекулы растворенного вещества настолько равномерно перемешаны с молекулами растворителя, что весь раствор в оптическом отношении представляет собой почти столь же однородную среду, что и чистый растворитель. *Флуктуации концентрации* растворенного вещества в растворителе могут быть причиной рассеяния света. В обычных усло-

виях флуктуации концентрации и обусловленное ими рассеяние света малы. Однако существует много комбинаций веществ, которые ниже определенной температуры $T_к$, называемой *критической температурой смешения*, растворяются друг в друге только частично, а выше этой температуры смешиваются в любых пропорциях (см. т. II, § 123). При критической температуре смешения две жидкости полностью смешиваются друг с другом только *при вполне определенных весовых отношениях.* Такова, например, смесь 20 частей по весу сероуглерода и 80 частей метилового спирта при температуре $T_к = 40$ °С. При критической температуре смешения надо ожидать особенно больших флуктуаций концентрации, подобно тому, что имеет место для флуктуаций плотности вблизи критической точки. Действительно, при критической температуре смешения наблюдается очень сильное рассеяние света, аналогичное критической опалесценции.

10. Вторым примером может служить *рассеяние света на границе раздела двух несмешивающихся жидкостей или на свободной поверхности жидкости.* Из-за теплового движения поверхность жидкости не бывает абсолютно гладкой. Она всегда *неровная.* На этих неровностях свет претерпевает дифракцию, т.е. происходит *поверхностное молекулярное рассеяние.* Если высота неровностей мала по сравнению с длиной волны, как это имеет место в обычных условиях, то амплитуда рассеянного света обратно пропорциональна *первой*, а его интенсивность *второй степени длины волны.* Поверхностное натяжение сглаживает неровности, появившиеся из-за тепловых флуктуаций. Поэтому молекулярное поверхностное рассеяние света наиболее интенсивно на границе двух несмешивающихся жидкостей с близким коэффициентом поверхностного натяжения.

11. Молекулярное рассеяние света в кристаллах впервые было надежно установлено Г.С. Ландсбергом в 1926–1927 гг. Трудность состояла не только в том, что интенсивность рассеянного света в хороших кристаллах по предварительной оценке должна составлять всего около 10^{-8} от интенсивности падающего света. В то время вообще было не ясно, существуют ли кристаллы, в которых основную долю рассеянного света составляет свет *молекулярного рассеяния*, а не паразитный свет, возникающий при рассеянии на различных вкраплениях, микротрещинах и других дефектах кристалла. Метод, с помощью которого удалось отделить одно рассеяние от другого, состоял в исследовании *температурной зависимости интенсивности рассеянного света.* Интенсивность паразитно рассеянного света *не должна зависеть от температуры*, а молекулярно рассеянного — *возрастать с температурой.* Г.С. Ландсберг нашел, что в лучших кристаллах кварца только 25 % рассеянного света не зависит от темпе-

ратуры и, следовательно, вызвано посторонними включениями, а остальные 75 % зависят от температуры линейно, что и указывает на их молекулярное происхождение.

§ 99. Явление Мандельштама–Бриллюэна

1. Как показано в предыдущем параграфе, электромагнитное поле в оптически неоднородной среде с диэлектрической проницаемостью $\varepsilon = \varepsilon_0 + \delta\varepsilon$ может быть представлено в виде $\mathbf{E} = \mathbf{E}_0 + \mathbf{E}'$, $\mathbf{H} = \mathbf{H}_0 + \mathbf{H}'$, где $\mathbf{E}_0$, $\mathbf{H}_0$ — поле падающей, а $\mathbf{E}'$, $\mathbf{H}'$ — рассеянной волн. При слабой неоднородности можно ограничиться линейным приближением и написать

$$\begin{gathered}\operatorname{rot}\mathbf{H}' - \frac{\varepsilon_0}{c}\frac{\partial\mathbf{E}'}{\partial t} = \frac{4\pi}{c}\frac{\partial}{\partial t}\delta\mathbf{P}, \quad \operatorname{div}(\varepsilon_0\mathbf{E}') = -4\pi\operatorname{div}(\delta\mathbf{P}), \\ \operatorname{rot}\mathbf{E}' - \frac{1}{c}\frac{\partial\mathbf{H}'}{\partial t} = 0, \quad \operatorname{div}\mathbf{H}' = 0,\end{gathered} \tag{99.1}$$

где

$$\delta\mathbf{P} = \frac{\delta\varepsilon}{4\pi}\mathbf{E}_0. \tag{99.2}$$

Эти уравнения показывают, что среда может рассматриваться как *однородная* с диэлектрической проницаемостью ε_0. Влияние фактически имеющихся неоднородностей эквивалентно наличию в среде *дополнительных источников волн*: каждый элемент объема среды dV дает дополнительное излучение как *диполь Герца* с дипольным моментом $\delta\mathbf{P}\,dV$. Это дополнительное излучение и есть *рассеянный свет*.

Уравнения (99.1) *линейны и однородны* как относительно полей $\mathbf{E}'$, $\mathbf{H}'$, так и относительно $\delta\varepsilon$. Отсюда следует, что если представить $\delta\varepsilon$ в виде $\delta\varepsilon = \sum\delta_i\varepsilon$, то в линейном приближении рассеянное излучение может быть получено простой суперпозицией полей, рассеянных на неоднородностях $\delta_i\varepsilon$. Таким образом, можно рассмотреть задачу о рассеянии падающей волны сначала для случая, когда в среде имеется всего *одна неоднородность* $\delta_i\varepsilon$ какого-либо специального вида. При этом неоднородности $\delta_i\varepsilon$, суперпозицией которых представляется $\delta\varepsilon$, можно выбирать произвольно. Рассмотрим сначала случай, когда $\delta\varepsilon$ состоит всего из одного слагаемого $\delta\varepsilon = a\exp(-i\mathbf{Kr})$, где a и $\mathbf{K}$ — постоянные. Пусть падающая волна плоская и представляется выражениями

$$\mathbf{E}_0 = \mathbf{A}e^{i(\omega t - \mathbf{kr})}, \quad \mathbf{H}_0 = \mathbf{B}e^{i(\omega t - \mathbf{kr})}.$$

Посмотрим, при каких длинах волн $\lambda = 2\pi/k$ и в каких направлениях будет наблюдаться рассеянное излучение.

Разобьем среду равноотстоящими плоскостями, перпендикулярными к вектору $\mathbf{K}$ (рис. 322). Выберем расстояние между

плоскостями равным $\Lambda = 2\pi/K$. Тогда, согласно (99.2), фазы вторичных источников на этих равноотстоящих плоскостях будут одинаковы. Если бы неоднородность была только в слое I, а дальше среда была однородна, то падающая волна претерпела бы отражение от этого слоя и частично прошла бы дальше. При наличии неоднородности только в слое II мы получили бы другую отраженную волну с той же амплитудой, но иной фазой. При наличии неоднородности в слое III получилась бы третья отраженная волна, и т.д. В линейном приближении поле рассеяния всей среды равно простой суперпозиции этих отраженных волн. Чтобы они не гасили, а усиливали друг друга, необходимо выполнение *условия Брэгга–Вульфа*: $2\Lambda \sin(\theta/2) = m\lambda$, где θ — угол рассеяния, т.е. угол между направлениями падающего и рассеянного излучений, а m — целое число (порядок дифракционного спектра).

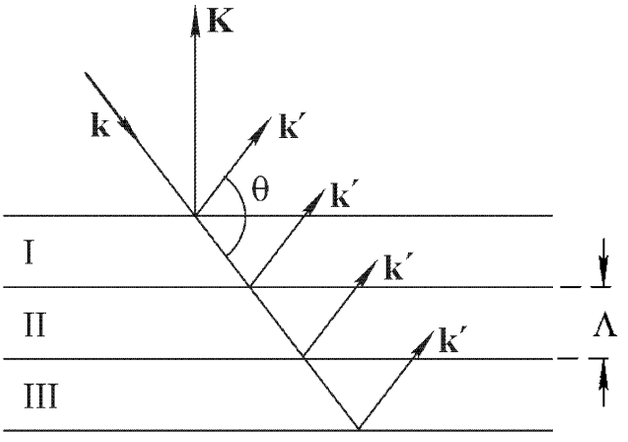


Рис. 322

Покажем, что $m = 1$. Все плоские волны, отраженные различными слоями, складываясь, дают волну вида $\mathbf{E}' = \mathbf{A}'e^{i(\omega t - \mathbf{k}'\mathbf{r}')}$, где волновой вектор $\mathbf{k}'$ определяет направление распространения отраженных волн. С другой стороны, дополнительная поляризация среды

$$\delta\mathbf{P} = \frac{\mathbf{E}_0}{4\pi}\,\delta\varepsilon = \frac{a\mathbf{A}}{4\pi}\,e^{i[\omega t - (\mathbf{k}+\mathbf{K})\mathbf{r}]}.$$

Подставляя эти выражения во второе уравнение (99.1) и сравнивая показатели, легко получить $\mathbf{k}' - \mathbf{k} = \mathbf{K}$, откуда

$$2\Lambda \sin(\theta/2) = \lambda. \tag{99.3}$$

Таким образом, при дифракции волны на синусоидальной неоднородности диэлектрической проницаемости в линейном приближении получается дифракционный спектр *только первого порядка.*

Любую неоднородность в среде можно по теореме Фурье представить *в виде суперпозиции плоских синусоидальных неоднородностей различных направлений.* Согласно доказанному выше такие синусоидальные неоднородности рассеивают свет *независимо* друг от друга. Но при фиксированном направлении рассеянного излучения эффективны не все синусоидальные неоднородности, а только такие, волновой вектор $\mathbf{K}$ которых

направлен *по биссектрисе угла, дополнительного к* θ до 180° (рис. 322). Остальные синусоидальные неоднородности для рассеяния в рассматриваемом направлении не играют роли. Мы видим, что механизм рассеяния света на неоднородностях диэлектрической проницаемости вполне аналогичен механизму рассеяния рентгеновских лучей в кристаллах в той форме, в какой он был представлен Вульфом и Брэггом (см. § 61).

2. До сих пор мы принимали во внимание изменения функции $\delta\varepsilon$ *в пространстве*, но не учитывали ее изменения *во времени*. Учет последнего обстоятельства приводит к новому явлению в рассеянии света. Считая, как и в предыдущем параграфе, ε функцией только плотности ρ, напишем в линейном приближении $\Delta\varepsilon = (d\varepsilon/d\rho)\Delta\rho$. Всякая неоднородность плотности, возникшая в среде, является *источником звуковых волн*. Разложим $\Delta\rho$ в интеграл или ряд Фурье и возьмем в этом разложении только те звуковые волны, которые существенны для рассеяния волн в рассматриваемом направлении. Их волновой вектор $\mathbf{K}$ был определен выше. Этому значению $\mathbf{K}$ соответствует определенная звуковая частота Ω и два направления распространения звуковой волны: вдоль $\mathbf{K}$ и против $\mathbf{K}$. Неоднородность $\delta\varepsilon$, вызывающая рассеяние света в рассматриваемом направлении, представится суммой $\delta\varepsilon = \delta\varepsilon_1 + \delta\varepsilon_2$, где $\delta\varepsilon_1$ и $\delta\varepsilon_2$ имеют вид плоских звуковых волн:

$$\delta\varepsilon_1 = a_1 e^{i(\Omega t - \mathbf{Kr})} \quad \text{и} \quad \delta\varepsilon_2 = a_2 e^{-i(\Omega t + \mathbf{Kr})}.$$

Им соответствуют векторы дополнительной поляризации среды:

$$\delta\mathbf{P}_1 = \frac{\mathbf{E}_0}{4\pi}\,\delta\varepsilon_1 = \frac{a_1\mathbf{A}}{4\pi}\,e^{i[(\omega+\Omega)t - (\mathbf{k}+\mathbf{K})\mathbf{r}]},$$

$$\delta\mathbf{P}_2 = \frac{a_2\mathbf{A}}{4\pi}\,e^{i[(\omega-\Omega)t - (\mathbf{k}+\mathbf{K})\mathbf{r}]}.$$

Таким образом, источники рассеянного излучения, а значит и само рассеянное излучение, будут меняться во времени с частотами $\omega + \Omega$ и $\omega - \Omega$ (модуляция световой волны акустической волной). В спектре рассеянного излучения должен наблюдаться дублет с теми же частотами. Это явление называется *тонкой структурой линий рэлеевского рассеяния* или *рассеянием Мандельштама–Бриллюэна*. Смещение частоты равно $\Omega = Kv = (2\pi/\Lambda)v$, где v — скорость звука, а Λ — длина звуковой волны. На основании (99.3)

$$\Omega = \frac{4\pi v}{\lambda}\sin\frac{\theta}{2} = 2\omega n\frac{v}{c}\sin\frac{\theta}{2}, \tag{99.4}$$

где c — скорость света в вакууме, а n — показатель преломления среды.

Дублет Мандельштама–Бриллюэна можно трактовать как *доплеровское изменение частоты света* при отражении от аку-

стической волны. Когда акустическая волна распространяется навстречу световой, происходит увеличение частоты света, в противоположном случае — уменьшение. Доплеровское изменение частоты определяется формулой

$$\frac{\Omega}{\omega} = \frac{2v \sin \theta/2}{c/n},$$

откуда и получается формула (99.4).

Представление о тепловом движении как о звуковых волнах всевозможных частот и направлений распространения было введено Дебаем (1884–1966) в его теории теплоемкости твердых тел. К ним Дебай применял методы статистической физики. Это — те же волны, которые вызывают рассеяние света и дублет Мандельштама–Бриллюэна.

3. Тонкая структура линий рэлеевского рассеяния была предсказана независимо друг от друга Л.И. Мандельштамом и Л. Бриллюэном. По свидетельству Г.С. Ландсберга, Л.И. Мандельштам выполнил свою работу еще в 1918 г., хотя краткая заметка о ней появилась значительно позже, в 1926 г., когда часть найденных Д.И. Мандельштамом результатов была уже опубликована Бриллюэном (1922 г.). Мандельштам и Ландсберг пытались на опыте обнаружить предсказанное явление при рассеянии света в кварце. Качественно им удалось констатировать существование явления. Однако недостаточная разрешающая способность их спектральной аппаратуры не позволяла исследовать его количественно. Кроме того, эти опыты привели их к открытию *комбинационного рассеяния света* (см. § 100). Естественно, что их внимание переключилось на исследование этого более важного явления. По их предложению исследованием тонкой структуры рэлеевского рассеяния занялся Е.Ф. Гросс (1897–1972) в Ленинграде.

Гросс обнаружил явление при рассеянии света *в жидкостях* [1]). Однако оказалось, что в жидкостях, наряду с двумя смещенными компонентами, наблюдается также и *несмещенная компонента*. Происхождение несмещенной компоненты было объяснено Ландау (1908–1968) и Плачеком (1905–1955).

[1]) Это было неожиданно, так как, согласно гидродинамической теории, поглощение звука в жидкостях пропорционально квадрату частоты ω. Если бы гидродинамическая теория была верна без ограничений, то звуковые волны оптических частот в жидкостях распространяться не могли бы. Обнаружение тонкой структуры в жидкостях послужило поводом Л.И. Мандельштаму и М.А. Леонтовичу (р. 1903) к разработке релаксационной молекулярной теории вязкости жидкостей и основанной на ней теории поглощения звука.

Рассматривая удельный объем жидкости V как функцию давления и энтропии, можно написать

$$\delta V = (\partial V/\partial P)_S\,\delta P + (\partial V/\partial S)_P\,\delta S. \qquad (99.5)$$

Отсюда видно, что существует два вида флуктуаций удельного объема: одни вызваны *флуктуациями давления* при постоянной энтропии, другие — *флуктуациями энтропии* при постоянном давлении. Флуктуации первого типа распространяются в виде *акустических волн* и ведут к появлению *смещенных компонент*. Флуктуационные неоднородности второго типа выравниваются посредством *теплопроводности*, а следовательно, распространяются значительно более медленно, — они и ведут к появлению в рассеянном свете *несмещенной компоненты*.

Для количественного исследования заметим, что процессы рассеяния света на флуктуациях давления и энтропии *некогерентны*. Поэтому интегральные интенсивности несмещенной I_ω и смещенных $I_{\omega-\delta\omega}$, $I_{\omega+\delta\omega}$ компонент связаны соотношением

$$\frac{I_\omega}{I_{\omega-\delta\omega}+I_{\omega+\delta\omega}} = \frac{(\partial V/\partial s)_P^2\overline{\Delta s^2}}{(\partial V/\partial P)_s^2\overline{\Delta P^2}} = -\left(\frac{\partial V}{\partial s}\right)_P^2\left(\frac{\partial s}{\partial T}\right)_P\left(\frac{\partial P}{\partial V}\right)_s,$$

где использованы выражения (97.22) и (97.24), а также формула $c_P = T(\partial s/\partial T)_P$, причем малой буквой s обозначена удельная энтропия. Так как дифференциал удельной энтальпии $di = T\,ds + V\,dP$ — полный дифференциал, то $(\partial T/\partial P)_s = (\partial V/\partial s)_P$. Поэтому

$$\frac{I_\omega}{I_{\omega-\delta\omega}+I_{\omega+\delta\omega}} = -\left(\frac{\partial T}{\partial P}\right)_s\left(\frac{\partial P}{\partial V}\right)_s\left(\frac{\partial V}{\partial s}\right)_P\left(\frac{\partial s}{\partial T}\right)_P = -\left(\frac{\partial T}{\partial V}\right)_s\left(\frac{\partial V}{\partial T}\right)_P,$$

или, на основании тождества $(\partial T/\partial V)_s(\partial V/\partial s)_T(\partial s/\partial T)_V = -1$,

$$\frac{I_\omega}{I_{\omega-\delta\omega}+I_{\omega+\delta\omega}} = \left(\frac{\partial V}{\partial T}\right)_P\left(\frac{\partial s}{\partial V}\right)_T\left(\frac{\partial T}{\partial s}\right)_V = \frac{(\partial V/\partial T)_P(\partial s/\partial V)_T}{(\partial s/\partial T)_V}.$$

Рассматривая энтропию s как функцию T и V, получим

$$\left(\frac{\partial s}{\partial T}\right)_P = \left(\frac{\partial s}{\partial T}\right)_V + \left(\frac{\partial s}{\partial V}\right)_T\left(\frac{\partial V}{\partial T}\right)_P.$$

Окончательно:

$$\frac{I_\omega}{I_{\omega-\delta\omega}+I_{\omega+\delta\omega}} = \frac{(\partial s/\partial T)_P-(\partial s/\partial T)_V}{(\partial s/\partial T)_V} = \frac{c_P-c_v}{c_v}. \qquad (99.6)$$

Эта формула была получена Ландау и Плачеком.

В аморфных твердых телах звуковые волны могут быть *продольными* и *поперечными*. Они распространяются с различными скоростями. Поэтому в рассеянном свете спектральная линия должна расщепляться на *пять компонент*: одну несмещенную и две пары смещенных компонент, из которых одна пара получается от рассеяния на продольных акустических волнах, а другая — на поперечных.

В.В. Владимирский (р. 1915) указал, что в кристаллах в общем случае спектральная линия неполяризованного света должна расщепляться на 25 *компонент*: одну несмещенную и 24 смещенные. Дело в том, что в кристалле в каждом направлении могут распространяться одна продольная акустическая волна и две поперечные. В том же направлении могут распространяться две световые волны, поляризованные во взаимно перпендикулярных плоскостях. Каждая из этих световых волн в свою очередь расщепляется на две волны при отражении от акустических волн соответствующих направлений распространения. Это и приводит к появлению в рассеянном свете 24 смещенных компонент. Однако из-за слабой анизотропии всех исследованных кристаллов эти 24 компоненты обычно группируются в шесть групп по четыре линии в каждой и не разрешаются спектральными приборами. На опыте наблюдаются *шесть смещенных компонент.*

4. С изобретением лазеров стала возможной генерация мощных (так называемых гигантских) световых импульсов, оказывающих существенное воздействие на среду, в которой распространяется свет. В переменном электрическом поле $\mathbf{E}$ возникает электрострикционное давление

$$\mathcal{P} = \frac{1}{8\pi}\left(\rho\frac{d\varepsilon}{d\rho}\right)E^2 \tag{99.7}$$

(см. т. III, § 32). Величина $\rho\, d\varepsilon/d\rho$ порядка единицы. В слабых световых полях, с которыми имеет дело линейная оптика, давление $\mathcal{P}$ ничтожно и его влиянием на среду можно полностью пренебречь. Но в световом поле гигантского лазерного импульса это давление может достигать сотни тысяч атмосфер. Тогда световые и акустические волны в среде надо рассматривать *совместно.* Они описываются сложной системой взаимосвязанных *нелинейных уравнений* электродинамики и акустики. Это приводит к ряду *нелинейных оптических явлений.* Одним из них является *вынужденное рассеяние Мандельштама–Бриллюэна.* Хотя нелинейные оптические явления будут разбираться в главе XI, возникновение вынужденного рассеяния Мандельштама–Бриллюэна удобнее разобрать уже здесь.

Пусть $\mathbf{E}_0 = \mathbf{A}_0\cos(\omega t - \mathbf{kr})$, $\mathbf{E}_1 = \mathbf{A}_1\cos[(\omega+\Omega)t - \mathbf{k}'\mathbf{r}]$, $\mathbf{E}_2 = \mathbf{A}_2\cos[(\omega-\Omega)t - \mathbf{k}'\mathbf{r}]$, $\mathbf{E}_3 = \mathbf{A}_3\cos[\omega t - \mathbf{k}'\mathbf{r}]$ — напряженности электрического поля падающей и трех рассеянных волн Мандельштама–Бриллюэна. Последние три волны возникают при рассеянии на тепловых флуктуациях. Интенсивности их сначала малы, но в дальнейшем могут усилиться за счет взаимодействия с падающей волной. Электрострикционное давление $\mathcal{P}$ определяется квадратом суммы всех полей, т.е. $(\mathbf{E}_0+\mathbf{E}_1+{}$ $+\mathbf{E}_2+\mathbf{E}_3)^2$. При возведении в квадрат представим по известным формулам тригонометрии квадраты и произведения косинусов в

виде сумм постоянных членов и косинусов суммарных и разностных аргументов. Постоянные члены для возбуждения звуковых волн не играют роли. Не имеют значения и члены с косинусами от суммарных аргументов. Это — *высокочастотные члены*, меняющиеся во времени с оптическими частотами, а звуковые волны быстро затухают с увеличением частоты. Возбуждение звуковых волн связано только с *низкочастотными членами*, содержащими косинусы разностных аргументов. Выпишем все эти члены, опуская при этом числовые коэффициенты и принимая во внимание соотношение $\mathbf{K} = \mathbf{k}' - \mathbf{k}$ (см. рис. 322). Получим

$$\mathbf{A}_0\mathbf{A}_1 \cos(\Omega t - \mathbf{Kr}), \quad \mathbf{A}_0\mathbf{A}_2 \cos(\Omega t + \mathbf{Kr}), \quad \mathbf{A}_0\mathbf{A}_3 \cos \mathbf{Kr},$$
$$\mathbf{A}_1\mathbf{A}_2 \cos 2\Omega t, \quad \mathbf{A}_1\mathbf{A}_3 \cos \Omega t, \quad \mathbf{A}_2\mathbf{A}_3 \cos \Omega t.$$

Из этих членов имеет значение *только первый*. Он представляет волну, распространяющуюся в том же направлении и с той же фазой, что и первичная звуковая волна, возникшая из-за тепловых флуктуаций. Поэтому будет происходить *параметрическое усиление* этой акустической волны и всех световых волн, рассеянных на ней. Такой процесс усиления будет продолжаться до тех пор, пока интенсивность рассеянного света не станет сравнимой с интенсивностью падающего. Это действительно и наблюдается на опыте. В отличие от некогерентного рассеяния на тепловых флуктуациях, *вынужденное рассеяние Мандельштама–Бриллюэна когерентно.*

ЗАДАЧА

При рассеянии света резонансной линии ртутной лампы ($\lambda = 253{,}65$ нм) в кристалле алмаза под углом $\theta = 90^\circ$ к направлению падающего пучка были найдены две пары смещенных компонент с $\delta\lambda = 0{,}052$ нм и $\delta\lambda = 0{,}032$ нм. (Речь идет о смещении относительно центральной — несмещенной — компоненты.) Определить скорости продольной и поперечной акустических волн в алмазе. Показатель преломления алмаза $n = 2{,}42$.

О т в е т. $v = \dfrac{c}{2n \sin(\theta/2)} \dfrac{\delta\lambda}{\lambda}$; $v_{\text{прод}} = 18\,000$ м/с; $v_{\text{попереч}} = 11\,000$ м/с.

Благодаря большой скорости звука в алмазе тонкую структуру линий рэлеевского рассеяния удается исследовать даже с помощью призменных спектрографов.

§ 100. Комбинационное рассеяние света

1. При спектральных исследованиях рассеяния света в кварце и исландском шпате (февраль 1928 г.) Мандельштам и Ландсберг обнаружили, что каждая спектральная линия падающего света сопровождается появлением системы линий измененной

частоты, называемых *сателлитами* (спутниками). Практически одновременно то же явление было открыто в Индии Раманом (1888–1970) и Кришнаном (1898–1961) при исследовании рассеяния света в жидкостях. Изменение длины волны оказалось значительно больше, чем при рассеянии Мандельштама–Бриллюэна. Это явление называется *комбинационным рассеянием света* или *эффектом Рамана* (который раньше послал в печать сообщение о своем открытии). Приведем основные законы комбинационного рассеяния, установленные экспериментально.

1) Частоты сателлитов отличаются от частоты возбуждающей их линии на $\Delta\omega^j_{\text{комб}}$, где j — номер сателлита, так что различным сателлитам соответствуют различные $\Delta\omega_{\text{комб}}$. При переходе от одной спектральной линии первичного пучка к другой совокупность значений $\Delta\omega^j_{\text{комб}}$ остается *одной и той же*. Она характерна для рассматриваемого вещества и меняется только при переходе от одного вещества к другому. Это обстоятельство используется в спектральном анализе методом комбинационного рассеяния света. Обычно частоты спектральных линий и их изменения принято характеризовать числом волн, укладывающихся в 1 см. Их мы будем обозначать через ν и $\Delta\nu$, полагая $\nu = 1/\lambda$. Таблица 11 показывает, насколько хорошо по измерени-

Таблица 11

Частота основной линии ν, см^{-1}	Бензол C_6H_6	Хлороформ $CHCl_3$		Четыреххлористый углерод CCl_4	
	$\Delta\nu_{\text{красн}}$	$\Delta\nu_{\text{красн}}$	$\Delta\nu_{\text{фиол}}$	$\Delta\nu_{\text{красн}}$	$\Delta\nu_{\text{фиол}}$
33 692	992,9	—	—	—	—
33 086	991,3	667,7	—	459,7	—
31 985	992,9	667,7	669,3	458,3	459,4
31 923	992,3	667,3	669,5	457,8	460,4
29 917	—	668,1	670,5	458,7	458,8
27 388	991,2	666,6	—	458,4	461,2
27 354	991,3	666,8	669,9	459,4	461,8
27 290	991,1	—	669,8	459,1	461,8
24 705	992,2	669,2	669,5	457,9	459,7
24 515	—	667,5	—	457,8	—
23 034	992,9	667,3	—	—	—
22 994	992,1	670,8	—	457,8	—
22 938	992,3	669,4	667,1	459,6	457,8
Среднее	992,0	668,0	669,4	458,5	460,2

ям Г.С. Ландсберга соблюдается постоянство $\Delta\nu$ для различных длин волн в комбинационном рассеянии света.

2) Каждому сателлиту с частотой $\omega - \Delta\omega_{\text{комб}}$, смещенной в красную сторону спектра, соответствует сателлит с частотой $\omega + \Delta\omega_{\text{комб}}$, смещенной в фиолетовую сторону. Первые сателлиты называются *красными* или *стоксовыми*, вторые — *фиолетовыми* или *антистоксовыми*.

3) Число различных сателлитов (т.е. число значений индекса j) и их относительная интенсивность при одной и той же температуре также зависят от рассеивающего вещества. Но *интенсивности фиолетовых сателлитов значительно меньше интенсивностей соответствующих им красных сателлитов и проявляют общую тенденцию к ослаблению по мере увеличения* $\Delta\nu_{\text{комб}}$.

4) Постоянные $\Delta\omega_{\text{комб}}$, *характерные для рассматриваемого вещества, совпадают с собственными частотами* $\Omega_{\text{инфр}}$ *инфракрасных колебаний того же вещества*, хотя не всем известным $\Omega_{\text{инфр}}$ находятся соответствующие $\Delta\omega_{\text{комб}}$, и наоборот. Кроме того, нередко интенсивной линии комбинационного рассеяния соответствует слабая линия инфракрасного поглощения и наоборот.

5) Линии комбинационного рассеяния света более или менее *поляризованы*. Степень поляризации различных сателлитов одной и той же линии не одинакова и не находится в прямой связи с поляризацией основной линии рассеянного света. Характер поляризации красных и фиолетовых сателлитов, соответствующих данному значению $\Delta\omega_{\text{комб}}$, *всегда одинаков и не зависит от частоты основной линии*.

2. Явление комбинационного рассеяния света было объяснено сразу же Мандельштамом и Ландсбергом, когда они открыли это явление.

В поле световой волны $\mathbf{E}$ электроны внутри молекулы приходят в колебания, и молекула приобретает индуцированный дипольный момент $\mathbf{p} = \beta\mathbf{E}$. С классической точки зрения тензор поляризуемости молекулы β определяется мгновенными положениями ее атомных ядер. Но сами ядра не находятся в покое, а совершают беспорядочное тепловое движение. По этой причине поляризуемость β не остается постоянной, а меняется во времени. Ее можно представить наложением гармонических колебаний, частоты которых определяются колебаниями атомных ядер, т.е. совпадают с собственными частотами инфракрасных колебаний молекулы. Возникает *модуляция колебаний* индуцированных дипольных моментов $\mathbf{p}$.

Если внешнее электрическое поле $\mathbf{E}$ меняется во времени гармонически с частотой ω, то в колебаниях дипольного момента $\mathbf{p}$ появятся комбинационные частоты $\omega \pm \Omega_{\text{инфр}}$. Такие же частоты

появятся и в излучении этих дипольных моментов, т.е. в рассеянном свете. Таково с классической точки зрения происхождение комбинационного рассеяния.

Изложенные соображения можно облечь в математическую форму. Если число ядер в молекуле равно s, то они обладают $3s$ степенями свободы. Из них три степени свободы *поступательные*, а три *вращательные*. Остальные $f = 3s - 6$ степеней свободы приходятся на *внутреннее движение ядер молекулы*. Для описания внутреннего движения ядер требуется f координат q_1, q_2, ..., q_j. Выбор координат произволен. Удобнее всего взять *нормальные обобщенные координаты*.

В положении равновесия ядер все координаты равны нулю. При малых отклонениях из положений равновесия каждая координата q_m при тепловом движении совершает свободное гармоническое колебание $q_m = a_m \cos(\Omega_m t + \delta_m)$ с инфракрасной частотой Ω_m и хаотически меняющейся фазой δ_m. В силу малости колебаний тензор β можно разложить в ряд и ограничиться первыми степенями по q. Считая для простоты β скаляром, получим

$$\beta = \beta_0 + \sum \left(\frac{\partial \beta}{\partial q_m}\right) q_m,$$

или после подстановки значений q_m

$$\beta = \beta_0 + \sum \left(\frac{\partial \beta}{\partial q_m}\right) a_m \cos(\Omega_m t + \delta_m) =$$

$$= \beta_0 + \frac{1}{2} \sum \left(\frac{\partial \beta}{\partial q_m}\right) a_m e^{i(\Omega_m t + \delta_m)} + \frac{1}{2} \sum \left(\frac{\partial \beta}{\partial q_m}\right) a_m e^{-i(\Omega_m t + \delta_m)}.$$

Падающую волну запишем в комплексной форме $\mathbf{E} = \mathbf{E}_0 e^{i\omega t}$. Тогда

$$\mathbf{p} = \beta_0 \mathbf{E}_0 e^{i\omega t} +$$

$$+ \frac{\mathbf{E}_0}{2} \sum \left(\frac{\partial \beta}{\partial q_m}\right) a_m e^{i[(\omega + \Omega_m)t + \delta_m]} + \frac{\mathbf{E}_0}{2} \sum \left(\frac{\partial \beta}{\partial q_m}\right) a_m e^{i[(\omega - \Omega_m)t - \delta_m]}.$$

Отсюда видно, что в рассеянном излучении появляется свет *не только с исходной частотой* ω, *но и с комбинированными частотами* $\omega \pm \Omega_m$. Ясно, что волны, рассеиваемые отдельными молекулами, *некогерентны*, так как при тепловом возбуждении колебаний ядер фазы δ_m меняются нерегулярно при переходе от одной молекулы к другой и от одного колебания к другому.

3. Классическая теория комбинационного рассеяния, как и всякого явления, связанного с излучением и поглощением света, конечно, недостаточна. Успехи этой теории связаны с *относительно большими* массами атомных ядер, благодаря чему многие особенности их колебаний правильно передаются уравнениями классической механики. Однако классическая теория

объясняет не все закономерности комбинационного рассеяния света. В частности, она не может объяснить соотношения между интенсивностями соответственных красных и фиолетовых сателлитов. По классической теории эти интенсивности должны быть *практически одинаковы.* Опыт же показывает, что *интенсивность красных сателлитов всегда больше интенсивности соответственных фиолетовых сателлитов.*

Квантовая теория дает естественное объяснение этой закономерности. Согласно квантовой механике энергетические уровни молекулы образуют *дискретный ряд* $\mathcal{E}_1$, $\mathcal{E}_2$, $\mathcal{E}_3$, ... Рассеяние фотона на молекуле аналогично процессу столкновения его с молекулой, к которому применим *закон сохранения энергии.* В таком процессе фотон может либо передать часть своей энергии молекуле, либо, наоборот, получить энергию от возбужденной молекулы. Согласно соотношению $\mathcal{E}_{\text{фот}} = \hbar\omega$, изменение энергии фотона проявляется в *изменении его частоты.* Пусть фотон с энергией $\hbar\omega$ рассеивается на молекуле с энергией $\mathcal{E}_n$. В результате получается рассеянный фотон с энергией $\hbar\omega'$, а молекула переходит на энергетический уровень $\mathcal{E}_m$. По закону сохранения энергии $\hbar\omega + \mathcal{E}_n = \hbar\omega' + \mathcal{E}_m$, откуда

$$\omega' = \omega + \Omega_{nm}, \tag{100.1}$$

где

$$\Omega_{nm} = (\mathcal{E}_n - \mathcal{E}_m)/\hbar. \tag{100.2}$$

Если $\mathcal{E}_n > \mathcal{E}_m$, то частота рассеянного фотона больше, чем падающего, т.е. в рассеянном свете появляется *фиолетовый сателлит.* Если же $\mathcal{E}_n < \mathcal{E}_m$, то при рассеянии возникает *красный сателлит.*

Пусть $\mathcal{E}_n < \mathcal{E}_m$. Тогда красный сателлит появится, когда исходным уровнем является уровень $\mathcal{E}_n$, а фиолетовый — когда исходным будет уровень $\mathcal{E}_m$. Для отношения интенсивностей сателлитов можно написать $I_{\text{кр}}/I_{\text{фиол}} = N_n/N_m$, где N_n — число молекул на уровне $\mathcal{E}_n$, а N_m — на уровне $\mathcal{E}_m$. Согласно формуле Больцмана, при тепловом равновесии

$$\frac{N_n}{N_m} = \exp\left(-\frac{\mathcal{E}_n - \mathcal{E}_m}{kT}\right) = \exp\frac{\hbar|\Omega_{nm}|}{kT},$$

а потому

$$\frac{I_{\text{кр}}}{I_{\text{фиол}}} = \exp\frac{\hbar|\Omega_{nm}|}{kT}. \tag{100.3}$$

Эта формула вполне объясняет наблюдаемое соотношение интенсивностей.

4. В мощных импульсах лазерного излучения наблюдается нелинейное явление, называемое *вынужденным комбинационным рассеянием света.* Оно возникает из-за *обратного воздействия световой волны на молекулы среды.* В неоднородном

электрическом поле $\mathbf{E}$ на молекулу с дипольным моментом $\mathbf{p}$ действует сила $\mathbf{F} = (\mathbf{p}\nabla)\mathbf{E}$. Силы такого рода действуют и на *части молекулы*, так как всякая электрически нейтральная часть молекулы, состоящая, например, из ядра и электрона, обладает дипольным моментом. Индуцированные дипольные моменты $\mathbf{p}$ пропорциональны полю $\mathbf{E}$, так что все эти силы *квадратично* зависят от поля. Поле $\mathbf{E}$ слагается из поля падающей волны $\mathbf{E}_0$ и поля рассеянных волн $\mathbf{E}'$. Первоначально поле $\mathbf{E}'$ слабое, так как оно возникает из-за тепловых флуктуаций в среде. Но затем оно может усиливаться из-за взаимодействия с падающей волной. Среди слагающих сил

$$[(\mathbf{E}_0 + \mathbf{E}')\nabla](\mathbf{E}_0 + \mathbf{E}')$$

присутствуют члены с произведением полей $\mathbf{E}_0$ и $\mathbf{E}'$, частоты которых совпадают с соответствующими частотами инфракрасных колебаний молекулы. Они вызовут *резонансное усиление* таких инфракрасных колебаний, что поведет к увеличению интенсивности соответствующих линий комбинационного рассеяния. Вынужденные колебания ядер молекулы происходят в фазе с падающей волной, а потому, в отличие от обычного (теплового) комбинационного рассеяния, *вынужденное комбинационное рассеяние когерентно с падающей волной.*

ГЛАВА IX

ТЕОРИЯ ОТНОСИТЕЛЬНОСТИ

§ 101. Принцип относительности в ньютоновской и релятивистской механике

1. *Теория относительности — фундаментальная физическая теория, охватывающая всю физику.* Она возникла в начале XX века в результате преодоления принципиальных трудностей, с которыми столкнулась электродинамика и оптика движущихся тел. Настойчивые и мучительные попытки преодолеть эти трудности на основе гипотезы мирового эфира окончились неудачей. Теория относительности отказалась от использования гипотезы эфира. В ее основе лежат не гипотезы, а *принципы* или *постулаты*, твердо установленные экспериментально. В этом сила теории относительности, причина ее успехов. В нашу задачу не входит подробное изложение истории теории относительности и экспериментального обоснования ее постулатов. Мы коснемся этих вопросов предельно кратко, чтобы лучше оттенить принципиальные моменты, глубже уяснить смысл постулатов и содержание теории относительности.

Творцом теории относительности является Альберт Эйнштейн. Теория относительности была изложена им в 1905 г. в основополагающей работе «К электродинамике движущихся тел». Многие результаты этой работы были получены ранее Лармором (1857–1942), Лоренцом и Пуанкаре (1854–1912). Однако и Лармор и Лоренц принципиально стояли на точке зрения неподвижного эфира, с которым связывалась преимущественная система отсчета. Ближе всего к теории относительности был Пуанкаре, который еще в 1898 г. дал критику понятия одновременности пространственно разделенных событий, повторенную в дальнейшем Эйнштейном. Пуанкаре заполнил также математические пробелы и устранил ошибки, допущенные Лоренцом. Однако принципиально новое и глубокое физическое понимание всей проблемы и последовательное построение тео-

рии относительности с единой точки зрения содержится лишь в упомянутой выше работе Эйнштейна, написанной к тому же без всякого влияния своих предшественников.

2. Начнем с повторения того, что было сказано в первом томе о принципе относительности в ньютоновской механике. Основной закон ньютоновской механики выражается уравнением

$$m\ddot{\mathbf{r}} = \mathbf{F}, \tag{101.1}$$

где $\mathbf{F}$ — сила, действующая на материальную точку, а $\mathbf{r}$ — радиус-вектор, определяющий положение последней относительно какой-либо инерциальной системы отсчета. Возьмем две инерциальные системы отсчета, одну из которых будем обозначать через S и называть *неподвижной* (или *нештрихованной*), а другую — через S', называя ее *движущейся* (или *штрихованной*). Пусть система S' движется относительно системы S равномерно и прямолинейно со скоростью $\mathbf{V}$. Если $\mathbf{r}$ и $\mathbf{r}'$ — радиусы-векторы, определяющие положения движущейся материальной точки относительно этих систем отсчета в момент времени t, то они связаны между собой *преобразованием Галилея*

$$\mathbf{r} = \mathbf{r}' + \mathbf{V}t. \tag{101.2}$$

При этом в ньютоновской механике предполагается, что время t *абсолютно*, т.е. одинаково во всех системах отсчета. Для простоты отсчет времени ведется с того момента, когда начала координат систем S и S' совмещаются между собой. Это не отражается на общности рассуждений.

Дважды дифференцируя соотношение (101.2) по времени, находим *формулы преобразования скорости и ускорения*:

$$\mathbf{v} = \mathbf{v}' + \mathbf{V}, \quad \mathbf{a} = \mathbf{a}'. \tag{101.3}$$

Таким образом, *ускорение инвариантно относительно преобразования Галилея*, т.е. одинаково в обеих системах отсчета S и S'. Радиусы-векторы $\mathbf{r}$ и $\mathbf{r}'$, скорости $\mathbf{v}$ и $\mathbf{v}'$ не одинаковы. Однако *разности радиусов-векторов, а также скоростей двух любых материальных точек одни и те же*, поскольку они определяют *относительные положения и скорость* одной точки относительно другой. Сила $\mathbf{F}$ в ньютоновской механике зависит только от *разностей радиусов-векторов и скоростей* взаимодействующих материальных точек. Поэтому она, а с ней и уравнение Ньютона (101.1) не меняются при преобразовании Галилея. То же относится и к дифференциальным уравнениям движения систем материальных точек в механике Ньютона. Таким образом, получается следующий результат:

Уравнения механики Ньютона, определяющие изменения состояния движения механических систем, инвариантны относительно преобразования Галилея.

Это положение называется *принципом относительности Галилея*. Ему можно дать также следующую формулировку:

Законы природы, определяющие изменения состояния движения механических систем, не зависят от того, к какой из двух инерциальных систем отсчета, движущихся одна относительно другой прямолинейно и равномерно, они относятся.

Дорелятивистская физика считала обе формулировки тождественными, поскольку при равномерном поступательном движении инерциальных систем отсчета относительно друг друга она не допускала никакого другого преобразования $\mathbf{r}$ и t, кроме преобразования Галилея. На самом деле вторая формулировка *более общая*, чем предыдущая, так как в ней не конкретизирован вид того преобразования координат и времени, относительно которого инвариантны уравнения механики. В дальнейшем, если не сделано специальной оговорки, принцип относительности понимается во *второй формулировке*.

3. Принцип относительности отнюдь не утверждает, что одно и то же физическое явление выглядит одинаково в различных инерциальных системах отсчета. Дело в том, что одни только дифференциальные уравнения механики не определяют движение системы. К ним необходимо присоединить еще *начальные условия*, например задать координаты и скорости всех взаимодействующих частиц в определенный момент времени. А эти начальные условия меняются при переходе от одной системы отсчета к другой. Именно из-за различия начальных условий движение предмета, свалившегося с полки равномерно движущегося вагона, происходит вниз по прямой линии, если его рассматривать относительно самого вагона, тогда как относительно полотна железной дороги то же движение совершается по параболе. Вот почему в формулировке принципа относительности говорится не об одинаковости явлений, а об *одинаковости законов*, определяющих изменение состояний движения механических систем.

Смысл принципа относительности можно также уяснить на следующем примере. Возьмем замкнутую систему тел A и зададим их начальные положения и скорости относительно инерциальной системы отсчета S. Пусть имеется тождественная с A другая замкнутая система тел A', в которой созданы в точности такие же начальные условия, но уже относительно инерциальной системы отсчета S'. Тогда движение в системе тел A относительно S будет тождественным с движением в системе тел A' относительно S'. В этом и состоит *равноправие* инерциальных систем отсчета, устанавливаемое принципом относительности.

Принципиально взаимодействие имеет место между любыми телами Вселенной. Поэтому, если бы и существовали замкнутые системы, то единственной из таких систем могла бы быть, строго

говоря, только вся Вселенная. Но тогда принцип относительности в том смысле, какой ему придан в последнем примере, был бы бессодержательным, так как двух Вселенных не существует. Принцип относительности только потому сохраняет содержание, что многими взаимодействиями, ввиду их слабости, можно пренебречь и таким путем выделить практически бесконечное множество ограниченных систем тел, каждая из которых приближенно ведет себя как *замкнутая*. В какой степени та или иная система удовлетворяет этому условию, зависит от исследуемого вопроса и от точности, какая предъявляется при его изучении. Достаточно ли для (приближенной) замкнутости системы отдаленности всех масс, не входящих в рассматриваемую систему? Ответ в соответствии с опытом гласит, что в случае *равномерного и прямолинейного движения* (относительно инерциальной системы отсчета) этого *достаточно*, а для всех остальных видов движения *недостаточно*.

4. Явления природы не представляется возможным разделить на чисто «механические» и «немеханические». Если бы даже это и можно было сделать, то принцип относительности не мог бы относиться к одним только «механическим» явлениям. Действительно, всякое «механическое» явление связано с множеством «физических» явлений и обусловлено ими. И если бы принцип относительности не был справедлив для этих «физических» явлений, то он не мог бы оставаться справедливым и для «чисто механических» явлений. Поэтому принцип относительности необходимо распространить на все явления природы и дать ему следующую формулировку:

Законы природы, определяющие изменение состояний физических систем, не зависят от того, к какой из двух инерциальных систем отсчета, движущихся одна относительно другой прямолинейно и равномерно, они относятся.

Это положение называется *частным*, или *специальным*, *принципом относительности Эйнштейна*. Он устанавливает равноправие *только инерциальных систем отсчета*. На основе этого принципа Эйнштейн создал в 1905 г. *частную*, или *специальную*, *теорию относительности*. Через 10 лет он обобщил принцип относительности на случай *произвольных неинерциальных систем отсчета* и создал *общую теорию относительности*, иначе называемую *релятивистской теорией тяготения*. Эта фундаментальная теория приобрела особое значение в связи с астрофизическими открытиями последнего времени. Общая теория относительности стала *основной теорией* в астрофизике, в частности в космологии. Однако в нашем курсе мы можем ограничиться изложением только специальной теории относительности.

§ 102. Опыт Майкельсона

1. Основные уравнения электродинамики Максвелла–Лорентца *не инвариантны относительно преобразования Галилея.* Действительно, скорость света в вакууме, вычисленная из этих уравнений, равна постоянной c. Такой результат оставался бы верным во всех инерциальных системах отсчета, если бы в них уравнения Максвелла имели один и тот же вид. Но это несовместимо с законом сложения скоростей (101.3), который является следствием преобразования Галилея.

Неинвариантность уравнений Максвелла относительно преобразования Галилея, как думали физики конца XIX и начала XX веков, не противоречит, однако, принципу относительности. Дело в том, что во всех электродинамических явлениях, помимо обыкновенного вещества, по господствовавшим в то время воззрениям, участвует еще *мировой эфир.* Уравнения Максвелла принимают обычный вид только в системе отсчета, в которой эфир покоится. Такая система является преимущественной или выделенной среди всех прочих инерциальных систем отсчета. Во всех остальных системах должно наблюдаться движение эфира — *эфирный ветер.* Его обнаружение считалось в XIX и начале XX веков одной из центральных проблем физики. Опыт привел к результату, что гипотеза эфира, как она понималась в физике XIX века, несостоятельна, а эфирного ветра не существует. Однако опыты по его обнаружению не совсем потеряли свое значение. Не вводя представления об эфире, можно поставить вопрос о существовании *преимущественной инерциальной системы отсчета.* На этот вопрос и отвечают опыты по обнаружению эфирного ветра.

2. Если принять гипотезу о существовании преимущественной системы отсчета, то прежде всего возникает вопрос, как движется Земля относительно этой системы? Обозначим скорость такого движения через $\mathbf{V}$. Тогда, согласно дорелятивистской кинематике, в направлении вектора $\mathbf{V}$ скорость света относительно Земли будет $c - V$, а в прямо противоположном направлении — $c + V$, где c — скорость света относительно преимущественной системы отсчета.

Этот результат можно подвергнуть экспериментальной проверке. Для этого, казалось бы, можно воспользоваться двумя одинаково идущими часами, установленными на Земле в различных точках A и B на расстоянии $l = AB$ друг от друга (рис. 323). Если прямая AB ориентирована вдоль скорости $\mathbf{V}$, то время, затрачиваемое светом на прохождение расстояния AB, будет $t_{AB} = l/(c-V)$, а на возвращение обратно $t_{BA} = l/(c+V)$. Разность этих времен, если ограничиться точностью до первого

порядка относительно V/c, равна $t_{AB}-t_{BA}\approx 2\,\frac{l}{c}\,\frac{V}{c}$. Если прямая AB составляет с направлением скорости $\mathbf{V}$ угол α, то

$$t_{AB}-t_{BA}\approx 2\,\frac{l}{c}\,\frac{V}{c}\cos\alpha.$$

Время t_{AB} находится по разности показаний часов в моменты прихода светового сигнала в B и его отправления из A. Таким же путем находится время t_{BA}.

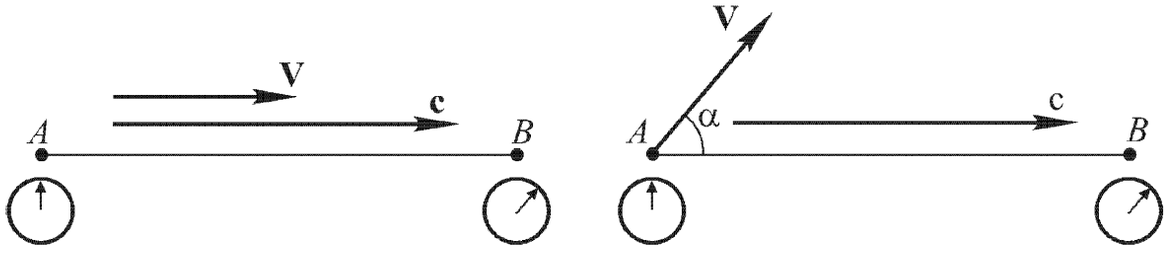


Рис. 323

Не было недостатка в предложениях, часто остроумных, как фактически осуществить опыт. Но в принципиальном отношении все они не отличались от схемы опыта, описанной выше. Предложения такого рода были оставлены после следующего замечания Майкельсона. Времена t_{AB} и t_{BA} измеряются с помощью *двух часов*, находящихся в различных точках пространства. Для измерения необходимо эти часы *синхронизовать*, что практически невозможно сделать с требуемой точностью. (Эйнштейн позднее заметил, что эта трудность *принципиальная*, а не только практическая, см. § 104.) Для осуществления опыта надо обойтись без синхронизации, т.е. все измерения производить с помощью *только одних часов*. Для этого из точки A, где помещены часы, надо послать световой сигнал и отразить его обратно в A с помощью зеркала, помещенного в B. Полное время распространения света t вдоль прямой AB туда и обратно можно измерить с помощью лишь одних часов. Делением удвоенного расстояния между точками A и B на это время можно найти скорость света c. Но это будет не скорость света в каком-либо одном определенном направлении, а *средняя скорость в двух прямо противоположных направлениях*. Майкельсон исследовал зависимость полного времени t от ориентации прямой AB.

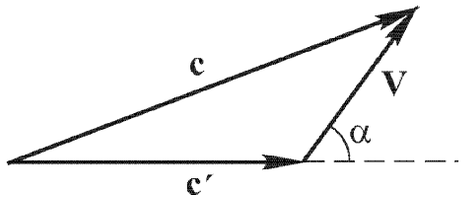


Рис. 324

Как видно из рис. 324, $c^2 = c'^2 + V^2 + 2Vc'\cos\alpha$, где $\mathbf{c}$ и $\mathbf{V}$ — скорости света и Земли относительно преимущественной системы отсчета, $\mathbf{c}'$ — скорость света относительно Земли, а α — угол между скоростями $\mathbf{V}$ и $\mathbf{c}'$. Решая полученное квадратное

уравнение относительно c', найдем

$$c' = -V\cos\alpha + c\sqrt{1-\beta^2\sin^2\alpha},$$

где введено обозначение $\beta = V/c$. Ограничиваясь квадратами β, получим

$$c' = c\left(1 - \beta\cos\alpha - \frac{1}{2}\beta^2\sin^2\alpha\right).$$

3. Майкельсон воспользовался сконструированным им интерферометром, описанным в § 35 (см. рис. 140). Пластинка P_2 была введена для уравнивания оптических длин плеч интерферометра. Без этого интерференция в белом свете была бы невозможна. В принципиальных рассуждениях от наличия пластинки P_2 можно отвлечься и считать разделительную пластинку P_1 бесконечно тонкой. Пусть плечо интерферометра *2* составляет с направлением скорости $\mathbf{V}$ угол α. Тогда на прохождение расстояния от A до зеркала M_2 и обратно свет затратит время

$$t_2 = \frac{l}{c}\frac{1}{1-\beta\cos\alpha-{}^1\!/_2\beta^2\sin^2\alpha} + \frac{l}{c}\frac{1}{1+\beta\cos\alpha-{}^1\!/_2\beta^2\sin^2\alpha} =$$
$$= \frac{2l}{c}\frac{1-{}^1\!/_2\beta^2\sin^2\alpha}{(1-{}^1\!/_2\beta^2\sin^2\alpha)^2-\beta^2\cos^2\alpha} \approx \frac{2l}{c}\frac{1-{}^1\!/_2\beta^2\sin^2\alpha}{1-\beta^2}.$$

Найдем теперь время t_1, которое требуется свету, чтобы пройти туда и обратно перпендикулярное плечо интерферометра. Оно найдется из предыдущего выражения заменой угла α на угол $\alpha \pm \frac{\pi}{2}$. Это дает

$$t_1 = \frac{2l}{c}\frac{1-{}^1\!/_2\beta^2\cos^2\alpha}{1-\beta^2}.$$

Таким образом,

$$t_2 - t_1 = \frac{l}{c}\frac{\beta^2}{1-\beta^2}\cos 2\alpha \approx \frac{l}{c}\beta^2\cos 2\alpha.$$

Этой разности времен соответствует разность фаз интерферирующих лучей

$$\varphi_2 - \varphi_1 = \frac{2\pi c}{\lambda}(t_2 - t_1) = \frac{2\pi l}{\lambda}\beta^2\cos 2\alpha.$$

При вращении интерферометра угол α будет меняться, а интерференционные полосы должны смещаться. Этот эффект и пытался обнаружить Майкельсон.

Если за единицу принять ширину интерференционной полосы, то смещение полос при повороте интерферометра будет

$$\Delta N = \frac{\Delta(\varphi_2-\varphi_1)}{2\pi} = \frac{l}{\lambda}\beta^2\Delta(\cos 2\alpha). \qquad (102.1)$$

Максимального значения

$$\Delta N_{\text{макс}} = \frac{2l}{\lambda}\beta^2 \qquad (102.2)$$

оно достигает, когда при повороте интерферометра $\cos 2\alpha$ меняется от ± 1 до ∓ 1. Отсюда видно, что ожидаемое смещение очень мало, так как оно *второго порядка* по $\beta \equiv V/c$. Точно оценить его невозможно, поскольку скорость Земли относительно преимущественной инерциальной системы отсчета неизвестна. Можно, однако, оценить нижний предел $\Delta N_{\text{макс}}$. Действительно, допустим сначала, что Солнце покоится относительно преимущественной системы отсчета. Тогда под V надо понимать скорость движения Земли по своей орбите, которая равна около 30 км/с, так что $\beta = 10^{-4}$, $\beta^2 = 10^{-8}$. В первом опыте Майкельсона (1881 г.) длина плеча интерферометра была $l = 1{,}2$ м. Если взять $\lambda = 600$ нм, то получится $\Delta N_{\text{макс}} = 0{,}04$ полосы.

При учете движения Солнца можно ожидать и большего смещения. Действительно, разложим скорость движения Солнца на составляющую $\mathbf{V}_{||}$ в плоскости земной орбиты и составляющую $\mathbf{V}_{\perp}$, перпендикулярную к этой плоскости. Полная скорость Земли будет $\mathbf{V} = \mathbf{V}_{\text{орб}} + \mathbf{V}_{||} + \mathbf{V}_{\perp}$, где $\mathbf{V}_{\text{орб}}$ — скорость орбитального движения Земли. Если опыт производится в момент, когда направления скоростей $\mathbf{V}_{\text{орб}}$ и $\mathbf{V}_{||}$ совпадают между собой, то для полной скорости Земли можно написать

$$V^2 = (V_{\text{орб}} + V_{||})^2 + V_{\perp}^2,$$

так что V получится больше $V_{\text{орб}}$. Следовательно, и ожидаемое смещение полос будет больше вычисленного выше.

Конечно, минимальное ожидаемое смещение в 0,04 полосы, вычисленное выше, могло бы наблюдаться только два раза в год, когда направление эфирного ветра совпадает или прямо противоположно направлению движения Земли. Поэтому необходимо производить наблюдения в течение по крайней мере *полугода*, что фактически и делалось Майкельсоном и всеми физиками, повторявшими его опыт.

Опыт был поставлен настолько тщательно, что позволял обнаружить смещение на 0,01 интерференционной полосы. Ожидаемое смещение на 0,04 полосы могло быть уверенно обнаружено на фоне неизбежных случайных ошибок измерений. Оказалось, что смещения нет. Точнее, результат опыта позволял утверждать, что скорость эфирного ветра, если бы он существовал, не может превышать 18 км/с.

4. В 1887 г. Майкельсон повторил свой опыт совместно с Морли (1837–1925). Для того чтобы при вращении интерферометра не возникали деформации, вызывавшие небольшое смещение интерференционных полос, интерферометр вместе с остальной аппаратурой монтировался на тяжелой цементной плите, которая плавала в сосуде со ртутью. С той же целью прибор все время находился в медленном вращении, даже в моменты на-

блюдения. Путем применения многократных отражений от 16 зеркал, установленных на той же плите, эффективная длина плеча интерферометра была доведена до 11 м. Ожидаемое смещение из-за орбитального движения Земли составляло уже 0,4 интерференционной полосы. Максимально наблюдавшееся смещение (из-за неизбежных случайных помех) составляло 0,02, а среднее — много меньше 0,01 интерференционной полосы. Опыт позволил утверждать, что скорость эфирного ветра не сможет превышать 7 км/с. Сам Майкельсон пришел к заключению, что результат опыта определенно отрицательный — *эфирного ветра не существует.*

Опыт Майкельсона неоднократно повторялся другими физиками с улучшенной аппаратурой и улучшенными методами наблюдения. Не останавливаясь на истории вопроса, укажем только, что к началу 30-х годов было с достоверностью установлено, что скорость эфирного ветра, если он и существует, не может превышать 1–1,5 км/с.

5. Изобретение лазеров, благодаря исключительно высокой степени монохроматичности и пространственной когерентности этих источников света, позволило чрезвычайно упростить схему опыта Майкельсона и повысить его точность. Опыт был поставлен в 1964 г. американскими физиками Джасея, Джаваном, Мурреем и Таунсом. Идея опыта заключалась в следующем.

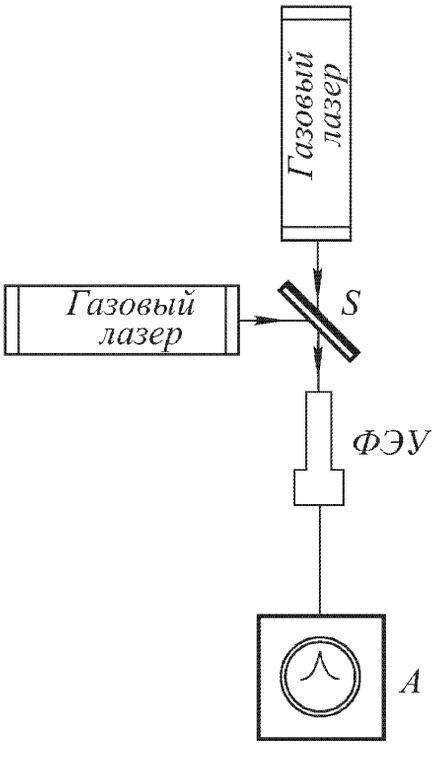


Рис. 325

Если скорость света зависит от направления его распространения относительно Земли, то частота, на которой генерирует лазер, должна изменяться при его повороте. Это изменение пропорционально β^2, т.е. второго порядка по β. Два одинаковых гелий-неоновых лазера были установлены на поворотной платформе перпендикулярно друг к другу (рис. 325). Световой пучок от одного из лазеров проходил полупосеребренное зеркало S, а пучок от другого лазера отражался от того же зеркала. Далее пучки шли в одном и том же направлении и попадали в фотоэлектронный умножитель, установленный на той же поворотной платформе. Если бы частоты этих пучков немного отличались друг от друга, то должны были бы возникнуть биения фототока с частотой, лежащей в области радиодиапазона, которые можно было бы наблюдать обычными

радиотехническими приемами с помощью анализатора A. Частота биений должна была бы меняться при вращении прибора. Если в исходном положении один лазер был ориентирован вдоль, а другой — перпендикулярно к направлению движения Земли, то при повороте прибора на 90°, согласно теории неподвижного эфира, из-за орбитального движения Земли должно было бы наблюдаться изменение разности частот лазеров примерно на 3 МГц, тогда как возможная ошибка опыта не превосходила нескольких герц. На опыте такое смещение обнаружено не было. На основании своих наблюдений экспериментаторы пришли к заключению, что скорость эфирного ветра, если бы таковой существовал, не может превышать 30 м/с. По сравнению с предыдущими результатами точность была повышена примерно в 50 раз.

6. Отрицательный результат опыта Майкельсона доказывает, что в земной системе отсчета скорость света в вакууме *изотропна*, т.е. не зависит от направления его распространения. Под скоростью света здесь понимается *средняя скорость* в двух прямо противоположных направлениях — она и только она определяет результат опыта с «одними часами». Обобщая этот результат, можно сказать, что он справедлив *во всех инерциальных системах отсчета.*

Остается выяснить, одинаковы ли числовые значения скорости света в различных системах отсчета? Утвердительный ответ можно дать на основании опыта Кеннеди и Торндайка, которые в 1932 г. изменили опыт Майкельсона. В их опыте использовался интерферометр Майкельсона, в котором длины плеч были *разными* и отличались на 16 см. Они наблюдали интерференционные полосы (при неподвижном относительно Земли интерферометре) в течение шести месяцев. За это время Земля из своего исходного положения непрерывно перемещалась в диаметрально противоположную точку своей орбиты, участвуя вместе с тем в движении Солнца. Тем самым реализовалось множество систем отсчета, отличающихся друг от друга положением и ориентацией в пространстве. Если бы в них скорость света была разной, то должно было бы наблюдаться регулярное смещение интерференционных полос. На самом деле смещение носило случайный характер и не превосходило 3/1000 ширины полосы. На этом основании можно прийти к заключению, что с точностью примерно 2 м/с скорость света в вакууме относительно Земли одинакова во всех ее положениях в пространстве.

Все это подтверждает равноправие различных инерциальных систем отсчета по отношению ко всем физическим явлениям — *преимущественной инерциальной системы отсчета не существует.*

§ 103. Независимость скорости света от движения источника

1. Помимо принципа или постулата относительности, Эйнштейн положил в основу теории относительности *постулат о независимости скорости света в вакууме от движения источника*. Его обычно называют *принципом «постоянства скорости света»*, хотя, как заметил еще Паули (1900–1958), такое название может привести к недоразумениям. Об универсальности постоянства скорости света в вакууме не может быть речи уже потому, что эта скорость постоянна *только в инерциальных системах отсчета*, а независимость скорости света от движения источника сохраняется и *в общей теории относительности*, где на выбор систем отсчета не накладывается никаких ограничений.

Независимость скорости света от движения источника можно было бы не выдвигать в качестве самостоятельного постулата, если с самого начала принять волновую теорию света. Однако столь фундаментальную теорию, какой является теория относительности, лучше строить, не связывая ее ни с какими гипотетическими представлениями о природе и механизме физических явлений.

2. Предположение о зависимости скорости света от движения источника ввел Ритц (1878–1909). Согласно этому предположению скорость света в вакууме постоянна и равна c только относительно источника. Если же источник света движется со скоростью $\mathbf{v}$ относительно какой-либо системы отсчета, то скорость света $\mathbf{c}'$ в этой системе векторно складывается из $\mathbf{c}$ и $\mathbf{v}$, т.е. $\mathbf{c}' = \mathbf{c} + \mathbf{v}$, как это происходит со скоростью снаряда при стрельбе из движущегося орудия. Такая гипотеза получила название *баллистической*. Ритц перестроил электродинамику Максвелла, приведя ее в соответствие с этой гипотезой. Ясно, что баллистическая гипотеза объясняет отрицательный результат опыта Майкельсона, если только источник света неподвижен относительно интерферометра.

Но баллистическая гипотеза, как и любое предположение о зависимости скорости света от движения источника, не согласуется с молекулярным объяснением отражения и преломления света (см. § 68 и 69). В основе такого объяснения лежит представление, что отраженная и преломленная волны возникают в результате интерференции падающей волны с вторичными волнами, излучаемыми молекулами и атомами среды. Действительно, если среду считать неподвижной, а источник света взять движущимся, то волны, излучаемые источником, будут распространяться с иными скоростями, чем волны, излучаемые молекулами н атомами среды. Интерференция между ними невозможна. Од-

нако это замечание есть аргументация, основанная на волновой теории света, а не экспериментальное доказательство, свободное от гипотетических представлений о природе света.

Убедительное доказательство несостоятельности баллистической гипотезы, как показал в 1913 г. голландский астроном де Ситтер (1872–1934), дают астрономические наблюдения над движением двойных звезд. Действительно, допустим, что баллистическая гипотеза верна. Предположим для простоты, что компоненты двойной звезды вращаются вокруг их центра масс по круговым орбитам в той же плоскости, в которой расположена Земля. Рассмотрим движение одной из этих двух звезд. Пусть v — скорость движения ее по круговой орбите. В положении звезды, когда она удаляется от Земли вдоль соединяющей их прямой, скорость света равна $(c-v)$, а в положении, когда звезда приближается, равна $(c+v)$. Если отсчитывать время от момента, когда звезда находилась в первом положении, то свет из этого положения дойдет до Земли в момент $t_1 = L/(c-v)$, а из второго положения — в момент $t_2 = T/2 + L/(c+v)$, где T — период обращения звезды, а L — расстояние до нее. При громадных расстояниях до звезд наблюдаемые движения звезды могли бы заметно отступать от законов Кеплера. В частности, при очень больших L могло бы случиться, что $t_2 \leqslant t_1$, т.е. звезда одновременно была бы видна в двух (и даже нескольких) положениях или обращалась бы в противоположном направлении. Ничего подобного, как показали астрономические наблюдения, не происходит.

Если бы баллистическая гипотеза была верна, то опыт Майкельсона должен был бы дать положительный результат, если его произвести не с земным источником света, а со светом от звезд. Томашек в 1926 г. так и поступил, но получил отрицательный результат.

А.М. Бонч-Бруевич и В.А. Молчанов в 1956 г., применив современные методы измерения скорости света, сравнили скорости света от правого и левого краев Солнца. Из-за осевого вращения Солнца один из этих краев приближался к нам со скоростью 2,3 км/с, а другой удалялся с той же скоростью. Обе скорости света с достаточной точностью совпали между собой.

Наконец, ставились специальные опыты, в которых сравнивались скорости γ-квантов, испускаемых движущимися возбужденными ядрами углерода ($^*C^{12}$) и неподвижными возбужденными ядрами кислорода ($^*O^{16}$). Ставились также опыты по аннигиляции электрона с позитроном, специально приспособленные для проверки независимости скорости света от движения источника. По сравнению с астрономическими опытами, в которых используются космические источники света со сравнитель-

но малыми скоростями, в опытах с атомными ядрами и элементарными частицами скорость источников гораздо выше (сравнима со скоростью самого света). Опыты подтвердили, что с точностью около 10 % скорость γ-квантов не зависит от движения источников.

Все приведенные опыты подтверждают постулат о независимости скорости света в вакууме от движения источника.

3. Заметим в заключение, что теория относительности вообще была бы невозможна, если бы не был установлен фундаментальный факт *конечности скорости распространения света*. Изучение методов и результатов измерения скорости света представляет громадный, не только исторический интерес. В частности, уточнение числового значения этой постоянной необходимо для точных измерений астрономических расстояний методами радиолокации. Это в свою очередь необходимо для целей космонавтики. Однако мы не можем касаться этих вопросов. Ограничимся замечанием, что в 1972 г. скорость света была определена на основе независимых измерений длины волны λ и частоты света ν. Источником света служил гелий-неоновый лазер, генерировавший излучение с длиной волны 3,39 мкм. Длина волны измерялась интерферометрически сравнением ее с эталоном длины, т.е. с длиной волны в вакууме оранжевой линии изотопа криптона-86. Ошибка таких измерений $\sim 10^{-5}$ нм. Частота лазерного излучения измерялась путем сравнения ее с атомным стандартом частоты, т.е. с частотой перехода между двумя сверхтонкими квантовыми уровнями атома цезия-133 в нулевом магнитном поле. При этом использовались методы нелинейной оптики — генерация излучений с суммарной и разностной частотами. В итоге для скорости света $c = \lambda\nu$ было найдено значение

$$c = 299\,792\,458 \pm 1{,}2 \text{ м/с},$$

превосходящее по точности все ранее полученные значения более чем на два порядка.

§ 104. Понятие одновременности

1. До теории относительности считалось, что принцип относительности и принцип независимости скорости света от движения источника противоречат друг другу. Действительно, согласно принципу относительности скорость света в вакууме во всех инерциальных системах отсчета должна быть одной и той же. Обозначим ее, как всегда, через c. Возьмем две инерциальные системы отсчета S и S', движущиеся относительно друг друга прямолинейно и равномерно. Пусть O и O' — какие-либо две точки, неподвижные в системах S и S' соответственно и про-

странственно совпадающие в момент времени $t = 0$. Произведем в этих точках в тот же момент времени кратковременную световую вспышку. Рассмотрим распространение этой вспышки сначала в системе S. Так как свет во всех направлениях распространяется с одной и той же скоростью c, то к моменту времени t вспышка достигнет сферы Σ радиуса ct с центром в точке O. Рассмотрим теперь распространение той же вспышки в системе S'. Обе системы совершенно равноправны. Они отличаются друг от друга только тем, что источник света относительно них движется по-разному. Но это в рассматриваемом вопросе не имеет никакого значения, так как скорость света не зависит от движения источника. Поэтому можно утверждать, что к моменту t та же вспышка достигнет сферы Σ' с тем же радиусом ct, но с центром в точке O'. Сферы Σ и Σ' разные, так как к моменту t их центры O и O' разойдутся. Таким образом, одна и та же световая вспышка *одновременно* должна находиться и на сфере Σ, и на сфере Σ', что, очевидно, невозможно.

2. Однако приведенное рассуждение несостоятельно, так как в его основе лежит *интуитивное представление об одновременности.* Если два события происходят в одном и том же месте пространства, то не возникает трудностей при суждении о том, одновременны эти события или нет. Но если события происходят *в разных местах пространства*, то без дальнейших разъяснений утверждение об их одновременности или неодновременности просто бессодержательно, так как *никакой априорной одновременности пространственно разделенных событий не существует.*

В принципе пространственно-временное описание явлений природы можно дать и без использования понятия одновременности. Действительно, возьмем какую-либо инерциальную систему отсчета. Для определения положения точки в ней будем пользоваться координатными осями, реализованными, например, с помощью трех взаимно перпендикулярных твердых стержней. Расстояние между двумя точками будем измерять путем откладывания неподвижного масштабного стержня, т.е. твердого стержня определенной длины, разделенного на метры и более мелкие единицы. Для измерения больших расстояний этот способ непригоден. Тогда надо применять косвенные методы измерения длин, о которых говорилось в § 1 первого тома. Для определения времени расставим в нашей координатной системе достаточно часто какие-либо *одинаково устроенные часы.* Стрелки часов в различных точках пространства установим совершенно произвольно и независимо друг от друга. После этого получится пространственно-временная система отсчета. Всякое событие в такой системе можно характеризовать местом, где оно произошло, и показаниями часов в том же месте.

Пока что введено только *местное время*, т.е. время в каждой точке пространства, причем времена в различных точках никак не связаны между собой. Это обстоятельство делает нашу систему практически совершенно непригодной для пространственно-временного описания явлений природы. Практической может быть только система с *единым временем*, в которой показания часов в различных точках не независимы, а определенным образом связаны или синхронизированы друг с другом. Способ синхронизации в принципе произволен. Требуется только, чтобы он был *внутренне непротиворечив*. Однако соображения целесообразности, основанные на принципе относительности и независимости скорости света от движения источника, почти однозначно вынуждают принять способ синхронизации, предложенный Эйнштейном, и основанное на нем определение одновременности.

3. Эйнштейн предложил устанавливать синхронизацию пространственно разделенных часов с помощью световых сигналов в вакууме. Пока синхронизация часов не установлена, не имеет смысла говорить о скорости в любом определенном направлении. Имеет смысл говорить лишь о *средней скорости* света в двух прямо противоположных направлениях, так как эту величину можно измерить с помощью *только одних часов*. Для этого из неподвижной точки A в момент времени t_1^A пошлем световой сигнал к неподвижному зеркалу, установленному на расстоянии l от A. Пусть отраженный сигнал вернется в A в момент t_2^A. Время отсчитывается по одним и тем же часам, так что никакой синхронизации не требуется. Разделив пройденное расстояние $2l$ на затраченное время $t_2^A - t_1^A$, найдем искомую среднюю скорость света $c = 2l/(t_2^A - t_1^A)$. Опыт Майкельсона и другие опыты показали, что эта величина есть *универсальная постоянная*. Теперь *полагаем по определению, что скорости света во всяких двух взаимно противоположных направлениях одинаковы*. Тем самым постулативно будет установлена *синхронизация часов* и введено *понятие одновременности*. После этого уже имеет смысл говорить о скорости любого движения в *одном направлении*, так как время, затрачиваемое на такое движение, можно измерить по двум пространственно разделенным часам, синхронизованным между собой.

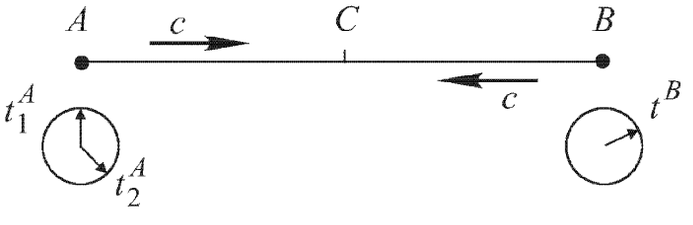


Рис. 326

Сформулируем теперь окончательно *правило Эйнштейна синхронизации часов. Пусть в неподвижных точках A и B* (рис. 326) *установлены одинаковые часы. Из точки A в момент t_1^A по часам A пошлем световой сиг-*

нал к зеркалу, установленному в точке B. Пусть отраженный сигнал возвращается в A в момент t_2^A (по часам A). В момент прихода сигнала в B поставим на часах в B время $t^B = {}^1/_2(t_1^A + t_2^A)$. Тогда по определению часы в A и B будут синхронизованы между собой.

Можно дать другой, эквивалентный, способ синхронизации. *В середине отрезка AB произведем световую вспышку. В момент прихода света от вспышки к часам A и B поставим их одинаково. Тогда часы будут синхронизованы.*

Но часы A и B можно синхронизовать между собой не только непосредственно, но также посредством третьих часов C. Синхронизуем сначала одним из указанных способов часы A и C, затем часы B и C. Мы утверждаем и рассматриваем это утверждение как постулат, что при этом часы A и B окажутся синхронизованными. Такое утверждение, как показывают все следствия из него, не приводит к противоречиям и согласуется с опытом. Таким образом, получается *пространственно-временная система отсчета с единым временем*, в которой часы синхронизованы между собой по правилу Эйнштейна. *Два пространственно разделенных события в этой системе отсчета называются одновременными, если часы, находящиеся в тех точках, где происходят эти события, показывают одно и то же время.*

4. *Одновременность есть понятие относительное*, т.е. два события, одновременные в одной инерциальной системе отсчета, могут оказаться не одновременными в другой системе. Поясним это на следующем мысленном опыте. По поверхности Земли равномерно и прямолинейно со скоростью V движется поезд $A'B'$ (рис. 327). Примем его за движущуюся систему отсчета S'. Неподвижной системой S будем считать полотно железной дороги. Пусть в концы поезда A' и B' ударяют две молнии, оставляющие следы A и B на полотне железной дороги. Отметим в поезде среднюю точку C', а на полотне — среднюю точку C между A и B. Пусть в точку C вспышки от молний приходят одновременно. Тогда с точки зрения системы S удары молний в концы поезда будут событиями одновременными. Но в момент встречи обеих вспышек точка C' окажется правее C. В этот момент вспышка от B' уже прошла через C', а от A' еще не дошла. Значит, с точки зрения системы S' удар молнии в B' произошел раньше удара в A'.

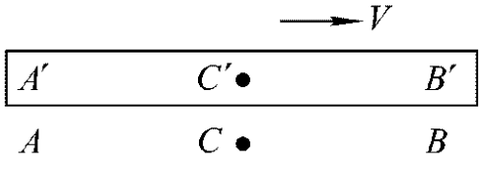


Рис. 327

Теперь ясно, как разрешается парадокс, о котором говорилось в п. 1. Свет доходит до сферы Σ одновременно в системе отсчета S, но не одновременно в системе S'. Аналогично, до

сферы Σ' свет доходит одновременно с точки зрения системы отсчета S', но не одновременно с точки зрения системы S.

Дорелятивистская физика развивалась, и вполне успешно, считая время и одновременность *абсолютными*, т.е. одинаковыми во всех системах отсчета. Но так происходило лишь до тех пор, пока рассматривались *медленные движения*. А распространение света есть быстрый процесс. Вот почему именно в оптике физика встретилась раньше всего с принципиальными трудностями, преодоленными теорией относительности.

§ 105. Преобразование координат и времени в теории относительности

1. Рассмотрим две инерциальные системы отсчета S и S', из которых вторая движется относительно первой прямолинейно и равномерно со скоростью $\mathbf{V}$, а следовательно, первая движется относительно второй со скоростью $-\mathbf{V}$. В каждой системе отсчета расставлены достаточно часто одинаковые часы, неподвижные в этой системе и синхронизованные по правилу Эйнштейна. Пусть x, y, z, t — координаты и время какого-либо события (например, столкновения двух шаров) в системе отсчета S, а x', y', z', t' — координаты и время того же события в системе отсчета S'. Возникает вопрос, как по значениям x, y, z, t найти значения x', y', z', t' и наоборот. Решение этого вопроса основано на предположении, что пространство *однородно* и *изотропно*, а время *однородно* [1]). Однородность пространства и времени означает, что все точки пространства и все моменты времени, как в системе S, так и в системе S', абсолютно эквивалентны. Изотропия же пространства означает полную эквивалентность всех пространственных направлений в системе S, а также в системе S'. В силу указанной однородности и изотропии пространства и времени связь между x, y, z, t и x', y', z', t' должна быть *линейной*.

Вопрос о чисто пространственных преобразованиях координат в этих предположениях решается в аналитической геометрии. Поэтому можно отвлечься от этого вопроса и сосредоточить все внимание на том, что нового вносит в преобразование координат и времени равномерное движение одной системы отсчета относительно другой. Для этой цели достаточно рассмотреть частный случай, когда начала O и O' координатных систем S

[1]) Эти свойства нельзя рассматривать как априорные. Они установлены и проверены экспериментально. Их объяснение, по-видимому, должна дать космология при рассмотрении происхождения и эволюции Вселенной в целом.

и S' в некоторый момент времени совмещаются. Этот момент мы примем за начало отсчета времени как в системе S, так и в системе S'. Тогда связь между x, y, z, t и x', y', z', t' будет не только линейной, но и *однородной*, так как нулевым значениям нештрихованных параметров соответствуют также нулевые значения штрихованных. Кроме того, оси X', Y', Z' координатной системы S' можно выбрать так, чтобы они были параллельны осям X, Y, Z координатной системы S и, следовательно, ось X' все время совмещалась с осью X (рис. 328).

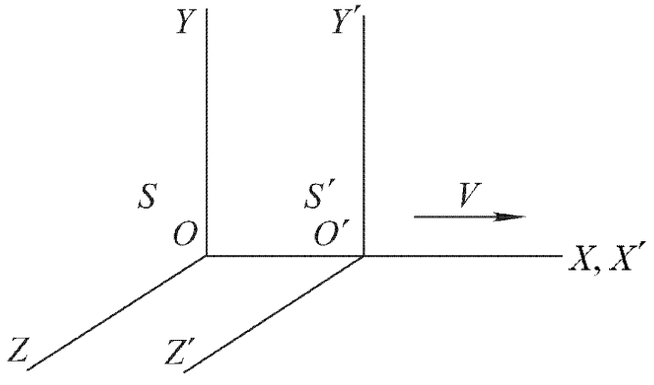


Рис. 328

2. Предположим, что относительная скорость V координатных систем S и S' меньше скорости света c ($V < c$). Пусть при этом условии из начала координат O в момент времени $t_1 > 0$ (по часам системы S) послан световой сигнал в положительном направлении оси X. Пусть этот сигнал приходит в точку O' в момент времени t' (по часам системы S'). Тогда, ввиду линейности связи между координатами и временем в системах S и S', время t' должно выражаться также линейно через t_1. При этом нулевому значению t_1 соответствует нулевое значение t', так как в момент $t = t'$ начала координат O и O' совмещаются между собой. Следовательно, должно быть

$$t' = kt_1, \tag{105.1}$$

где k — некоторый коэффициент. В силу изотропии пространства он может зависеть только от абсолютного значения скорости $\mathbf{V}$, но не от ее направления.

Выразим коэффициент k через скорость V. Пусть в точке O' световой сигнал отражается и, распространяясь в обратном направлении, возвращается в точку O в момент t_2 (по часам системы S). Тогда, ввиду полной эквивалентности систем отсчета S и S',

$$t_2 = kt', \tag{105.2}$$

где k имеет то же значение, что и в формуле (105.1). Исключив t', получим

$$t_2 = k^2 t_1. \tag{105.3}$$

Согласно эйнштейновскому правилу синхронизации часов сигнал, посланный из O в момент t_1 и возвратившийся обратно в

момент t_2, отражается в точке O' в момент времени

$$t = \frac{t_1 + t_2}{2} = \frac{1 + k^2}{2} t_1 \tag{105.4}$$

(по часам системы S). Расстояние x, проходимое точкой O' за время t, равно $x = Vt$. То же расстояние свет проходит за время $t - t_1$, и следовательно, $x = c(t - t_1)$. Таким образом, $Vt = c(t - t_1)$. Подставив сюда значение t из предыдущей формулы, получим

$$V(k^2 + 1) = c(k^2 - 1), \tag{105.5}$$

откуда

$$\frac{V}{c} = \frac{k^2 - 1}{k^2 + 1}; \tag{105.6}$$

$$k^2 = \frac{c + V}{c - V} = \frac{1 + V/c}{1 - V/c}. \tag{105.7}$$

Извлекая из последнего выражения квадратный корень, найдем k. После этого получаем две вспомогательные формулы:

$$k + \frac{1}{k} = \frac{2}{\sqrt{1 - (V/c)^2}}, \quad k - \frac{1}{k} = \frac{2V/c}{\sqrt{1 - (V/c)^2}}, \tag{105.8}$$

которые понадобятся нам в дальнейшем. Все эти соотношения, разумеется, имеют смысл только при условии $V < c$.

Отметим, в частности, что из формул (105.1) и (105.4) следует

$$\frac{t'}{t} = \frac{2k}{k^2 + 1} = \sqrt{1 - (V/c)^2}, \tag{105.9}$$

откуда $t' < t$. Величина t есть время движения точки O' из неподвижного начала O до точки, в которой его догонит световой сигнал, измеренное по «неподвижным часам», т.е. часам системы S. Величина t' имеет смысл того же времени, но измеренного уже по «движущимся часам», т.е. часам системы S'. Таким образом, величины t и t' представляют собой времена между одними и теми же событиями, измеренные соответственно в «неподвижной» S и «движущейся» S' системах отсчета. Формула (105.9) показывает, что движущиеся часы идут медленнее неподвижных. Это явление будет обсуждено ниже с различных точек зрения, а сейчас мы вернемся к вопросу о преобразовании координат и времени.

3. Пусть на оси X произошло какое-то событие A, например столкновение двух шаров (рис. 329). В системе отсчета S это событие характеризуется абсциссой x и моментом времени t. В системе отсчета S' абсцисса и время того же события будут x', t'. Рассмотрим световой сигнал, отправляющийся из начала координат O к месту события A (рис. 329 *а*). На прохождение расстояния OA сигнал затрачивает время x/c. Чтобы он пришел

в A в момент t, его надо отправить из O в момент $t_1 = t - x/c$. На прохождение расстояния $O'A$ тот же сигнал затрачивает время x'/c, так что он проходит через точку O' в момент $t'_1 = t' - x'/c$. По определению коэффициента k $t'_1 = kt_1$ или $t' - x'/c = k(t - x/c)$. Отразим теперь световой сигнал в A в обратном направлении (рис. 329 *б*). Через точку O' он пройдет в момент времени $t'_2 = t' + x'/c$, а через точку O — в момент $t_2 = t + x/c$. Теперь роли штрихованных и нештрихованных величин поменялись местами, и на основании того же определения коэффициента k можно написать $t_2 = kt'_2$, или $t + x/c = k(t' + x'/c)$. Таким образом,

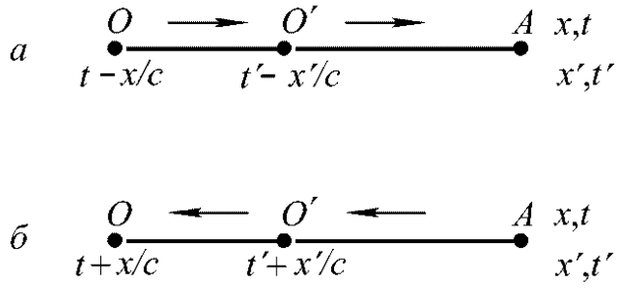


Рис. 329

$$t' - x'/c = k(t - x/c), \quad k(t' + x'/c) = t + x/c. \tag{105.10}$$

Полученные соотношения справедливы, в каком бы месте оси X ни произошло событие A. Пусть, например, оно произошло левее точек O и O', как указано на рис. 330 *а*. Отправим от места события световой сигнал вправо. Так как теперь x и x' отрицательны, то сигнал достигнет точек O и O' в более поздние моменты $t - x/c$ и $t' - x'/c$. На основании определения коэффициента k напишем $t' - x'/c = k(t - x/c)$, а это есть первое соотношение (105.10). Возьмем теперь световой сигнал, распространяющийся справа налево (рис. 330 *б*). Пусть он достигает точки A в моменты t и t' по часам в системах S и S' соответственно. Через точки O и O' сигнал пройдет в моменты времени $t + x/c$ и $t' + x'/c$, а потому $t' + x'/c = k(t' - x'/c)$, т.е. получается и второе соотношение (105.10).

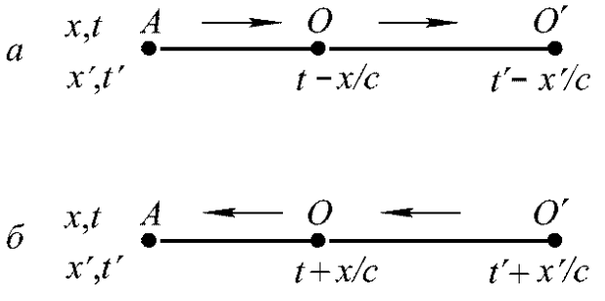


Рис. 330

Очень полезно рассмотреть все случаи взаимного расположения точек A, O, O' и убедиться, что во всех случаях справедливы соотношения (105.10).

Из соотношений (105.10) находим

$$\begin{aligned} x' &= \frac{1}{2}\left(k + \frac{1}{k}\right)x - \frac{1}{2}\left(k - \frac{1}{k}\right)ct, \\ t' &= \frac{1}{2}\left(k + \frac{1}{k}\right)t - \frac{1}{2}\left(k - \frac{1}{k}\right)\frac{x}{c}, \end{aligned} \tag{105.11}$$

или на основании (105.8)

$$x' = \frac{x - Vt}{\sqrt{1-\beta^2}}, \quad y' = y, \quad z' = z, \quad t' = \frac{t - Vx/c^2}{\sqrt{1-\beta^2}}, \tag{105.12}$$

где введено обозначение

$$\beta = V/c. \tag{105.13}$$

Мы добавили формулы $y' = y$, $z' = z$, которые показывают, что поперечные координаты события y и z не преобразуются. Для доказательства последнего утверждения изготовим в системах отсчета S и S' одним и тем же способом два одинаковых твердых стержня, каждый из которых неподвижен в своей системе отсчета. Установим их своими концами на оси X параллельно осям Y и Y' (рис. 331). Длины стержней y и y', измеренные соответственно в системах S и S', конечно, будут одинаковы. Но это еще не означает, что справедливо второе уравнение (105.12). Надо еще показать, что y и y' можно рассматривать как координаты *одного и того же события*. Для доказательства этого к свободным концам стержней прикрепим маленькие шарики A и A'. Пусть сначала шарик A' расположен левее шарика A. Мы утверждаем, что при движении шарик A' обязательно столкнется с шариком A. Действительно, если бы движущийся шарик прошел выше или ниже неподвижного, то системы отсчета S и S' не были бы эквивалентны. Но если шарики столкнутся, то y и y' становятся координатами одного и того же события — столкновения шариков, и второе уравнение (105.12) может считаться доказанным. Так же доказывается и третье уравнение (105.12).

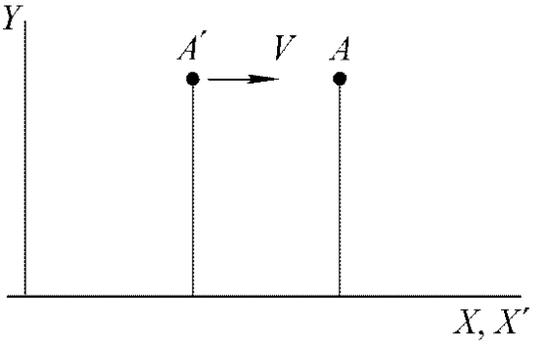


Рис. 331

Формулы (105.12) и решают задачу о преобразовании координат и времени при переходе от одной системы отсчета к другой. Они называются *преобразованием Лоренца* (этот термин был введен Пуанкаре). Лоренц получил их в 1904 г. К тем же формулам несколько раньше (в 1900 г.) пришел Лармор. И Лармор, и Лоренц, однако, принципиально стояли на точке зрения неподвижного эфира. У них истинным было только время t в системе отсчета, в которой эфир покоится. Величина же t' лишь формально играла роль времени — это была математическая переменная, вводимая таким образом, чтобы соблюдалась инвариантность уравнений электродинамики при переходе от переменных x, y, z, t к переменным x', y', z', t'. Настоящий вывод формул преобразования Лоренца и установление их истинного

смысла дал Эйнштейн в 1905 г. В его теории все инерциальные системы отсчета совершенно эквивалентны, а t' является таким же «истинным временем», как и t. Это проявляется, в частности, если уравнения (105.12) разрешить относительно x, y, z, t. Таким путем получатся формулы «обратного преобразования»

$$x = \frac{x' + Vt}{\sqrt{1-\beta^2}}, \quad y = y', \quad z = z', \quad t = \frac{t' + Vx'/c^2}{\sqrt{1-\beta^2}}, \tag{105.14}$$

имеющие тот же вид, что и формулы «прямого преобразования» (105.12). Как и следовало ожидать, они получаются из формул (105.12) простой заменой V на $-V$.

Формулы (105.12) и (105.14) при $\beta > 1$ дали бы мнимые значения для координат и времени. Поэтому нет смысла говорить о движении одной системы отсчета со скоростью V, превышающей скорость света c. Отсюда следует, что скорость любого тела не может превышать c, так как с каждым телом можно связать систему отсчета.

При медленных движениях, когда $(V/c)^2 \ll 1$ и $Vv/c^2 \ll 1$ (v — скорость движения тела), преобразование Лорентца, как и следовало ожидать, в пределе переходит в преобразование Галилея.

4. В дорелятивистской физике пространство и время считались *независимыми* друг от друга. Расстояния между двумя пространственными точками (точнее, расстояния между двумя материальными точками в один и тот же момент времени), а также промежутки времени между двумя событиями считались *одинаковыми во всех системах отсчета.* Иными словами, обе эта величины считались *инвариантными* при переходе от одной системы отсчета к другой. В теории относительности такая инвариантность была утрачена. Вместо двух инвариантов — пространственного и временного — в ней сохранился только *один, пространственно-временной, инвариант.* Его легко найти, перемножая почленно уравнения (105.11). Это дает

$$c^2t^2 - x^2 = c^2t'^2 - x'^2 = \mathrm{Inv}\ . \tag{105.15}$$

Для удобства запишем этот инвариант через новые временны́е переменные

$$\tau = ct, \quad \tau' = ct', \tag{105.16}$$

имеющие размерность длины. Введение таких переменных означает, что промежутки времени теперь измеряются теми же единицами, что и пространственные расстояния: за единицу времени принимается время, в течение которого свет проходит единицу расстояния. В новых переменных

$$\tau^2 - x^2 = \tau'^2 - x'^2 = \mathrm{Inv}\ . \tag{105.17}$$

Учтем теперь, что при выбранной нами ориентации координатных осей $y' = y$, $z' = z$. Поэтому

$$\tau^2 - (x^2 + y^2 + z^2) = \tau'^2 - (x'^2 + y'^2 + z'^2) = \mathrm{Inv}\ . \qquad (105.18)$$

В этом виде пространственно-временной инвариант (105.18) уже не зависит от ориентации систем отсчета S и S' относительно друг друга, а также от направления скорости V. Однако в формуле (105.18) предполагается, что одно из событий фиксировано. Таким событием является совмещение начал координат O и O'. Чтобы освободиться от этого ограничения, запишем (105.18) в виде

$$\Delta\tau^2 - (\Delta x^2 + \Delta y^2 + \Delta z^2) = \Delta\tau'^2 - (\Delta x'^2 + \Delta y'^2 + \Delta z'^2) = \mathrm{Inv}\,, \qquad (105.19)$$

где $\Delta\tau$, Δx, ..., $\Delta\tau'$, $\Delta x'$, ... означают разности между временами и пространственными координатами двух событий 1 и 2 в системах отсчета S и S':

$$\Delta\tau = \tau_2 - \tau_1, \quad \Delta x = x_2 - x_1, \quad \Delta y = y_2 - y_1, \quad \Delta z = z_2 - z_1,$$
$$\Delta\tau' = \tau'_2 - \tau'_1, \quad \Delta x' = x'_2 - x'_1, \quad \Delta y' = y'_2 - y'_1, \quad \Delta z' = z'_2 - z'_1.$$

Квадратный корень из инварианта (105.19) называется *интервалом между рассматриваемыми событиями* и в дальнейшем обозначается через s_{12} или Δs. Очевидно, квадрат интервала между событиями 1 и 2 можно представить в виде

$$s_{12}^2 = (\tau_2 - \tau_1)^2 - l_{12}^2 = (\tau'_2 - \tau'_1)^2 - l_{12}'^2, \qquad (105.20)$$

где l_{12} и l'_{12} — расстояния между точками, в которых произошли события, в системах S и S' соответственно.

Можно также в формулы преобразования Лоренца ввести параметры τ и τ' вместо t и t'. Ограничиваясь частным случаем, представленным на рис. 328, перепишем формулы (105.12) в виде

$$x' = \frac{x - \beta\tau}{\sqrt{1-\beta^2}}, \quad \tau' = \frac{\tau - \beta x}{\sqrt{1-\beta^2}}. \qquad (105.21)$$

Величина β имеет смысл скорости движущейся системы координат, если за единицу принять скорость света.

Минковский (1864–1909) для описания пространственно-временных событий ввел геометрическую терминологию. Совокупность значений τ, x, y, z, характеризующую время и место события, он назвал *мировой точкой*. Многообразие мировых точек есть четырехмерное пространство, называемое *миром* или *пространством Минковского*. Линия в пространстве Минковского называется *мировой линией*. Интервал между двумя событиями принимается за *инвариантное расстояние* между соответствующими мировыми точками. На основе таких представ-

лений было создано тензорное исчисление в пространстве Минковского, аналогичное тензорному исчислению обычной эвклидовой геометрии. Оно является адекватным математическим аппаратом специальной теории относительности.

5. Интервалы между событиями можно разделить на *вещественные* и *чисто мнимые*. Ввиду инвариантности интервала, это деление *не зависит от выбора системы отсчета*.

Поставим вопрос, можно ли выбрать такую систему отсчета, в которой события 1 и 2 были бы *одноместны*, т.е. происходили бы в одной и той же точке пространства? Если S' — такая система, то в ней $l'_{12} = 0$, и на основании (105.20) квадрат интервала может быть представлен в виде $s^2_{12} = (\tau'_2 - \tau'_1)^2$. Отсюда видно, что $s^2_{12} > 0$, т.е. необходимо, чтобы интервал s_{12} был вещественным. Для доказательства достаточности этого условия можно без нарушения общности ограничиться частным преобразованием Лоренца (105.21). Чтобы рассматриваемые события в системе S' пространственно совпадали, достаточно, чтобы выполнялось условие $\Delta x' = 0$, т.е. $\Delta x = \beta\,\Delta\tau$. Отсюда видно, что система S' должна двигаться со скоростью $\beta = \Delta x/\Delta\tau$. Но для вещественных интервалов $|\Delta x| < \Delta\tau$, так что $|\beta| < 1$. Значит, система S' должна двигаться со скоростью, меньшей скорости света, а потому ее можно реализовать. Промежуток времени между одноместными событиями в системе отсчета S' будет равен $\Delta\tau' = |s_{12}|$, или в обычных единицах $\Delta t' = |s_{12}|/c$. Вещественные интервалы называются *времениподобными*.

Поставим теперь вопрос о существовании системы отсчета, в которой события 1 и 2 были бы *одновременны*. Если S' — такая система, то $\tau'_2 - \tau'_1 = 0$ и, следовательно, на основании (105.20) должно быть $s^2_{12} = -l'^2_{12}$. Значит, необходимо, чтобы интервал s_{12} был чисто мнимым. Достаточность этого условия доказывается совершенно так же, как в предыдущем случае. Расстояние между точками, в которых произошли одновременные события 1 и 2, в системе S' равно $l'_{12} = |s_{12}|$. Чисто мнимые интервалы называются *пространственноподобными*.

Рассмотрим, наконец, особый случай, когда интервал между событиями равен нулю. В этом случае, чтобы сделать события одноместными, надо перейти к системе отсчета, движущейся со скоростью света. То же требуется, чтобы события стали одновременными. И то и другое невозможно. Нулевые интервалы называются *световыми*. Такими интервалами связаны отправление светового сигнала из некоторой точки и приход его в другую точку пространства.

6. В случае частного преобразования Лоренца (105.21), соответствующего рис. 328, все изложенное можно наглядно ин-

терпретировать графически. Это возможно потому, что достаточно ограничиться рассмотрением преобразования только одной пространственной координаты x и времени τ, т.е. рассуждать так, как если бы пространство Минковского было *двумерной плоскостью* (x, τ). Произвольное событие (мировую точку) O в этой плоскости примем за начало прямоугольной системы координат (рис. 332). Проведем через O взаимно перпендикулярные оси, одну из которых примем за пространственную ось x, а другую — за временную ось τ в системе отсчета S. Штриховые прямые $\tau = x$ и $\tau = -x$ будут мировыми линиями световых сигналов, распространяющихся соответственно в положительном и отрицательном направлениях оси X. Пусть A — произвольное событие. Соединим точки A и O прямой мировой линией $\tau = x \operatorname{tg} \alpha$, описывающей равномерное движение какого-то тела вдоль оси X со скоростью β. Если $\beta < 1$, то с этим телом можно связать систему отсчета S', а прямую OA принять за пространственную ось X'. Ось времени τ' найдется из условия, что на ней $x' = 0$. Как видно из (105.21), это будет мировая линия $x = \beta\tau$, т.е. прямая OB, наклоненная под тем же углом $\alpha = \operatorname{arctg} \beta$, но уже к временной оси τ. В новой системе отсчета события O и A будут одновременными, но не одноместными. События же O и B одноместны, но не одновременны.

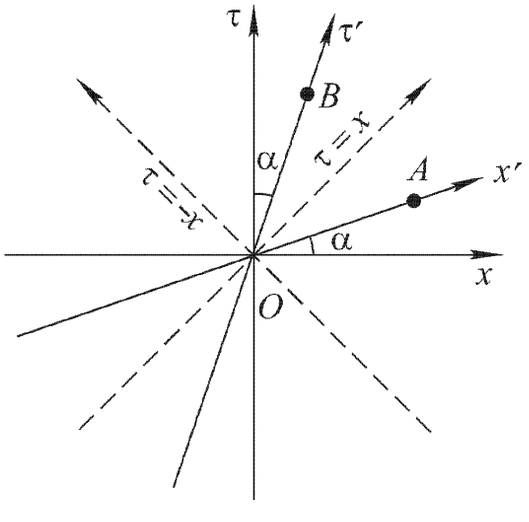


Рис. 332

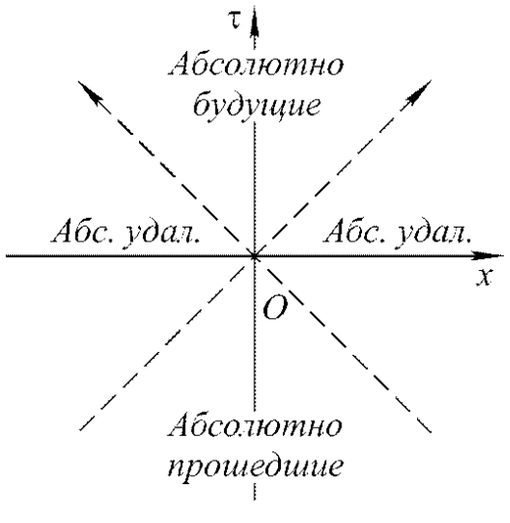


Рис. 333

Вообще, все события по отношению к событию O можно разделить на *абсолютно удаленные* и *абсолютно неодновременные* (т.е. *абсолютно прошедшие* и *абсолютно будущие*), как это показано на рис. 333. Для первых интервал между событиями чисто мнимый, для вторых — вещественный. Границей между такими событиями служат штриховые мировые линии световых сигналов, распространяющихся в положительном и отрицательном направлениях пространственной оси. В четырехмерном про-

странстве Минковского такой границей будет трехмерное многообразие, а именно конус $\tau^2 - x^2 - y^2 - z^2 = 0$, осью которого является ось времени τ. Он называется *световым конусом.*

Причинно связанными могут быть только события, интервал между которыми времениподобный. Например, событие O (см. рис. 332) могло бы быть причиной события B, так как в любой системе отсчета событие B наступает позже события O. Но события не могут быть причинно связанными, когда интервал между ними пространственноподобный. Таковы, например, события O и A (рис. 332). В неподвижной системе отсчета S событие A происходит позже события O. В штрихованной системе (x', τ') оба эти события одновременны. Если же взять систему отсчета, движущуюся быстрее системы (x', τ'), но все еще медленнее света, то ее пространственная ось будет на рис. 332 наклонена круче оси X'. В такой системе отсчета событие A произойдет раньше события O. Таким образом, нельзя удовлетворить требованию, чтобы в любой системе отсчета «причина» предшествовала «следствию». Это и доказывает наше утверждение.

Прямая OA с уравнением $\tau = \beta x$ есть мировая линия некоторого движения, происходящего со скоростью $x/\tau = 1/\beta$, т.е. со *сверхсветовой скоростью.* Существование сверхсветовых скоростей не противоречит теории относительности. Последняя допускает любые скорости. Однако в случае распространения состояний со сверхсветовыми скоростями интервал между любыми двумя состояниями будет пространственноподобным, а потому каждое из этих состояний не может быть причиной другого. Такие процессы не могут служить «сигналами» для передачи информации. *Все тела и сигналы, передающие воздействие, не могут распространяться со скоростью, превышающей скорость света в вакууме. Скорость света в вакууме есть максимально возможная скорость распространения взаимодействий* [1]).

§ 106. Лоренцово сокращение длины и замедление времени

1. Если твердый стержень покоится в какой-то системе отсчета, то его длина l_0 определяется сравнением с масштабным

[1]) За последние примерно 10 лет обсуждается вопрос о существовании гипотетических сверхсветовых частиц — *тахионов.* Ясно, что если бы такие частицы существовали, то они не могли бы служить «сигналами» для передачи взаимодействий. Подробнее см. *Угаров В.А.* Специальная теория относительности.—М.: Наука, 1977.

стержнем, покоящимся в той же системе отсчета. Величину l_0 можно назвать *собственной длиной стержня*, поскольку она не зависит от выбора системы отсчета, в которой покоится стержень. Но если стержень движется, то необходимо условиться, что понимать под его длиной в покоящейся системе отсчета. Во избежание недоразумений специально подчеркнем, что все измерения расстояний и промежутков времени во всякой системе отсчета должны производиться с помощью масштабных стержней и часов, *неподвижных* в этой системе.

Длиной l движущегося стержня в покоящейся системе отсчета называется расстояние между двумя точками в этой системе, мимо которых концы стержня проходят одновременно. Для нахождения связи между l и l_0 воспользуемся частной формой преобразования Лоренца (105.12). Пусть стержень покоится в системе S' и лежит на оси X'. Тогда разность координат его концов $\Delta x'$ в системе S' и будет длиной l_0 покоящегося стержня. Разность же координат тех же концов Δx в системе S, взятая в один и тот же момент t, будет длиной l движущегося стержня. Но из первой формулы (105.12) при $t = \text{const}$ следует $\Delta x' = \Delta x/\sqrt{1-\beta^2}$, а потому

$$l = l_0\sqrt{1-\beta^2}. \tag{106.1}$$

Таким образом, *длина движущегося стержня короче, чем покоящегося.* Это явление называется *лоренцовым сокращением длины.* Первоначально оно было введено независимо друг от друга Фицжеральдом (1851–1901) и Лоренцом. Но у них это была гипотеза *ad hoc*, специально придуманная для объяснения отрицательного результата опыта Майкельсона, хотя Лоренц и пытался обосновать ее с точки зрения электронной теории. В теории относительности лоренцово сокращение получается, а его истинный физический смысл устанавливается без каких бы то ни было добавочных предположений. Это, конечно, не исключает возможности атомистического объяснения лоренцова сокращения, а также явления замедления хода движущихся часов, о котором говорится дальше. Но для этого надо располагать не только уравнениями электродинамики, но и пока еще не известными законами, определяющими строение вещества. Когда такие законы будут установлены, можно на их основе рассмотреть и вопрос об изменении длин и времен. Однако если теория относительности верна, то результат такого рассмотрения заранее известен. Правильные законы природы должны быть *инвариантны относительно преобразования Лоренца*, а потому в вопросе о сокращении длин и замедлении времени они не могут привести к иным результатам, чем теория относительности.

2. Рассмотрим теперь какие-либо два события, интервал между которыми времениподобный. Промежутки времени Δt и $\Delta t'$ между этими событиями, измеренные в «неподвижной» S и «движущейся» S' системах отсчета, вообще говоря, будут разными. Конечно, обе системы S и S' совершенно равноправны, так что между Δt и $\Delta t'$ может существовать любое соотношение $\Delta t \lessgtr \Delta t'$. Допустим теперь, что в качестве движущейся системы S' взята такая, в которой оба события происходят *в одном и том же месте пространства.* (В случае времениподобных интервалов такая система существует, см. п. 5 предыдущего параграфа.) Тогда система S' становится выделенной среди всех прочих инерциальных систем отсчета. Время, измеренное в такой системе, будем обозначать через t_0, а в «неподвижной» системе S — по-прежнему через t. Для нахождения соотношения между Δt и Δt_0 будем предполагать, что в (105.14) координата x' постоянна, т.е. одна и та же для обоих событий. Тогда из последней формулы (105.14) получим $\Delta t = \Delta t'/\sqrt{1-\beta^2}$. Но, согласно нашему определению, $\Delta t'$ и есть Δt_0. Следовательно,

$$\Delta t_0 = \Delta t\sqrt{1-\beta^2}. \tag{106.2}$$

Из (106.2) видно, что $\Delta t_0 < \Delta t$, т.е. *промежуток времени между двумя событиями минимален в той системе отсчета, в которой эти события одноместны.* Это явление называется *замедлением хода движущихся часов.* Причина такого названия заключается в следующем. Допустим, что взяты какие-то одни часы, помещенные в одну из точек системы S'. Такие часы называются «движущимися», поскольку они движутся вместе с системой S'. Все же часы, покоящиеся в системе S, называются «неподвижными». Ясно, что в системе S' любые два события, происходящие в месте нахождения «движущихся» часов, будут одноместны, так что эти часы измерят промежуток времени Δt_0. Здесь моменты наступления обоих событий отмечаются по *одним и тем же* («движущимся») часам, так что никакой синхронизации не требуется. Напротив, в «неподвижной» системе S моменты наступления событий отмечаются по часам, находящимся в *различных местах пространства.* Для измерения Δt необходимо иметь *двое часов*, синхронизованных между собой по правилу Эйнштейна. С этим и связана *асимметрия* при измерении времени по «неподвижным» и «движущимся» часам, проявляющаяся в неравенстве $\Delta t_0 < \Delta t$.

3. Допустим теперь, что какая-либо частица движется относительно «неподвижной» системы отсчета S по криволинейной траектории с переменной по величине скоростью v. В специальной теории относительности допускаются только такие пространственно-временные системы отсчета, которые движут-

ся относительно S равномерно и прямолинейно. Возьмем бесконечное множество таких систем, движущихся со всевозможными скоростями и во всевозможных направлениях. Система отсчета, относительно которой мгновенная скорость частицы равна нулю, и связанные с ней часы называются *сопутствующими*. При движении частица непрерывно переходит из одной сопутствующей системы отсчета в другую. Разобьем траекторию частицы в системе S на бесконечно короткие отрезки. Пусть dt — время, затрачиваемое в системе S на прохождение одного из таких отрезков. Согласно (106.2), по сопутствующим часам на то же движение потребуется время $dt_0 = dt\sqrt{1 - v^2/c^2}$. Конечный промежуток времени, измеренный по неподвижным часам, представится интегралом

$$t_2 - t_1 = \int \frac{dt_0}{\sqrt{1 - v^2/c^2}}, \tag{106.3}$$

взятым по всей траектории движения.

Если часы движутся вместе с частицей, то возникает вопрос, как связана длительность, измеренная по таким часам, с длительностью, измеренной по сопутствующим часам? На этот вопрос нельзя ответить *без рассмотрения конкретного устройства часов*, так как при ускоренном движении появляются силы инерции, влияющие на ход часов. Например, в случае движения пружинных часов силы инерции могут деформировать балансир или другие части их. Деформации могут быть настолько большими, что часы остановятся или сломаются. Маятниковые часы вообще станут непригодными для измерения времени в отсутствие поля тяготения. Сердце живого организма (а оно в принципе может рассматриваться тоже как «часы») перестанет биться, а сам организм погибнет, если ему сообщить достаточно большое ускорение, и т.д. Однако в принципе допустимы и такие часы, на ход которых сила тяжести и силы инерции оказывают ничтожное влияние, в пределе совсем исчезающее. Такие часы условимся называть *идеальными*. Измеряемая ими длительность совпадает с длительностью, измеренной по сопутствующим часам. Этот признак в принципе позволяет экспериментально проверить, являются ли те или иные конкретные часы идеальными или не являются. Ясно, что в формуле (106.3) под t_0 можно понимать время, отсчитываемое *по идеальным часам, движущимся вместе с частицей*. Это время называется *собственным временем*, поскольку оно инвариантно, т.е. не зависит от движения частицы. Из (106.3) видно, что длительности по «неподвижным» и «движущимся» идеальным часам связаны соотношением

$$t_2 - t_1 > t_{02} - t_{01}. \tag{106.4}$$

В частности, если часы вернутся в исходное положение (относи-

тельно системы S), то они покажут время, меньшее времени по «неподвижным» часам.

4. Наилучшим приближением к идеальным часам являются *атомные* и в особенности *ядерные часы*. Силы инерции по своему действию эквивалентны гравитационным силам (см. т. I, § 71). В обычных условиях как те, так и другие силы пренебрежимо малы по сравнению с электрическими и еще более мощными ядерными силами, определяющими процессы в электронных оболочках и ядрах атомов. В таких процессах гравитационные силы и силы инерции практически не играют никакой роли.

Рассмотрим, например, атом цезия-133 в *цезиевых эталонных часах*, с помощью которых устанавливается эталон времени — *секунда*. По принятому соглашению электромагнитное излучение такого атома в отсутствие внешних полей совершает $\nu = 9\,192\,631\,770 \approx 10^{10}$ колебаний в секунду. Радиус атома $\sim 10^{-8}$ см. Если атом уподобить гармоническому осциллятору, то при колебаниях будут развиваться громадные ускорения $\sim (2\pi\nu)^2 r \sim 10^{13}$ см/с^2. Если часам сообщить ускорение ~ 1 см/с^2, то для этого потребуются силы в 10^{13} раз слабее. Ясно, что никакого существенного действия на процессы внутри атома они оказать не могут. Только при измерениях времени с относительной точностью $\sim 10^{-12}$–10^{-13} такие ускорения и соответствующие им гравитационные поля, возможно, могут сказаться на ходе часов и результатах измерений. Тогда их надо учитывать. Из сказанного ясно, что из всех часов атомные и ядерные часы являются, пожалуй, единственными, которые пригодны для обнаружения и исследования релятивистского замедления времени, по крайней мере при скоростях, малых по сравнению со скоростью света.

Явление радиоактивного распада атомов или других нестабильных частиц может выполнять роль идеальных часов. Радиоактивный распад подчиняется закону

$$n = n_0 e^{-t/\tau}, \tag{106.5}$$

где n_0 — начальное число частиц, а n — число их через время t. Постоянная τ называется *временем жизни* рассматриваемой нестабильной частицы. О времени t можно судить по отношению n/n_0. Релятивистское замедление времени было подтверждено в явлении распада *мюонов* (μ-мезонов). Так называются нестабильные заряженные частицы, масса которых в 207 раз превышает массу электрона. Заряд мюона равен заряду электрона, но может быть и положительным, и отрицательным. Мюоны образуются в космических лучах в верхних слоях атмосферы (на высотах порядка 10 км). Сравнение интенсивностей потока мюонов в космических лучах на горе и у ее основания показало,

что среднее время жизни мюона в лабораторной системе отсчета $\tau \approx 10^{-5}$ с. С другой стороны, космические мюоны можно было замедлить в свинцовом блоке и с помощью специального устройства отфильтровать медленные мюоны. Измерения показали, что время жизни медленного (покоящегося) мюона $\tau_0 \approx \approx 2{,}20 \cdot 10^{-6}$ с. Если бы не было релятивистского замедления времени, то поток космических мюонов, даже если бы они двигались со скоростью света, уменьшался бы в e раз при прохождении расстояния ≈ 600 м. На расстоянии уже 5 км их интенсивность уменьшилась бы в $e^{50/6} \approx 4000$ раз, т.е. мюоны вообще не могли бы достигать поверхности земли. В действительности их интенсивность при прохождении такого расстояния уменьшается примерно в $e^{5/3} \approx 5$ раз. Учет релятивистского замедления времени устраняет это противоречие. Действительно, время жизни мюона в лабораторной системе отсчета τ связано с собственным временем жизни соотношением $\tau = \tau_0/\sqrt{1-\beta^2}$. Измерение средней кинетической энергии космических мюонов показало, что она $\approx 10^9$ эВ. По этой энергии нетрудно рассчитать $\sqrt{1-\beta^2} \approx 0{,}1$. Поэтому следует ожидать, что $\tau \approx 2{,}2 \cdot 10^{-6}/0{,}1 \approx \approx 2 \cdot 10^{-5}$ с. Это по порядку величины согласуется со значением, полученным на опыте.

5. С построением мощных ускорителей заряженных частиц опыты подобного рода производились в более определенных и лучше контролируемых условиях. Наиболее подходящими частицами для таких опытов оказались заряженные *пионы* (иначе называемые π-мезонами). Их масса в 273 раза больше массы электрона, а заряд равен заряду электрона. Пионы во множестве образуются при взаимодействии протонов высоких энергий с веществом. Среднее время жизни пиона в системе отсчета, где он покоится, $\tau_0 = 2{,}60 \cdot 10^{-8}$ с. На циклотроне Колумбийского университета был получен пучок пионов со скоростью $v = 0{,}75\,c$ ($\beta = 0{,}75$). Если бы не было релятивистского замедления времени, то за время τ_0 пучок проходил бы расстояние $0{,}75 \cdot 3 \cdot 10^{10} \times \times 2{,}60 \cdot 10^{-8} = 5{,}85$ м. На самом деле, как показали измерения, расстояние, на котором интенсивность пучка уменьшается в e раз, равно $8{,}5 \pm 0{,}6$ м, т.е. в лабораторной системе время жизни пиона τ в $1{,}45 \pm 0{,}11$ раза больше собственного времени жизни τ_0. Но это хорошо согласуется с формулой $\tau/\tau_0 = 1/\sqrt{1-\beta^2}$, которая дает $\tau/\tau_0 = 1{,}51$.

Результат опыта можно интерпретировать и как проявление *релятивистского сокращения длины*. Действительно, в системе отсчета, где пион покоится, его время жизни равно τ_0. Лаборатория движется относительно пиона со скоростью v. За время

τ_0 она проходит в системе пиона расстояние $l = v\tau_0$. Если то же расстояние измерять масштабным стержнем, который покоится в лаборатории, то оно окажется равным $l_0 = l/\sqrt{1-\beta^2} = 8{,}8$ м, что согласуется с опытом.

6. Хафель и Китинг поставили опыт для обнаружения релятивистского замедления хода часов уже в *макроскопических условиях*. В этом опыте четыре экземпляра цезиевых часов в октябре 1971 г. были помещены на реактивных самолетах, облетевших вокруг земного шара в восточном и западном направлениях. Временные интервалы $t_{\text{вост}}$ и $t_{\text{зап}}$, измеренные по часам, двигавшимся соответственно на восток и на запад, сравнивались с интервалами $t_{\text{неп}}$, измеренными эталонными неподвижными часами, находившимися в Морской обсерватории в Вашингтоне. После усреднения по четырем движущимся часам получились такие результаты:

$$t_{\text{вост}} - t_{\text{неп}} = (-59 \pm 10)\cdot 10^{-9}\ \text{с}, \quad t_{\text{зап}} - t_{\text{неп}} = (273 \pm 7)\cdot 10^{-9}\ \text{с}.$$

Посмотрим, что следует ожидать согласно теории относительности. Ускорение центра Земли, вызванное гравитационным притяжением Солнца, составляет примерно 0,18 ускорения, которое получают точки земного экватора из-за осевого вращения Земли. Хотя это ускорение и немало, его можно не принимать во внимание, ввиду *принципа эквивалентности сил тяготения и сил инерции* (см. т. I, § 71). Если не учитывать неоднородность гравитационного поля Солнца, то это поле будет полностью компенсировано центробежной силой инерции, обусловленной вращением центра Земли вокруг Солнца. Таким образом, можно считать, что относительно инерциальной системы отсчета центр Земли движется прямолинейно и равномерно. Поэтому невращающаяся система отсчета с началом координат в центре Земли практически будет также инерциальной системой отсчета. Ее мы и используем в последующих рассуждениях. Пусть самолеты облетают земной шар по параллели со скоростью v относительно Земли на постоянной высоте h. Их скорости относительно инерциальной системы отсчета будут $V_{\text{вост}} = V_0 + v$, $V_{\text{зап}} = V_0 - v$, где V_0 — скорость наземной лаборатории в той же системе. Сначала будем считать, что $h = 0$. Тогда, если t_0 — собственное время, то

$$t_{\text{неп}} = \int dt_0 \Big/ \sqrt{1 - V_0^2/c^2} \approx t_0 - \frac{V_0^2}{2c^2}\, t_0.$$

Аналогично,

$$t_{\text{вост}} \approx t_0 - \frac{(V_0+v)^2}{2c^2}\, t_0, \quad t_{\text{зап}} \approx t_0 - \frac{(V_0-v)^2}{2c^2}\, t_0.$$

Вычтем отсюда предыдущие выражения и учтем, что в окончательных формулах в пределах точности расчета время t_0 можно

заменить на $t_{\text{неп}}$. Тогда получим

$$t_{\text{вост}} - t_{\text{неп}} = -\frac{2V_0 v + v^2}{2c^2}\, t_{\text{неп}}, \quad t_{\text{зап}} - t_{\text{неп}} = \frac{2V_0 v - v^2}{2c^2}\, t_{\text{неп}}. \quad (106.6)$$

Естественно, под $t_{\text{неп}}$ надо понимать время нахождения самолетов в воздухе *без учета времени остановок*, так как время остановок не влияет на величину разностей $t_{\text{вост}} - t_{\text{неп}}$ и $t_{\text{зап}} - t_{\text{неп}}$.

Допустим, что самолет облетает земной шар в течение суток. Тогда для широты Вашингтона потребуется скорость $v = V_0 \approx \approx 1000$ км/ч ≈ 300 м/с. По формулам (106.6) найдем $t_{\text{вост}} - t_{\text{неп}} \approx \approx -130 \cdot 10^{-9}$ с, $t_{\text{зап}} - t_{\text{неп}} \approx +43 \cdot 10^{-9}$ с.

В приведенном вычислении не учтено влияние *гравитационного потенциала Земли* на течение времени (см. § 109, а также т. I, § 72). С учетом этого обстоятельства оба выражения (106.6) надо увеличить на

$$\Delta t_{\text{грав}} = \frac{gh}{c^2}\, t_{\text{неп}}, \quad (106.7)$$

где h — высота полета, а g — ускорение свободного падения на поверхности Земли. Если положить $h = 10$ км, то в приведенном выше примере получится $\Delta t_{\text{грав}} = 94 \cdot 10^{-9}$ с, т.е. «гравитационный эффект» того же порядка, что и учтенный выше «кинематический эффект». Хафель и Китинг провели вычисление с учетом этой поправки, выполнив интегрирование по фактически совершенным маршрутам самолетов. Они получили теоретические значения:

$$t_{\text{вост}} - t_{\text{неп}} = (-40 \pm 23) \cdot 10^{-9} \text{ с}, \quad t_{\text{зап}} - t_{\text{неп}} = (275 \pm 21) \cdot 10^{-9} \text{ с},$$

удовлетворительно согласующиеся с полученными на опыте. Заметим, что из разности $(t_{\text{зап}} - t_{\text{вост}})$ поправка $\Delta t_{\text{грав}}$ исключается. Эта разность обусловлена только «кинематическим эффектом» замедления времени. Совпадение теоретических результатов с измеренными на опыте рассматривается как доказательство не вызывающего сомнения релятивистского замедления хода движущихся часов.

7. Остановимся в заключение на так называемом *парадоксе близнецов*. Из двух братьев-близнецов A остается на Земле, а B отправляется в кругосветное путешествие на межзвездном корабле, двигаясь со скоростью, близкой к скорости света. Через 5 лет по своим часам брат B возвращается обратно и находит брата A глубоким стариком. Оказалось, что за время путешествия по часам на Земле прошло 50 лет. Таким образом, открывается возможность за время человеческой жизни совершать не только путешествия к далеким звездным мирам, но и *путешествия в будущее*. Если отвлечься от технической и практической стороны вопроса, то такие путешествия принципиально возможны.

В самом деле, биологические процессы не представляют собой какую-то обособленную группу явлений природы. Как и прочие явления природы, они подчиняются законам физики. Если на межзвездном корабле создать условия, близкие к условиям на Земле, то и жизненные процессы на нем будут протекать практически так же, как и на Земле. Биения сердца в человеческом организме выполняют роль часов. Если за время жизни сердце человека на Земле совершает $2 \cdot 10^9$ ударов, то столько же ударов оно совершит и на корабле. Но движущиеся часы идут медленнее неподвижных. Если за время путешествия сердце брата B совершит $1{,}5 \cdot 10^8$ ударов, то на Земле к моменту встречи сердце брата A успеет совершить ударов в 10 раз больше. Но это и есть «парадокс близнецов».

В 1974 г. парадокс близнецов был подтвержден экспериментально на ускорителе в ЦЕРНе (Европейский центр по ядерным исследованиям, Швейцария). Ускоренные мю-мезоны удерживались магнитным полем на круговой орбите радиуса 5 м в течение свыше 150 мкс. За это время они совершали более 10^5 оборотов. Энергия мезонов превышала энергию покоя примерно в 12 раз, так что $1/\sqrt{1-\beta^2} = 12$. Поэтому ожидаемое время жизни мезона в лабораторной системе должно составлять $2{,}2 \cdot 12 = 26{,}4$ мкс. Опыт дал для этого времени $26{,}37 \pm 0{,}05$ мкс.

§ 107. Эффект Доплера и аберрация света

1. Если в «неподвижной» системе S распространяется монохроматическая волна с определенной частотой ω и в определенном направлении, то в «движущейся» системе S' та же волна будет иметь другую частоту ω' и распространяться в другом направлении. Изменение частоты волны при переходе от одной системы отсчета к другой называется *эффектом Доплера* (1803–1853), а изменение направления — *аберрацией света*. При изложении теории этих явлений мы ограничимся случаем, когда световая волна распространяется *в вакууме*. Тогда в обеих системах отсчета пространство, в котором распространяется волна, будет *изотропным*. При наличии же среды изотропия сохранится только в той системе отсчета, в которой среда покоится. В системе отсчета, относительно которой среда движется, появляется анизотропия, вносящая осложнения в теорию эффекта Доплера и аберрации в материальных средах.

В релятивистской теории изменение частоты и направления распространения волны проще всего можно определить из условия *равенства фаз* одной и той же волны в обеих системах отсчета. Предполагая, что волна плоская, это условие можно записать

в виде

$$\omega t - \mathbf{kr} = \omega' t' - \mathbf{k}'\mathbf{r}', \qquad (107.1)$$

где $\mathbf{r}$, t и $\mathbf{r}'$, t' — координаты и время одного и того же события в рассматриваемых системах отсчета, причем начало отсчета времени выбрано так, что в момент совмещения начал координат обеих систем $t = t' = 0$.

Для доказательства сформулированного условия предположим, что в началах координат в момент $t = t' = 0$ на волне сделана метка, переносимая волной с фазовой скоростью, так что фаза колебания в месте нахождения метки в обеих системах отсчета будет $\varphi_1 = \varphi_1' = 0$. Вообразим теперь произвольно движущихся наблюдателей A и A', из которых A для определения своего положения в пространстве и времени пользуется координатами и часами системы S, а A' — координатами и часами системы S'. Можно, например, не нарушая общности, предположить, что наблюдатель A покоится в системе отсчета S, а наблюдатель A' — в системе отсчета S'. Однако это предположение в такой специальной форме не обязательно. (Обычно при рассмотрении эффекта Доплера систему S связывают с источником, а систему S' — с наблюдателем.) Оба наблюдателя начинают счет проходящих мимо них волн с того момента, когда мимо них проходит сделанная метка. Пусть оба наблюдателя встречаются друг с другом. Обозначим через $\mathbf{r}$, t координаты и время этого события в системе S, а через $\mathbf{r}'$, t' — в системе S'. Наблюдатель A найдет, что к моменту встречи фаза пришедшего к нему колебания будет $\varphi_2 = \omega t - \mathbf{kr}$, и сосчитает к этому моменту

$$N = \frac{\varphi_2 - \varphi_1}{2\pi} = \frac{\omega t - \mathbf{kr}}{2\pi}$$

волн. Наблюдатель A' найдет, что к моменту встречи мимо него прошло

$$N' = \frac{\omega' t' - \mathbf{k}'\mathbf{r}'}{2\pi}$$

волн. Но число прошедших волн не может зависеть от того, кто их считает, а потому

$$N = N', \qquad \text{или} \qquad \omega t - \mathbf{kr} = \omega' t' - \mathbf{k}'\mathbf{r}',$$

что и требовалось доказать.

Для нахождения формул, определяющих преобразование частоты и направления распространения волны в релятивистской теории, достаточно в формуле (107.1) выразить переменные $\mathbf{r}(x, y, z)$ и t через переменные $\mathbf{r}'(x', y', z')$ и t', воспользовавшись для этого формулами преобразования Лоренца (105.14). Сделав это и сравнив коэффициенты при одинаковых переменных,

получим

$$\omega' = \frac{\omega - k_x V}{\sqrt{1-\beta^2}}, \tag{107.2}$$

$$k'_{x'} = \frac{k_x - \omega V/c^2}{\sqrt{1-\beta^2}}, \quad k'_{y'} = k_y, \quad k'_{z'} = k_z. \tag{107.3}$$

Аналогично получаются формулы обратного преобразования:

$$\omega = \frac{\omega' + k'_{x'} V}{\sqrt{1-\beta^2}}, \tag{107.4}$$

$$k_x = \frac{k'_{x'} + \omega V/c^2}{\sqrt{1-\beta^2}}, \quad k_y = k'_{y'}, \quad k_z = k'_{z'}. \tag{107.5}$$

Формулы (107.2) и (107.4) описывают эффект Доплера, а формулы (107.3) и (107.5) — аберрацию света.

2. Рассмотрим частный случай, когда волна распространяется вдоль оси X. В этом случае полагаем в формуле (107.2) $k_x = \omega/c$ и получаем

$$\omega' = \omega\sqrt{\frac{1-\beta}{1+\beta}}. \tag{107.6}$$

Изменение частоты в этом случае называется *продольным эффектом Доплера.* Понятно, что результат (107.6) можно получить и из общей формулы (107.4). Частота волны при наблюдении из системы S' уменьшается, когда эта система движется в направлении распространения волны, и увеличивается, когда она движется навстречу волне. При медленных движениях, когда можно пренебречь квадратами отношения $\beta \equiv V/c$, формула (107.6) переходит в

$$\omega' \approx \omega(1-\beta) = \omega(1 - V/c). \tag{107.7}$$

3. Рассмотрим теперь второй частный случай, когда наблюдение ведется в системе отсчета S' и притом перпендикулярно к направлению распространяющейся в ней волны. В этом случае удобнее воспользоваться формулой (107.4), полагая в ней $k'_{x'} = 0$. Это дает

$$\omega' = \omega\sqrt{1-\beta^2} \approx \omega(1 - \beta^2/2). \tag{107.8}$$

Происходит смещение частоты *в длинноволновую область спектра* во втором порядке по β. Это — чисто релятивистский эффект, невозможный в классической (дорелятивистской) теории. Он называется *поперечным эффектом Доплера.*

Для установления физического смысла поперечного эффекта Доплера предположим, что мимо наблюдателя прошел цуг из N волн. Наблюдатель измерил по своим часам время прохождения этого цуга и нашел для него значение t'. Очевидно, $\omega' = 2\pi N/t'$.

Найдем длительность того же процесса по часам системы S. Для этого надо отметить показания таких часов в моменты прохождения начала и конца цуга мимо наблюдателя. Разность этих показаний и определит длительность цуга t по часам системы S. Опять можно написать $\omega = 2\pi N/t$ и, следовательно, $\omega'/\omega = t/t'$. В этих измерениях время t' измеряется с помощью *одних и тех же часов*, неподвижных относительно наблюдателя, а для измерения времени t надо пользоваться *различными часами* системы S, синхронизованными между собой. Поэтому часы наблюдателя надо рассматривать как «неподвижные», а все часы системы S — как «движущиеся» (см. § 106, п. 2). Но движущиеся часы идут медленнее неподвижных, причем $t/t' = \sqrt{1-\beta^2}$, а потому $\omega' = \omega\sqrt{1-\beta^2}$. Таким образом, поперечный эффект Доплера есть не что иное, как проявление релятивистского эффекта *замедления хода движущихся часов*. Тем же эффектом объясняется появление квадратного корня $\sqrt{1-\beta^2}$ в формулах (107.2) и (107.4).

4. Поперечный эффект Доплера впервые удалось наблюдать Айвсу и Стилуэллу в 1938 г., а затем в 1941 г. Источником света служил пучок каналовых лучей атомов водорода, несущихся со значительной скоростью $v \sim 10^8$ см/с. Специальная конструкция трубки обеспечивала высокую однородность каналовых лучей по скоростям. Скорость v вычислялась по приложенному к электродам напряжению и по величине продольного доплеровского эффекта. Свет, испущенный движущимися атомами перпендикулярно к направлению их движения, попадал в спектрограф. По оценке авторов спектрограф давал возможность измерять длину волны с точностью 0,00025 нм. Главная трудность опыта состояла в том, что при отклонении направления пучка каналовых лучей от строгой перпендикулярности к линии наблюдения поперечное доплеровское смещение, пропорциональное $(v/c)^2$, могло оказаться перекрытым более сильным продольным эффектом Доплера, поскольку этот эффект — первого порядка по $(v\cos\vartheta)/c$ (ϑ — угол между скоростью атома и линией наблюдения). При скорости атома $v \approx 10^8$ см/с оба эффекта сравнимы между собой уже при $\vartheta \sim 5'$. Эта трудность была преодолена тем, что ось спектрографа устанавливалась перпендикулярно к плоскости зеркала. В спектрограф попадал не только свет, непосредственно излученный атомом, но и свет, отраженный от зеркала. Последний можно рассматривать как свет, излученный движущимся изображением атома. Если атом движется под углом ϑ к линии наблюдения, то его изображение в зеркале будет двигаться под углом $\pi - \vartheta$. Доплеровские смещения первого порядка для атома и его изображения в зеркале

будут равны по величине, но противоположны по знаку. Если бы не было смещения второго порядка, то в спектрографе наблюдались бы две линии $\lambda_1 = \lambda - \Delta\lambda$ и $\lambda_2 = \lambda + \Delta\lambda$, симметрично расположенные относительно исходной линии λ. Эта симметрия будет нарушена поперечным доплеровским смещением $\delta\lambda$, которое происходит всегда в длинноволновую сторону спектра. С учетом этого $\lambda_1 = \lambda - \Delta\lambda + \delta\lambda$, $\lambda_2 = \lambda + \Delta\lambda + \delta\lambda$. Отсюда

$$\delta\lambda = \frac{\lambda_1 + \lambda_2}{2} - \lambda.$$

Таким путем удается отделить квадратичный (поперечный) эффект Доплера от линейного (продольного). Согласно (107.8), изменение длины волны при поперечном эффекте Доплера определяется выражением

$$\delta\lambda = \frac{1}{2}\lambda\beta^2. \tag{107.9}$$

В одном из опытов при напряжении на трубке 6788 В ожидаемое смещение, вычисленное по напряжению, должно было быть $\delta\lambda = 0{,}00116$ нм, а вычисленное по линейному доплеровскому смещению 0,00109 нм. Опыт дал $\delta\lambda = 0{,}0011$ нм. Подобное согласие получалось и при других напряжениях на трубке.

Результаты Айвса и Стилуэлла были подтверждены и другими учеными. Это были первые опыты, на которых впервые было экспериментально доказано *релятивистское замедление времени*, хотя сами Айвс и Стилуэлл, ничего не говоря о теории относительности, пытались интерпретировать их с точки зрения старой теории эфира.

5. Перейдем теперь к аберрации света. Введем единичные векторы $\mathbf{N}$ и $\mathbf{N}'$ в направлении распространения волны в системах отсчета S и S'. Тогда $\mathbf{k} = \omega\mathbf{N}/c$, $\mathbf{k}' = \omega'\mathbf{N}'/c$, и из формул (107.2) и (107.3) получим

$$N'_{x'} = \frac{N_x - \beta}{1 - \beta N_x}, \quad N'_{y'} = \frac{\sqrt{1-\beta^2}}{1 - \beta N_x}, \quad N'_{z'} = \frac{\sqrt{1-\beta^2}}{1 - \beta N_x}. \tag{107.10}$$

Аберрация света была открыта в 1727 г. королевским астрономом Брадлеем (1692–1762) при наблюдении звезд в телескоп. По этой причине ее называют также *астрономической аберрацией*. При описании этого явления в качестве «неподвижной» системы отсчета S возьмем систему Коперника (в которой Солнце считается неподвижным), а в качестве «движущейся» системы S' — систему, связанную с Землей. При этом мы полностью отвлечемся от вращения Земли вокруг собственной оси и будем рассуждать так, как если бы этого вращения совсем не было. Оси X и X' выберем в направлении орбитального движения Земли, скорость которого обозначим через V. Ось Y направим произвольно, но перпендикулярно к направлению движения Земли.

Обозначим далее через α угол между направлением на звезду (направлением оси телескопа) в системе S и положительным направлением оси Y (рис. 334). Тогда $N_x = -\sin\alpha$, $N_y = -\cos\alpha$, $N_z = 0$. Соответствующие величины в системе S' обозначим теми же буквами, но со штрихами. После подстановки в первую формулу (107.10) получим

$$\sin\alpha' = \frac{\sin\alpha + \beta}{1 + \beta\sin\alpha}, \qquad (107.11)$$

или

$$\sin\alpha = \frac{\sin\alpha' - \beta}{1 - \beta\sin\alpha'}. \qquad (107.12)$$

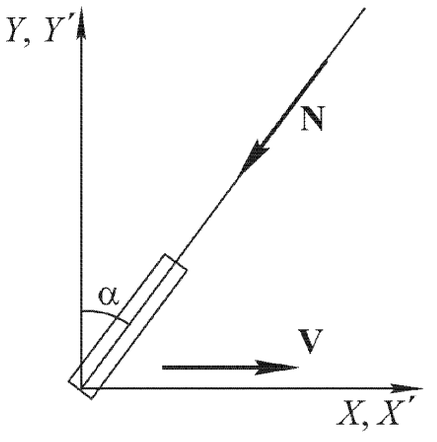


Рис. 334

Формулами (107.11) и (107.12) и определяется направление распространения волны в одной из систем отсчета, если известно это направление в другой системе отсчета.

Рассмотрим частный случай, когда свет в системе S распространяется вдоль оси Y. В этом случае $\alpha = 0$, и следовательно,

$$\sin\alpha' = \beta. \qquad (107.13)$$

Для того чтобы увидеть звезду в телескоп, его необходимо из перпендикулярного положения наклонить вперед на угол α', определяемый формулой (107.13). Угол α' называется *углом аберрации*. В аналогичном случае, когда свет в системе S' распространяется вдоль оси Y', т.е. когда $\alpha' = 0$, получим

$$\sin\alpha = -\beta. \qquad (107.14)$$

Телескоп надо наклонить на такой же угол, но уже назад. Скорость орбитального движения Земли $V \approx 30$ км/с, так что $\alpha \approx \approx V/c \approx 10^{-4}$ рад $\approx 20{,}5''$.

Если бы скорость $\mathbf{V}$ оставалась постоянной по величине и направлению, то видимое положение звезд на небесном своде оставалось бы неизменным и заметить аберрацию было бы нельзя. В действительности направление движения Земли по ее орбите непрерывно изменяется. Вследствие этого из-за аберрации будет изменяться и видимое положение звезд на небесном своде. Через полгода направление движения Земли изменится на противоположное, а угол аберрации изменит знак. Если звезда находится в плоскости земной орбиты, то ее видимое положение будет совершать в той же плоскости прямолинейное колебание с угловой амплитудой $\alpha = V/c = 20{,}5''$ и периодом в один год. Если направление на звезду (в системе S) перпендикулярно к плоскости земной орбиты, то видимое положение ее на небесном своде будет описывать окружность с угловым радиусом $20{,}5''$. Во вся-

ком промежуточном положении траекторией видимого положения звезды будет эллипс, большая полуось которого имеет такие же угловые размеры. Так выглядело бы явление аберрации света в воображаемом идеальном случае, если бы единственным движением Земли было только ее движение по круговой орбите вокруг Солнца. В действительности аберрация света есть ничтожный эффект, накладывающийся на движение звезд по небесному своду, обусловленное вращением Земли вокруг собственной оси. Требовалась исключительная наблюдательность и экспериментальное искусство, чтобы выделить и измерить этот эффект. Это и было сделано Брадлеем.

Подчеркнем особо, что явление аберрации света не имеет никакого отношения к самой скорости движения звезд относительно Земли. Оно отражает только *изменение скорости* этого относительного движения, обусловленное движением Земли. Вот почему аберрация света одна и та же *для всех звезд*, хотя их скорости относительно Земли весьма различны.

6. Брадлей видел в явлении аберрации доказательство конечности скорости распространения света. Измерив угол аберрации α, он вычислил эту скорость но формуле $\alpha = V/c$. По времени это было второе (после Рёмера) измерение скорости света. Сам Брадлей истолковал открытое им явление с точки зрения ньютоновской корпускулярной теории света. В нерелятивистской кинематике скорости световой корпускулы $\mathbf{v}$ и $\mathbf{v}'$ в системах отсчета S и S' связаны соотношением $\mathbf{v} = \mathbf{v}' + \mathbf{V}$. Ограничимся случаем, когда в системе S световая корпускула движется по оси Y со скоростью $v_y = -c$. В системе S' появляется составляющая скорости корпускулы в перпендикулярном направлении $v'_x = v_x - V_x = -V$. Из-за этого траектория корпускулы в системе S' будет наклонена к оси Y под углом α', определяемым формулой

$$\operatorname{tg}\alpha' = \frac{v'_x}{v'_y} = \frac{V}{c} = \beta. \qquad (107.15)$$

Этой формулой в ньютоновской корпускулярной теории и определяется угол аберрации. При малых углах α' она совпадает с релятивистской формулой (107.13), если пренебречь квадратичными членами по β. Если бы к движению световой корпускулы применить релятивистский закон сложения скоростей (с учетом того, что скорость корпускулы по абсолютной величине равна c), то получились бы в точности прежние релятивистские формулы (107.10). Наблюдение аберрации света не позволяет, следовательно, сделать выбор между волновой и корпускулярной теориями света.

7. Истолкование аберрации в волновых теориях света, даже в их электромагнитной форме, встретило большие трудно-

сти, пока эти теории оставались механистическими. Все они были неудовлетворительны уже по той причине, что вводили произвольные предположения относительно механических свойств эфира и его движения относительно Земли. Впервые безупречная и в идейном отношении необычайно простая теория аберрации света (и оптического эффекта Доплера) была дана только в теории относительности.

В связи с этим затронем один вопрос, поставленный в эфирной теории света. Что будет с углом аберрации, если телескоп для его измерения заполнить водой? На этот вопрос эфирные теории света давали различный ответ, в зависимости от того, какие предположения вводили они о движении эфира относительно Земли и т.п. Опыт был поставлен Эйри в 1871 г. Оказалось, что при заполнении телескопа водой угол аберрации не изменяется. Объяснение этого результата в теории относительности не представляет затруднений. Для простоты рассуждений обратим направление распространения света, предположив, что источник света помещен в главном фокусе объектива телескопа. Поскольку нет никакого эфира, в системе отсчета, где телескоп покоится, вода или воздух, заполняющие его, а также стекло самого объектива телескопа оптически изотропны. В этой системе отсчета лучи выйдут из телескопа параллельно главной оптической оси, независимо от того, заполнен ли телескоп водой или не заполнен. Для определения угла аберрации надо выполнить переход к движущейся системе отсчета S'. Но это можно сделать для волны, *уже вышедшей из телескопа*. Направление этой волны совершенно не зависит от того, какой средой заполнен сам телескоп. Наличие телескопа и этой среды на таком переходе никак не отразится. Следовательно, и угол аберрации не будет зависеть от среды, заполняющей телескоп.

§ 108. Эффект Доплера в акустике и теории эфира

Эффект Доплера был открыт самим Доплером в 1842 г. на акустических волнах. В дальнейшем теория этого эффекта была перенесена без всяких изменений в оптику. При этом предполагалась справедливость волновой эфирной теории света. Место воздуха, в котором распространяются звуковые волны, в оптике занял световой эфир. В остальном все рассуждения в акустике и оптике были абсолютно тождественны. Эфирная теория безвозвратно ушла в область истории. Но акустический эффект Доплера полностью сохранил свое значение. Поэтому имеет смысл изложить здесь теорию этого эффекта. Поскольку, однако, этот том посвящен оптике, мы по-прежнему будем говорить о световых волнах в эфире. Для перехода к акустике слово «эфир» надо заменить словом «воздух», а световые волны — звуковыми.

В эфирной теории, помимо источника и наблюдателя, в явлении принимает участие промежуточная среда — световой эфир. С этим связаны

усложнение и неопределенность эфирной теории эффекта Доплера, поскольку в каждом конкретном случае мы не можем сказать, как движутся источник и наблюдатель относительно «неуловимого» эфира. Различные теории эфира отличались друг от друга прежде всего тем, как они выбирали систему отсчета, в которой эфир покоится и, следовательно, ведет себя как оптически изотропная среда. Пусть в эфире распространяется плоская монохроматическая волна. Ее частоту в системе отсчета, в которой эфир покоится, обозначим через ω, а волновой вектор — через $\mathbf{k} = \omega\mathbf{N}/c$, где $\mathbf{N}$ — единичный вектор, в направлении которого волна распространяется. Фаза волны в указанной системе отсчета представится выражением $\varphi = \omega t - \mathbf{kr}$. Пусть относительно эфира равномерно движутся источник со скоростью $\mathbf{V}_{\text{и}}$ и наблюдатель со скоростью $\mathbf{V}_{\text{н}}$. Их радиусы-векторы будут соответственно $\mathbf{r}_{\text{и}} = \mathbf{V}_{\text{и}}t$ и $\mathbf{r}_{\text{н}} = \mathbf{V}_{\text{н}}t$. Фазы колебаний в этих движущихся точках определятся выражениями $\varphi_{\text{и}} = (\omega - \mathbf{kV}_{\text{и}})t$ и $\varphi_{\text{н}} = (\omega - \mathbf{kV}_{\text{н}})t$. Отсюда следует, что источник в системе отсчета, где он покоится, излучает волны с частотой $\omega_{\text{и}} = \omega - \mathbf{kV}_{\text{и}}$, а частота, воспринимаемая наблюдателем, определяется выражением $\omega_{\text{н}} = \omega - \mathbf{kV}_{\text{н}}$. Почленным делением с учетом соотношения $\mathbf{k} = \omega\mathbf{N}/c$ исключаем промежуточную частоту ω и находим

$$\frac{\omega_{\text{н}}}{\omega_{\text{и}}} = \frac{1 - (\mathbf{NV}_{\text{н}})/c}{1 - (\mathbf{NV}_{\text{и}})/c}. \tag{108.1}$$

Это и есть основная формула, определяющая доплеровское изменение частоты в теории эфира и в акустике. Мы видим, что в этой теории частота $\omega_{\text{н}}$ определяется движением как источника, так и наблюдателя *относительно эфира*, а также направлением распространения $\mathbf{N}$ волны в самом «неподвижном» эфире. В этом ее отличие от теории относительности, в которой эфира нет, а потому эффект Доплера зависит только от скорости источника *относительно наблюдателя* $(\mathbf{V}_{\text{и}} - \mathbf{V}_{\text{н}})$. В частности, при одной и той же относительной скорости $(\mathbf{V}_{\text{и}} - \mathbf{V}_{\text{н}})$ формула (108.1) приводит к разным результатам, в зависимости от того, что движется: источник или наблюдатель. Когда движется источник, а наблюдатель неподвижен, она дает

$$\omega_{\text{н}} = \frac{\omega_{\text{и}}}{1 - \mathbf{NV}_{\text{и}}/c}. \tag{108.2}$$

Если же движется наблюдатель, а источник остается в покое, то

$$\omega_{\text{н}} = \omega_{\text{и}}(1 - \mathbf{NV}_{\text{н}}/c). \tag{108.3}$$

В линейном приближении, когда в формуле (108.1) можно пренебречь квадратами обоих отношений $V_{\text{и}}/c$ и $V_{\text{н}}/c$, она переходит в

$$\frac{\omega_{\text{н}}}{\omega_{\text{и}}} = 1 + \frac{\mathbf{N}}{c}(\mathbf{V}_{\text{и}} - \mathbf{V}_{\text{н}}). \tag{108.4}$$

В эту формулу входит лишь относительная скорость $(\mathbf{V}_{\text{и}} - \mathbf{V}_{\text{н}})$, а не скорости $\mathbf{V}_{\text{и}}$ и $\mathbf{V}_{\text{н}}$ в отдельности. Однако формула еще не определяет изменения частоты, поскольку в нее входит также направление распространения волны $\mathbf{N}$ в «неподвижном» эфире. Действительно, хотя волна и посылается источником к наблюдателю, ее направление из-за аберрации будет изменяться с изменением движения эфира. Только в частном случае, когда источник или наблюдатель покоятся относительно эфира, эта неопределенность исчезает, а формула (108.4) совпадает с тем, что дает в первом порядке теория относительности. Тогда ей можно придать вид

$$\omega_{\text{н}} = (1 + V_N/c), \tag{108.5}$$

где V_N — скорость источника относительно наблюдателя по лучу зрения. Она считается положительной, когда источник приближается к наблюдателю, и отрицательной, когда он удаляется.

Все изложенное показывает, какие существенные упрощения и определенность внесла теория относительности в теорию эффекта Доплера и аберрации света.

ЗАДАЧА

На рис. 335 приведена схема опыта Фуко по определению скорости света в материальной среде. Лучи от источника S, пройдя через стеклянную пластинку M и линзу L, отражаются от плоского зеркала R, которое может вращаться вокруг оси, перпендикулярной к плоскости чертежа. Линза L дает изображение источника S на поверхности вогнутого зеркала C, центр кривизны которого совпадает с осью вращения зеркала R. Сосуд P наполняют исследуемым веществом, в котором измеряется скорость света. Если зеркало R неподвижно, то лучи, отраженные от C и R, снова сойдутся в точке S. Зеркало M отклонит часть лучей в сторону и даст действительное изображение источника в S_1. При вращении зеркала R изображение S_1 смещается в S_1'. По величине смещения S_1S_1' можно вычислить скорость света в исследуемом веществе. При отражении от движущегося зеркала происходит доплеровское изменение частоты света. Учитывая это, показать, что метод Фуко дает *групповую скорость света.*

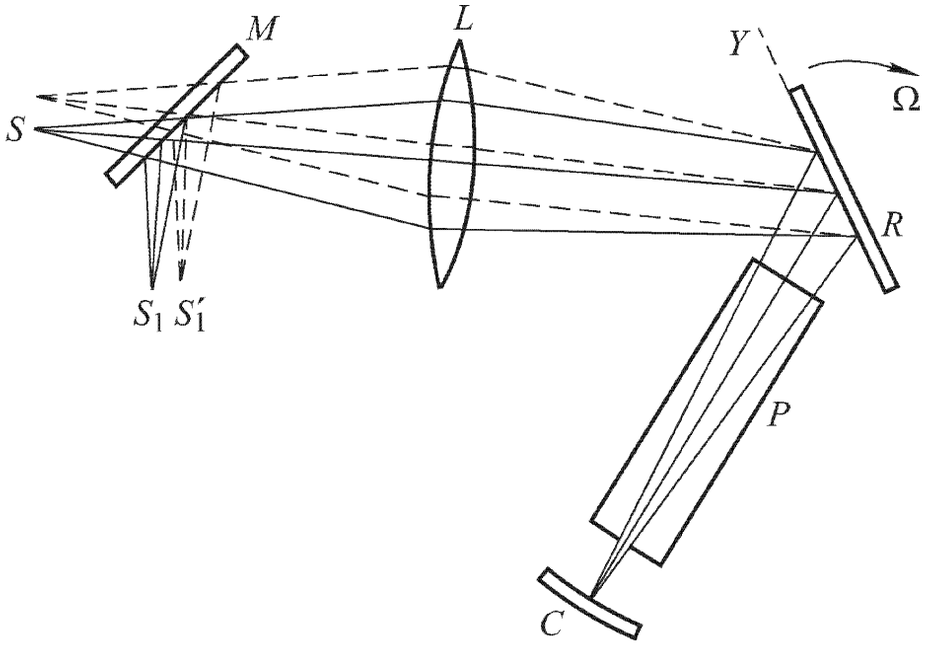


Рис. 335

Р е ш е н и е. Не теряя общности, воздушные зазоры между R и P, а также между P и C можно считать бесконечно тонкими и во всех расчетах пренебречь толщиной этих зазоров.

Обычный расчет проводится следующим образом. На прохождение расстояния от R до C и обратно волновой фронт, распространяющийся с фазовой скоростью v, затрачивает время $T = 2D/v$, где D — расстояние между зеркалами R и C. За это время R повернется на угол $\varphi = T\Omega$, если Ω — угловая скорость вращения зеркала. Луч, отраженный от зеркала R, вращается с вдвое большей скоростью. За то же время он повернется на угол

$\alpha = 2\varphi = 2T\Omega = (4D/v)\Omega$. Угол α легко рассчитать по величине смещения S_1S_1'. Таким образом,

$$v = 4D\Omega/\alpha. \tag{108.6}$$

В этом рассуждении не принято во внимание доплеровское изменение частоты при отражении света от вращающегося зеркала R. Поэтому оно не дает ответа, что за скорость вычисляется по формуле (108.6).

Поместим начало координат на оси вращающегося зеркала R и направим ось Y по линии пересечения плоскости этого зеркала с плоскостью чертежа. Так как линейные скорости различных точек вращающегося зеркала различны, то и изменение частоты волны при ее отражении от зеркала будет разным в зависимости от того, в каком месте зеркала произошло отражение. Благодаря этому различные точки волнового фронта будут распространяться в среде с различными фазовыми скоростями. Это поведет к вращению волнового фронта в среде. Если за направление положительного вращения принять направление вращения зеркала R, то для угловой скорости вращения волнового фронта в среде можно написать

$$\Omega' = \frac{1}{\cos\varphi}\frac{dv}{dy} = \frac{1}{\cos\varphi}\frac{dv}{d\omega}\frac{d\omega}{dy},$$

где φ — угол падения светового луча на зеркало R. Так как $v = \omega/k$, то

$$\frac{dv}{d\omega} = \frac{1}{k} - \frac{\omega}{k^2}\frac{dk}{d\omega} = \frac{v}{\omega} - \frac{v^2}{\omega u},$$

где u — групповая скорость. Остается определить $d\omega/dy$. Если ω — частота волны, отраженной от зеркала в точке с координатой y, а $\omega + d\omega$ — с координатой $y + dy$, то в первом порядке $d\omega/\omega = -\dfrac{2}{c}\,\Omega\cos\varphi\,dy$, откуда

$$\frac{1}{\cos\varphi}\frac{d\omega}{dy} = -2\Omega\frac{\omega}{c} = -2\Omega\frac{\omega}{nv},$$

где n — показатель преломления. Таким образом,

$$\Omega' = \left(\frac{v^2}{\omega u} - \frac{v}{u}\right)\cdot 2\Omega\frac{\omega}{nv} = \frac{2\Omega}{n}\left(\frac{v}{u} - 1\right).$$

Отраженный от зеркала C волновой фронт будет также поворачиваться при распространении в веществе с угловой скоростью Ω' и притом, как легко сообразить, в том же направлении, что и падающий волновой фронт. С другой стороны, на прохождение слоя вещества толщиной $2D$ волновой фронт затрачивает время $T = 2D/v$. За это время он повернется в среде на угол $\Omega'T = \dfrac{4D\Omega}{nv}\left(\dfrac{v}{u} - 1\right)$. По выходе из сосуда P в вакуум волновой фронт преломляется, вследствие чего угол поворота увеличивается в n раз и становится равным

$$n\Omega'T = \frac{4D\Omega}{v}\left(\frac{v}{u} - 1\right) = \frac{4D\Omega}{u} - \frac{4D\Omega}{v}.$$

Этот поворот надо прибавить к повороту $4D\Omega/v$, найденному ранее без учета эффекта Доплера. Таким образом, измеряемый угол поворота α в действительности равен

$$\alpha = \frac{4D\Omega}{v} + \left(\frac{4D\Omega}{u} - \frac{4D\Omega}{v}\right) = \frac{4D\Omega}{u},$$

так что вместо формулы (108.6) получится

$$u = 4D\Omega/\alpha. \tag{108.7}$$

Следовательно, метод вращающегося зеркала Фуко дает групповую скорость.

§ 109. Замедление хода часов в гравитационном поле

1. Вопросы релятивистской теории тяготения относятся к компетенции *общей теории относительности*, излагать которую мы не предполагаем. Затронем лишь кратко вопрос о влиянии поля тяготения на течение времени. Будем исходить из *принципа эквивалентности гравитационных сил и сил инерции*. Он был подробно изложен в § 71 первого тома. Гравитационное поле будем предполагать слабым (критерий слабости поля указывается ниже).

Возьмем какую-либо инерциальную систему отсчета S, в различных точках которой установлены неподвижные часы, синхронизованные по правилу Эйнштейна. Как всегда, время, определяемое по таким синхронизованным часам, будем обозначать через t. Пусть относительно S ускоренно движутся какие-то одни часы A. Время, отмечаемое часами A в точках, через которые они проходят, условимся обозначать через t_0. Часы A предполагаются идеальными, так что ускорение их само по себе не оказывает никакого влияния на физические процессы в часах. Такие часы идут одинаково быстро с часами в сопутствующих инерциальных системах отсчета. Промежутки времени по неподвижным и движущимся часам связаны соотношением

$$dt = dt_0/\sqrt{1-(v/c)^2}, \tag{109.1}$$

где v — скорость часов A относительно системы S. В инерциальной системе отсчета S нет никакого гравитационного поля. Единственной причиной замедления времени t_0 по сравнению с t может быть только движение часов A относительно инерциальной системы отсчета S.

2. Однако возможна и другая точка зрения. Допустим сначала, что часы A относительно системы S движутся с постоянным ускорением $\mathbf{a}$. Будем отсчитывать время t от того момента, когда скорость v была равна нулю. Тогда $v = \sqrt{2ax}$, где x — расстояние, которое прошли часы A за время t. Следовательно,

$$dt = dt_0/\sqrt{1-2ax/c^2}. \tag{109.2}$$

Теперь введем ускоренную систему отсчета S_0, движущуюся вместе с часами A. В этой системе часы A неподвижны, зато есть силы инерции. Если все явления описывать, приняв S_0 за систему отсчета, то причиной замедления времени t_0 следует считать

силы инерции. Сила инерции, отнесенная к единице массы движущегося тела, равна $-\mathbf{a}$. Но, согласно принципу эквивалентности, силы инерции по своим физическим действиям неотличимы от гравитационного поля, напряженность которого в рассматриваемом нами случае равна $\mathbf{g} = -\mathbf{a}$. Введем еще *гравитационный потенциал* $\varphi = \pm gx$. Тогда предыдущая формула примет вид

$$dt = dt_0/\sqrt{1 - 2\varphi/c^2} \approx dt_0\left(1 - \frac{\varphi}{c^2}\right), \tag{109.3}$$

или

$$\frac{dt - dt_0}{dt_0} = -\frac{\varphi}{c^2}. \tag{109.4}$$

За нуль гравитационного потенциала принят потенциал такой точки, в которой движущиеся и неподвижные часы идут одинаково быстро. Поэтому в формулах (109.3) и (109.4) интервал времени dt можно отсчитывать не по часам инерциальной системы S, а по тем часам, покоящимся в системе S_0, которые находятся в точке B с нулевым потенциалом. Можно вообще начало отсчета гравитационного потенциала поместить в любой точке, если формулу (109.4) записать в виде

$$\frac{dt_{0A} - dt_{0B}}{dt_{0A}} = \frac{\varphi_B - \varphi_A}{c^2}, \tag{109.5}$$

где интервалы времени dt_{0A} и dt_{0B} отсчитываются по двум часам, покоящимся в ускоренной системе отсчета S_0 и помещенным в точках A и B с гравитационными потенциалами φ_A и φ_B. При этом на потенциалы должно быть наложено ограничение «слабого поля»:

$$\left|\frac{\varphi_B - \varphi_A}{c^2}\right| \ll 1, \tag{109.6}$$

так как только при этом условии приближенно справедлива ньютоновская теория тяготения и имеет смысл понятие гравитационного потенциала. (На поверхности Земли $\varphi/c^2 \approx -7 \cdot 10^{-10}$, на поверхности Солнца $\varphi/c^2 \approx -2{,}12 \cdot 10^{-6}$, если за нуль принять потенциал в бесконечности.) Поэтому в формуле (109.5) безразлично, писать ли в знаменателе dt_{0A} или dt_{0B}.

Формула (109.5) имеет следующий смысл. Пусть в точках A и B гравитационного поля помещены тождественные часы, покоящиеся в системе отсчета S_0. Для сравнения темпа хода этих часов будем после каждого удара часов A посылать световые сигналы к часам B. Пусть N_A — число посланных сигналов. За время прихода сигналов часы в B совершат N_B ударов, причем

$$\frac{N_A - N_B}{N_A} = \frac{\varphi_B - \varphi_A}{c^2}. \tag{109.7}$$

Если $\varphi_A < \varphi_B$ (световой сигнал распространяется против направления гравитационного поля), то $N_A > N_B$, т.е. часы B идут

медленнее часов A. В противоположном случае, когда $\varphi_B < \varphi_A$, медленнее будут идти часы A. Вообще, *часы в гравитационном поле идут тем медленнее, чем выше гравитационный потенциал в том месте, где они находятся*. Этот эффект, как и замедление хода движущихся часов, следует рассматривать как *свойство пространства-времени*, а не как влияние гравитационного поля на ход физических процессов в часах. Такое влияние, как уже указывалось в § 106, в случае реальных часов несомненно существует, но оно существенно зависит от того, какие физические явления положены в основу устройства часов. Для идеальных часов такого влияния нет — в этом случае эффект гравитационного замедления хода часов проявляется в чистом виде. Отмеченное обстоятельство проявляется и в формуле (109.5): гравитационное замедление времени есть *взаимное свойство двух часов*, определяющееся разностью гравитационных потенциалов между ними, тогда как для физических процессов в часах, очевидно, существен не потенциал, а напряженность гравитационного поля.

Конечно, формула (109.5) верна независимо от того, является ли гравитационное поле «фиктивным» полем сил инерции или «истинным» гравитационным полем, создаваемым массами вещества. Но эта формула была получена для однородного гравитационного поля. Применяя обычный прием, распространим ее на случай произвольного *неоднородного гравитационного поля*. Для этого вообразим, что в системе отсчета S_0 вдоль произвольной кривой AB, соединяющей часы A и B, расставлены достаточно часто тождественные часы 1, 2, 3, ..., n. Гравитационное поле между каждыми соседними часами можно считать однородным и применить к ним формулу (109.5). Перемножив почленно все эти формулы и учтя, что в пределах их точности различием знаменателей можно пренебречь, мы снова придем к результату (109.5) и тем самым докажем его справедливость и для неоднородного поля.

Так как при наличии гравитационного поля темп хода даже тождественных часов в разных точках пространства разный, то теряет смысл тот способ синхронизации часов и установления одновременности пространственно разделенных событий, который применялся в инерциальных системах отсчета, т.е. при отсутствии гравитационного поля. При наличии гравитационного поля сохраняет смысл лишь *местное время*, устанавливаемое по тождественным часам в каждой точке пространства. Следует говорить не о синхронизации пространственно разделенных часов, а только о сравнении *темпа их хода*, как это описано выше.

В свете изложенного парадокс близнецов можно рассматривать с иной точки зрения. Часы брата A все время находятся в

инерциальной системе отсчета, а часы брата B испытывают вместе с ним ускорение, что эквивалентно тому, что они неподвижны, но находятся в гравитационном поле. Это поле замедляет ход часов B. При возвращении к A часы B будут показывать меньшее время, чем A.

В последнее время (1976 г.) гравитационное замедление времени было подтверждено группой американских физиков Мерилендского университета. Измерялась разность показаний атомных часов на самолете и в наземной лаборатории. Самолет курсировал на высоте 10 км с небольшой скоростью, чтобы уменьшить кинематический эффект замедления времени. Время полета было около 15 часов. Ожидаемый гравитационный эффект составлял примерно +50 нс, кинематический −7 нс. Летающие часы сравнивались с наземными до, после и во время полета. Таким образом, можно было следить за монотонным возрастанием разности показаний сравниваемых часов. Опыт подтвердил теорию с точностью до 1,6 %.

3. Применим полученные результаты к излучению света. Пусть два тождественных неподвижных атома находятся в точках A и B гравитационного поля. Каждый из них в том месте, где он находится, излучает свет с одной и той же частотой ω_0. Рассматривая излучающие атомы как часы и применяя к ним формулу (109.5) или (109.7), получим

$$\frac{\omega_0 - \omega}{\omega_0} = \frac{\varphi_B - \varphi_A}{c^2}, \tag{109.8}$$

где ω — частота света, приходящего от A в точку наблюдения B. Если свет распространяется от низшего потенциала φ_A к высшему φ_B (т.е. против гравитационного поля), то наблюдается изменение частоты *в красную сторону*. Для света, приходящего от поверхности Солнца к Земле, это изменение частоты составляет $(\omega_0 - \omega)/\omega_0 = 2{,}12 \cdot 10^{-6}$. Напротив, если свет распространяется от высшего гравитационного потенциала к низшему (т.е. в направлении гравитационного поля), то наблюдаемая частота света *увеличивается*. Это явление, называемое *гравитационным смещением спектральных линий*, было теоретически предсказано Эйнштейном, а затем подтверждено экспериментально. Несколько иначе описанное явление было рассмотрено в § 72 т. I.

§ 110. Сложение скоростей в теории относительности

1. Задача (физического) сложения скоростей формулируется следующим образом (см. т. I, § 7). Пусть скорость движущейся частицы в системе отсчета S' равна $\mathbf{v}'$, причем сама система S' равномерно движется относительно «неподвижной» системы

S со скоростью $\mathbf{V}$ в направлении оси X. Требуется определить скорость $\mathbf{v}$ той же частицы в системе отсчета S. Обозначим координаты частицы в системе S в момент времени t через x, y, z. Те же величины в системе S' обозначим штрихованными буквами t', x', y', z'. Так как обе эти группы переменных характеризуют *одно и то же событие* (прохождение частицы через одну и ту же пространственно-временную точку), то они связаны между собой формулами преобразования Лорентца (105.12). Поскольку $\mathbf{V} = \text{const}$, координаты и время при движении частицы получают приращения

$$dx = \frac{dx' + V\,dt}{\sqrt{1-\beta^2}}, \quad dy = dy', \quad dz = dz', \quad dt = \frac{dt' + Vx'/c^2}{\sqrt{1-\beta^2}}.$$

Разделив первые три соотношения на последнее и замечая, что $dx/dt = v_x$, $dx'/dt' = v'_x$, получим

$$v_x = \frac{v'_x + V}{1 + v'_x V/c^2}, \quad v_y = \frac{v'_y\sqrt{1-\beta^2}}{1 + v'_x V/c^2}, \quad v_z = \frac{v'_z\sqrt{1-\beta^2}}{1 + v'_x V/c^2}. \qquad (110.1)$$

Эти формулы и выражают *правило сложения скоростей в релятивистской кинематике.* При медленных движениях, когда можно пренебречь квадратичными величинами V^2/c^2 и $v'_x V/c^2$, они переходят в нерелятивистские формулы

$$v_x = v'_x + V, \quad v_y = v'_y, \quad v_z = v'_z,$$

получающиеся из преобразования Галилея.

2. Для исследования формул (110.1) удобно принять за единицу скорость света c, введя безразмерные величины $\beta_x = v_x/c$, $\beta'_x = v'_x/c$, ... Тогда

$$\beta_x = \frac{\beta'_x + \beta}{1 + \beta\beta'_x}, \quad \beta_y = \frac{\beta'_y\sqrt{1-\beta^2}}{1 + \beta\beta'_x}, \quad \beta_z = \frac{\beta'_z\sqrt{1-\beta^2}}{1 + \beta\beta'_x}. \qquad (110.2)$$

Все скорости по величине не могут превосходить c, а соответствующие им безразмерные величины β, β'_x, ... — единицы. Возникает, однако, вопрос, не может ли в результате (физического) сложения двух скоростей получиться скорость, превосходящая c? Ответ на этот вопрос должен быть отрицательным.

Для доказательства рассмотрим сначала случай сложения параллельных скоростей, когда направления векторов $\mathbf{v}'$ и $\mathbf{V}$ совпадают или прямо противоположны. Результирующая скорость в этом случае полностью определяется первой формулой (110.1) или эквивалентной ей первой формулой (110.2). Так как β и β'_x не превосходят единицы, то их можно представить в виде $\beta = \operatorname{th}\vartheta$, $\beta'_x = \operatorname{th}\vartheta'_x$, где аргументы гиперболических тангенсов ϑ и ϑ'_x могут принимать любые вещественные значения. Тогда

первая формула (110.2) запишется в виде

$$\beta_x = \frac{\operatorname{th}\vartheta'_x + \operatorname{th}\vartheta}{1 + \operatorname{th}\vartheta \operatorname{th}\vartheta'_x}.$$

Сравнивая ее с формулой сложения гиперболических тангенсов

$$\operatorname{th}(\vartheta'_x + \vartheta) = \frac{\operatorname{th}\vartheta'_x + \operatorname{th}\vartheta}{1 + \operatorname{th}\vartheta \operatorname{th}\vartheta'_x},$$

видим, что результирующую безразмерную скорость также можно представить гиперболическим тангенсом $\beta_x = \operatorname{th}\vartheta_x$, причем

$$\vartheta_x = \vartheta'_x + \vartheta. \tag{110.3}$$

Это значит, что (физическое) сложение параллельных скоростей сводится к алгебраическому сложению аргументов соответствующих им гиперболических тангенсов. Но при любых значениях аргументов гиперболический тангенс по абсолютной величине не может превосходить единицы. Следовательно, каковы бы ни были (параллельные) скорости составляющих движений, при их (физическом) сложении не может получиться скорость, превосходящая c. Если одна из складываемых скоростей равна c, то при сложении ее с любой другой скоростью получится также c. В этом и проявляется специфичность скорости света как предельной скорости, которую нельзя превзойти, сколько бы ни ускорять движение тела.

Для завершения доказательства остается исследовать случай, когда складываемые скорости взаимно перпендикулярны. Без ограничения общности можно считать, что скорость $\mathbf{v}'$ направлена вдоль оси Y'. Тогда, полагая в формуле (110.2) $\beta'_x = \beta'_z = 0$, получим

$$\beta_x = \beta, \quad \beta_y = \beta'_y\sqrt{1-\beta^2}, \quad \beta_z = 0.$$

Квадрат (безразмерной) скорости в системе S будет

$$\left(\frac{v}{c}\right)^2 = \beta_x^2 + \beta_y^2 = \beta^2 + \beta_y'^2(1-\beta^2).$$

Придав β постоянное, но произвольное значение, будем рассматривать величину $(v/c)^2$ как функцию аргумента $\beta_y'^2$. Производная этой величины $(1-\beta^2) \geqslant 0$, а потому функция $(v/c)^2$ возрастает с возрастанием $\beta_y'^2$. Она достигает максимума при $\beta_y'^2 = 1$, который равен 1. При всех остальных значениях β'_y отношение $(v/c)^2$ не может превосходить единицы. Это и доказывает, что при сложении перпендикулярных скоростей также не может получиться скорость, превосходящая c.

3. Когда складываемые скорости параллельны, то индекс x можно опустить и написать

$$v = \frac{v' + V}{1 + v'V/c^2}. \tag{110.4}$$

Если $V \ll c$, то в этой формуле можно ограничиться первыми степенями по V. В этом приближении

$$v \approx (v' + V)\left(1 - \frac{v'V}{c^2}\right) \approx v' + \left(1 - \frac{v'^2}{c^2}\right) V. \tag{110.5}$$

Применим последнюю формулу к распространению света в жидкости, равномерно движущейся со скоростью V. Скорость света относительно неподвижной жидкости будет $v' = c/n$, где n — показатель преломления жидкости. Предполагая, что свет распространяется в направлении течения жидкости, найдем его скорость v относительно «неподвижной» системы отсчета. Для этого подставляем в формулу (110.5) $v'/c = 1/n$ и находим

$$v = \frac{c}{n} + \left(1 - \frac{1}{n^2}\right) V. \tag{110.6}$$

Формула (110.6) была получена еще Френелем в 1818 г. Он исходил из представления, что эфир увлекается движущимися телами, однако не полностью, а частично. Лоренц в электронной теории вновь пришел к формуле (110.6). Но истинный смысл ее был установлен только в теории относительности.

Формула (110.6) была экспериментально подтверждена Физо в 1851 г. Схема его опыта в усовершенствованном виде, какой ему придал Майкельсон в 1886 г., показана на рис. 336. Луч

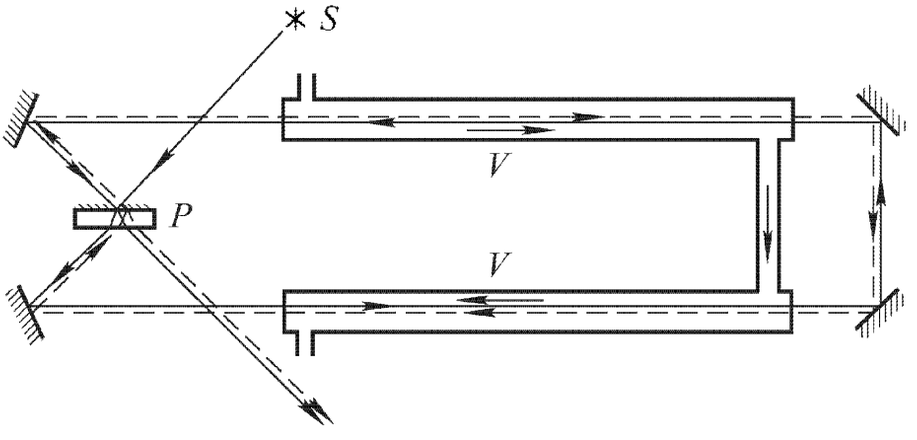


Рис. 336

от источника S раздваивается разделительной пластинкой P. Один луч на рис. 336 изображен сплошной, а другой — штриховой линией. Затем лучи проходят через трубки, в которых течет вода. Один луч идет в направлении, а другой против течения воды. Из-за различия скоростей лучей относительно неподвижных стенок трубки между ними при выходе из прибора

возникает разность хода, изменяющаяся с изменением скорости течения V. Сначала наблюдается интерференция между лучами при неподвижной, а затем при текущей воде. По смещению интерференционных полос можно измерить разность хода, возникающую при течении, а по ней и разность скоростей $v - c/n$.

4. Лоренц в электронной теории обобщил формулу (110.6), учтя дисперсию света. Формулу Лорентца легко получить и в теории относительности. Надо только учесть доплеровское изменение длины волны, возникающее при течении воды. Обозначим через λ длину световой волны (в вакууме) в «неподвижной» системе отсчета S, а через $\lambda + \delta\lambda$ — в системе S', относительно которой вода неподвижна. Доплеровское изменение длины волны $\delta\lambda$ в первом порядке по V определяется выражением

$$\frac{\delta\lambda}{\lambda} = \frac{V}{v'} = \frac{Vn}{c}.$$

Следовательно,

$$v'(\lambda + \delta\lambda) = v'(\lambda) + \frac{dv'}{d\lambda}\frac{Vn}{c}\lambda,$$

или

$$v'(\lambda + \delta\lambda) = v'(\lambda) + \frac{d}{d\lambda}\left(\frac{c}{n}\right)\frac{Vn\lambda}{c} = \frac{c}{n(\lambda)} - \frac{\lambda}{n}\frac{dn}{d\lambda}V.$$

В формуле (110.5) под v' надо понимать, конечно, не $v'(\lambda)$, а $v'(\lambda+\delta\lambda)$. При этом второе слагаемое в правой части (110.5) в рассматриваемом приближении следует оставить без изменения. В результате получим

$$v = \frac{c}{n} + \left(1 - \frac{1}{n^2} - \frac{\lambda}{n}\frac{dn}{d\lambda}\right)V. \tag{110.7}$$

Наличие добавочного члена, обусловленного дисперсией, было экспериментально подтверждено Зееманом в 1914 г.

Нелишне подчеркнуть, что доплеровское изменение длины волны относится к системе отсчета S', в которой вода неподвижна. Течение воды не сказывается на длине волны в «неподвижной» системе S, так что оба интерферирующих луча при выходе из прибора (см. рис. 336) имеют *одну и ту же длину волны*. Только благодаря этому и возможна интерференция.

§ 111. Релятивистская механика

1. Если какой-либо закон природы представлен в виде $A = B$, причем при переходе от одной системы отсчета к другой величины A и B остаются *неизменными*, то эти величины и самый закон называются *инвариантными* относительно этого перехода. Более общим является понятие *ковариантности*. Если при переходе от одной системы отсчета к другой величины A и B хотя и не остаются неизменными, но *преобразуются одинаково*, то закон $A = B$ сохраняется и в новой системе отсчета. В этом случае говорят, что закон $A = B$ *ковариантен* относительно рассматриваемого преобразования систем отсчета. Часто термин «инвариантность закона» употребляют в смысле его ковариантности.

До теории относительности допустимыми считались только *галилеевы преобразования координат*. Относительно этих преобразований уравнения механики Ньютона были ковариантны (инвариантны), тогда как уравнения электродинамики Максвелла–Лорентца — не ковариантны. Теория относительности показала, что от галилеева преобразования надо отказаться и заменить его *преобразованием Лорентца*. Тогда принцип относительности требует, чтобы *законы природы были ковариантны относительно преобразования Лорентца*. Этому требованию уравнения электродинамики удовлетворяют, а уравнения механики Ньютона не удовлетворяют. Поэтому механика Ньютона должна быть изменена.

В ньютоновской механике сила, действующая на тело в какой-то момент времени, определяется положением всех взаимодействующих тел *в тот же момент*. Но в теории относительности понятие «тот же момент времени» зависит от выбора системы отсчета. Невозможно автоматически преобразовать каждый закон сил ньютоновской механики в лоренц-ковариантную форму. Допустимы только такие теории, из которых *может быть исключено понятие действия на расстоянии*. Такая возможность существует в *теории столкновений*. Последняя исходит из идеализированного представления, что взаимодействие имеет место только в продолжение того промежутка времени, когда расстояние между телами или точечными частицами бесконечно мало по сравнению с размерами самих тел или другими характерными расстояниями, определяющими характер процессов столкновения. До и после этого бесконечно малого промежутка времени тела движутся свободно. К процессам столкновений применимы законы сохранения импульса и энергии, но им надо придать лоренц-ковариантную форму. Это и является целью настоящего параграфа. Дальнодействие можно исключить также при рассмотрении движения электрически заряженных частиц в электромагнитных полях. Однако изложение относящихся сюда вопросов электродинамики потребовало бы слишком много места, а потому мы ограничимся только рассмотрением процессов столкновения.

2. Для решения поставленной задачи проще и логичнее всего воспользоваться понятием *четырехмерного вектора* (или, короче, *4-вектора*) в пространстве Минковского. Каждое «точечное» событие в таком пространстве характеризуется совокупностью четырех координат x, y, z, $\tau \equiv ct$. При переходе от системы отсчета S к системе отсчета S' разности координат двух точек преобразуются по формулам

$$\Delta x' = \frac{\Delta x - \beta\,\Delta\tau}{\sqrt{1-\beta^2}}, \quad \Delta y' = \Delta y, \quad \Delta z' = \Delta z, \quad \Delta\tau' = \frac{\Delta\tau - \beta\,\Delta x}{\sqrt{1-\beta^2}}, \tag{111.1}$$

как это следует из (105.12). Напомним, что квадрат интервала между рассматриваемыми точками есть инвариант:

$$\Delta\tau^2 - \Delta x^2 - \Delta y^2 - \Delta z^2 = \mathrm{Inv}\ . \qquad (111.2)$$

Мы воспользовались частным преобразованием Лоренца (105.12), в котором предполагается, что координатные оси X, Y, Z параллельны осям X', Y', Z', а система S' движется относительно S вдоль оси X. Можно было бы взять любую ориентацию осей и любое направление движения, но это только усложнило бы запись, ничего не меняя по существу.

Назовем четырехмерным вектором совокупность четырех величин A_x, A_y, A_z, A_τ, которые при переходе от одной системы отсчета к другой преобразуются так же, как разности координат двух точек в пространстве Минковского, т.е.

$$A'_{x'} = \frac{A_x - \beta A_\tau}{\sqrt{1-\beta^2}}, \quad A'_{y'} = A_y, \quad A'_{z'} = A_z, \quad A'_{\tau'} = \frac{A_\tau - \beta A_x}{\sqrt{1-\beta^2}}. \qquad (111.3)$$

Величины A_x, A_y, A_z называются *пространственными*, а A_τ — *временной* составляющей четырехмерного вектора. Пространственные составляющие мы объединим в обычный трехмерный вектор $\mathbf{A}$ и будем обозначать четырехмерный вектор через $(\mathbf{A}, A_\tau)$. Из тождественности законов преобразования (111.1) и (111.3) следует, что четырехмерный вектор $(\mathbf{A}, A_\tau)$ обладает *инвариантом*:

$$A_\tau^2 - \mathbf{A}^2 = \mathrm{Inv}\ . \qquad (111.4)$$

Если какой-либо закон природы записан в четырехмерной векторной форме $(\mathbf{A}, A_\tau) = (\mathbf{B}, B_\tau)$, то он лоренц-ковариантен, так как обе части написанного равенства при преобразовании Лоренца преобразуются одинаково. Четырехмерная векторная форма эквивалентна двум уравнениям: $\mathbf{A} = \mathbf{B}$ и $A_\tau = B_\tau$. Этим замечанием мы и воспользуемся для решения поставленной задачи. Именно, мы постулируем, что *закон сохранения импульса и энергии можно записать в виде равенства четырехмерных векторов.* Задача состоит в том, чтобы найти вид этих векторов.

3. Рассмотрим частицу, движущуюся со скоростью $\mathbf{v}$ относительно неподвижной системы отсчета S. Пусть $d\mathbf{r}$ — ее перемещение за время $dt = d\tau/c$. Эти величины образуют четырехмерный вектор $(d\mathbf{r}, c\,dt)$. Очевидно, он останется четырехмерным вектором и после умножения его составляющих на одну и ту же постоянную. Возьмем в качестве таковой m_0/dt_0, где m_0 — некоторая постоянная, а $dt_0 = dt\sqrt{1 - v^2/c^2}$ — собственное время, которое, как известно, является инвариантом. Тогда

получим четырехмерный вектор

$$m_0\left(\frac{d\mathbf{r}}{dt_0}, c\,\frac{dt}{dt_0}\right) = m(\mathbf{v}, c), \tag{111.5}$$

где введено обозначение

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}}. \tag{111.6}$$

Допустим теперь, что частица движется *медленно*, так что величиной v^2/c^2 можно пренебречь. Возьмем в качестве m_0 массу частицы, как она определяется в нерелятивистской механике. Тогда пространственная составляющая четырехмерного вектора (111.5) будет $m_0\mathbf{v}$. Вектор $m_0\mathbf{v}$ в ньютоновской механике называется *импульсом*. Поэтому в релятивистской механике естественно определить импульс выражением

$$\mathbf{P} = m\mathbf{v} = \frac{m_0\mathbf{v}}{\sqrt{1 - v^2/c^2}}, \tag{111.7}$$

поскольку в пределе при малых скоростях оно переходит в нерелятивистское выражение $m_0\mathbf{v}$. Величина m_0 называется *массой покоя*, а m — *массой движения* или *релятивистской массой*. Таким образом, $(\mathbf{P}, mc)$ есть четырехмерный вектор, а величина $(mc)^2 - P^2$ — его инвариант. Значение этого инварианта легко найти: при $v = 0$ он обращается в $(m_0c)^2$, а потому

$$(mc)^2 - P^2 = (m_0c)^2. \tag{111.8}$$

Остается выяснить физический смысл временной части четырехмерного вектора (111.5). Для этого замечаем, что $d\mathbf{P}/dt$ есть сила, действующая на частицу. Работа этой силы на перемещении $\mathbf{v}\,dt$ равна $dA = \mathbf{v}\,d\mathbf{P} = \mathbf{P}\,d\mathbf{P}/m$ или на основании (111.8) $dA = c^2\,dm$. Энергия частицы найдется интегрированием этого выражения по m. Если постоянную интегрирования положить равной нулю, то получится формула

$$\mathcal{E} = mc^2 = \frac{m_0c^2}{\sqrt{1 - v^2/c^2}}. \tag{111.9}$$

Величина $\mathcal{E}$ называется *полной энергией частицы*. Для покоящейся частицы $m = m_0$, так что (111.9) переходит в

$$\mathcal{E}_0 = m_0c^2. \tag{111.10}$$

Величина $\mathcal{E}_0$ называется *энергией покоя частицы*.

Формула (111.9), впервые в общем виде полученная Эйнштейном, устанавливает *взаимосвязь между массой и энергией*. Кинетическая энергия частицы определяется выражением

$$K = \mathcal{E} - \mathcal{E}_0 = (m - m_0)c^2. \tag{111.11}$$

При медленных движениях это выражение переходит в обычную формулу $K = {}^1\!/_2 m_0 v^2$.

Импульс и энергия теперь объединены в четырехмерный вектор

$$(\mathbf{P}, mc) = (\mathbf{P}, \mathcal{E}/c), \tag{111.12}$$

называемый *вектором импульса–энергии.* Его инвариантом относительно преобразования Лорентца является

$$\left(\frac{\mathcal{E}}{c}\right)^2 - P^2 = (m_0 c)^2 = \mathrm{Inv}\ . \tag{111.13}$$

Тем самым в теории относительности законы сохранения импульса и энергии перестают быть независимыми законами, а объединяются в *единый закон сохранения четырехмерного вектора импульса–энергии.* Его называют также *законом сохранения импульса–энергии.*

Остается ответить на два вопроса. Во-первых, почему при вычислении работы сила была определена так же, как в нерелятивистской механике, т.е. как производная $d\mathbf{P}/dt$? Во-вторых, почему энергия всегда определяется с точностью до несущественной произвольной постоянной, здесь же она определена *однозначно*? Ответ на оба вопроса, в сущности, один и тот же. Он состоит в том, что на величину $\mathcal{E}$, вычисленную выше и названную полной энергией, было наложено требование, чтобы она (после деления на c) была *временной компонентой четырехмерного вектора* (111.5). Если энергию не определить однозначно, то она этому требованию удовлетворять не будет.

Для системы невзаимодействующих частиц, а также частиц, взаимодействующих только при столкновениях, четырехмерный вектор импульса–энергии определяется как сумма четырехмерных векторов импульса–энергии этих частиц. При этом в теории относительности достигается однообразная трактовка упругих и неупругих столкновений. Независимо от характера столкновения сохраняется трехмерный вектор импульса системы. Следовательно, должна сохраняться и энергия, как (умноженная на c) временная компонента четырехмерного вектора. Вместе с энергией сохраняется и релятивистская масса. Только при упругих и неупругих столкновениях она по-разному распределяется между массой покоя и массой, связанной с кинетической энергией макроскопического движения. Например, при столкновении двух одинаковых неупругих шаров, движущихся с одинаковыми скоростями навстречу друг другу, исчезновение кинетической энергии макроскопического движения (т.е. переход его во внутреннее молекулярное движение) проявляется в эквивалентном увеличении массы покоя системы: масса нагретого шара больше, чем масса такого же холодного шара. При упругих же столкновениях остаются неизменными и масса покоя, в масса, связанная с кинетической энергией макроскопического движения.

Существуют частицы (*фотоны, нейтрино*), для которых масса покоя равна нулю. Для них связь (111.13) между энергией и импульсом имеет вид

$$P = \mathcal{E}/c. \tag{111.14}$$

Такие частицы всегда движутся со скоростью c. Иначе, как видно из формул (111.7) и (111.9), импульс и энергия таких частиц обращались бы в нуль.

ЗАДАЧИ

1. Две одинаковые частицы движутся в лабораторной системе навстречу друг другу с одной и той же скоростью v. Найти относительную скорость V каждой из них относительно другой. Какой энергией $\mathcal{E}'$ в лабораторной системе отсчета должна обладать одна из частиц, чтобы получить ту же относительную скорость, если вторая частица (мишень) неподвижна? (Принцип действия *ускорителя на встречных пучках.*)

Р е ш е н и е. По теореме сложения скоростей

$$V = \frac{2v}{1 + v^2/c^2}.$$

Искомая полная энергия, которую надо было бы сообщить одной частице, равна

$$\mathcal{E}' = \frac{\mathcal{E}_0}{\sqrt{1 - V^2/c^2}},$$

где $\mathcal{E}_0$ — энергия покоя частицы. Фактическая энергия, которой обладает частица, $\mathcal{E} = \mathcal{E}_0/\sqrt{1 - v^2/c^2}$. Отсюда нетрудно получить

$$\mathcal{E}' = 2\frac{\mathcal{E}^2}{\mathcal{E}_0} - \mathcal{E}_0, \tag{111.15}$$

а для кинетической энергии

$$K' = 2\left(\frac{\mathcal{E}^2}{\mathcal{E}_0} - \mathcal{E}_0\right). \tag{111.16}$$

(Другое решение см. в т. I, § 28.)

2. Вывести формулу, являющуюся релятивистским обобщением формулы Циолковского (см. т. I, § 21) для движения ракеты. Считать, что скорости ракеты и газовой струи направлены вдоль одной прямой.

Р е ш е н и е. На основании законов сохранения импульса и энергии

$$mv + m_{\text{газ}}v_{\text{газ}} = \text{const}, \quad m + m_{\text{газ}} = \text{const},$$

где m и $m_{\text{газ}}$ — релятивистские массы ракеты и газов, а v и $v_{\text{газ}}$ — их скорости в произвольный момент времени. Газы, уже покинувшие ракету, не влияют на ее движение. Поэтому можно считать, что в рассматриваемый момент времени $m_{\text{газ}} = 0$. Тогда не возникает неопределенности, что следует понимать под $v_{\text{газ}}$. Однако, поскольку газы непрерывно образуются, $dm_{\text{газ}} \neq 0$. Дифференцируя предыдущие уравнения, получим

$$m\,dv + (v - v_{\text{газ}})\,dm = 0.$$

По релятивистскому закону сложения скоростей

$$v_{\text{газ}} = \frac{v-u}{1-vu/c^2},$$

где u — скорость газовой струи относительно ракеты. Исключение $v_{\text{газ}}$ приводит к уравнению

$$dv + u\frac{1-v^2/c^2}{1-vu/c^2}\frac{dm}{m} = 0.$$

Воспользовавшись формулой (111.6), после несложных преобразований найдем

$$\frac{dv}{v^2-c^2} = \frac{u}{c^2}\frac{dm_0}{m_0}.$$

Предполагая скорость u газовой струи постоянной и интегрируя, получим искомый результат:

$$\frac{(m_0)_{\text{нач}}}{m_0} = \left(\frac{1+v/c}{1-v/c}\right)^{c/2u}. \qquad (111.17)$$

ГЛАВА X

ТЕПЛОВОЕ ИЗЛУЧЕНИЕ

§ 112. Равновесное излучение в полости

1. Излучение света происходит в результате переходов атомов, молекул и других атомных систем из состояний с большей в состояния с меньшей энергией. Так называемое *тепловое* или *температурное излучение* отличается от других видов излучения (люминесценции) только способом перехода излучающих систем в возбужденные состояния. В явлениях теплового излучения такой переход осуществляется в результате теплового движения атомов и молекул.

В проблемах теплового излучения особо важное значение имеет понятие так называемого *равновесного излучения*. Для установления этого понятия рассмотрим полость с неподвижными и непрозрачными стенками, температура которых поддерживается постоянной. Атомы и молекулы стенок переходят в возбужденные состояния за счет энергии теплового движения и при обратных переходах в невозбужденные состояния дают излучение, заполняющее полость. Падая на стенки полости, лучистая энергия частично отражается, частично поглощается. Происходит изменение направления распространения, спектрального состава, поляризации, интенсивности излучения. В результате всех этих процессов, как это следует из общего начала термодинамики, в полости в конце концов устанавливается макроскопически вполне определенное состояние излучения, при котором за каждый промежуток времени количество излученной лучистой энергии определенного цвета, направления распространения и поляризации в среднем равно количеству поглощенной энергии того же цвета, направления распространения и поляризации. Как и всякое равновесное состояние, оно характеризуется тем, что каждому микропроцессу, происходящему в системе, с той же вероятностью соответствует микропроцесс, идущий в обратном направлении (*принцип детального равновесия*). Благодаря это-

му состояние излучения в полости и остается *макроскопически неизменным* во времени. Переход в равновесное состояние, как и всякий статистический процесс, управляется *вероятностными законами*. В полости устанавливается хаотическое состояние излучения, которому соответствует *наибольшая вероятность*. Оно и называется равновесным излучением.

Свойства равновесного излучения: плотность лучистой энергии, ее распределение по спектру частот и направлениям распространения, а также поляризация излучения *совершенно не зависят от формы и материала стенок полости*. Эти свойства, подобно состоянию газа в сосуде, определяются *только температурой стенок полости*. Равновесное излучение *однородно*, т.е. его плотность одна и та же во всех точках внутри полости. Оно *изотропно и неполяризовано*: все возможные направления распространения излучения представлены с одинаковой вероятностью, а направления векторов $\mathbf{E}$ и $\mathbf{B}$ и каждой точке пространства хаотически меняются во времени. Поскольку излучение находится в тепловом равновесии со стенками, можно говорить о температуре не только стенок, но и о *температуре самого излучения*, считая по определению обе температуры равными. Надо, однако, подчеркнуть, что температура равновесного излучения есть *свойство самого излучения*, а не стенки, с которой оно находится в тепловом равновесии. О ней имеет смысл говорить и тогда, когда вообще нет никакой стенки. В частности, например, плотность энергии равновесного излучения однозначно определяет и его температуру (см. § 115).

Если стенки полости совершенно непоглощающие, например идеально зеркальные, то в такой полости не будет поглощения и испускания света. В полость можно ввести излучение произвольного спектрального состава. Отражаясь от стенок, излучение меняет направление распространения, но его спектральный состав сохраняется неизменным. При надлежащей геометрической форме полости с зеркальными стенками возможны и такие случаи, когда сохраняются также направление распространения и поляризация излучения. Так будет, например, когда полость имеет форму прямого цилиндра с абсолютно зеркальными основаниями. Тогда луч света произвольной частоты и поляризации может распространяться туда и обратно параллельно оси цилиндра, последовательно отражаясь от зеркальных оснований. Но все подобные случаи являются идеальными и никогда точно не реализуются в действительности. Излучение в полости в этих случаях *неравновесно* и *неустойчиво*. Уже при сколь угодно малых отклонениях от идеальности, если только подождать достаточно долго, в полости обязательно установится равновесное излучение. Идеальные системы, однако, имеют большое значение в теоретических рассуждениях. Можно брать стенки абсолют-

но зеркальными и в то же время считать, что в полости всегда устанавливается равновесное излучение. Для этого достаточно, например, ввести в полость сколь угодно малое поглощающее и излучающее тельце — *пылинку*, по выражению Планка. Такая пылинка, практически не играя никакой роли в энергетическом балансе системы, переводит, однако, любое неравновесное состояние, возникшее в полости, в равновесное.

2. Введем теперь некоторые величины, характеризующие состояние излучения в пространстве. Эти величины имеют смысл для произвольного излучения, а не только для равновесного.

Обозначим через u плотность энергии излучения, т.е. количество такой энергии в единице объема пространства. Ее можно разложить по частотам или длинам волн, т.е. представить в виде

$$u = \int_0^\infty u_\omega \, d\omega = \int_0^\infty u_\lambda \, d\lambda. \tag{112.1}$$

Величины $u_\omega \, d\omega$ и $u_\lambda \, d\lambda$ имеют смысл объемной плотности лучистой энергии, приходящейся на интервал частот ω, $\omega + d\omega$ или интервал длин волн λ, $\lambda + d\lambda$. Коэффициенты u_ω и u_λ называются *спектральными плотностями лучистой энергии*. Если речь идет об одном и том же спектральном интервале, но представленном в различных формах, то $u_\omega \, d\omega = u_\lambda \, d\lambda$. При этом $\lambda = 2\pi c/\omega$, и следовательно, $d\lambda/\lambda = -d\omega/\omega$. Знак минус мы опустим — он означает только, что с возрастанием частоты длина волны убывает. Таким образом, считая величины $d\omega$ и $d\lambda$ существенно положительными, можно написать

$$u_\lambda = \frac{\omega}{\lambda} u_\omega, \quad u_\omega = \frac{\lambda}{\omega} u_\lambda. \tag{112.2}$$

В теоретической физике обычно пользуются величиной u_ω, в экспериментальной отдают предпочтение u_λ. В случае равновесного излучения функции u_ω и u_λ зависят только от частоты ω (или длины волны λ) и от температуры излучения T, но не зависят от формы и материала стенок полости. Они зависят от среды, занимающей полость. Эта зависимость будет установлена в § 114. Пока же будем предполагать, что в полости вакуум. В этом случае u_ω будет *универсальной функцией только* ω и T, а u — *универсальной функцией только* T. Нахождение функции $u_\omega(\omega, T)$ является *основной задачей* теории теплового излучения.

Поток лучистой энергии, проходящей за время dt через малую площадку ds в пределах телесного угла $d\Omega$, ось которого перпендикулярна к площадке ds, можно представить выражением

$$d\Phi = I \, ds \, d\Omega \, dt. \tag{112.3}$$

Величина I называется *удельной интенсивностью излучения*. Если ее разложить в спектр, т.е. представить интегралом

$$I = \int_0^\infty I_\omega \, d\omega, \tag{112.4}$$

то величину I_ω называют *удельной интенсивностью излучения частоты* ω.

Найдем связь между u и I (а также между u_ω и I_ω) для поля излучения в вакууме. Возьмем в пространстве бесконечно малый прямоугольный параллелепипед с площадью основания ds и высотой dl (рис. 337). Выделим пучок лучей, вступающих через площадку ds внутрь параллелепипеда, направления которых лежат в пределах телесного угла $d\Omega$, а ось телесного угла нормальна к основанию ds. Каждый из этих лучей доходит до второго основания параллелепипеда за время $dt = dl/c$. За это время из выделенного пучка лучей через площадку ds внутрь параллелепипеда вступает лучистая энергия $I\,ds\,d\Omega\,dt = (I/c)\,d\Omega\,dV$, где $dV = dl\,ds$ — объем параллелепипеда. Разделив на dV, найдем $(I/c)\,d\Omega$. Эта величина есть плотность du лучистой энергии, распространяющейся в пределах телесного угла $d\Omega$. По своему смыслу она может быть функцией точки в пространстве, но не может зависеть от формы параллелепипеда. Поэтому для нахождения полной плотности u лучистой энергии в рассматриваемой точке пространства надо проинтегрировать выражение du по всем направлениям в пространстве. В общем случае при таком интегрировании надо принять во внимание, что удельная интенсивность I зависит от направления излучения. Но в случае изотропного излучения, каким, в частности, является равновесное излучение, такой зависимости нет. В этом случае интегрирование сводится к простой замене элементарного телесного угла $d\Omega$ на полный телесный угол $\Omega = 4\pi$. Тогда получится

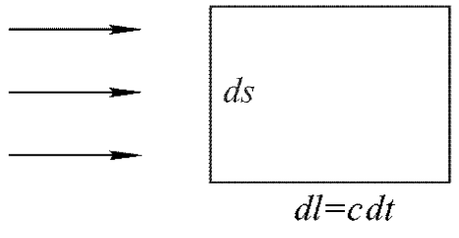


Рис. 337

$$u = \frac{4\pi}{c} I. \tag{112.5}$$

Аналогично,

$$u_\omega = \frac{4\pi}{c} I_\omega. \tag{112.6}$$

Такие же соотношения справедливы и в том случае, когда полость заполнена однородной изотропной непоглощающей средой. Надо только величину c заменить групповой скоростью света в рассматриваемой среде.

§ 113. Закон Кирхгофа

1. Перейдем теперь к рассмотрению *законов теплового излучения и поглощения тел.* Ограничимся частным случаем, когда излучающее и поглощающее тело *непрозрачно.* Этому условию можно удовлетворить, если толщина тела достаточна, чтобы всякий луч, вступивший в него, успел поглотиться, не достигнув противоположной границы тела. Поток лучистой энергии с частотами между ω и $\omega + d\omega$, излучаемый за время dt площадкой ds поверхности тела внутрь телесного угла $d\Omega$, можно представить выражением

$$d\Phi = E_\omega \, ds \cos\varphi \, d\Omega \, d\omega \, dt, \tag{113.1}$$

где φ — угол между направлением излучения и нормалью к площадке ds. Здесь введен множитель $\cos\varphi$, так как излучение удобнее относить не к величине полной площади ds, а к ее «видимой» части, т.е. проекции площадки ds на плоскость, перпендикулярную к направлению излучения (см. § 22). Величина E_ω называется *излучательной способностью* поверхности тела в направлении, определяемом углом φ [1]).

Поглощательной способностью A_ω поверхности для излучения той же частоты, направления распространения и поляризации называется безразмерная величина, показывающая, какая доля энергии падающего излучения поглощается рассматриваемой поверхностью.

Речь идет, конечно, только о *чисто температурном излучении*, когда все свойства излучающего и поглощающего тела определяются *только температурой тела.* Исключаются из рассмотрения все случаи «холодного» или неравновесного свечения, когда светящееся тело излучает не потому, что оно нагрето, а потому, что определенные уровни энергии его возбуждены и тело постепенно «высвечивается», переходя в нормальное, невозбужденное состояние.

Что касается зависимости излучательной способности тела от окружающей однородной среды, то к этому вопросу можно подходить с двух точек зрения. Можно определить E_ω через *полный поток* лучистой энергии выделенного направления и частоты, исходящий от поверхности тела в *окружающую среду.* При такой точке зрения излучательная способность характеризует свойства *двух сред*, граничащих между собой вдоль рас-

[1]) Здесь мы отступаем от традиции. По историческим причинам излучающую способность тела принято характеризовать *полным потоком* лучистой энергии, посылаемым излучающей площадкой ds наружу по *всем направлениям*, т.е. в пределах телесного угла 2π. При более детальном рассмотрении надо разложить ее по разным направлениям и частотам, что мы и делам.

сматриваемой поверхности. Но можно встать на другую точку зрения, развивавшуюся Прево (1751–1839) в начале прошлого века. Согласно этой точке зрения, результирующий поток лучистой энергии через поверхность тела есть *разность двух потоков*, которые оба излучаются в вакуум: потока, *излучаемого телом* в окружающее пространство, и потока, *излучаемого средой* в пространство, занятое телом. Первый поток определяет излучательную способность *тела*, второй — излучательную способность *среды*. Мы будем придерживаться второй точки зрения, так как тогда величины E_ω и A_ω будут характеристиками только *самого тела*, не зависящими от окружающей среды. При прочих равных условиях они определяются *только температурой тела*.

2. Установим связь между излучательной и поглощательной способностью тела. Так как эти величины характеризуют только поверхность самого тела и совершенно не зависят от окружающего излучения, то в рассуждениях относительно этого излучения можно вводить любые предположения. Предположим, что излучающее тело со всех сторон окружено равновесным излучением, температура которого равна температуре тела. Это, например, можно осуществить, если в качестве рассматриваемого участка поверхности тела взять часть внутренней поверхности замкнутой полости, температура стенок которой поддерживается постоянной. Выделим из всего излучения часть, заполняющую интервал частот ω, $\omega + d\omega$, и рассмотрим превращения ее при излучении и отражении от стенок полости. На площадку ds стенки (рис. 338) за время dt в пределах телесного угла $d\Omega$ падает лучистый поток

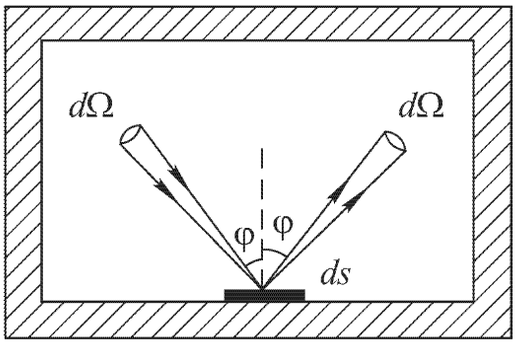


Рис. 338

$$I_\omega \, ds \cos\varphi \, d\omega \, d\Omega \, dt.$$

Часть его

$$(1 - A_\omega) I_\omega \, ds \cos\varphi \, d\omega \, d\Omega \, dt$$

отражается, остальная часть поглощается. На отраженный поток накладывается поток

$$E_\omega \, ds \cos\varphi \, d\omega \, d\Omega \, dt$$

собственного излучения площадки. Таким образом, от площадки ds внутрь полости исходит лучистый поток

$$[(1 - A_\omega) I_\omega + E_\omega] \, ds \cos\varphi \, d\omega \, d\Omega \, dt.$$

Абсолютно черных тел, как и других идеализированных объектов, в природе не существует. Наилучшим приближением к абсолютно черному телу является замкнутая полость, в стенке которой сделано малое отверстие, через которое излучение из полости может выходить наружу. Если стенки полости непрозрачны, то при достаточно малых размерах отверстия в полости установится излучение, лишь бесконечно мало отличающееся от равновесного. Через отверстие будет выходить практически такое же излучение, какое испускалось бы абсолютно черной площадкой той же формы и размеров.

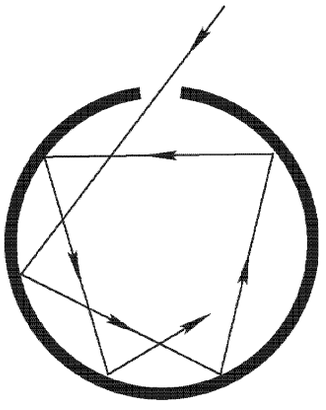


Рис. 339

Убедиться в этом можно и другим способом. Луч света, вступивший снаружи в полость через ее отверстие, будет претерпевать многократные отражения от стенок полости (рис. 339) При каждом отражении часть лучистой энергии поглощается. После многократных отражений луч либо совсем не выйдет наружу через отверстие, либо выйдет лишь ничтожная часть лучистой энергии, вступившей в полость. Почти вся энергия поглотится стенками полости. Это значит, что полость с малым отверстием в отношении поглощения, а потому по закону Кирхгофа и в отношении испускания, ведет себя практически *как абсолютно черное тело.* Изложенный способ всегда применяется при точных количественных измерениях излучения абсолютно черного тела.

Поясним изложенное примером. Если стенки полости с малым отверстием ярко осветить снаружи, то отверстие будет выделяться своей чернотой на светлом фоне освещенных стенок. Таким представляется, например, открытое окно здания. Так будет происходить даже в том случае, когда наружная поверхность стенок полости закрашена черной краской. Если же раскалить стенки полости, сделанные из материала с малой поглощательной способностью (например, из белого фарфора), то отверстие будет ярко светиться на более тусклом фоне прямого излучения самих стенок. При закрашивании стенок снаружи в черный цвет стенки светятся ярче, но все же их яркость остается меньше яркости отверстия. Закрашивание стенок изнутри совсем не сказывается на яркости отверстия.

4. Так как для абсолютно черного тела $e_\omega = I_\omega$, а равновесное излучение изотропно, то излучательная способность абсолютно черного тела одинакова по всем направлениям. При тепловом излучении она, очевидно, совпадает с поверхностной яркостью тела (см. § 22). Значит, *излучательная способность абсолютно черного тела подчиняется закону Ламберта* (см. § 22).

Обратное, конечно, не справедливо. Закон Ламберта может выполняться и для не абсолютно черных тел. Закон Кирхгофа позволяет полнее исследовать этот вопрос. Согласно этому закону, $E_\omega = A_\omega e_\omega$, а потому тело будет излучать по закону Ламберта, если его поглощательная способность не зависит от направления поглощаемого излучения. Так как для непрозрачного тела энергия падающего света равна сумме энергий поглощенного и рассеянного света, то, очевидно, закону Ламберта будет подчиняться излучение всяких тел, равномерно рассеивающих падающий свет во все стороны, независимо от его направления. Такие тела называются *абсолютно матовыми*. Итак, *тепловое излучение абсолютно матовых тел следует закону Ламберта*. На эту связь закона Ламберта с законом Кирхгофа указал В.А. Ульянин (1863–1930) в 1897 г., хотя его рассуждения и были несколько сложнее приведенных здесь.

§ 114. Формула Кирхгофа–Клаузиуса

Найдем теперь плотность энергии и удельную интенсивность равновесного излучения в прозрачной однородной изотропной среде с показателем преломления n. Такое излучение устанавливается в замкнутой полости, заполненной рассматриваемой средой, стенки которой поддерживаются при постоянной температуре. В точности такое же излучение установится в среде и в том случае, когда она заполняет только часть полости. Предположим, что часть полости заполнена рассматриваемой средой, а в другой находится вакуум. Равновесное излучение как в среде, так и в вакууме совершенно не зависит от формы и свойств поверхности, вдоль которой среда граничит с вакуумом. Не меняя окончательного результата, можно принять, что эта граница плоская и гладкая.

Излучение в среде не зависит и от ее размеров. Поэтому среду можно считать настолько протяженной, чтобы при сколь угодно малом коэффициенте поглощения световой луч, вступивший в среду, успел полностью поглотиться, не достигнув стенок полости. Тогда обмен энергией между средой и вакуумом будет происходить только в результате отражения и преломления излучения на рассматриваемой границе. Такой обмен подчиняется *принципу детального равновесия* и не может нарушить состояние равновесия излучения как в среде, так и в вакууме. Из этого условия и можно найти соотношение между плотностью энергии u^0 и удельной интенсивностью I^0 излучения в вакууме с такими же величинами u и I в среде. (В этом параграфе все величины, относящиеся к вакууму, снабжены нулем в индексе, а все величины в среде оставлены без индекса.)

Ввиду выполнения принципа детального равновесия, рассуждение достаточно провести не для всего излучения, а только для произвольной части его, заполняющей интервал частот ω, $\omega + d\omega$. Через единичную площадку границы раздела в единицу времени из вакуума в пределах телесного угла Ω_0 падает лучистый поток $\Phi_1 = I^0_\omega\, d\Omega_0\, d\omega \cos\varphi_0$, где φ_0 — угол падения (рис. 340). Согласно принципу детального равновесия, должен существовать такой же лучистый поток, распространяющийся в обратном направлении. Он состоит из двух потоков. Первый поток возникает в результате отражения потока Φ_2 и равен $(1 - A_\omega) I^0_\omega\, d\Omega_0\, d\omega \cos\varphi_0$. Второй поток возникает в результате преломления потока Φ_3, приходящего снизу из среды. Так как в силу формул Френеля коэффициенты отражения на границе раздела при прямом и обратном ходе лучей одинаковы, то второй поток будет $A_\omega I_\omega\, d\Omega\, d\omega \cos\varphi$, где φ — угол преломления, а $d\Omega$ — телесный угол в среде, переходящий при преломлении в $d\Omega_0$. Сокращая на $d\omega$, запишем условие детального равновесия в виде

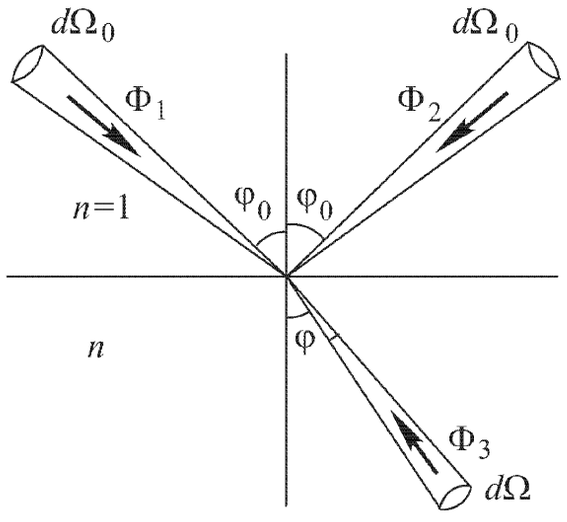


Рис. 340

$$(1 - A_\omega) I^0_\omega \cos\varphi_0\, d\Omega_0 + A_\omega I_\omega \cos\varphi\, d\Omega = I^0_\omega \cos\varphi_0\, d\Omega_0,$$

или

$$I_\omega \cos\varphi\, d\Omega = I^0_\omega \cos\varphi_0\, d\Omega_0.$$

В качестве $d\Omega_0$ возьмем (не изображенный на рис. 340) телесный угол, заключенный между коническими поверхностями, образующие которых составляют с нормалью к границе раздела углы φ_0 и $\varphi_0 + d\varphi_0$, т.е. $d\Omega_0 = 2\pi \sin\varphi_0\, d\varphi_0$. Аналогично, $d\Omega = 2\pi \sin\varphi\, d\varphi$. Тогда

$$I_\omega \cos\varphi \sin\varphi\, d\varphi = I^0_\omega \cos\varphi_0 \sin\varphi_0\, d\varphi_0.$$

В силу закона преломления $\sin\varphi_0 = n \sin\varphi$, так что $\cos\varphi_0\, d\varphi_0 = n \cos\varphi\, d\varphi$. Поэтому

$$I_\omega \sin\varphi = n I^0_\omega \sin\varphi_0.$$

Отсюда в силу того же закона преломления

$$I_\omega = n^2 I^0_\omega. \tag{114.1}$$

Эта формула и решает поставленную задачу. Она была получена Кирхгофом в 1860 г. и независимо от него Клаузиусом в 1864 г. Ее называют *формулой Кирхгофа–Клаузиуса.*

В вакууме величины I^0_ω и u^0_ω связаны соотношением (112.6), т.е. $cu^0_\omega = 4\pi I^0_\omega$. В среде оно заменяется на $v_{\text{гр}} u_\omega = 4\pi I_\omega$, где $v_{\text{гр}}$ — групповая скорость. Следовательно,

$$u_\omega = u^0_\omega n^2 \frac{c}{v_{\text{гр}}}. \tag{114.2}$$

Для недиспергирующих сред

$$u_\omega = n^3 u^0_\omega. \tag{114.3}$$

Соотношениям (114.1) и (114.3) можно придать другую, легче запоминаемую форму, глубокий смысл которой выяснится при рассмотрении проблемы методами статистической физики (см § 117). Для этого заметим, что $n = \lambda_0/\lambda$. Тогда (114.1) и (114.3) преобразуются в

$$\lambda^2 I_\omega = \lambda_0^2 I^0_\omega, \tag{114.4}$$

$$\lambda^3 u_\omega = \lambda_0^3 u^0_\omega. \tag{114.5}$$

Соотношение (114.4) означает, что *в равновесном излучении потоки лучистой энергии через квадратную площадку со стороной, равной длине волны λ (в рассматриваемой среде), отнесенные к единице телесного угла, в направлении нормали к площадке одинаковы для всех сред и зависят только от их температуры.* Это утверждение справедливо как для полного потока, так и для его монохроматических составляющих. Аналогично, соотношение (114.5) утверждает, что *энергия равновесного излучения, локализованная в кубике с ребром, равным длине волны λ, одинакова во всех изотропных недиспергирующих средах и определяется только температурой среды.* Это утверждение относится также не только к полной плотности энергии, но и к ее монохроматическим составляющим.

§ 115. Закон Стефана–Больцмана

1. В XIX веке производились многочисленные исследования зависимости *интегральной лучеиспускательной способности* нагретых тел от температуры, т.е. величины, которая определяет *суммарную энергию* всех длин волн, излучаемых телами. Эти исследования приводили к противоречивым результатам. Основная причина расхождений была окончательно выяснена после установления закона Кирхгофа, так как излучение определяется не только температурой, но также составом тела и физическими свойствами излучающей поверхности. А на эту сторону дела в экспериментальных исследованиях не обращалось должного внимания. Из эмпирически установленных законов следует отметить только результат, найденный в 1879 г.

Стефаном (1835–1898). Он нашел, что *для черных тел излучательная способность пропорциональна четвертой степени температуры.* Через пять лет Больцман получил этот результат теоретически из термодинамических соображений и показал, что он *абсолютно верен для абсолютно черных тел.* Этот результат, получивший название *закона Стефана–Больцмана*, был подтвержден последующими опытами по излучению абсолютно черного тела.

Вывод Больцмана и все последующие работы по теории теплового излучения существенно используют результаты Максвелла, предсказавшего и рассчитавшего давление света (см. т. II, § 61; т. III, § 145, а также задачу 2 к § 84 этого тома). Для изотропного излучения это давление равно $\mathcal{P} = {}^1\!/_3 u$, где u — интегральная плотность лучистой энергии. К такому выражению должна приводить всякая релятивистская теория света, независимо от того, является ли она корпускулярной или волновой. До теории относительности этот результат, разумеется, не был известен, а результаты Максвелла не считались общепризнанными. В частности, согласно нерелятивистской корпускулярной теории должно было бы быть $\mathcal{P} = {}^2\!/_3 u$, как это предсказывает кинетическая теория газов (см. т. II, § 59). Поэтому опыты П.Н. Лебедева, впервые измерившего в 1900 г. световое давление, подтвердившие результаты Максвелла, имели основополагающее значение для всей термодинамики лучистой энергии.

2. Доказательство закона Стефана–Больцмана мы проведем *методом циклов*, так как таким путем попутно будут получены важные соотношения, используемые в дальнейшем. Допустим, что изотропное излучение произвольного спектрального состава заключено в адиабатическую оболочку с абсолютно зеркальными стенками. Произведем над ним адиабатический квазистатический процесс, при котором объем V, ограниченный оболочкой, меняется бесконечно медленно. Чтобы быть уверенным, что во время этого процесса излучение все время остается изотропным, можно взять оболочку сферической формы (в следующем параграфе будет показано, что эта предосторожность является лишней). Внутренняя энергия излучения в оболочке равна uV. При увеличении объема оболочки на dV за счет этой энергии совершается работа $\mathcal{P}\,dV$, так что $\mathcal{P}\,dV = -d(uV)$. А так как для изотропного излучения $\mathcal{P} = {}^1\!/_3 u$, то этому уравнению можно придать вид ${}^4\!/_3 u\,dV + V\,du = 0$. Отсюда следует, что во время процесса

$$uV^{4/3} = \text{const}, \tag{115.1}$$

или

$$\mathcal{P}V^{4/3} = \text{const}. \tag{115.2}$$

Это — *уравнение адиабаты* для изотропного излучения, совершенно аналогичное уравнению адиабаты Пуассона для идеального газа. Постоянная адиабаты равна $\gamma = {}^4/_3$.

В силу эффекта Доплера при адиабатическом сжатии или расширении излучения должен меняться его спектральный состав. Допустим, например, что изотропное излучение занимает спектральный интервал ω, $\omega + d\omega$. В результате отражения от движущейся стенки частота ω и ширина интервала $d\omega$ изменятся и сделаются равными ω' и $d\omega'$. При этом будет выполняться соотношение

$$u_\omega\, d\omega \cdot V^{4/3} = u'_{\omega'}\, d\omega' \cdot V'^{4/3} = \text{const}, \qquad (115.3)$$

где V' и $u'_{\omega'}$ — объем и спектральная плотность энергии излучения частоты ω' в конце процесса.

3. Докажем теперь закон Стефана–Больцмана. Для этого произведем над черным излучением *цикл Карно*. Допустим, что излучение заключено в цилиндре, боковые стенки и поршень которого идеально отражающие, а дно черное и может приводиться в тепловой контакт с нагревателем температуры T_1 и холодильником температуры T_2. Излучение можно адиабатически изолировать с помощью идеально отражающей задвижки, вводимой сбоку для прикрытия черного дна цилиндра. В отсутствие задвижки, когда черное дно цилиндра приведено в тепловой контакт с нагревателем или холодильником, излучение в цилиндре, конечно, будет равновесным. Чтобы быть уверенным, что оно сохранится равновесным и во время адиабатического процесса, когда задвижка вставлена, введем внутрь цилиндра черную пылинку, роль которой была выяснена в п. 1 § 112. На изотерме *1–2* (рис. 341) дно цилиндра контактирует с нагревателем. Количество тепла, переданное нагревателем на этой изотерме, равно

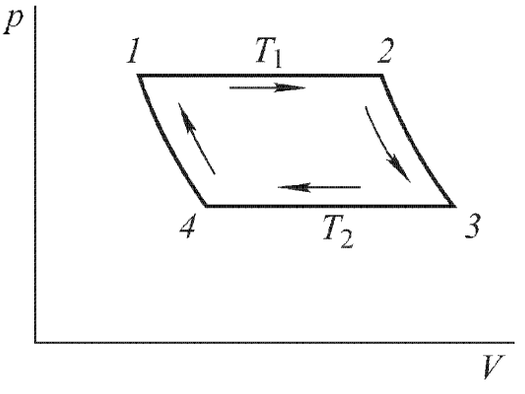


Рис. 341

$$Q_1 = u_1(V_2 - V_1) + \mathcal{P}_1(V_2 - V_1) = \frac{4}{3}\, u_1(V_2 - V_1).$$

Количество тепла, отданное холодильнику на изотерме *3–4*, $Q_2 = {}^4/_3 u_2(V_3 - V_4)$. По теореме Карно

$$\frac{Q_1}{Q_2} = \frac{u_1(V_2 - V_1)}{u_2(V_3 - V_4)} = \frac{T_1}{T_2}.$$

На адиабатах *2–3* и *4–1* в силу (115.1) выполняются соотношения

$$u_1^{3/4} V_2 = u_2^{3/4} V_3, \quad u_1^{3/4} V_1 = u_2^{3/4} V_4.$$

Отсюда

$$\frac{V_2 - V_1}{V_3 - V_4} = \left(\frac{u_2}{u_1}\right)^{3/4},$$

а потому $u_1/T_1^4 = u_2 T_2^4 = \text{const}$. Следовательно,

$$u = aT^4, \tag{115.4}$$

где a — универсальная постоянная. Но это есть иная форма закона Стефана–Больцмана.

Результат (115.4) можно получить короче, если к равновесному излучению применить общую термодинамическую формулу

$$\left(\frac{\partial U}{\partial V}\right)_T = T\left(\frac{\partial \mathcal{P}}{\partial T}\right)_V - \mathcal{P}. \tag{115.5}$$

Подставив сюда $U = Vu(T)$, $\mathcal{P} = {}^1\!/_3 u(T)$, придем к тому же дифференциальному уравнению, интегрированием которого была получена формула (115.4). Однако мы не хотели пользоваться формулой (115.5).

§ 116. Теорема и закон смещения Вина

1. Важные результаты в термодинамике излучения были получены Вильгельмом Вином (1864–1928) в 1893–1894 гг. Вин доказал, что *равновесное излучение, заключенное в оболочке с идеально отражающими стенками, будет оставаться равновесным при квазистатическом сжатии или расширении оболочки.*

Для наших целей при доказательстве теоремы Вина достаточно ограничиться оболочкой сферической формы. В этом случае, ввиду сферической симметрии системы, отпадает необходимость специально доказывать, что в ходе процесса изотропия излучения все время сохраняется. Сожмем излучение квазистатически от начального объема V_1 до конечного V_2. При этом будет совершена работа против сил светового давления, и энергия излучения в оболочке увеличится. Спектральный состав излучения также изменится, из-за эффекта Доплера. Допустим, что в результате этого излучение перестанет быть равновесным. Введем внутрь оболочки в конечном состоянии бесконечно малую черную пылинку, поглощающую и излучающую свет. По истечении достаточно длительного времени она превратит неравновесное излучение в оболочке в равновесное. Это — необратимый процесс, идущий самопроизвольно. Обратный процесс превращения равновесного излучения в неравновесное, разумеется, сам собою идти не может.

Когда излучение внутри оболочки станет равновесным, не убирая пылинку, начнем бесконечно медленно адиабатически

расширять оболочку, доведя объем излучения до исходного значения V_1. После этого удалим пылинку. Энергия пылинки бесконечно мала, а потому ее наличие может сказаться на общей энергии излучения в полости также бесконечно мало. С другой стороны, давление изотропного излучения зависит только от интегральной плотности энергии излучения u, но не от его спектрального состава Поэтому работа, которую совершит световое давление при расширении оболочки, будет с точностью до бесконечно малой величины равна внешней работе, совершенной над излучением при его сжатии. Отсюда следует, что в результате сжатия и последующего расширения энергия, а с ней и температура излучения не изменятся.

Система совершила круговой процесс, в ходе которого она не получала и не отдавала тепло, а общая работа, произведенная ею, равна нулю. Значит, в окружающих телах не произошло никаких изменений, а потому рассматриваемый круговой процесс обратим. Но это невозможно, так как одна из стадий этого кругового процесса по нашему предположению необратима. Следовательно, это предположение неверно, и теорема Вина доказана.

2. Значение теоремы Вина — методическое. Действительно, адиабатически и квазистатически меняя объем равновесного излучения в оболочке с идеально зеркальными стенками, можно получить равновесное излучение произвольной плотности, а следовательно, и температуры. Энергию (и температуру) этого излучения можно найти, вычислив работу, совершенную над ним в этом процессе. Его спектральный состав найдется, если вычислить доплеровское изменение частоты излучения при его отражении от движущейся оболочки. Тем самым будет установлено определенное соответствие между параметрами равновесного излучения в начале процесса и на любой стадии его.

Применим этот метод к равновесному излучению в сферической оболочке с идеально зеркальными стенками. При бесконечно медленном адиабатическом расширении или сжатии оболочки излучение в ней все время будет оставаться равновесным, так что его можно в любой момент времени характеризовать определенной температурой T. Выделим внутри оболочки произвольный луч, падающий на оболочку под углом ϑ (рис. 342).

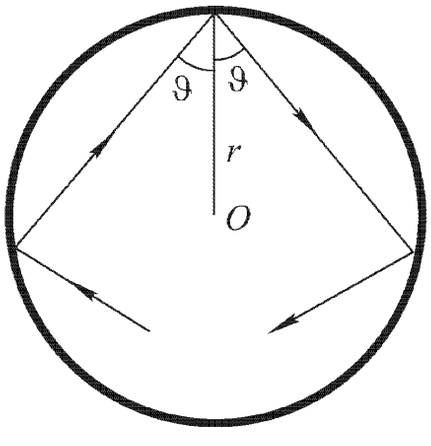


Рис. 342

Время между двумя последовательными отражениями этого луча равно $\Delta t = (2r/c)\cos\vartheta$. За это время радиус оболочки r полу-

чает приращение $\Delta r = \dot{r}\,\Delta t$. При каждом отражении происходит доплеровское изменение частоты, определяемое формулой

$$\frac{\Delta\omega}{\omega} = -\frac{2\dot{r}\cos\vartheta}{c} = -\frac{2\Delta r\cos\vartheta}{c\,\Delta t} = -\frac{\Delta r}{r},$$

если пренебречь квадратом бесконечно малой радиальной скорости $\dot{r}$ расширения оболочки. Относительное изменение частоты $\Delta\omega/\omega$ определяется только относительным изменением $\Delta r/r$ радиуса оболочки. Такая же формула получится и в том случае, когда за время изменения радиуса оболочки на Δr произойдет не одно, а много отражений светового луча. Требуется только, чтобы выполнялось условие $\Delta r \ll r$. При бесконечно медленном расширении величины Δr и $\Delta\omega$ можно заменить их дифференциалами, т.е. написать

$$\frac{d\omega}{\omega} + \frac{dr}{r} = 0. \tag{116.1}$$

Это означает, что реальный процесс, в котором последовательные отражения отделены друг от друга малыми, но все же конечными промежутками времени, при расчетах заменяется идеализированной схемой, в которой эти отражения следуют друг за другом *непрерывно во времени*. Интегрируя уравнение (116.1), получим

$$\omega r = \text{const}. \tag{116.2}$$

Так как $r \sim V^{1/3}$, то этот результат можно записать также в виде

$$\omega^3 V = \text{const}. \tag{116.3}$$

В таком виде он справедлив для полости произвольной формы. А поскольку он получен для бесконечно медленного процесса, величина $\omega^3 V$ является *адиабатическим инвариантом*. Комбинируя его с ранее полученными адиабатическими инвариантами (115.1) и (115.3), получим новые адиабатические инварианты. Так, из формул (115.1) и (116.3) следует

$$\frac{\omega^4}{u} = \text{const}, \tag{116.4}$$

или на основании закона Стефана–Больцмана

$$\frac{\omega}{T} = \text{const}. \tag{116.5}$$

Аналогично, формулы (115.3) и (116.3) дают

$$\frac{u_\omega\,d\omega}{\omega^4} = \text{const}. \tag{116.6}$$

Таким образом, при квазистатическом расширении или сжатии равновесного излучения в полости с зеркальными стенками каждая квазимонохроматическая составляющая излучения ведет себя *независимо от остальных составляющих* и меняется

так, что величины $\omega^3 V$, u/ω^4 и $u_\omega\, d\omega/\omega^4$ остаются постоянными, т.е. являются адиабатическими инвариантами. По теореме Вина при таком процессе излучение все время остается равновесным. Такое же излучение можно было бы получить в неподвижной оболочке, нагревая или охлаждая ее стенки. Поэтому полученные результаты можно представить как *свойства только самого равновесного излучения*, не связывая их ни с каким конкретным процессом. Сформулируем их следующим образом. *Изменим любым способом температуру равновесного излучения от T до T', чтобы излучение оставалось равновесным. Каждой частоте ω излучения в начальном состоянии приведем в соответствие такую частоту ω' в конечном состоянии, чтобы $\omega/T = \omega'/T'$ и, следовательно, $d\omega/T = d\omega'/T'$. Тогда плотности лучистой энергии в этих состояниях будут связаны соотношениями*

$$\frac{u}{\omega^4} = \frac{u'}{\omega'^4}, \tag{116.7}$$

$$\frac{u_\omega\, d\omega}{\omega^4} = \frac{u'_{\omega'}\, d\omega'}{\omega'^4}. \tag{116.8}$$

Эти результаты составляют содержание так называемого *закона смещения Вина в его наиболее общей форме.*

3. Из формулы (116.8) получаем

$$u_\omega(\omega, T) = \frac{\omega^4}{\omega'^4}\frac{d\omega'}{d\omega}u'_{\omega'}(\omega', T) = \frac{T^3}{T'^3}u'_{\omega'}\left(\frac{T'}{T}\,\omega, T'\right).$$

Это соотношение справедливо при любом значении температуры T', а потому величина справа от T' не зависит. Величине T' можно придать *любое значение*, представив полученное соотношение в виде

$$u_\omega(\omega, T) = T^3\varphi\left(\frac{\omega}{T}\right), \tag{116.9}$$

где $\varphi(\omega/T)$ — *универсальная функция* аргумента ω/T. Ввиду соотношения $\omega/T = \omega'/T'$, тот же результат можно записать в виде

$$u_\omega(\omega, T) = \omega^3 f\left(\frac{\omega}{T}\right), \tag{116.10}$$

где $f(\omega/T) \equiv (T'/\omega')^3\varphi(\omega/T)$ — новая *универсальная функция* того же аргумента ω/T. Тем самым определение универсальной функции $u_\omega(\omega, T)$ *двух аргументов* сведено к задаче нахождения универсальной функции $\varphi(\omega, T)$ или $f(\omega/T)$ только *одного аргумента* ω/T. Отсюда следует, что *если известно спектральное распределение в равновесном излучении при какой-либо произвольной температуре T', то с помощью формулы* (116.9) *или* (116.10) *можно найти это распределение при всякой другой температуре T.*

В переменных λ, T, как это следует из соотношения (112.2), формулам (116.9) и (116.10) можно придать вид

$$u_\lambda = T^5 \varphi_1(\lambda T), \tag{116.11}$$

$$u_\lambda = \frac{1}{\lambda^5} f_1(\lambda T), \tag{116.12}$$

где $\varphi_1(\lambda T)$ и $f_1(\lambda T)$ — новые *универсальные функции*. При фиксированной температуре T величина u_λ становится функцией только длины волны λ. Эта функция не может возрастать монотонно, а должна иметь максимум. В противном случае интегральная плотность излучения

$$u = \int_0^\infty (u_\lambda)_{T=\text{const}}\, d\lambda \tag{116.13}$$

не могла бы оставаться конечной. Длину волны в максимуме обозначим через λ_m. Если ввести обозначение $x = \lambda T$, то для определения положения максимума получится уравнение

$$d\varphi_1/d\lambda = T\, d\varphi_1/dx = 0, \quad \text{т.е.} \quad d\varphi_1/dx = 0.$$

Таким образом, при всех температурах максимум получается при одном и том же значении аргумента x. Отсюда следует, что *при повышении температуры максимум функции* $u_{\lambda,T=\text{const}}$ *смещается в сторону более коротких волн и притом так, что выполняется соотношение*

$$\lambda_m T = b = \text{const}. \tag{116.14}$$

Измерения дают: $b = 0{,}2898$ см·К. Этот результат, определяющий смещение максимума излучения при изменении температуры T, называют *законом Вина в его специальной форме*. В точках максимума функции $u_\lambda(T)$ и $u_{\lambda'}(T')$ относятся как *пятые степени* абсолютных температур T и T'.

Функция $u_\omega(\omega, T)$ при постоянной температуре T также обращается в максимум при какой-то частоте $\omega = \omega_m$. Частота ω_m не равна $2\pi c/\lambda_m$, так как речь идет о максимумах *различных функций* $u_\omega(\omega)$ и $u_\lambda(\lambda)$. При изменении температуры излучения положение максимума смещается, но при этом имеет место соотношение

$$\frac{\omega_m}{T} = \text{const}, \tag{116.15}$$

которое является *другой формой закона смещения Вина* (116.14).

4. Выведенный ранее закон Стефана–Больцмана является следствием общей формулы (116.9). Действительно, вводя обо-

значение $x = \omega/T$, из этой формулы при постоянной T находим

$$u = \int_0^\infty u_\omega \, d\omega = T^3 \int_0^\infty \varphi\left(\frac{\omega}{T}\right) d\omega = T^4 \int_0^\infty \varphi(x)\, dx = aT^4,$$

где $a \equiv \int_0^\infty \varphi(x)\, dx$ — универсальная постоянная.

§ 117. Формула Рэлея–Джинса

1. Результатами, изложенными в предыдущих параграфах, исчерпывается все, что могла дать феноменологическая термодинамика в проблеме теплового излучения. Ее оказалось недостаточно для решения основной проблемы теории теплового излучения: определения функции $u_\omega(\omega, T)$ или функции $I_\omega(\omega, T)$, связанной с ней соотношением (112.6). Для этого оказалось необходимым привлечь *статистические методы* и учесть *квантовые свойства вещества и излучения.* Первая попытка теоретического решения указанной проблемы была предпринята в 1887 г. В.А. Михельсоном (1860–1927). В то время, как показало последующее развитие физики, правильное решение рассматриваемой проблемы было, конечно, невозможно. Заслуга Михельсона состоит в том, что он привлек внимание физиков к одной из важнейших проблем, решение который положило начало *квантовой физики.*

Общий метод теоретического определения функции $u_\omega(\omega, T)$ в рамках классической физики, не связанный с модельными представлениями, был указан в 1900 г. Рэлеем и через пять лет более подробно развит Джинсом (1877–1946). Рэлей и Джинс применили к равновесному излучению в полости теорему классической статистической механики о *равномерном распределении кинетической энергии по степеням свободы.* Согласно этой теореме, *в состоянии статистического равновесия на каждую степень свободы приходится в среднем кинетическая энергия* $\frac{1}{2}kT$, где $k = 1{,}38 \cdot 10^{-16}$ эрг/К — постоянная Больцмана.

Если степень свободы колебательная, то надо учесть еще *потенциальную энергию.* В случае *гармонических колебаний* среднее значение потенциальной энергии равно также $\frac{1}{2}kT$ (см. т. II, § 63). Таким образом, *в состоянии статистического равновесия на каждую колебательную степень свободы приходится средняя энергия, равная* kT.

Эта теорема сводит задачу нахождения функции $u_\omega(\omega, T)$ к определению *числа степеней свободы излучения в полости.* Поскольку равновесное излучение в полости не зависит от ее

формы и материала стенок, можно предположить, что полость имеет форму куба с идеально отражающими стенками. А чтобы излучение в полости было равновесным, можно ввести в нее бесконечно малую черную пылинку, как это делалось в предыдущем параграфе при доказательстве теоремы Вина.

2. Чтобы лучше уяснить метод определения числа степеней свободы, рассмотрим этот вопрос сначала не для векторного электромагнитного поля, а для *скалярного волнового поля*, например для продольных акустических волн. В такой постановке этот вопрос имеет и самостоятельный интерес, например в теории теплоемкости твердых тел Дебая. Волновое поле будем характеризовать какой-то функцией $V(\mathbf{r}, t)$, удовлетворяющей волновому уравнению

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 V}{\partial t^2} = 0. \tag{117.1}$$

Предположим, что на стенках полости функция V обращается в нуль. Тогда (а также и при других граничных условиях) по теореме Фурье функция $V(\mathbf{r}, t)$ может быть представлена суперпозицией стоячих волн. Координатные оси X, Y, Z направим параллельно ребрам кубической полости. Так как граничные условия должны удовлетворяться на гранях $x = 0$, $y = 0$, $z = 0$ и на параллельных им гранях полости, то каждая стоячая волна должна представляться функцией с *разделяющимися переменными*, т.е.

$$V = X(x)Y(y)Z(z)\exp(i\omega t).$$

Отсюда

$$\frac{1}{V}\frac{\partial^2 V}{\partial x^2} = \frac{1}{X}\frac{d^2 X}{dx^2}, \ldots, \frac{1}{V}\frac{\partial^2 V}{\partial t^2} = \frac{\omega^2}{c^2}.$$

Подставив эти выражения в уравнение (117.1), получим

$$\frac{1}{X}\frac{d^2 X}{dx^2} + \frac{1}{Y}\frac{d^2 Y}{dy^2} + \frac{1}{Z}\frac{d^2 Z}{dz^2} = \frac{\omega^2}{c^2}.$$

Левая часть этого уравнения есть сумма функций от x, y, z соответственно, т.е. от различных переменных. Она равна постоянной ω^2/c^2. Это может быть тогда и только тогда, когда каждая из этих функций сама постоянна, т.е.

$$\frac{1}{X}\frac{d^2 X}{dx^2} = q_x^2, \quad \frac{1}{Y}\frac{d^2 Y}{dy^2} = q_y^2, \quad \frac{1}{Z}\frac{d^2 Z}{dz^2} = q_z^2,$$

где q_x, q_y, q_z — постоянные, удовлетворяющие условию

$$q_x^2 + q_y^2 + q_z^2 = q^2 \equiv \frac{\omega^2}{c^2}. \tag{117.2}$$

Решение первого уравнения запишем в виде $X = \sin(q_x x + \delta_x)$, где δx — постоянная интегрирования. Второй постоянной интегрирования служит амплитуда, которую без потери общности

можно опустить. На грани полости $x = 0$ функция X должна обращаться в нуль, т.е. $\sin \delta_x = 0$. Также не теряя общности, можно положить $\delta_x = 0$. На противоположной грани $x = l$, где l — длина ребра кубической полости, функция X также должна обращаться в нуль, а потому $\sin k_x l = 0$. Такие же результаты получаются и для координат y и z. Следовательно,

$$V = \sin q_x x \, \sin q_y y \, \sin q_z z \, e^{i\omega t}, \tag{117.3}$$

где

$$q_x l = m_x \pi, \quad q_y l = m_y \pi, \quad q_z l = m_z \pi, \tag{117.4}$$

а m_x, m_y, m_z — целые числа. Все их можно считать положительными, так как введение отрицательных чисел не приводит к новым, линейно независимым решениям. В выражение (117.3) можно было бы ввести еще постоянную амплитуду, зависящую от m_x, m_y, m_z, но для наших целей в этом нет надобности. Это выражение и представляет общий вид *стоячей волны* в полости. Каждой тройке целых положительных чисел m_x, m_y, m_z, удовлетворяющей условиям (117.4), соответствует одна стоячая волна. Число возможных стоячих волн бесконечно велико.

Будем рассматривать q_x, q_y, q_z как прямоугольные координаты точки трехмерного «пространства волновых векторов» $\mathbf{q}$. Эти «изображающие точки» расположатся в узлах кубической решетки, элементарная ячейка которой есть кубик с длиной стороны $\Delta q_x = (\pi/l)\Delta m_x = \pi/l$ и объемом $(\pi/l)^3$. Решетка заполняет только положительный октант пространства волновых векторов, так как все координаты q_x, q_y, q_z положительны. Объем части шара радиуса q, лежащей в этом октанте, равен $\frac{1}{8}(4\pi/3)q^3 = (\pi/6)q^3$. Число изображающих точек в нем и будет равно числу Z стоячих волн, волновые числа которых не превосходят $q = 2\pi/\lambda$. Подавляющее число волн очень короткие, для них величина q очень велика по сравнению с длиной π/l ребра элементарного кубика. Поэтому число стоячих волн в указанном октанте шара найдется делением его объема на объем элементарного кубика. Таким путем получается асимптотическая формула

$$Z = \frac{(\pi/6)q^3}{\pi^3/l^3} = \frac{V}{6\pi^2}\, q^3 = \frac{V}{6\pi^2}\frac{\omega^3}{c^3}, \tag{117.5}$$

где $V = l^3$ — объем полости. Она справедлива, когда сторона кубической полости очень велика по сравнению с длиной волны $\lambda = 2\pi/q$. Можно доказать, что формула (117.5) остается верной и для полости произвольной формы, хотя в этом случае стоячие волны и не будут представляться выражениями вида (117.3).

Дифференцируя (117.5), получим

$$dZ = \frac{V}{2\pi^2}\, q^2\, dq = \frac{V}{2\pi^2 c^3}\, \omega^2\, d\omega. \tag{117.6}$$

Эта асимптотическая формула дает число стоячих волн в интервале частот ω, $\omega + d\omega$. Она, разумеется, справедлива только для достаточно широких интервалов $d\omega$, когда $dZ \gg 1$.

3. Рассуждения существенно не изменятся и в случае векторного (электромагнитного) поля. В этом случае вектор $\mathbf{E}$ и его прямоугольные составляющие E_x, E_y, E_z удовлетворяют прежнему волновому уравнению (117.1). Для полости с идеально зеркальными стенками граничные условия требуют обращения в нуль тангенциальных составляющих вектора $\mathbf{E}$ на стенках полости. Для этого необходимо и достаточно, чтобы выполнялись следующие условия:

$$\begin{aligned} E_y = E_z = 0 \quad &\text{при} \quad x = 0 \quad \text{и} \quad x = l, \\ E_z = E_x = 0 \quad &\text{при} \quad y = 0 \quad \text{и} \quad y = l, \\ E_x = E_y = 0 \quad &\text{при} \quad z = 0 \quad \text{и} \quad z = l. \end{aligned}$$

Стоячая волна, удовлетворяющая этим условиям, имеет вид

$$\begin{aligned} E_x &= \sin(q_x x + \varphi_x) \sin q_y y \, \sin q_z z \, e^{i\omega t}, \\ E_y &= \sin q_x x \, \sin(q_y y + \varphi_y) \sin q_z z \, e^{i\omega t}, \\ E_z &= \sin q_x x \, \sin q_y y \, \sin(q_z z + \varphi_z) \, e^{i\omega t}, \end{aligned}$$

где q_x, q_y, q_z определяются прежними формулами (117.4). Что касается фаз φ_x, φ_y, φ_z, то их можно было бы найти из граничных условий, которым на стенках полости удовлетворяют нормальные составляющие электрического вектора $\mathbf{E}$. Но для наших целей в этом нет необходимости, так как для определения величин Z и dZ достаточно знать только значения, которые могут принимать составляющие q_x, q_y, q_z. Однако по сравнению с предыдущим случаем выражения (117.5) и (117.6) надо удвоить, так как электромагнитные волны *векторные* и притом *поперечные*. Каждому направлению распространения соответствуют *две волны*, поляризованные во взаимно перпендикулярных плоскостях, в результате суперпозиции которых может быть получена волна любой поляризации, распространяющаяся в том же направлении. Итак, для электромагнитного поля

$$dZ = \frac{V}{\pi^2 c^3} \omega^2 \, d\omega. \tag{117.7}$$

На каждую стоячую волну в состоянии статистического равновесия приходится в среднем энергия $\bar{\mathcal{E}} = kT$: одна половина ее — электрическая, другая — магнитная. Записав энергию равновесного излучения в полости в спектральном интервале $d\omega$ в виде $V u_\omega \, d\omega$, из формулы (117.7) получим

$$u_\omega = \frac{\omega^2}{\pi^2 c^3} \bar{\mathcal{E}} = \frac{kT}{\pi^2 c^3} \omega^2. \tag{117.8}$$

Этот результат известен под названием *формулы Рэлея–Джинса*, хотя он независимо и практически одновременно с Рэлеем был получен также Планком из столь же общих, но несколько других соображений (см. п. 6).

4. Прежде чем обсуждать формулу Рэлея–Джинса, заметим, что в случае полости, заполненной изотропной средой, число стоячих волн будет определяться прежними формулами (117.5) и (117.6), если только в них величину c заменить скоростью света v в рассматриваемой среде (предполагается, что среда изотропная). Отсюда следует, что числа Z и dZ в одном и том же интервале частоты, а с ними и функция u_ω пропорциональны c^3/v^3, т.е. кубу показателя преломления среды n. Но это есть *закон Кирхгофа–Клаузиуса*, доказанный в § 114. Вывод справедлив при более общих предположениях, чем это сделано в тексте. Нет необходимости ссылаться на классическую теорему о равномерном распределении энергии по степеням свободы. Достаточно, чтобы средняя энергия гармонического осциллятора была функцией только частоты ω, как это имеет место в квантовой теории.

5. Формула Рэлея–Джинса согласуется с общей термодинамической формулой Вина (116.9) или (116.10). Более того, вид этой формулы может быть непосредственно установлен на основе одной только формулы Вина, если ограничиться предельным случаем низких частот. Для этого заметим, что в формуле (116.10) функция $f(\omega/T)$ должна быть возрастающей функцией температуры T и обращаться в нуль при $T = 0$. Ее удобнее рассматривать как функцию аргумента T/ω и разложить в ряд по степеням этого аргумента. Если ограничиться первым членом этого ряда, то получится

$$u_\omega(\omega, T) = CT\omega^2. \tag{117.9}$$

Но по своему виду этот результат совпадает с формулой (117.8). Только коэффициент C остается неопределенным. Зато формула (117.9) выведена в более общих предположениях, чем (117.8). Таким образом, в предельном случае *низких частот* нет сомнений в правильности вида формулы (117.9). И действительно, для достаточно длинных волн (точный критерий приводится в следующем параграфе) формула Рэлея–Джинса прекрасно согласуется с опытом. Она с успехом применяется в длинноволновой инфракрасной области спектра и в радиодиапазоне.

Но классическая статистика требует, чтобы формула Рэлея–Джинса (117.8) была верна *при любых частотах*. Однако это невозможно, так как тогда для интегральной плотности энергии

получилось бы бесконечное значение:

$$u = \frac{kT}{\pi^2 c^3} \int_0^\infty \omega^2 \, d\omega = \infty.$$

Отсюда следует, что *по теории Рэлея–Джинса тепловое равновесие между веществом и излучением невозможно.* Этот вывод противоречит опыту. П.С. Эренфест назвал его *ультрафиолетовой катастрофой.* Причина ультрафиолетовой катастрофы заключается в том, что в теории Рэлея–Джинса излучение в полости имеет *бесконечное*, а вещество *конечное* число степеней свободы. Поэтому, если бы было справедливо равномерное распределение энергии по степеням свободы, то при тепловом равновесии вся энергия должна была бы сосредоточиться в излучении.

6. Можно было бы возразить, что классическая статистическая механика, следствием которой является теорема о равномерном распределении кинетической энергии по степеням свободы, неприменима к системам с бесконечным числом степеней свободы. Но такое возражение неубедительно. В основе классической статистической механики лежат уравнения классической механики в форме Гамильтона (1805–1865). Хотя они и были установлены для механических систем с конечным числом степеней свободы, но можно показать, что *излучение в полости можно описывать бесконечным, но счетным числом обобщенных координат, также подчиняющихся уравнениям Гамильтона.* Следовательно, и вся система, состоящая из вещества и излучения, будет описываться уравнениями Гамильтона.

Поэтому было бы непонятно, почему теорема о равномерном распределении энергии справедлива для одних и не имеет места для других степеней свободы.

Кроме того, к формуле Рэлея–Джинса независимо пришел также Планк, применивший эту теорему *только к веществу, но не к излучению.* Он провел рассуждение для одномерного гармонического осциллятора, например, квазиупруго-связанного электрона, помещенного в полость с равновесным излучением. Под действием хаотически меняющегося электромагнитного поля излучения осциллятор будет совершать колебания с хаотически меняющимися амплитудами и фазами, излучая и поглощая при этом электромагнитные волны. Энергия осциллятора будет совершать беспорядочные флуктуации вокруг среднего значения $\bar{\mathcal{E}}$. В результате идейно-простых, но несколько длинных вычислений Планк пришел к формуле

$$u_\omega = \frac{\bar{\mathcal{E}}}{\pi^2 c^3} \omega^2. \qquad (117.10)$$

Если в этой формуле применить к осциллятору (вещество!) теорему о равномерном распределении кинетической энергии по степеням свободы, то получится формула Рэлея–Джинса.

§ 118. Формула Планка

1. Правильная формула для спектральной плотности энергии равновесного излучения, подтвержденная всеми экспериментальными исследованиями, была найдена Планком сначала полуэмпирическим путем. Спустя короткое время, Планк нашел теоретический вывод этой формулы, изложенный им 14 декабря 1900 г. на заседании Немецкого физического Общества. Этот день считается днем рождения новой — *квантовой* — физики. Идея о квантах, осторожно формулированная Планком, развилась в стройное и глубокое учение, покорившее всю физику. Решение проблемы равновесного излучения Планком было бы невозможно, если бы в своих исследованиях он не находился в тесном контакте с экспериментаторами. Из экспериментальных работ по излучению абсолютно черного тела особенно следует отметить исследования Лэнглея (1834–1906), Рубенса (1865–1922), Пашена (1865–1947), Вина, Луммера (1860–1925), Э. Принсхейма (1859–1917), Курльбаума.

2. Гипотеза Планка состоит в том, что излучение и поглощение света веществом происходит не непрерывно, а конечными порциями, называемыми *квантами света* или *квантами энергии*. Получим формулу Планка тем же методом, который применялся при выводе формулы Рэлея–Джинса. Тогда гипотезу Планка удобно взять в следующей форме: энергия гармонического осциллятора может принимать не произвольные, а только избранные значения, образующие дискретный ряд: 0, $\mathcal{E}_0$, $2\mathcal{E}_0$, $3\mathcal{E}_0$, ..., где $\mathcal{E}_0$ — определенная величина, зависящая только от собственной частоты ω осциллятора. Здесь под осциллятором понимается не только частица, могущая совершать гармонические свободные колебания, но, например, и стоячая волна определенной частоты в полости.

Если осциллятор изолирован, то по истечении достаточно длительного времени он потеряет всю свою энергию на излучение и перейдет на наинизший энергетический уровень с энергией $\mathcal{E} = 0$. Но если осциллятор находится в полости, стенки которой поддерживаются при постоянной температуре, то наряду с излучением будут происходить и акты поглощения, в результате которых возбуждаются и высшие энергетические уровни. Установится вполне определенное состояние детального равновесия, в котором число актов излучения в среднем равно числу обратных актов поглощения. В этом состоянии будут возбужде-

ны все энергетические уровни, но с различными вероятностями. И все, что требуется для нахождения функции $u_\omega(\omega, T)$, — это определить среднюю энергию $\bar{\mathcal{E}}$ осциллятора в этом состоянии статистического равновесия. Такая задача уже была решена нами (см. т. II, § 85). Приведем еще раз это решение в несколько измененной форме. По теореме Больцмана вероятности возбуждения энергетических уровней осциллятора пропорциональны величинам

$$1, e^{-\mathcal{E}_0/kT}, e^{-2\mathcal{E}_0/kT}, \ldots$$

Поэтому

$$\bar{\mathcal{E}} = \frac{\sum\limits_{n=1}^{\infty} n\mathcal{E}_0 e^{-n\mathcal{E}_0/kT}}{\sum\limits_{n=0}^{\infty} e^{-n\mathcal{E}_0/kT}} = \mathcal{E}_0 \frac{\sum\limits_{n=1}^{\infty} ne^{-nx}}{\sum\limits_{n=0}^{\infty} e^{-nx}},$$

где введено обозначение $x = \mathcal{E}_0/kT$. Значение знаменателя определяется формулой

$$\sum_{n=0}^{\infty} e^{-nx} = \frac{1}{1 - e^{-x}}.$$

Числитель находится дифференцированием этой формулы по x:

$$\sum_{n=1}^{\infty} ne^{-nx} = \frac{e^{-x}}{(1 - e^{-x})^2},$$

и следовательно,

$$\bar{\mathcal{E}} = \frac{\mathcal{E}_0}{e^x - 1} = \frac{\mathcal{E}_0}{e^{\mathcal{E}_0/kT} - 1}. \tag{118.1}$$

Подставив это значение в формулу (117.8), найдем

$$u_\omega = \frac{\omega^2}{\pi^2 c^3} \frac{\mathcal{E}_0}{e^{\mathcal{E}_0/kT} - 1}. \tag{118.2}$$

Перейдем в этой формуле к пределу $\mathcal{E}_0 \to 0$. Тогда $e^{\mathcal{E}_0/kT} \approx \approx 1 + \mathcal{E}_0/kT$, а потому $\bar{\mathcal{E}} = kT$. Получается классическое выражение для средней энергии осциллятора. Формула (118.2) в рассматриваемом предельном случае переходит, следовательно, в формулу Рэлея–Джинса. Этого и следовало ожидать, так как переход к пределу $\mathcal{E}_0 \to 0$ фактически означает возвращение к классическому рассмотрению, когда энергия осциллятора меняется не дискретными порциями, а непрерывно. Однако Планк, и это привело его к великому открытию, поставил вопрос, что получится, если предельный переход $\mathcal{E}_0 \to 0$ не производить. Величину $\mathcal{E}_0$ он определил из требования, чтобы выражение (118.2) удовлетворяло общей термодинамической формуле Вина

(116.10). Приравняв (118.2) и (116.10), видим, что это требование сводится к выполнению соотношения

$$\frac{1}{\pi^2 c^3}\frac{\mathcal{E}_0/\omega}{e^{\mathcal{E}_0/kT}-1}=f\left(\frac{\omega}{T}\right). \tag{118.3}$$

Но $\mathcal{E}_0$ есть характеристика *только самого осциллятора*, а потому не может зависеть от температуры T — макроскопического параметра, определяющего состояние вещества и излучения. Величина $\mathcal{E}_0$ может зависеть *только от собственной частоты ω осциллятора*. В таком случае, чтобы левая часть равенства (118.3) была функцией только аргумента ω/T, необходимо и достаточно, чтобы

$$\mathcal{E}_0=\hbar\omega, \tag{118.4}$$

где $\hbar$ — постоянная. Она *универсальна*, поскольку в правой части (118.3) стоит универсальная функция $f(\omega/T)$ аргумента ω/T. Величина $\hbar$ называется *постоянной Планка*. Ее числовое значение по современным данным равно

$$\hbar=(1{,}0545887\pm 0{,}0000057)\cdot 10^{-27}\text{ эрг}\cdot\text{с}=$$
$$=(1{,}0545887\pm 0{,}0000057)\cdot 10^{-34}\text{ Дж}\cdot\text{с}. \tag{118.5}$$

Сам Планк пользовался постоянной

$$h\equiv 2\pi\hbar=(6{,}626176\pm 0{,}000036)\cdot 10^{-27}\text{ эрг}\cdot\text{с}=$$
$$=(6{,}626176\pm 0{,}000036)\cdot 10^{-34}\text{ Дж}\cdot\text{с}. \tag{118.5а}$$

Если теперь выражение (118.4) подставить в (118.2), то и получится формула Планка

$$u_\omega=\frac{\hbar\omega^3}{\pi^2c^3}\frac{1}{e^{\hbar\omega/kT}-1}. \tag{118.6}$$

Формулу Планка обычно пишут в переменных ν, T:

$$u_\nu=\frac{8\pi h\nu^3}{c^3}\frac{1}{e^{h\nu/kT}-1}, \tag{118.7}$$

а также в переменных λ, T:

$$u_\lambda=\frac{8\pi hc}{\lambda^5}\frac{1}{e^{hc/\lambda kT}-1}. \tag{118.8}$$

При низких частотах, когда

$$\hbar\omega/kT\ll 1, \tag{118.9}$$

формула (118.6) переходит в формулу Рэлея–Джинса (117.8). В другом предельном случае высоких частот, когда

$$\hbar\omega/kT\gg 1, \tag{118.10}$$

получается формула

$$u_\omega=\frac{\hbar\omega^3}{\pi^2c^3}e^{-\hbar\omega/kT}. \tag{118.11}$$

К формуле такого вида в 1896 г. пришел Вин на основе некоторых произвольных допущений.

Для высоких частот (ультрафиолет) формула Вина прекрасно согласуется с опытом. Однако в области низких частот (инфракрасная область спектра) она дает совершенно неверные результаты, здесь применима формула Рэлея–Джинса. Первоначально Планк и искал эмпирическую формулу, которая бы при низких частотах совпадала с формулой Рэлея–Джинса и непрерывно переходила в формулу Вина в области высоких частот.

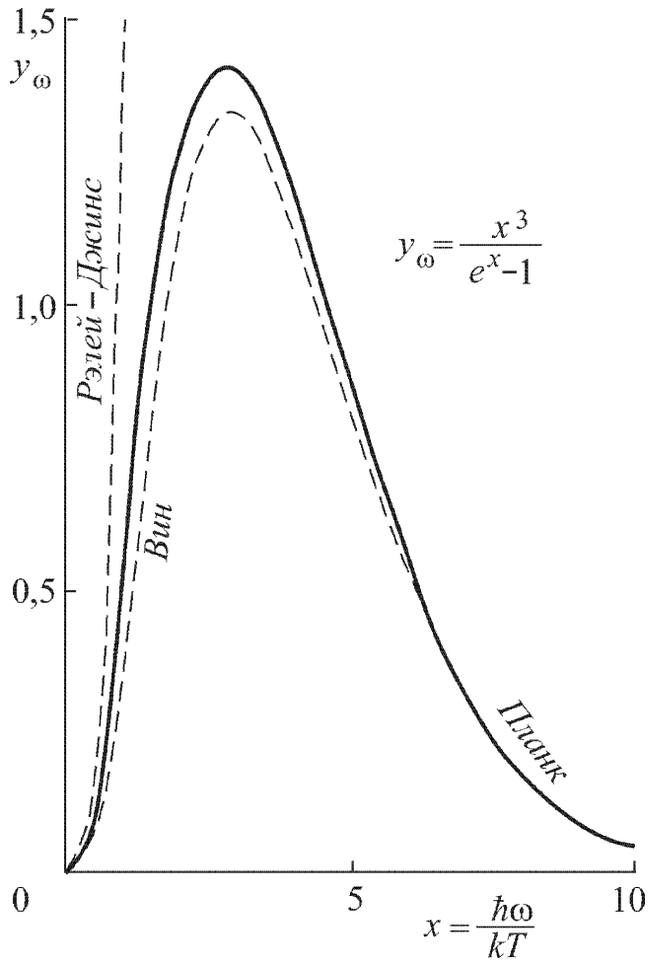


Рис. 343

Если ввести безразмерную переменную $x = \hbar\omega/kT$, то формулу (118.6) можно записать в виде

$$u_\omega = \frac{k^3T^3}{\pi^2c^3\hbar^2}\frac{x^3}{e^x-1}. \quad (118.6а)$$

При введении другой безразмерной переменной $x = \lambda kT/hc$ получится

$$u_\lambda = \frac{8\pi k^5T^5}{h^4c^4}\frac{1/x^5}{e^{1/x}-1}. \quad (118.8а)$$

Графики функций

$$y_\omega = \frac{x^3}{e^x-1} \quad \text{и} \quad y_\lambda = \frac{1/x^5}{e^{1/x-1}}$$

представлены на рисунках 343 и 344. Штриховые линии на тех же рисунках представляют те же функции, если их аппроксимировать по формулам Вина и Рэлея–Джинса.

3. Пользуясь формулой Планка, уточним значения постоянных в законах Стефана–Больцмана (115.4) и Вина (116.14), а также (116.15). Очевидно, что эти законы должны быть следствиями формулы Планка, так как последняя является частным случаем общей термодинамической формулы Вина (116.9). Согласно формуле Планка, интегральная плотность энергии равновесного излучения в вакууме равна

$$u = \frac{\hbar}{\pi^2c^3}\int\limits_0^\infty \frac{\omega^3\,d\omega}{\exp(\hbar\omega/kT)-1} = \frac{k^4T^4}{\pi^2c^3\hbar^3}\int\limits_0^\infty \frac{x^3e^{-x}\,dx}{1-e^{-x}}$$

(введена безразмерная переменная интегрирования $x = \hbar\omega/kT$).

Разложив знаменатель последнего интеграла $1 - e^{-x}$ в ряд и интегрируя, получим для последнего интеграла

$$\int_0^\infty x^3 e^{-x}(1 + e^{-x} + e^{-2x} + \ldots)\,dx = 6\left(1 + \frac{1}{2^4} + \frac{1}{3^4} + \ldots\right) = \frac{\pi^4}{15},$$

так как по известной формуле сумма последнего ряда в скобках равна $\pi^4/90$. Таким образом,

$$u = \frac{\pi^2 k^4}{15c^3\hbar^3}\,T^4 = \frac{8}{15}\,\frac{\pi^5 k^4}{c^3 h^3}\,T^4, \qquad (118.12)$$

т.е. получается закон Стефана–Больцмана (115.4), в котором постоянная a выражена через одни только фундаментальные постоянные c, h и k.

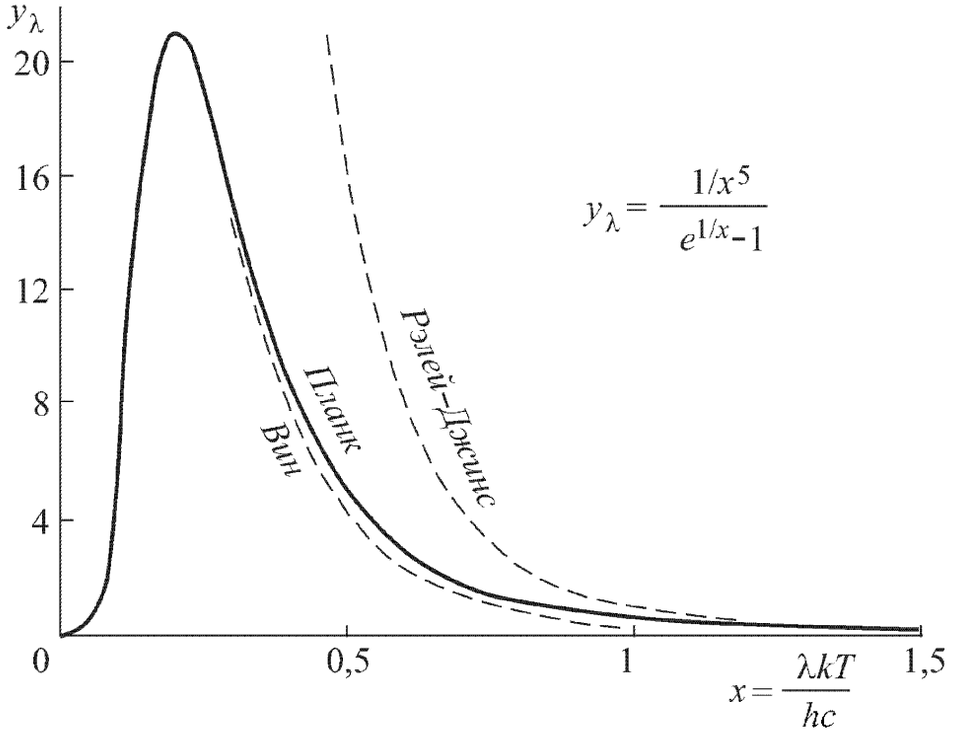


Рис. 344

На практике более удобна формула для энергетической светимости S излучающей абсолютно черной поверхности. Это есть интегральный лучистый поток, излучаемый наружу во всех направлениях (т.е. в телесный угол 2π) единицей площади такой поверхности в единицу времени. Она связана с яркостью B излучающей поверхности соотношением $S = \pi B = \pi I$ (см. § 22) или, ввиду формулы (112.5), $S = cu/4$. (Эта формула вполне аналогична выражению для среднего числа молекул газа, ударяющихся в единицу времени об единицу площади стенки сосуда, в который газ заключен, см. т. II, § 75.) Подставив сюда

выражения (118.12), получим

$$S = \sigma T^4, \tag{118.13}$$

где

$$\sigma = \frac{\pi^2 k^4}{60c^2\hbar^3} = \frac{2\pi^5 k^4}{15c^2 h^3} = (5{,}67032 \pm 0{,}00071) \cdot 10^{-8}\ \text{Вт}\cdot\text{м}^{-2}\cdot\text{К}^{-4}. \tag{118.14}$$

Величина σ называется *постоянной Стефана–Больцмана.*

4. Найдем теперь постоянную b в законе смещения Вина (116.14). Для этого надо найти значение $\lambda = \lambda_m$, для которого функция (118.8) при постоянном T обращается в максимум. Введем безразмерную переменную $\beta = hc/\lambda kT$ и выразим через нее функцию (118.8). Тогда, как легко убедиться, задача сводится к отысканию минимума функции $(e^\beta - 1)/\beta^5$. Приравняв нулю первую производную этой функции по β, получим уравнение

$$e^{-\beta} + \frac{\beta}{5} - 1 = 0, \tag{118.15}$$

корень которого $\beta = 4{,}9651142$. Поэтому

$$b = \lambda_m T = hc/(k\beta) = (2{,}897790 \pm 0{,}000090) \cdot 10^{-3}\ \text{м}\cdot\text{К}. \tag{118.16}$$

Если вместо λ пользоваться частотой ω, то закон Вина надо писать в виде (116.15). Тогда положение максимума, как нетрудно убедиться, будет определяться уравнением

$$(3 - \beta)e^\beta - 3 = 0, \tag{118.17}$$

где $\beta = \hbar\omega/kT = hc/(\lambda kT)$, т.е. β имеет тот же смысл, что и в предыдущем случае. Корень этого уравнения $\beta' = 2{,}8214393$. Если снова перейти к длинам волн, то максимуму функции u_ω соответствует длина волны λ'_m, определяемая условием

$$\lambda'_m T = hc/(k\beta'). \tag{118.18}$$

Таким образом, максимум на кривой частот сдвинут в длинноволновую сторону относительно максимума на кривой длин волн и притом так, что $\lambda'_m/\lambda_m = \beta/\beta' \approx 1{,}76$.

Заметим еще, что, измерив на опыте величины c, σ и b, можно по формулам (118.14) и (118.16) вычислить универсальные постоянные k и h, что и было впервые сделано Планком. После этого можно найти число Авогадро $N = R/k$ и элементарный заряд $e = F/N$, где R — универсальная газовая постоянная, а F — число Фарадея. Когда Планк производил эти вычисления, величины k, N и e были известны с малой точностью. Планк получил для них, а также для постоянной h значения, мало отличающиеся от современных.

5. Формулу Планка можно также получить, рассматривая равновесное излучение в полости как *фотонный газ*, к которому применима *статистика Бозе–Эйнштейна* (см. т. II, § 82).

Особенность этого газа состоит в том, что в результате взаимодействия с веществом фотоны могут рождаться и уничтожаться. Число их N в полости не остается постоянным. При равновесии оно устанавливается таким, что свободная энергия $F(T, V, N)$ при заданных T и V обращается в минимум, а потому $\partial F/\partial N = 0$. Но $\partial F/\partial N$ есть *химический потенциал* μ газа. Таким образом, для фотонов должно быть $\mu = 0$. Поэтому общая формула Бозе–Эйнштейна

$$\bar{n}_i = \frac{1}{\exp\left[(\mathcal{E}_i - \mu)/kT\right] - 1}, \tag{118.19}$$

определяющая среднее число частиц в i-м квантовом состоянии, в случае фотонов переходит в формулу

$$\bar{n} = \frac{1}{e^{\hbar\omega/kT} - 1}. \tag{118.20}$$

Число квантовых состояний в интервале частот $(\omega, \omega + d\omega)$ определяется выражением (117.7). Умножив $\bar{n}$ на это число, на энергию фотона $\hbar\omega$ и разделив на объем полости V, получим формулу Планка (118.6).

§ 119. Спонтанное и индуцированное излучение

1. В 1916 г. Эйнштейн дал новый вывод формулы Планка, основанный на представлениях Бора о механизме излучения. В этой работе было введено понятие *индуцированного излучения* — явления, на котором основан принцип действия лазера.

Пусть $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, \ldots$ — значения энергии, которые может принимать атом или вообще любая атомная система. Атом может самопроизвольно перейти из высшего энергетического состояния $\mathcal{E}_n$ в низшее $\mathcal{E}_m$ с испусканием света. Такое излучение называется *спонтанным*. Если атом находится в световом поле, то последнее может вызывать переходы как с высшего уровня $\mathcal{E}_n$ на низший $\mathcal{E}_m$, так и обратно с низшего $\mathcal{E}_m$ на высший $\mathcal{E}_n$. Первые переходы сопровождаются излучением света. Оно и называется *индуцированным* (*вынужденным*) излучением. Обратные переходы сопровождаются поглощением света атомом.

Имеются аналоги описанных явлений и в классической физике. Если атом рассматривать как колебательную систему, то в поле световой волны она будет совершать вынужденные колебания. В зависимости от соотношения фаз между колебаниями этой системы и светового поля амплитуда колебаний атома может как увеличиваться (поглощение света), так и уменьшаться (вынужденное излучение).

Эйнштейн применил к описанию процессов спонтанного и вынужденного излучения *вероятностные методы*. При этом

для проблемы равновесного излучения не имеет значения, присуща ли вероятность ансамблю физических объектов или самим элементарным законам, управляющим их поведением.

Рассмотрим теперь много одинаковых атомов в световом поле. Последнее будем предполагать изотропным и неполяризованным. Тогда отпадает вопрос о зависимости коэффициентов, вводимых ниже, от поляризации и направления излучения. Пусть N_m и N_n — числа атомов в состояниях $\mathcal{E}_m$ и $\mathcal{E}_n$, причем состояния $\mathcal{E}_m$ и $\mathcal{E}_n$ могут быть взяты какими угодно из ряда допустимых состояний $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, \ldots$ Среднее число переходов атомов из состояния $\mathcal{E}_n$ в состояние $\mathcal{E}_m$ в единицу времени из-за спонтанного излучения будет пропорционально исходному числу атомов N_n. Представим его в виде $A_n^m N_n$. Эйнштейн постулировал, что из-за индуцированного излучения среднее число переходов между теми же уровнями будет по-прежнему пропорционально N_n, а также спектральной плотности излучения $u(\omega_{mn})$ при частоте испускаемого света, соответствующей рассматриваемому переходу. Обозначим это число через $B_n^m N_n u(\omega_{mn})$. Аналогично, среднее число переходов с уровня $\mathcal{E}_m$ на уровень $\mathcal{E}_n$ из-за поглощения света представится как $B_m^n N_m u(\omega_{mn})$. Величины A_n^m, B_n^m, B_m^n называются *коэффициентами Эйнштейна.* Они являются характеристиками *только самого атома* и могут зависеть лишь от частоты ω_{mn}.

2. Допустим теперь, что поле излучения, в котором находятся атомы, равновесное и имеет температуру T. Тогда имеет место детальное равновесие, а потому

$$A_n^m N_n + B_n^m N_n u(\omega_{mn}) = B_m^n N_m u(\omega_{mn}). \tag{119.1}$$

Если уровни энергии $\mathcal{E}_m$ и $\mathcal{E}_n$ *простые*, а не кратные, то коэффициенты Эйнштейна связаны соотношением

$$B_m^n = B_n^m. \tag{119.2}$$

Действительно, будем повышать температуру системы. Коэффициенты Эйнштейна при этом меняться не будут, так как они от температуры не зависят. Спонтанное же излучение будет играть все меньшую и меньшую роль по сравнению с индуцированным. Если им пренебречь, то условие детального равновесия примет вид $B_n^m N_n = B_m^n N_m$. Но, согласно формуле Больцмана, при $T \to \infty$ населенности уровней N_n и N_m должны сравняться. Отсюда и следует, что $B_m^n = B_n^m$.

Допустим теперь, что уровень $\mathcal{E}_m$ состоит из g_m, а уровень $\mathcal{E}_n$ — из g_n слившихся простых энергетических подуровней. Такие уровни называются *кратными*, а целые числа g_m и g_n — их *кратностями.* Вероятность перехода атома с уровня $\mathcal{E}_m$ на каждый простой подуровень $\mathcal{E}_n$ меньше вероятности перехода на кратный уровень $\mathcal{E}_n$ в g_n раз, т.е. она равна uB_m^n/g_n. Аналогично, вероятность перехода с уровня $\mathcal{E}_n$ на простой подуровень $\mathcal{E}_m$

будет uB_n^m/g_m. Но по доказанному эти вероятности равны между собой. Поэтому для кратных уровней соотношение (119.2) заменится на $B_m^n/g_n = B_n^m/g_m$ или

$$g_m B_m^n = g_n B_n^m. \tag{119.3}$$

Так как спектральная плотность излучения $u(\omega)$ не зависит от того, какие атомы использованы для ее вычисления, то для простоты расчета энергетические уровни $\mathcal{E}_n$, $\mathcal{E}_m$ можно считать простыми. Тогда с учетом формулы Больцмана

$$\frac{N_m}{N_n} = \exp\frac{\mathcal{E}_n - \mathcal{E}_m}{kT}$$

из (119.1) получим

$$u(\omega_{mn}) = \frac{\alpha(\omega_{mn})}{\exp\left[(\mathcal{E}_n - \mathcal{E}_m)/(kT)\right] - 1}, \tag{119.4}$$

где $\alpha(\omega_{mn}) = A_n^m/B_n^m$. Сравнение этого выражения с формулой Вина (116.10) показывает, что разность $\mathcal{E}_n - \mathcal{E}_m$ должна быть линейной функцией частоты, т.е.

$$\mathcal{E}_n - \mathcal{E}_m = \hbar\omega_{mn}. \tag{119.5}$$

В предельном случае низких частот (или высоких температур)

$$u(\omega_{mn}) = \frac{\alpha(\omega_{mn})kT}{\hbar\omega_{mn}}.$$

Но в этом случае применима формула Рэлея–Джинса, так что

$$\alpha(\omega_{mn}) = \frac{\hbar\omega_{mn}^3}{\pi^2 c^3}. \tag{119.6}$$

Подставим теперь выражения (119.5) и (119.6) в формулу (119.4). При этом индексы m и n можно опустить, так как в нашем выводе уровни энергии $\mathcal{E}_m$ и $\mathcal{E}_n$ можно взять произвольными. Тогда получится формула Планка (118.6).

Из универсальности функции (118.6) следует, что и постоянная $\hbar$, введенная посредством соотношения (119.5), также универсальна. Формула (119.5), определяющая частоту излучаемого света при квантовых переходах между энергетическими уровнями, называется *правилом частот Бора.*

Заметим еще, что если бы не учитывать индуцированное излучение, т.е. положить $B_n^m u(\omega_{mn}) = 0$, то вместо формулы Планка получился бы ее предельный случай — формула Вина (118.11). Отсюда следует, что *формула Планка с неизбежностью приводит к заключению о существовании индуцированного излучения.* При низких температурах индуцированное излучение, очевидно, несущественно по сравнению со спонтанным. Вот почему в области низких температур, когда $\hbar\omega/kT \gg 1$, справедлива формула Вина.

ЗАДАЧИ

1. Определить среднее число фотонов n в единице объема полости, заполненной равновесным (черным) излучением, при температуре T.

О т в е т.

$$n = \frac{k^3T^3}{\hbar^3\pi^2c^3}\int\limits_0^\infty \frac{x^2e^{-x}}{1-e^{-x}}\,dx =$$

$$= \frac{2k^3T^3}{\hbar^3\pi^2c^3}\left(1+\frac{1}{2^3}+\frac{1}{3^3}+\ldots\right) \approx 1{,}202\frac{2k^3T^3}{\hbar^3\pi^2c^3} = 20{,}5T^3,$$

где T — температура в кельвинах.

2. В полости с зеркальными стенками находится изотропное излучение с средней плотностью n фотонов в единице объема. Определить среднее число фотонов z, ежесекундно ударяющихся об единицу площади стенки.

О т в е т. $z = \dfrac{1}{4}nc$.

3. Решить предыдущую задачу в предположении, что излучение равновесное, а температура стенок полости равна T. Какое среднее число фотонов N будет выходить ежесекундно из полости через отверстие в стенке полости площадью $s = 1$ см^2, если $T = 1000$ К?

О т в е т. $n \approx 2{,}404\dfrac{k^3T^3}{\hbar^3\pi^2c^3}$, $N = 0{,}6\dfrac{k^3T^3}{\hbar^3\pi^2c^2} \approx 1{,}5\cdot 10^{20}$.

4. В какой области спектра равновесного (черного) излучения при температуре $T = 300$ К интенсивность индуцированного излучения превосходит интенсивность спонтанного?

О т в е т. Если $\lambda \geqslant hc/(kT\ln 2)$, то интенсивность индуцированного излучения становится равной или большей интенсивности спонтанного. При $T = 300$ К получаем $\lambda \geqslant 692$ мкм.

5. При какой температуре равновесного (черного) излучения индуцированное излучение в видимой области ($\lambda = 550$ нм) превосходит спонтанное?

О т в е т. $T > hc/(k\lambda\ln 2) \approx 3{,}8\cdot 10^4$ К.

6. Среднее число фотонов в единице объема равновесного (черного) излучения, приходящееся на интервал частот $(\omega, \omega + d\omega)$ или на соответствующий ему интервал длин волн $(\lambda, \lambda + d\lambda)$, можно представить в виде $dN = f(\omega, T)\,d\omega = \varphi(\lambda, T)\,d\lambda$, где T — температура излучения. Найти положение максимумов функций $f(\omega, T)$ и $\varphi(\omega, T)$ при фиксированной температуре T.

О т в е т. Для функции $f(\omega, T)$ максимум получается при $h\omega/(kT) = x$, где x — корень уравнения $(x-2)e^x + 2 = 0$, т.е. $x = 1{,}593624$. Соответствующая длина волны λ_1 находится из соотношения

$$\lambda_1 T = \frac{2\pi\hbar c}{kx} = \frac{hc}{kx} = 0{,}902867 \text{ см·К}.$$

Для функции $\varphi(\lambda, T)$ максимум получается при $\lambda = \lambda_2$, причем $\dfrac{hc}{\lambda kT} = \xi$, где ξ — корень уравнения $(\xi - 4)e^\xi + 4 = 0$, т.е. $\xi = 3{,}920690$. Таким образом, $\lambda_2 T = 0{,}368967$ см·К. Следовательно, $\lambda_2 < \lambda_1$.

7. Равновесное излучение заключено в полости, стенки которой поддерживаются при постоянной (абсолютной) температуре T. Вычислить флуктуации энергии $\mathcal{E}$ такого излучения в объеме V в спектральном интервале

$(\omega, \omega + d\omega)$, пользуясь формулами 1) Вина, 2) Рэлея–Джинса, 3) Планка, и интерпретировать полученные результаты с точки зрения корпускулярных и волновых представлений о свете.

Р е ш е н и е. Применяя к (97.15) формулы Вина, Рэлея–Джинса и Планка, получим

$$\overline{\Delta\mathcal{E}^2} = \bar{\mathcal{E}}\hbar\omega \qquad \text{(Вин)}, \tag{119.7}$$

$$\overline{\Delta\mathcal{E}^2} = \bar{\mathcal{E}}kT = \frac{\pi^2 c^3}{V\omega^2\, d\omega}\,\overline{\mathcal{E}^2} \qquad \text{(Рэлей– Джинс)}, \tag{119.8}$$

$$\overline{\Delta\mathcal{E}^2} = \bar{\mathcal{E}}\hbar\omega + \frac{\pi^2 c^3}{V\omega^2\, d\omega}\,\overline{\mathcal{E}^2} \qquad \text{(Планк)}. \tag{119.9}$$

Формула (119.7) имеет такой же вид, что и формула (97.7) для флуктуации числа частиц идеального газа. Ее можно было бы получить из корпускулярных представлений, рассматривая излучение как газ независимых частиц. Напротив, формула (119.8) соответствует волновым представлениям о свете. Здесь флуктуации возникают из-за суперпозиции волн различных частот. Формула (119.9) соответствует синтезу обоих представлений.

8. Освещенность E, создаваемая звездой первой величины на поверхности Земли при нормальном падении света, составляет около 10^{-6} лк. Можно ли объяснить мерцание звезд квантовыми флуктуациями света?

Р е ш е н и е. Среднее число фотонов, попадающее от звезды в зрачок глаза в одну секунду,

$$\bar{N} = \frac{EAs}{h\lambda/c} \approx 6 \cdot 10^{34} \sim 10^{34},$$

где A — механический эквивалент света, площадь зрачка глаза s принята равной 0,5 см^2, а длина волны $\lambda = 550$ нм. В рассматриваемой области спектра с хорошим приближением можно пользоваться формулой Вина, а потому

$$\overline{\Delta N^2} = \bar{N}, \quad \frac{1}{\bar{N}}\sqrt{\overline{\Delta N^2}} \sim 10^{-17}.$$

Отсюда видно, что квантовые флуктуации света к мерцанию звезд не имеют отношения.

ГЛАВА XI

ЛАЗЕРЫ И НЕЛИНЕЙНАЯ ОПТИКА

§ 120. Принципы работы лазера

1. Спонтанное излучение некогерентно. В этом случае атомы источника излучают свет независимо друг от друга. Фазы волн, испускаемых различными атомами, их поляризация и направления распространения никак не связаны между собой. Обычные источники света — пламена, лампы накаливания, газоразрядные трубки, люминесцентные лампы и пр. — излучают некогерентно. В них свечение вызывается либо столкновениями между атомами, совершающими тепловое движение, либо электронными ударами. Правда, в таких источниках наряду со спонтанным происходит и индуцированное излучение. Однако оно возбуждается некогерентным спонтанным излучением, а потому и само некогерентно. Испускаемый свет характеризуется большей или меньшей степенью беспорядка. Максимальный беспорядок достигается в равновесном излучении в полости. В нем представлены всевозможные фазы и частоты, всевозможные направления колебаний, всевозможные направления распространения света. Если заимствовать терминологию из акустики и радиотехники, то можно сказать, что указанные источники света генерируют не правильные или упорядоченные волны, а шумы, пригодные только для освещения, грубой сигнализации, получения изображений, фотографирования и пр., но не для передачи речи, телевидения и т.д., осуществляющихся посредством радиоволн, излучаемых радиостанциями.

Однако можно создать и когерентно излучающие источники света, в которых бы различные атомы излучали волны согласованно, подобно радиостанциям, т.е. с одинаковыми частотами, фазами, поляризацией и направлением распространения. Такие источники открыли широкие возможности для разнообразных научных и технических применений. Они называются *оптическими квантовыми генераторами* или *лазерами*. Слово

«лазер» образовалось из первых букв полного английского названия «Light amplification by stimulated emission of radiation», что в переводе означает: усиление света посредством индуцированного излучения. Созданию лазеров предшествовало изобретение *мазеров*, т.е. усилителей микроволн, работающих также на принципе индуцированного излучения. Поэтому первоначально лазеры назывались *оптическими мазерами*. Подробное рассмотрение устройства и работы лазеров и мазеров дается в квантовой электронике. Здесь, в оптике, мы ограничимся только кратким изложением принципов работы лазеров.

Лазер работает на принципе *индуцированного излучения*. Допустим, что на атом падает фотон с энергией $\hbar\omega = \mathcal{E}_2 - \mathcal{E}_1$, где $\mathcal{E}_1$ и $\mathcal{E}_2$ — какие-либо два энергетических уровня атома. Если атом находится на нижнем уровне $\mathcal{E}_1$, то падающий фотон может поглотиться. Если же атом находится на верхнем уровне $\mathcal{E}_2$, то может произойти вынужденный переход на нижний уровень $\mathcal{E}_1$ с испусканием второго фотона. Индуцированно излученный фотон характеризуется не только *той же частотой* ω (как и при спонтанном излучении), но также *теми же фазой, поляризацией* и *направлением распространения*. Вместо одного падающего фотона получается *два тождественных фотона*. Эта особенность индуцированного излучения и используется в лазерах.

2. Рассмотрим теперь не единичный атом, а среду из атомов. Обозначим через N_1 и N_2 числа атомов в единице объема на уровнях $\mathcal{E}_1$ и $\mathcal{E}_2$ соответственно. Допустим, что в среде распространяется плоская монохроматическая волна, частота которой определяется условием $\hbar\omega = \mathcal{E}_2 - \mathcal{E}_1$. За время $dt = dx/v$, где v — скорость распространения, а dx — расстояние, пройденное волной, с нижнего уровня на верхний переходит в среднем $u(\omega)B_1^2N_1\,dt$ атомов и такое же число фотонов поглощается. Из-за индуцированного излучения с верхнего уровня на нижний перейдет $u(\omega)B_2^1N_2\,dt$ атомов и родится такое же число фотонов той же поляризации и направления распространения, что и у рассматриваемой волны. Фотоны, излученные спонтанно, а также фотоны, индуцированные другими волнами, можно не учитывать, так как среди них только ничтожная часть распространяется в нужном направлении и обладает нужной поляризацией. Увеличение числа фотонов в единице объема при прохождении волной расстояния $dx = v\,dt$ представится выражением

$$dN_{\text{фот}} = (B_2^1N_2 - B_1^2N_1)u(\omega)\frac{dx}{v}.$$

Коэффициенты Эйнштейна B_2^1 и B_1^2 связаны соотношением $g_2B_2^1 = g_1B_1^2$, где g_1 и g_2 — кратности уровней $\mathcal{E}_1$ и $\mathcal{E}_2$. Используя это соотношение, перепишем предыдущее уравнение в

виде

$$dN_{\text{фот}} = \left(\frac{N_2}{g_2} - \frac{N_1}{g_1}\right) g_1 B_1^2 u(\omega) \frac{dx}{v}. \tag{120.1}$$

Чтобы при распространении в среде волна усиливалась, необходимо выполнение условия

$$\frac{N_2}{g_2} > \frac{N_1}{g_1}. \tag{120.2}$$

Его можно записать в виде

$$n_2 > n_1, \tag{120.3}$$

где $n_1 = N_1/g_1$ и $n_2 = N_2/g_2$ — числа атомов на каждом из простых уровней, из которых состоят сложные уровни $\mathcal{E}_1$ и $\mathcal{E}_2$.

В обычных условиях, когда среда находится в *термодинамическом равновесии*, $n_2 < n_1$, т.е. *на каждом простом верхнем уровне находится меньше атомов, чем на нижнем.* Это непосредственно следует из формулы Больцмана

$$n = n_0 e^{-\mathcal{E}/kT}.$$

Можно искусственно получить *термодинамически неравновесную среду*, у которой выполняется соотношение, обратное (120.3) или (120.2). Такая среда называется *активной* или *средой с инверсной заселенностью по отношению к энергетическим уровням* $\mathcal{E}_1$ *и* $\mathcal{E}_2$. Следовательно, *для усиления световой волны необходимо, чтобы среда, в которой волна распространяется, была активной.* Идея использования индуцированного излучения для усиления волны была впервые высказана в 1939 г. в докторской диссертации В.А. Фабрикантом (р. 1907) и впоследствии (в 1951 г.) на нее было выдано авторское свидетельство. В то время на идею Фабриканта не было обращено должного внимания. Казалось, что создание систем с инверсной заселенностью энергетических уровней — дело бесперспективное.

3. Усиление света в активной среде обычно сравнивают с нарастанием лавины, изображая фотоны в виде шариков. Летящий фотон-шарик порождает второй фотон-шарик с переходом атома с верхнего уровня на нижний. Получаются два одинаковых шарика, летящих в прежнем направлении, затем четыре шарика и т.д. Но эта грубая иллюстрация не объясняет, как в результате наложения фотонов формируется монохроматическая волна строго определенного направления. Эта сторона дела становится понятной, если сравнить изучаемое нами явление с классической картиной распространения плоской монохроматической волны в однородной среде. Волна вызывает колебания в атомах и молекулах среды. Последние переизлучают шаровые волны, когерентные друг с другом и с падающей волной. Эти шаровые волны, интерферируя между собой, создают снова плоский вол-

новой фронт, распространяющийся в среде. Они влияют только на фазовую скорость волны. Если среда абсолютно прозрачна, то амплитуда волны должна оставаться постоянной, как того требует закон сохранения энергии. В поглощающих средах энергия волны частично переходит в тепло — амплитуда волны убывает. Но в активной среде молекулы и атомы находятся в возбужденных состояниях. За счет энергии возбуждения вторичные световые волны, излучаемые молекулами и атомами, усиливаются. Однако их фазы и поляризация остаются прежними. Поэтому остаются прежними поляризация и фаза также и результирующей волны, возникающей в результате интерференции таких вторичных волн. Усиливается только ее амплитуда.

4. Индуцированное излучение было использовано для *генерации когерентных световых волн*. Идея этого впервые была высказана в 1957 г. А.М. Прохоровым (р. 1916) и Н.Г. Басовым (р. 1922) и независимо от них Ч. Таунсом (р. 1915). Чтобы активное вещество превратить в генератор световых колебаний, надо осуществить *обратную связь*. Необходимо, чтобы часть излученного света все время находилась в зоне активного вещества и вызывала вынужденное излучение все новых и новых атомов. Для этого активное вещество помещают между двумя параллельными зеркалами. Допустим, например, что оно представляет собой цилиндр, а плоскости зеркал S_1 и S_2 перпендикулярны к оси этого цилиндра (рис. 345). Тогда луч света, претерпевая многократные отражения от зеркал S_1 и S_2, будет

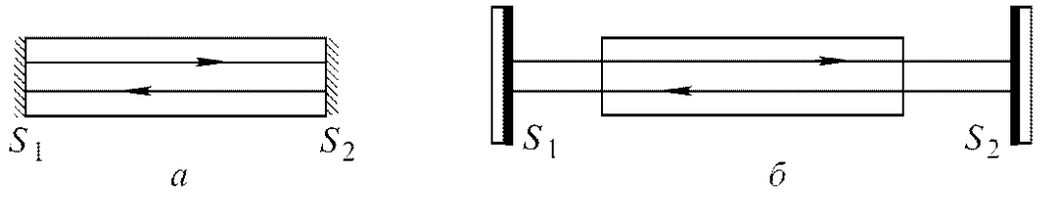


Рис. 345

проходить много раз через активное вещество, усиливаясь при этом в результате вынужденных переходов атомов с высшего энергетического уровня $\mathcal{E}_2$ на более низкий уровень $\mathcal{E}_1$. Получается *открытый резонатор*, представляющий собой в сущности интерферометр Фабри–Перо, только заполненный активной средой.

Такой резонатор будет *не только усиливать свет*, но также *коллимировать* и *монохроматизировать* его. Для простоты предположим сначала, что зеркала S_1 и S_2 идеальные. Тогда лучи, параллельные оси цилиндра, будут проходить через активное вещество туда и обратно неограниченное число раз. Все же лучи, идущие наклонно, в конце концов попадут на боковую стенку цилиндра, где они рассеются или выйдут наружу. Ясно

поэтому, что максимально усилятся лучи, распространяющиеся параллельно оси цилиндра. Этим и объясняется коллимация лучей. Конечно, строго параллельные лучи получить нельзя. Этому препятствует дифракция света. Угол расхождения лучей принципиально не может быть меньше *дифракционного предела* $\delta\vartheta \approx \lambda/D$, где D — ширина пучка. Однако в лучших газовых лазерах такой предел практически достигнут.

5. Объясним теперь, как происходит *монохроматизация света*. Для простоты проведем рассуждение применительно к рис. 345 *а*, когда роль зеркал S_1 и S_2 выполняют отполированные и посеребренные торцы цилиндра активного вещества, перпендикулярные к его геометрической оси. Пусть L — длина цилиндра. Если $2L = m\lambda$, т.е. на длине L укладывается целое число полуволн m, то световая волна, выйдя от S_1, после прохождения через цилиндр туда и обратно вернется к S_1 в той же фазе. Такая волна усилится при втором и всех следующих прохождениях через активное вещество в прямом и обратном направлениях. Ближайшая длина волны $\lambda \pm \Delta\lambda$, для которой должно происходить такое же усиление, найдется из условия $2L = (m \pm 1)(\lambda \mp \Delta\lambda)$. Следовательно, $\Delta\lambda = \lambda/m$, т.е. $\Delta\lambda$, как и следовало ожидать, совпадает со спектральной областью интерферометра Фабри–Перо.

Учтем теперь, что энергетические уровни $\mathcal{E}_1$ и $\mathcal{E}_2$ и спектральные линии, возникающие при переходах между ними, не бесконечно тонкие, а имеют *конечную ширину*. Предположим сначала, что ширина спектральной линии, излучаемой атомами, меньше дисперсионной области прибора. Тогда из всех длин волн, излучаемых атомами, условию $2L = m\lambda$ может удовлетворять только одна длина волны λ. Такая волна усилится максимально. Это и ведет к сужению спектральных линий, генерируемых лазером, т.е. к монохроматизации света.

Степень монохроматизации нетрудно определить. Пусть свет проходит через активное вещество туда и обратно N раз. Для длины волны λ имеем $2LN = Nm\lambda$. Возьмем ближайшую длину волны λ', удовлетворяющую условию $2LN = (Nm \pm 1)\lambda'$. Для такой длины волны каждый цуг волн, возникший при прохождении через активное вещество туда и обратно, будет отличаться по фазе от предыдущего и последующего цугов на $\pm 2\pi/N$. В результате все N цугов погасят друг друга — получится минимум интенсивности для λ'. Отсюда ясно, что ширина спектральной линии, усиливаемой лазером, будет $\delta\lambda = |\lambda - \lambda'| \sim \lambda/(Nm)$, т.е. она определяется разрешающей силой прибора. При $N \to \infty$ получится $\delta\lambda \to 0$, т.е. бесконечно тонкая спектральная линия. В действительности из-за неидеальности отражающих поверхностей зеркал $\delta\lambda$ остается конечной. Однако при хороших отражающих поверхностях лазер дает очень тонкую, практически мо-

нохроматическую линию. Допустим теперь, что спектральные линии, излучаемые атомами активной среды, шире дисперсионной области прибора $\Delta\lambda$. В этом случае вместо усиления одной спектральной линии может возникнуть усиление нескольких линий. Свет лазера будет состоять из близких, практически монохроматических линий.

6. В приведенных рассуждениях не принято во внимание, что в резонаторе Фабри–Перо могут усиливаться волны, распространяющиеся не только параллельно его оси, но и *под малыми углами* φ к ней. Условие усиления имеет вид $2L\cos\varphi = m\lambda$. Это, как и во всяком объемном резонаторе, создает систему стоячих волн с узлами и пучностями. Каждой такой волне соответствует определенный тип или, как принято говорить, *мода колебаний*. Это обстоятельство отражается на распределении интенсивности в поперечном сечении пучка лучей, генерируемых лазером. Здесь наблюдается система светлых пятен, разделенных темными узловыми линиями. Это, конечно, усложняет дело. Но мы не будем входить в подробное обсуждение этого вопроса. Заметим только, что наряду с плоскими зеркалами в резонаторах употребляются длиннофокусные *сферические зеркала*. Они могут быть, например, вогнутыми конфокальными зеркалами, т.е. зеркалами, у которых главные фокусы совпадают и находятся в середине системы. Применение подобных зеркал, как показывают опыт и численные расчеты, позволяет значительно уменьшить потери света в лазерах и упрощает их юстировку.

7. Учтем теперь, что в реальном лазере часть света, чтобы ее можно было использовать, должна быть выпущена из активной среды наружу. С этой целью одно из зеркал, например S_2, делается *полупрозрачным*. Кроме того, и зеркало S_1 лишь частично отражает свет, хотя коэффициент отражения его и близок к 100 %. Это приводит к ослаблению светового пучка. Чтобы лазер был генератором света, необходимо, чтобы усиление светового пучка в активной среде превосходило некоторое минимальное — *пороговое* — значение. Именно, должно быть выполнено следующее условие. Световой пучок, вышедший от S_1, после прохождения туда и обратно через активную среду и отражения от S_1 должен вернуться в исходное положение с неменьшей интенсивностью. Иначе в результате последовательного повторения этих процессов интенсивность пучка будет непрерывно убывать и лазер перестанет генерировать. Поэтому для генерации недостаточно выполнения простого неравенства (120.3). Оно должно быть выполнено с некоторым запасом, т.е. число атомов N_2 на верхнем уровне в единице объема активной среды должно превышать некоторое минимальное — пороговое — значение.

Конечно, нарастание интенсивности волны в активной среде не может продолжаться беспредельно, так как заселенность

верхнего энергетического уровня ограничена. По мере обеднения атомами верхнего уровня $\mathcal{E}_2$ скорость нарастания интенсивности волны будет уменьшаться и волна начнет затухать еще до того, как перестанет выполняться условие (120.3).

§ 121. Рубиновый лазер

1. Существует много материалов, применяемых в качестве активных сред в лазерах. Сюда относятся различные диэлектрические кристаллы, стекла, газы, полупроводники и даже плазма. Эта сторона дела подробно изучается в квантовой электронике. Мы ограничимся описанием устройства и принципа действия только двух лазеров: *рубинового* и *гелий-неонового*.

Первый квантовый генератор света был создан в 1961 г. Мейманом (р. 1927) на рубине. Рубин — это твердый кристалл, основой которого является корунд, т.е. кристалл окиси алюминия (Al_2O_3), в котором небольшая часть атомов алюминия (около 0,05 %) замещена ионами хрома Cr^{+++} [1]). Ионы хрома играют основную роль в работе квантового генератора. Корунд — это диэлектрик с широкой запрещенной энергетической зоной между валентной зоной и зоной проводимости (см. т. III, § 100). Энергетические уровни хрома в корунде лежат в этой запрещенной зоне. Они были изучены задолго до создания лазеров. Основным (невозбужденным) уровнем является уровень $\mathcal{E}_1$ (рис. 346). Он имеет сложную структуру, которая, однако, не играет роли в работе квантового генератора. Выше лежат близкие возбужденные уровни $\mathcal{E}_{2а}$ и $\mathcal{E}_{2б}$. Это — узкие уровни. При переходе с них на основной уровень $\mathcal{E}_1$ излучается красный свет с длинами волн $\lambda = 694{,}3$ и 692,9 нм. Он-то и придает рубину характерную розовую или

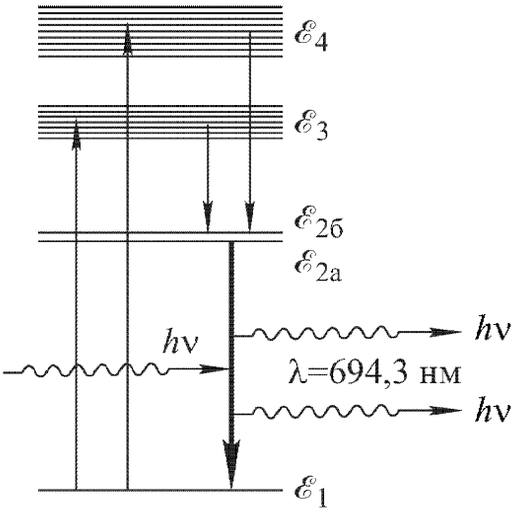


Рис. 346

[1]) Для лазеров монокристаллы рубина выращивают искусственно. Смесь Al_2O_3 и Cr_2O_3 в виде пудры сыплется сверху на выращиваемый кристалл рубина, верхняя кромка которого находится в кислородно-водородном пламени горелки с температурой 2050 °С, достаточной для плавления рубина. При медленном опускании кристалла расплавленный слой смеси выходит из пламени и кристаллизуется. Таким путем удается получать монокристаллы рубина в виде стержней длиной до 0,5 м и диаметром до 5 см.

красную окраску (в зависимости от концентрации ионов хрома). Более интенсивна линия $\lambda = 694{,}3$ нм. По этой причине только она и усиливается при работе лазера. Вторая линия не играет роли. Выше уровней $\mathcal{E}_{2а}$ и $\mathcal{E}_{2б}$ расположены две сравнительно широкие полосы энергий $\mathcal{E}_3$ и $\mathcal{E}_4$. Переходы между уровнями этих полос и основным уровнем $\mathcal{E}_1$ сопровождаются излучением зеленого и голубого света соответственно.

2. Инверсная заселенность создается между уровнями $\mathcal{E}_1$ и $\mathcal{E}_2$. Для этого используется так называемая *оптическая накачка*, т.е. освещение кристалла рубина мощной вспышкой света. Рубину придают форму цилиндрического стержня с диаметром 0,1–2 см и длиной от 2 до 20 см и больше. Концы стержня тщательно отполированы. Они могут служить зеркалами. Тогда их серебрят, как показано на рис. 345 *а*. Зеркала могут быть и внешними (рис. 345 *б*), тогда серебрение не нужно. Для освещения рубинового стержня применяют импульсные ксеноновые газоразрядные *лампы-вспышки*, через которые разряжаются батареи высоковольтных конденсаторов (напряжение 2–3 кВ). Длительность вспышки порядка одной миллисекунды. Лампа-вспышка имеет форму спиральной трубки, обвивающейся вокруг рубинового стержня (рис. 347). Она может быть и прямолинейной. Тогда применяют зеркальные осветители, имеющие форму эллиптических цилиндров с внутренними отражающими поверхностями. Лампа-вспышка располагается вдоль одной из фокальных линий цилиндра; отраженный свет концентрируется на рубиновом стержне, помещаемом вдоль другой фокальной линии.

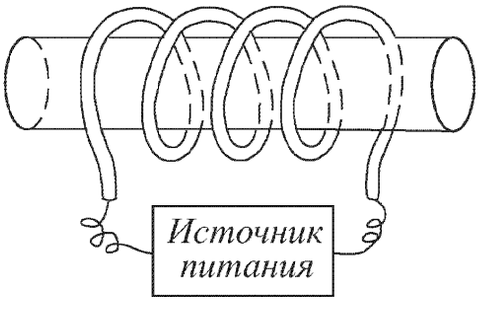


Рис. 347

Если бы энергетический спектр состоял только из двух уровней $\mathcal{E}_1$ и $\mathcal{E}_2$, то с помощью оптической накачки создать инверсную заселенность их было бы невозможно. Действительно, как видно из (119.2), индуцированные переходы атомов с нижнего уровня на верхний и обратно идут с одинаковыми скоростями. Поэтому лампа-вспышка самое большее могла бы лишь уравнять заселенности обоих уровней. Наличие же спонтанного излучения приводило бы к обеднению (атомами) верхнего уровня по сравнению с нижним.

Положение меняется благодаря наличию третьего, широкого энергетического уровня, состоящего из полос $\mathcal{E}_3$ и $\mathcal{E}_4$. Лампа-вспышка переводит атомы хрома из невозбужденного состояния в возбужденное, т.е. в энергетические полосы $\mathcal{E}_3$ и $\mathcal{E}_4$. Значительная ширина этих полос имеет большое практическое значение.

Лампа-вспышка излучает свет, близкий к белому. Если бы уровни $\mathcal{E}_3$ и $\mathcal{E}_4$ были очень узкими, то лишь ничтожная часть энергии лампы-вспышки могла бы быть использована на их возбуждение. Благодаря же значительной ширине полос $\mathcal{E}_3$ и $\mathcal{E}_4$ на их возбуждение уходит 10–15 % лучистой энергии лампы-вспышки. На уровнях $\mathcal{E}_3$ и $\mathcal{E}_4$ возбужденные атомы хрома живут очень короткое время ($\sim 10^{-8}$ с). За это время они переходят на один из уровней $\mathcal{E}_2$. При таком переходе атомы хрома не излучают, а расходуют свою энергию на возбуждение колебаний кристаллической решетки. Возможность обратного возвращения атома из полос $\mathcal{E}_3$ и $\mathcal{E}_4$ на уровень $\mathcal{E}_1$, хотя и существует, но вероятность такого процесса пренебрежимо мала по сравнению с вероятностью перехода на уровни $\mathcal{E}_2$. Уровни $\mathcal{E}_2$ метастабильны. Время жизни атома на них $\sim 10^{-3}$ с, что по атомным масштабам очень велико. Это позволяет накапливать атомы на уровнях $\mathcal{E}_2$. Если переводить атомы с уровня $\mathcal{E}_1$ на уровни $\mathcal{E}_3$ и $\mathcal{E}_4$ достаточно быстро, то на уровни $\mathcal{E}_2$ можно перевести больше половины атомов хрома. Тогда на уровнях $\mathcal{E}_2$ окажется больше атомов, чем на уровне $\mathcal{E}_1$, т.е. возникнет инверсная заселенность этих уровней.

3. Подсчитаем мощность лампы-вспышки, необходимую для создания инверсной заселенности уровней $\mathcal{E}_1$ и $\mathcal{E}_2$. Средняя длина волны, излучаемая при переходе из полос $\mathcal{E}_3$ и $\mathcal{E}_4$ на уровень $\mathcal{E}_1$, порядка 450 нм. Поэтому для перевода одного атома хрома с уровня $\mathcal{E}_1$ на уровни $\mathcal{E}_3$ и $\mathcal{E}_4$ требуется в среднем энергия $h\nu = hc/\lambda \approx 4{,}4 \cdot 10^{-12}$ эрг. В рубине, обычно употребляемом для лазеров, на каждый см^3 приходится около 10^{19} атомов хрома. Не менее половины из них надо перевести в полосы $\mathcal{E}_3$ и $\mathcal{E}_4$. На это требуется энергия $\approx 2{,}2 \cdot 10^7$ эрг $= 2{,}2$ Дж. Атомы должны быть переведены на возбужденные уровни $\mathcal{E}_3$ и $\mathcal{E}_4$ за время, не превышающее времени жизни атома хрома на уровне $\mathcal{E}_2$ ($\sim 2 \cdot 10^{-3}$ с). В противном случае инверсная заселенность получиться не может. Таким образом, минимальная мощность, идущая на возбуждение атомов, должна быть порядка $2{,}2/10^{-3} = 2200$ Вт/см^3 $= 2{,}2$ кВт/см^3. Если коэффициент использования лучистой энергии лампы принять равным 10–15 %, то потребляемая мощность лампы должна быть ~ 20 кВт/см^3. Для рубинового стержня с объемом 10 см^3 получится ~ 200 кВт. Это достигается сравнительно легко. Нетрудно подсчитать, что при напряжении 2000 В емкость батареи конденсаторов должна быть не меньше 100 мкФ. Обычно употребляются батареи с емкостью ~ 1000 мкФ.

Значительная доля лучистой энергии (более 50 %), поглощенной рубиновым стержнем, тратится на его нагревание. Но при температурах, превышающих примерно 1000 К, рубиновый ла-

зер перестает генерировать. Поэтому в некоторых конструкциях лазеров предусмотрено охлаждение рубина проточной водой и даже жидким азотом.

4. Рубин — *одноосный кристалл.* Обыкновенный и необыкновенный показатели преломления светло-красного рубина, употребляемого в лазерах, равны соответственно $n_o = 1{,}7653$ и $n_e = 1{,}7573$ (для $\lambda = 656$ нм). Рубиновый лазер может давать поляризованный свет без каких-либо специальных поляризационных приспособлений. Для этого оптическая ось рубинового стержня не должна совпадать с его геометрической осью. Возникновение линейной поляризации нельзя объяснить различием коэффициентов отражения обыкновенного и необыкновенного лучей, так как разность $n_o - n_e$ слишком мала. Поляризация возникает потому, что в обыкновенной волне направления волновой нормали и луча совпадают между собой, а в необыкновенной не совпадают. Чтобы зеркала лазера действовали эффективно, как резонатор, необходимо, чтобы волновые нормали были к ним перпендикулярны. Но тогда в рубине только обыкновенный луч будет распространяться параллельно геометрической оси стержня, а необыкновенный пойдет под углом к ней, попадая на боковую поверхность стержня. Поэтому резонатор будет эффективнее усиливать обыкновенные лучи, чем необыкновенные. Из лазера будут преимущественно выходить обыкновенные лучи, в которых электрический вектор перпендикулярен к оптической и геометрической осям кристалла. Если эти оси совпадают, то поляризация не возникает. Если же, как показали наблюдения, угол между ними заключен в пределах от 60 до 90°, то линейная поляризация полная.

5. Наиболее распространенным является *импульсный режим* работы рубинового лазера. Лампа-вспышка дает импульс света длительностью $\sim 10^{-3}$ с. Длительность когерентного импульса, излучаемого лазером, несколько короче. Это связано, во-первых, с тем, что требуется некоторое время, чтобы заселенность уровня $\mathcal{E}_2$ достигла порогового значения, после чего лазер начинает генерировать. Во-вторых, с тем, что генерация лазера прекращается, когда интенсивность вспышки лампы понижается настолько, что ее становится уже недостаточно для поддержания заселенности выше пороговой. Сам лазерный импульс имеет сложную структуру и состоит из множества отдельных импульсов длительностью до 10^{-6} с, следующих друг за другом с интервалом $\sim 3 \cdot 10^{-6}$–10^{-5} с. Мощность рубинового лазера в импульсе может достигать десятков кВт при длине стержня в 20–30 см и диаметре 1,5 см.

Эта мощность может быть повышена до нескольких десятков мегаватт за счет уменьшения длительности генерируемого

импульса. Для этого на уровень $\mathcal{E}_2$ надо перевести не половину, а большую часть атомов хрома, оставив уровень $\mathcal{E}_1$ практически незаполненным. Тогда при длине рубинового стержня 10–20 см импульс генерируемого света, как показывает расчет, может развиться за время $\sim 10^{-8}$ с. Лампа-вспышка легко могла бы произвести такой переход, если бы в рубине не возникала генерация, возвращающая атомы хрома снова на исходный уровень $\mathcal{E}_1$. А генерация начинается сразу же, как только заселенность уровня $\mathcal{E}_2$ станет выше пороговой. Задачу можно решить, если на короткое время выключить обратную связь, осуществляемую зеркалами. За это время можно перевести все атомы хрома на возбужденный уровень $\mathcal{E}_2$. При выключенных зеркалах среднее время жизни атома хрома на уровне $\mathcal{E}_2$ составляет около 0,002 с. Если в течение этого времени на мгновение снова включить зеркала, то происходит почти мгновенный вынужденный переход атомов на невозбужденный уровень $\mathcal{E}_1$ и связанная с ним мощная вспышка лазерного излучения.

6. Быстрое включение и выключение обратной связи, а также вообще изменение добротности резонатора можно осуществить, заменив одно из зеркал призмой P полного отражения (рис. 348), вращающейся с угловой скоростью ~ 500 об/с. Лазер может генерировать только в положении призмы, изображенном на рис. 348, или очень близком к нему. Во всех остальных положениях обратной связи нет и лазер не генерирует. В это время происходит оптическая накачка рубинового стержня R. Когда призма займет положение, указанное на рис. 348, происходит излучение мощного лазерного импульса. Хотя при этом призма и продолжает быстро вращаться, но за короткое время $\sim 10^{-8}$ с, требующееся для развития лазерного импульса, она не успевает сколько-нибудь заметно сместиться из указанного положения. После высвечивания запасенной энергии при дальнейшем вращении призмы снова начинается оптическая накачка, пока при повороте на 360° не произойдет излучение нового лазерного импульса. Однако при прохождении через стержень мощного импульса выключение призмы не всегда достаточно для прекращения действия обратной связи, так как она может сохраняться из-за отражения света от концов рубинового стержня (для рубина коэффициент отражения $(n-1)^2/(n+1)^2 \approx 7{,}6\,\%$). Во избежание этого концы рубинового

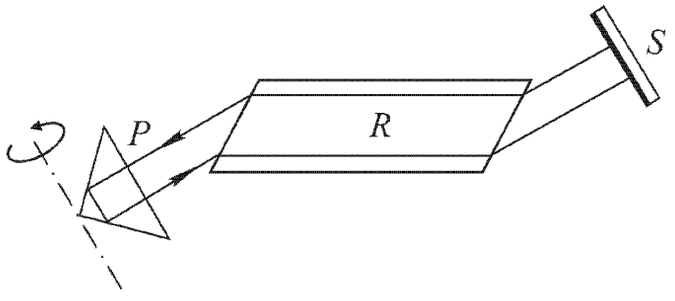


Рис. 348

стержня делают скошенными под углом к его геометрической оси. Обычно угол падения берут равным *углу Брюстера* (на границе рубин–воздух $\varphi_B \approx 30°$). Тогда генерируется линейно поляризованный свет, электрический вектор которого лежит в плоскости падения.

7. Более быстрое включение и выключение обратной связи производится с помощью оптического затвора — *ячейки Керра* с нитробензолом или *ячейки Поккельса* с анизотропным кристаллом (рис. 349). Ячейка вместе со своим конденсатором повернута вокруг оси установки на угол 45° (это на рис. 349 не показано). Поляризатор P (поляроид или поляризационная призма) ориентирован так, чтобы полностью пропускать линейно поляризованное излучение рубинового стержня, когда он генерирует. Перед включением лампы-вспышки на ячейку K подается такое напряжение, чтобы она поворачивала плоскость поляризации излучения лазера на угол 90°. Тогда свет через поляризатор P не пройдет, т.е. обратная связь будет выключена. В то время, когда лампа дает вспышку, происходит оптическая накачка рубинового стержня. Если затем быстро снять напряжение с конденсатора ячейки, то линейно поляризованное излучение рубина начнет свободно распространяться между зеркалами S_1 и S_2 и возникает мощный импульс индуцированного излучения лазера длительностью порядка 10^{-8} с.

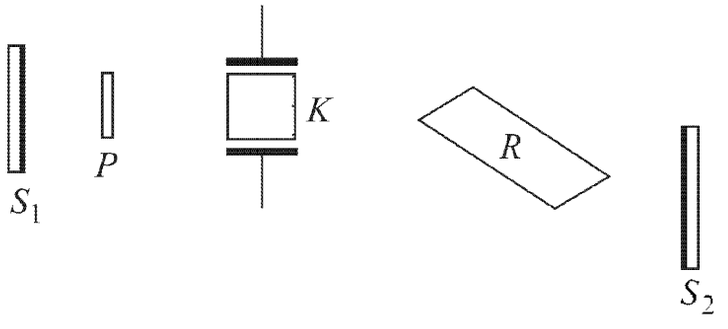


Рис. 349

8. Большое распространение в качестве оптических затворов получили также *просветляющиеся фильтры*, помещаемые вместо ячейки Керра или Поккельса. (Поляризатор P в этом случае не нужен.) Их действие основано на увеличении прозрачности вещества, когда интенсивность света становится достаточно большой, как это имеет место в случае излучения лазеров (см. § 89, п. 6). При малой интенсивности света фильтр поглощает свет, почти полностью устраняя обратную связь. С увеличением заселенности верхнего уровня возникает слабая генерация рубинового стержня, несколько уменьшающая поглощение фильтра. Это приводит к усилению обратной связи и вызывает лавину лазерного излучения. Последняя по мере нарастания все более и более просветляет фильтр. Когда интенсивность излучения начнет уменьшаться, поглощение фильтра будет быстро возрастать, а обратная связь ослабляться. Поскольку вся система работает автоматически, лазер с просветляющимися фильтрами

во время вспышки лампы накачки может генерировать серию импульсов, следующих друг за другом.

9. Быстрое включение обратной связи позволяет сократить длительность импульса лазерного излучения до 10^{-7}–10^{-8} с. Так как энергия импульса, снимаемая с рубинового стержня длиной 20 см и диаметром 1,5 см, составляет 1–2 Дж, то при этом развивается мощность 10–200 МВт. Дальнейшее повышение мощности лазерного импульса может быть достигнуто путем усиления света в каскаде из нескольких последовательно расположенных лазеров. Первый лазер в таком каскаде должен служить *генератором*, остальные — *усилителями света*. Если мощность генерируемого импульса достаточно велика, то уже небольшой его части вблизи передового фронта достаточно, чтобы вызвать вынужденные переходы в невозбужденное состояние всех возбужденных атомов хрома в рубине усилителей. Это сокращает длительность импульса и повышает его мощность. Таким путем удалось получить *гигантские импульсы света* длительностью в $2 \cdot 10^{-9}$ с при полной энергии импульса ~ 20 Дж. Это соответствует мощности $\sim 10^4$ МВт. Но и такая мощность еще не предел. Лазер с просветляющимся фильтром дает импульс длительностью $\sim 10^{-8}$ с, состоящий из последовательности импульсов, длительность которых может составлять 10^{-1}–10^{-12} с. Выделение одного такого импульса с последующим усилением его позволяет достигнуть мощности $\sim 10^7$ МВт.

10. Из других твердых материалов, употребляемых в лазерах, надо упомянуть прежде всего *неодимовое стекло* и *флюорит кальция* (CaF_2) с небольшой примесью атомов редкоземельных элементов: *диспрозия* (Dy), *самария* (Sm) и пр. Активность среды создается ионами неодима Nd^{+++}, диспрозия Dy^{++}, самария Sm^{++} и т.д. Неодимовое стекло генерирует инфракрасное излучение с длиной волны $\lambda = 1{,}06$ мкм, а флюорит кальция с диспрозием — еще более длинноволновое излучение с $\lambda = 2{,}36$ мкм. Стержни из неодимового стекла изготовить легче, чем из рубина. Стеклянные стержни могут достигать длины 0,5–1 м при диаметре до 5 см. Ширина полос у неодимового стекла больше, чем у рубина, что позволяет получить больший КПД. Зато у рубина более высокая прочность. При длительности импульса 10^{-8}–10^{-9} с рубин выдерживает мощности $(3\text{–}4)\cdot 10^3$ МВт на каждый $см^2$ поперечного сечения стержня, а неодимовое стекло — только $(1\text{–}1{,}5)\cdot 10^3$ МВт.

В отличие от рубинового лазера, работающего по *трехуровневой схеме*, лазер на флюорите кальция с примесью диспрозия работает по *четырехуровневой схеме*. Это позволяет создавать достаточно мощные лазеры, работающие *не импульсами*, *а*

непрерывно, что для некоторых применений очень существенно. Между основным уровнем $\mathcal{E}_1$ диспрозия в кристалле и энергетической полосой $\mathcal{E}_2$ расположены два промежуточных уровня $\mathcal{E}_3$ и $\mathcal{E}_4$ (рис. 350). Между ними и создается инверсная заселенность атомов диспрозия. Лампа-вспышка переводит атомы диспрозия через широкую энергетическую полосу $\mathcal{E}_2$ на уровень $\mathcal{E}_3$. Уровень $\mathcal{E}_4$ по энергии приподнят над уровнем $\mathcal{E}_1$ настолько высоко, что при температуре жидкого азота его заселенность N_4 практически равна нулю. Поэтому нет необходимости переводить на уровень $\mathcal{E}_3$ больше половины атомов диспрозия. Достаточно лишь, чтобы разность заселенностей $N_3 - N_4$ уровней $\mathcal{E}_3$ и $\mathcal{E}_4$ превысила пороговое значение. Оно составляет лишь небольшую долю атомов диспрозия. Достаточна сравнительно небольшая мощность лампы-накачки (≈ 15–20 Вт), чтобы осуществить непрерывную работу лазера. Кроме того, непрерывная накачка позволяет генерировать гигантские импульсы с большой частотой повторения (до нескольких кГц). Мощности света, получаемые при этом в каждом импульсе, около 1–2 МВт.

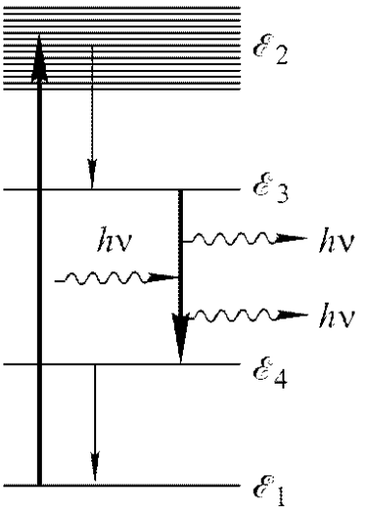


Рис. 350

ЗАДАЧИ

1. Мощность рубинового лазера в импульсе равна $W = 100$ МВт. Площадь поперечного сечения рубинового стержня $S = 1$ см^2. Оценить напряженность E электрического поля и давление света в лазерном пучке, считая его строго плоскопараллельным ($\lambda = 694{,}3$ нм). Как изменятся эти величины, если пучок сфокусировать идеальной линзой с фокусным расстоянием $f = 5$ см?

Р е ш е н и е. Напряженность электрического поля E_0 в несфокусированном пучке оценим по формуле

$$W = \frac{c}{4\pi}\,\overline{EH}S = \frac{c}{4\pi}\,\overline{E^2}S,$$

а давление излучения $\mathcal{P}_0$ — по формуле $\mathcal{P}_0 = W/(cS)$. Таким путем находим

$$E_0 \approx \sqrt{\overline{E^2}} = \sqrt{\frac{4\pi W}{cS}} = 647 \text{ СГСЭ} = 1{,}94 \cdot 10^5 \text{ В/см},$$

$$\mathcal{P}_0 \approx 3{,}3 \cdot 10^4 \text{ дин/см}^2 \approx 0{,}033 \text{ атм}.$$

Для расчета соответствующих величин в фокусе можно пользоваться теми же формулами, вычислив предварительно ширину пучка. Для оценки будем считать, что весь свет концентрируется в пределах центрального светлого кружка с радиусом $R = 0{,}61f\lambda/r$ и площадью

$$\pi R^2 = \pi(0{,}61f\lambda/r)^2 = (0{,}61\pi f\lambda)^2/S,$$

где r — радиус поперечного сечения падающего пучка. Эту площадь и надо подставить в предыдущие формулы вместо S. В результате получим

$$E \approx \sqrt{\overline{E^2}} \approx \frac{S}{0{,}61\pi f\lambda} E_0 = 1{,}5 \cdot 10^3 E_0 = 2{,}9 \cdot 10^8 \text{ В/см},$$

$$\mathcal{P} = \left(\frac{S}{0{,}61\pi f\lambda}\right)^2 \mathcal{P}_0 = 2{,}25 \cdot 10^6 \mathcal{P}_0 \approx 7{,}4 \cdot 10^4 \text{ атм}.$$

Такое давление было бы, если бы пучок лазерного света фокусировался в вакууме. Если же он падает на вещество, то из-за большой концентрации энергии вблизи фокуса происходит практически мгновенное испарение вещества (с превращением его в плазму). С этим связано еще большее повышение давления.

2. Эффективной температурой $T_{\text{эф}}$ лазерного излучения называется такая температура абсолютно черного тела, при которой оно дает излучение той же удельной интенсивности I_ν частоты ν, что и лазер. Оценить эффективную температуру гелий-неонового лазера, генерирующего в непрерывном режиме свет с длиной волны $\lambda = 632{,}8$ нм ($\nu \approx 5 \cdot 10^{14}$ Гц.) Ширина спектральной линии генерируемого света $\delta\nu \approx 10^4$ Гц, расходимость светового пучка $\delta\vartheta \approx 1' \approx 3 \cdot 10^{-4}$ рад, мощность излучения $W = 10$ мВт.

Р е ш е н и е. Удельная интенсивность лазерного излучения равна $I_\nu = W/(\delta\nu \cdot \delta\Omega)$, где $\delta\Omega = (\delta\vartheta)^2 \approx 10^{-7}$ ср — телесный угол, определяющий расходимость светового пучка. По формуле Рэлея–Джинса

$$I_\nu = \frac{2k\nu^2}{c^2} T_{\text{эф}}.$$

Приравнивая это предыдущему выражению, получим

$$T_{\text{эф}} \approx \frac{c^2 W}{2k\nu^2 \delta\nu \cdot \delta\Omega} \approx 10^{15} \text{ К},$$

что примерно в 10^{11} раз превосходит температуру Солнца.

§ 122. Гелий-неоновый лазер

1. Энергетический спектр газов, во всяком случае при тех давлениях, которые применяются в рекламных трубках и лазерах, весьма точно совпадает с энергетическим спектром *изолированных атомов и молекул*, из которых состоит газ. Поэтому спектральные линии газов *более узкие*, чем линии примесей в твердых телах. Газы отличаются *высокой оптической однородностью* и *малой плотностью*, а поэтому слабо рассеивают и искажают волны, распространяющиеся в них. Все это позволяет использовать в газовых лазерах большие расстояния между зеркалами и получать *острую направленность*, *высокую монохроматичность* и *стабильность частоты излучения*. Так, в гелий-неоновом лазере реально получаемая расходимость пучка лучей обычно составляет 1–$2'$, а в лучших случаях достигает теоретического предела $\delta\vartheta \sim \lambda/D \sim 20''$. Относительная ширина

на спектральной линии в излучении такого лазера будет равна $\delta\omega/\omega \sim 10^{-11}$–$10^{-12}$, а сама частота ω в оптимальных условиях поддерживается с относительной точностью $\sim 10^{-14}$ (хотя в обычных условиях относительная точность не выше 10^{-10}).

Однако при малой плотности газов в них нельзя получить столь же большие концентрации возбужденных атомов, а потому и столь же большие импульсные мощности излучения, как в твердых телах. Так, выходная мощность гелий-неонового лазера в непрерывном режиме обычно составляет от десяти до нескольких сот милливат. Однако, ввиду высокой монохроматичности и направленности излучения, эта величина все же громадна по сравнению с тем, что могут дать тепловые источники света. Она соответствует эффективной температуре излучения, превышающей температуру Солнца примерно в 10^{11}–10^{12} раз (см. задачу 2 к предыдущему параграфу). Впрочем, в непрерывном режиме инфракрасный лазер на CO_2 может генерировать до 10 кВт, а ионный аргоновый лазер в видимой области — до $\sim$ 1 кВт. В импульсном режиме мощность этих лазеров может составлять несколько сот киловатт.

2. Первым газовым лазером был *гелий-неоновый лазер*, созданный в конце 1960 г. Джаваном (р. 1926), Беннетом (р. 1903) и Эрриотом. Принципиальная схема гелий-неонового лазера в ее современном виде приведена на рис. 351. Лазер состоит из

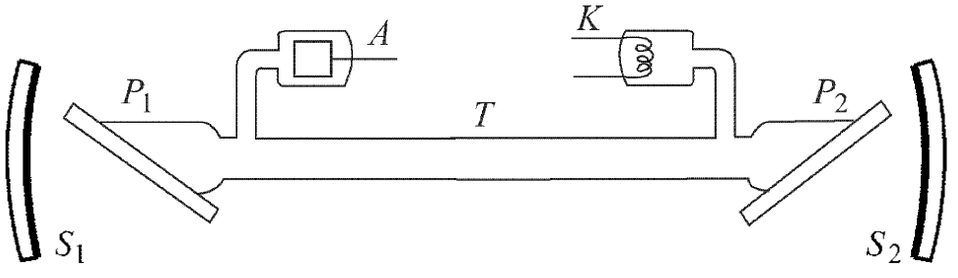


Рис. 351

газоразрядной трубки T длиной от нескольких десятков сантиметров до 1,5–2 м и внутренним диаметром 7–10 мм. Трубка наполнена смесью гелия (давление $\sim$ 1 мм рт. ст.) и неона ($\sim$ 0,1 мм рт. ст.). Концы трубки закрыты плоскопараллельными стеклянными или кварцевыми пластинками P_1 и P_2, установленными под углом Брюстера к ее оси. Это создает линейную поляризацию лазерного излучения с электрическим вектором, параллельным плоскости падения. Зеркала S_1 и S_2, между которыми помещается трубка, делаются обычно сферическими с многослойными диэлектрическими покрытиями. Они имеют высокие коэффициенты отражения и практически не поглощают свет. Пропускаемость зеркала, через которое преимущественно

выходит излучение лазера, составляет обычно 2 %, другого — менее 1 %. Между электродами трубки прикладывается постоянное напряжение 1–2 кВ. Катод K трубки может быть холодным, но для увеличения разрядного тока применяют также трубки с пустотелым цилиндрическим анодом, катод которых нагревается низковольтным источником тока. Разрядный ток в трубке составляет несколько десятков миллиампер. В первом варианте гелий-неонового лазера, построенном Мейманом, применялась безэлектродная трубка, в которой возбуждался высокочастотный разряд. Лазер может работать и в непрерывном, и в импульсном режиме. Он генерирует красный свет с длиной волны 632,8 нм и может генерировать также инфракрасное излучение с длинами волн 1,150 и 3,390 мкм. Но тогда необходимо иметь торцевые окна, прозрачные для инфракрасного света, и зеркала с высокими коэффициентами отражения в инфракрасной области спектра.

3. Рассмотрим теперь, как возникает инверсная заселенность атомов неона. Упрощенная схема уровней неона приведена справа на рис. 352. Выше уровня $\mathcal{E}_4$ у неона имеется еще 28 уровней с энергией, меньшей $\mathcal{E}_3$, но они для нас не имеют значения и на рисунке не изображены. Возбуждение атомов неона происходит в результате *столкновений их с электронами газоразрядной плазмы.* При определенном режиме разряда этот процесс может привести к инверсной заселенности уровней $\mathcal{E}_1$ и $\mathcal{E}_2$. Однако заселенность уровней $\mathcal{E}_1$ и $\mathcal{E}_3$, а также уровней $\mathcal{E}_4$ и $\mathcal{E}_3$ остается неинверсной. Инверсной заселенности препятствует долгоживущий метастабильный уровень $\mathcal{E}_5$, лежащий немного ниже короткоживущего уровня $\mathcal{E}_1$. Заселенность уровня $\mathcal{E}_5$ велика, за счет этого происходит пополнение быстро опустошающегося уровня $\mathcal{E}_1$, и инверсии заселенности между уровнями $\mathcal{E}_1$ и $\mathcal{E}_3$ не возникает.

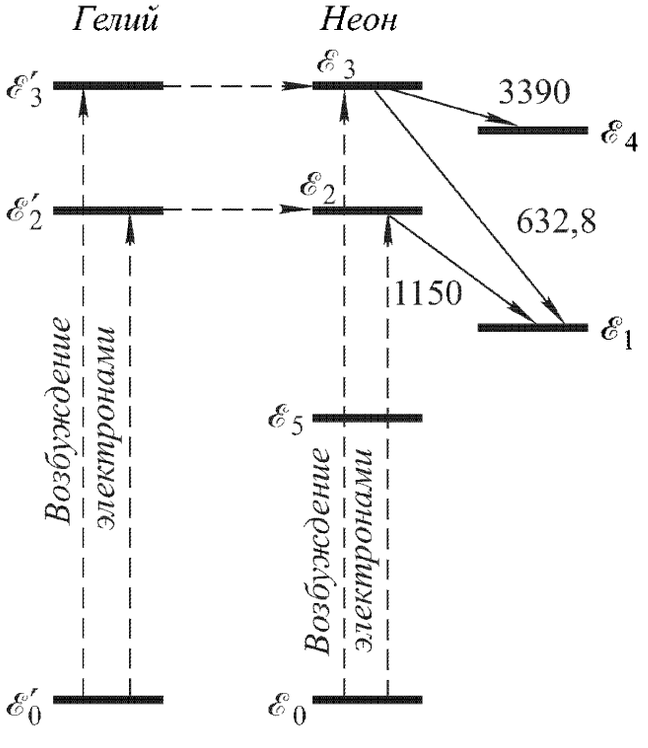


Рис. 352

Добавление гелия меняет дело. Из всех уровней гелия, помимо нормального $\mathcal{E}'_0$, для работы лазера имеют значение метастабильные уровни $\mathcal{E}'_2$ и $\mathcal{E}'_3$ с энергиями 19,82 и 20,61 эВ соответственно. Спонтанный радиационный переход с этих уровней на

основной уровень $\mathcal{E}'_0$ «запрещен», т.е. происходит с очень малой вероятностью. Поэтому время жизни атома на уровнях $\mathcal{E}'_2$ и $\mathcal{E}'_3$ очень велико. В результате электронных ударов на этих метастабильных уровнях накапливается очень много атомов гелия. Но уровни гелия $\mathcal{E}'_2$ и $\mathcal{E}'_3$ почти совпадают с уровнями $\mathcal{E}_2$ и $\mathcal{E}_3$ неона. Благодаря этому при столкновениях возбужденных атомов гелия с невозбужденными атомами неона интенсивно происходят *безызлучательные переходы* атомов гелия в невозбужденное состояние с резонансной передачей энергии атомам неона. Этот процесс возбуждения атомов неона на рис. 352 символически изображен горизонтальными штриховыми стрелками. В результате концентрации атомов неона на уровнях $\mathcal{E}_2$ и $\mathcal{E}_3$ сильно возрастают, и возникает инверсная заселенность по отношению к уровням $\mathcal{E}_1$ и $\mathcal{E}_3$, а разность заселенностей уровней $\mathcal{E}_2$ и $\mathcal{E}_1$ увеличивается в несколько раз.

Выясним в заключение влияние *столкновений* атомов неона со стенками трубки. Такие столкновения практически не влияют на заселенность уровней $\mathcal{E}_2$ и $\mathcal{E}_3$ и непосредственно уровня $\mathcal{E}_1$, так как все эти уровни *короткоживущие*. За время жизни в возбужденных состояниях на этих уровнях атомы неона практически не успевают доходить до стенок трубки. Указанные уровни разрушаются значительно раньше. Напротив, на уровне $\mathcal{E}_5$ возбужденные атомы живут долго, претерпевая в этих состояниях многочисленные столкновения со стенками трубки. Столкновения разгружают уровень $\mathcal{E}_5$, в результате чего атомы неона переходят с уровня $\mathcal{E}_1$ на уровень $\mathcal{E}_5$. Опустошение уровня $\mathcal{E}_1$ происходит быстрее, чем при заселенном уровне $\mathcal{E}_5$. Разница заселенностей уровней $\mathcal{E}_3$ и $\mathcal{E}_1$ увеличивается, что повышает эффективность работы лазера. Процесс опустошения уровня $\mathcal{E}_1$ происходит наиболее эффективно при некотором *оптимальном диаметре трубки*. Опыты показали, что максимальная мощность гелий-неонового лазера достигается при диаметре трубки ~ 7 мм. При бóльших диаметрах мощность лазера падает, несмотря на сильное увеличение объема рабочего газа (объем трубки пропорционален квадрату ее диаметра). Это связано с тем, что эффективное опустошение уровня $\mathcal{E}_1$ происходит у атомов, находящихся *вблизи стенок трубки*, а атомы, находящиеся вблизи ее центра, *практически выключаются из процесса генерации*.

§ 123. Нелинейная поляризация среды

1. Изобретение лазеров сделало возможным экспериментировать с интенсивными световыми пучками, в которых напряженность электрического поля не пренебрежимо мала по срав-

можно пользоваться моделью изотропной среды, полагая

$$\mathbf{P} = \alpha\mathbf{E} + \alpha_2 E\mathbf{E} + \alpha_3 E^2\mathbf{E} + \ldots, \tag{123.2}$$

где поляризуемости α, α_2, α_3, ... являются уже *скалярами*. Такое упрощение вполне допустимо *при качественном рассмотрении* возможных нелинейных оптических явлений. Надо только иметь в виду, что в кристаллах в выбранном направлении могут распространяться волны не всех, а только *избранных поляризаций*. Соотношение (123.2) приближенно применимо к каждой из таких волн, причем для различных волн поляризуемости α, α_2, ... имеют разные значения. Кроме того, волны разных поляризаций могут *нелинейно взаимодействовать*, обмениваясь энергией друг с другом. Такое взаимодействие должно иметь место при тензорной связи (123.1) между $\mathbf{P}$ и $\mathbf{E}$. Но оно было бы невозможно, если бы эта связь была скалярной типа (123.2). Понятно, что при нашем подходе влияние такого взаимодействия может быть учтено только качественно.

2. Разобьем поляризацию $\mathbf{P}$ на *линейную* и *нелинейную части*: $\mathbf{P} = \mathbf{P}_{\text{л}} + \mathbf{P}_{\text{нл}}$. Нелинейная часть определяется выражением

$$\mathbf{P}_{\text{нл}} = \alpha_2 E\mathbf{E} + \alpha_3 E^2\mathbf{E} + \ldots, \tag{123.3}$$

а линейная $\mathbf{P}_{\text{л}} = \alpha\mathbf{E}$. В соответствии с этим и индукция $\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P}$ представится суммой *линейной части* $\mathbf{D}_{\text{л}} = \mathbf{E} + 4\pi\mathbf{P}_{\text{л}}$ и *нелинейной* $\mathbf{D}_{\text{нл}} = 4\pi\mathbf{P}_{\text{нл}}$. Линейная часть, очевидно, равна $\mathbf{D}_{\text{л}} = \varepsilon\mathbf{E}$, где ε — обычная *диэлектрическая проницаемость среды*, как она определяется в линейной электродинамике. После этого запишем систему фундаментальных уравнений Максвелла в следующем виде:

$$\begin{gathered}
\operatorname{rot}\mathbf{H} - \frac{\varepsilon}{c}\frac{\partial\mathbf{E}}{\partial t} = \frac{4\pi}{c}\frac{\partial\mathbf{P}_{\text{нл}}}{\partial t}, \\
\operatorname{rot}\mathbf{E} - \frac{1}{c}\frac{\partial\mathbf{H}}{\partial t} = 0, \\
\operatorname{div}(\varepsilon\mathbf{E}) = -4\pi\operatorname{div}\mathbf{P}_{\text{нл}}, \\
\operatorname{div}\mathbf{H} = 0.
\end{gathered} \tag{123.4}$$

Для решения такой системы применяем *метод последовательных приближений*. В нулевом приближении в уравнении (123.4) отбрасываем правые части. Получатся *обычные уравнения линейной электродинамики*. В качестве нулевого приближения возьмем плоскую волну

$$\mathbf{E} = \mathbf{E}_0 = \mathbf{A}\cos(\omega t - \mathbf{kr}), \tag{123.5}$$

где волновой вектор $\mathbf{k}$ удовлетворяет обычному соотношению $\mathbf{k}^2 = \varepsilon\omega^2/c^2$. Для нахождения первого приближения в (123.3) отбросим кубичные и высшие члены, а в квадратичном члене

$\alpha_2 E\mathbf{E}$ поле $\mathbf{E}$ заменим его выражением (123.5) в нулевом приближении. После этого снова получатся *линейные уравнения*, но уже *неоднородные*, с известными правыми частями. Эти правые части могут быть истолкованы как *добавочные источники волн*, обусловленные нелинейной частью поляризации среды. Каждый элемент объема dV среды *переизлучает волны* как диполь Герца с добавочным дипольным моментом $\mathbf{P}_{\text{нл}}\,dV$. Эти излучения, накладываясь на волну (123.5), и создают волновое поле $\mathbf{E}_0+\mathbf{E}_1$ *в первом приближении.* Второе приближение находится так же. Для этого выражение (123.3) обрываем на членах третьей степени, заменяя в оборванном выражении вектор $\mathbf{E}$ на $\mathbf{E}_0+\mathbf{E}_1$, после чего вычисляем поле $\mathbf{E}_0+\mathbf{E}_1+\mathbf{E}_2$ *во втором приближении*, и т.д.

К изложенному надо еще добавить, что следует понимать под ε в уравнениях (123.4), когда среда *обладает дисперсией.* Ответ заключается в следующем. Если взять какое-либо приближение, то в правой части уравнений (123.4) появятся слагаемые не только с исходной частотой ω, но и с частотами 2ω, $3\omega,\ldots$ Могут появиться и другие частоты, как, например, при параметрической генерации света (см. § 126). Надо $\mathbf{E}$ и $\mathbf{H}$ искать в виде суммы монохроматических полей *с теми же частотами.* Уравнение (123.4) следует написать *для каждой частоты в отдельности*, сохранив в правой части только члены той же частоты и понимая под ε значение функции ε также *при той же частоте.* Именно такой *символический смысл* имеет система уравнений (123.4).

§ 124. Первое приближение. Оптическое детектирование. Генерация вторых гармоник, суммарной и разностной частот

1. Нелинейная добавка (123.3) к поляризации среды, вычисленная в нулевом приближении, равна

$$P_{\text{нл}}=\alpha_2 E_0^2=\frac{\alpha_2 A^2}{2}+\frac{\alpha_2 A^2}{2}\cos\left[2(\omega t-\mathbf{kr})\right]. \qquad (124.1)$$

Первое слагаемое в этом выражении не зависит от времени. С ним связано так называемое *оптическое детектирование*, т.е. возникновение в нелинейной среде постоянной электрической поляризации при прохождении через нее мощной световой волны. Это явление аналогично *выпрямлению синусоидального электрического тока.* Его можно наблюдать, если между обкладками конденсатора, одна из которых заземлена через большое сопротивление, поместить кристалл (например, кварца) и пропустить через него световой пучок от рубинового лазера.

Вследствие детектирования световой пучок возбуждает в цепи конденсатора *импульс электрического тока*, который можно обнаружить с помощью осциллографа.

2. Второе слагаемое в (124.1) гармонически меняется во времени. С ним связана *генерация в нелинейной среде второй гармоники*, т.е. волны с удвоенной частотой $\omega_2 = 2\omega$. Для нахождения поля этой гармоники поступаем так, как изложено в предыдущем параграфе. Переходя к комплексной форме, получаем систему уравнений

$$\begin{gathered} \operatorname{rot}\mathbf{H} - \frac{\varepsilon(2\omega)}{c}\frac{\partial \mathbf{E}}{\partial t} = i\omega\frac{4\pi\alpha_2}{c}AAe^{2i(\omega t-\mathbf{kr})}, \\ \operatorname{rot}\mathbf{E} - \frac{1}{c}\frac{\partial \mathbf{H}}{\partial t} = 0, \\ \operatorname{div}\mathbf{E} = \operatorname{div}\mathbf{H} = 0. \end{gathered} \tag{124.2}$$

Найдем сначала *частное решение* этой системы

$$\mathbf{E} = \mathbf{A}_1 e^{2i(\omega t-\mathbf{kr})}, \quad \mathbf{H} = \mathbf{B}_1 e^{2i(\omega t-\mathbf{kr})},$$

соответствующее *вынужденным колебаниям* с частотой 2ω. Из второго уравнения обычным путем находим, что векторы $\mathbf{E}$ и $\mathbf{H}$ *взаимно перпендикулярны*. Аналогично, из последних двух уравнений следует, что $(\mathbf{kA}_1) = (\mathbf{kB}_1) = 0$, т.е. рассматриваемая плоская волна *поперечна как в отношении вектора* $\mathbf{E}$, *так и в отношении вектора* $\mathbf{H}$. Учтя это, а также соотношение $k^2c^2 = \omega^2\varepsilon(\omega)$, из первых двух уравнений получим

$$\mathbf{A}_1 = \frac{2\pi\alpha_2}{\varepsilon(\omega)-\varepsilon(2\omega)}AA.$$

Надо еще удовлетворить условию, чтобы на входе в нелинейную среду (где мы поместим начало координат) интенсивность второй гармоники обращалась в нуль. Для этого к частному решению, найденному выше, надо добавить *общее решение* соответствующей однородной системы уравнений и подобрать амплитуду его так, чтобы указанное условие выполнялось. Возвращаясь снова к вещественной форме записи, таким путем получим

$$\mathbf{E}_1 = \frac{2\pi\alpha_2}{\varepsilon(\omega)-\varepsilon(2\omega)}AA[\cos(\omega_2 t - 2\mathbf{kr}) - \cos(\omega_2 t - \mathbf{k}_2\mathbf{r})], \tag{124.3}$$

где

$$k_2^2 = \omega_2^2\varepsilon(\omega_2)/c^2.$$

3. Таким образом, вторая гармоника представляет собой наложение двух волн одной и той же частоты $\omega_2 = 2\omega$: *вынужденной волны* $\cos(\omega_2 t - 2\mathbf{kr})$ и *свободно распространяющейся волны* $-\cos(\omega_2 t - \mathbf{k}_2\mathbf{r})$. Обе волны распространяются в одном и том же направлении, но *с различными фазовыми скоростями*. Поэтому по мере распространения будет меняться разность фаз между

ними и возникнет характерное в таких случаях явление *биений*. Интенсивность I_1 второй гармоники найдется возведением в квадрат и последующим усреднением по времени выражения (124.3). Опуская числовые коэффициенты и обозначая через I интенсивность исходной волны, таким путем найдем

$$I_1 \sim \frac{\alpha_2^2 \omega^2 x^2 I^2}{n^2 c^2} \left(\frac{\sin\beta}{\beta}\right)^2, \tag{124.4}$$

где x — расстояние, пройденное волной, и введено обозначение

$$\beta = \frac{(2\mathbf{k} - \mathbf{k}_2)\mathbf{r}}{2} = \frac{(2k - k_2)x}{2}. \tag{124.5}$$

При этом в знаменателе формулы (124.4) мы пренебрегли различием между показателями преломления $n(\omega)$ и $n(2\omega)$.

Когда $\beta = 0,\ \pi,\ 2\pi,\ \ldots$, интенсивность первой гармоники обращается в нуль. Максимумы интенсивности получаются примерно посередине между минимумами. Таким образом, с возрастанием x интенсивность второй гармоники возрастает, когда β лежит приблизительно между нулем и $\pi/2$, между π и $3\pi/2$ и т.д. В этих случаях *энергия переходит от исходной волны ко второй гармонике*. Если же β лежит приблизительно между $\pi/2$ и π, между $3\pi/2$ и 2π и т.д., то с возрастанием x интенсивность второй гармоники убывает. В этих случаях *энергия снова возвращается от второй гармоники к исходной волне*. Такой процесс *перекачки энергии* периодически повторяется по мере распространения исходной волны.

Условие $\beta = \pi/2$ определяет расстояние x, до которого происходит перекачка энергии от исходной волны ко второй гармонике с последующим возвращением ее опять в исходную волну. Это расстояние называется *когерентной длиной*. Для нее из указанного условия нетрудно получить

$$l_{\text{ког}} = \frac{\lambda}{4|n(\omega) - n(2\omega)|}, \tag{124.6}$$

где λ — длина исходной волны, $n(\omega)$ — ее показатель преломления, а $n(2\omega)$ — показатель преломления второй гармоники. *Чем больше когерентная длина, тем интенсивнее происходит перекачка энергии от исходной волны во вторую гармонику.*

Когда

$$n(\omega) = n(2\omega), \quad \text{а потому} \quad 2\mathbf{k} = \mathbf{k}_2, \tag{124.7}$$

то длина когерентности $l_{\text{ког}}$ обращается в бесконечность. В этом случае переход энергии особенно интенсивен и должен происходить от исходной волны к ее второй гармонике. Обе волны $\cos(\omega_2 t - 2\mathbf{kr})$ и $-\cos(\omega_2 t - \mathbf{k}_2\mathbf{r})$ распространяются с одинаковыми фазовыми скоростями, а потому фазовое соотношение между ними сохраняется все время при их распространении. С

этим, как и при всяком резонансе, и связана эффективность обмена энергией между взаимодействующими волнами. Поэтому условие (124.7) называется *фазовым синхронизмом* между рассматриваемыми волнами. В этом случае $\beta = 0$ при любом x, и из (124.4) предельным переходом $\beta \to 0$ получаем

$$I_1 = \frac{\alpha_2^2 \omega^2 x^2 I^2}{n^2 c^2}. \tag{124.8}$$

Конечно, при достаточно больших x формула (124.8) перестает быть верной, так как при $x \to \infty$ она дает $I_1 \to \infty$, а интенсивность второй гармоники не может превосходить интенсивности I исходной волны. В этом случае метод последовательных приближений, с помощью которого была получена формула (124.8), неприменим.

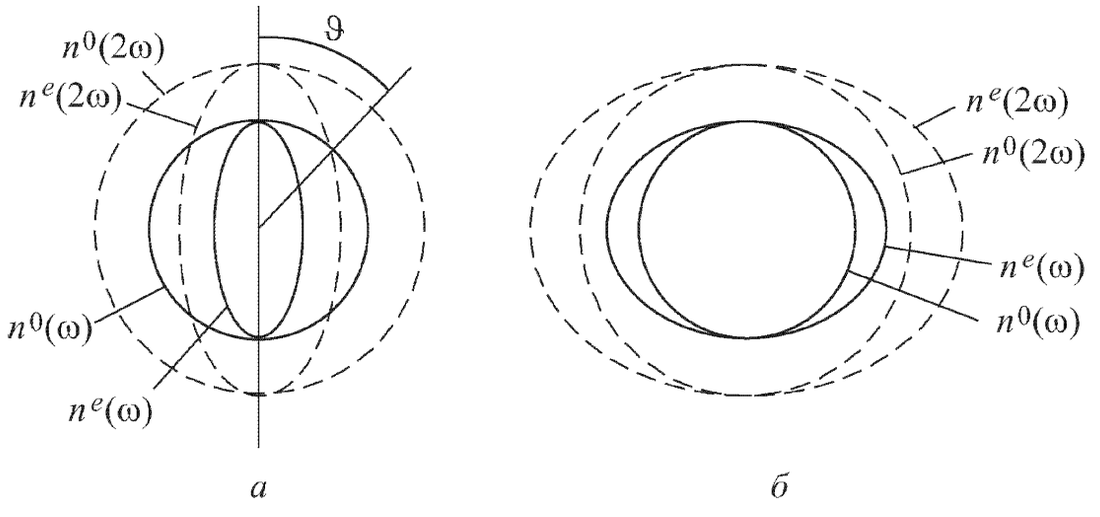
Когерентную длину $l_{\text{ког}}$ *можно определить как максимальное расстояние, на котором приближенно сохраняется фазовый синхронизм между падающей и переизлученной волнами, т.е. расстояние, на котором набег фазы одной из этих волн относительно другой не превышает* π.

4. Генерация второй гармоники впервые была обнаружена Франкеном в 1961 г. Излучение рубинового лазера фокусировалось на кристалле кварца, расположенном между двумя фильтрами, прозрачными только в узких окрестностях длин волн $\lambda_1 = 694{,}3$ нм и $\lambda_2 = \lambda_1/2 = 347{,}0$ нм. Из кристалла, помимо исходного красного излучения лазера ($\lambda_1 = 694{,}3$ нм), выходило также ультрафиолетовое излучение с половинной длиной волны 347,0 нм. В первых опытах Франкена в энергию второй гармоники превращалось $\sim 10^{-8}$ энергии первичного излучения.

Почему лишь столь ничтожная доля энергии переходила ко второй гармонике? Это объясняется *малостью когерентной длины* $l_{\text{ког}}$ в кварце. Для интенсивного обмена энергией надо удовлетворить *условию фазового синхронизма* $n(\omega) = n(2\omega)$. Но это невозможно сделать для *изотропных сред в прозрачной области* спектра, так как в этой области показатель преломления $n(\omega)$ *монотонно возрастает с частотой*. Равенство $n(\omega) = n(2\omega)$ может удовлетворяться только тогда, когда частота ω взята в прозрачной области, а 2ω — в области сильного поглощения или наоборот.

Однако, как указали в 1962 г. Джордмейн и Терхьюн и подтвердили это указание на опыте, *фазовый синхронизм на частотах* ω *и* 2ω *можно осуществить между обыкновенной и необыкновенной волнами в некоторых кристаллах*. На рис. 353 представлены обыкновенный n^o и необыкновенный n^e показатели преломления одноосного кристалла в зависимости от направления волновой нормали. Сплошные кривые относятся к частоте ω, штриховые — к удвоенной частоте 2ω. На первом из

этих двух рисунков кривые $n^o(\omega)$ и $n^e(2\omega)$ пересекаются между собой. Точкам их пересечения соответствуют направления, для которых между обыкновенной волной с частотой ω и ее гармоникой с частотой 2ω выполняется фазовый синхронизм. Эти направления называются *направлениями синхронизма*, а угол ϑ между ними и оптической осью кристалла — *углом синхронизма.*



а *б*

Рис. 353

Хотя обе волны и поляризованы в различных плоскостях, но они могут нелинейно взаимодействовать между собой, поскольку в кристаллах квадратичная поляризуемость есть не скаляр α_2, а *тензор* $(\alpha_2)_{jkl}$. Поэтому при наличии фазового синхронизма должна происходить перекачка энергии от исходной волны к ее второй гармонике, что и наблюдается на самом деле. Таким путем удается более половины падающего света превратить во вторую гармонику. Понятно, что при таких больших интенсивностях второй гармоники метод последовательных приближений может оказаться неточным и даже неприменимым. Однако качественное заключение о влиянии фазового синхронизма *остается в силе.*

Подходящим кристаллом может быть одноосный кристалл *дигидрофосфата калия* KH_2PO_4 (сокращенно KDP). Для этого кристалла при $\lambda = 1{,}15$ мкм, как показывает расчет, подтверждаемый наблюдениями, угол синхронизма равен $41°35'$. Существует красивый демонстрационный опыт. Кристалл KDP, вырезанный параллельно оптической оси, кладется на столик, который может вращаться вокруг вертикальной оси. Оптическая ось кристалла должна быть горизонтальна. На кристалл направляется мощный инфракрасный луч от лазера на неодимовом стекле ($\lambda = 1060$ нм). Луч лазера невидим, но его можно обнаружить с помощью листа черной бумаги. Бумага загорается, если ее поместить на пути луча. При произвольной ориентации

кристалла никакого видимого света не возникает. Но если кристалл медленно поворачивать, то из него выходит ослепительно яркий зеленый луч ($\lambda = 530$ нм), когда станет выполняться условие синхронизма.

Не всякий одноосный кристалл годится для опыта такого типа. Не годится, например, кристалл кварца, в котором, как видно из рис. 353 *б*, условие фазового синхронизма не выполняется ни для какого направления.

5. Остановимся теперь на *генерации волн с суммарной и разностной частотами.* Природа этого явления в точности такая же, что и генерация второй гармоники. Поэтому достаточно только указать, в чем состоит явление. Если на нелинейную среду направить два мощных пучка света с различными частотами ω_1 и ω_2, то из нее будет выходить свет *не только с первоначальными частотами* ω_1 и ω_2 *и их гармониками* $2\omega_1$ *и* $2\omega_2$, *но и свет с суммарной* $\omega_1 + \omega_2$ *и разностной* $\omega_1 - \omega_2$ *частотами.* Подобными методами генерации волн разных частот удается далеко проникнуть в инфракрасную и ультрафиолетовую области спектра. Например, удалось получить ультрафиолетовое излучение с длиной волны ≈ 80 нм.

§ 125. Второе приближение. Самофокусировка

1. Для нахождения *второго приближения* надо использовать вектор $\mathbf{P}_{\text{нл}}$ в первом приближении, т.е.

$$\mathbf{P}_{\text{нл}} = \alpha_2(E_0 + E_1)(\mathbf{E}_0 + \mathbf{E}_1) + \alpha_3 E_0^2 \mathbf{E}_0.$$

Однако мы ограничимся *только изотропными средами или кристаллами, обладающими центром симметрии.* Для них, как было показано в § 123, $\alpha_2 = 0$, и следовательно, в нужном приближении $\mathbf{P}_{\text{нл}} = \alpha_3 E_0^2 \mathbf{E}_0$. Подставив сюда выражение (123.5), получим

$$\mathbf{P}_{\text{нл}} = \frac{3\alpha_3 A^2}{4}\mathbf{A}\cos(\omega t - \mathbf{kr}) + \frac{\alpha_3 A^2}{4}\mathbf{A}\cos[3(\omega t - \mathbf{kr})]. \quad (125.1)$$

Слагаемое с тройной частотой $\omega_3 = 3\omega$, очевидно, приводит к *генерации третьей гармоники.* Разность показателей преломления $n(3\omega) - n(\omega)$ здесь еще больше, чем в случае второй гармоники. Это ограничивает выбор кристаллов, для которых можно удовлетворить *условию фазового синхронизма.* Основная трудность опыта связана с малыми значениями кубичной поляризуемости α_3, что вынуждает применять большие кристаллы и большие освещенности, часто приводящие к разрушению кристаллов. Несмотря на это, генерация третьей гармоники наблюдалась еще в 1962 г. группой американских физиков на кри-

сталле исландского шпата при освещении его светом рубинового лазера. На выходе кристалла удалось зарегистрировать излучение с длиной волны 231,3 нм. Позднее генерация наблюдалась в некоторых оптически изотропных кристаллах (например, LiF, NaCl), жидкостях и газах.

В третьем приближении, помимо гармоник более низкого порядка, очевидно, должна появиться *четвертая*, а в следующем приближении — *пятая* гармоники и т.д. Экспериментально четвертая гармоника наблюдалась С.А. Ахмановым в 1974 г., а пятая — Харрисом в 1973 г.

2. Посмотрим теперь, какие явления связаны с первым слагаемым в выражении (125.1). Множитель $\mathbf{A}\cos(\omega t - \mathbf{kr})$ есть исходная падающая волна $\mathbf{E}_0$. Ясно, что в рассматриваемом приближении его можно заменить на $\mathbf{E}$, так как такая замена сказывается лишь на членах *более высокого порядка малости*, не учитываемых в рассматриваемом приближении. После этого (123.4) запишется в виде

$$\operatorname{rot}\mathbf{H} - \frac{1}{c}[\varepsilon(\omega) + 3\pi\alpha_3(\omega)A^2]\frac{\partial\mathbf{E}}{\partial t} = 0.$$

Отсюда видно, что влияние рассматриваемого слагаемого эквивалентно *изменению диэлектрической проницаемости или показателя преломления среды.* Учитывая малость поправки к $\varepsilon(\omega)$, для показателя преломления n в поле интенсивной световой волны можно написать

$$n = n_0 + n_2 A^2, \tag{125.2}$$

где n_0 — значение показателя преломления среды в линейной оптике, а $n_2(\omega)$ — некоторый коэффициент, зависящий от свойств среды.

Помимо рассмотренной, есть и другие причины изменения показателя преломления в электрическом поле. В нелинейной среде из-за *электрострикции* световая волна вызывает появление *постоянного давления*, аналогично тому как появляется постоянное слагаемое в формуле (124.1). Это приводит *к изменению плотности и показателя преломления среды.* В жидкостях с анизотропными молекулами показатель преломления изменяется из-за *высокочастотного эффекта Керра* (см. § 90). Показатель преломления всегда изменяется из-за *нагревания среды световой волной.* Во всех этих случаях изменение показателя преломления пропорционально *квадрату амплитуды*, а потому может быть также представлено формулой (125.2).

Постоянная n_2 может быть и отрицательной, и положительной. Она особенно велика у нитробензола и имеет для него положительный знак.

3. Из изложенного следует, что если через однородную среду проходит интенсивный пучок света, то среда становится *оптиче-*

ски неоднородной. Луч света в такой среде загибается в сторону большего показателя преломления. С этим связано явление *самофокусировки* (когда $n_2 > 0$) и *дефокусировки* (когда $n_2 < 0$) света, предсказанное теоретически Г.А. Аскарьяном в 1962 г. и впервые наблюдавшееся Н.Ф. Пилипецким и А.Р. Рустамовым в 1965 г. Затем самофокусировка наблюдалась для многих газов, жидкостей и твердых тел.

Чтобы простейшим путем понять сущность явления самофокусировки, предположим, что в однородную среду с показателем преломления n_0 вступает плоскопараллельный пучок лучей кругового поперечного сечения с диаметром D (рис. 354). Допустим сначала, что амплитуда пучка постоянна по всему сечению. Показатель преломления в пространстве, занятом пучком, сделается равным $n = n_0 + n_2 A^2$, причем мы предположим, что $n_2 > 0$. Из-за дифракции пучок расширяется. Практически все направления лучей сосредоточатся в пределах конуса с углом при вершине $2\vartheta_{\text{диф}}$, где $\vartheta_{\text{диф}} = 1{,}22\lambda/(Dn_0)$, а λ — длина волны в вакууме. (Направления лучей относятся к пространству внутри цилиндра.) Предельный угол скольжения ϑ_0 для полного отражения от боковой стенки цилиндра определяется соотношением

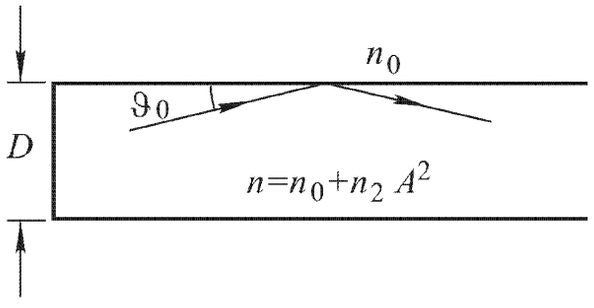


Рис. 354

$$\cos\vartheta_0 = n_0/(n_0 + n_2 A^2).$$

Ввиду малости этого угла отсюда находим: $1 - \cos\vartheta_0 \approx A^2 n_2/n_0$ и, следовательно,

$$\vartheta_0^2 \approx 2A^2 n_2/n_0.$$

Если $\vartheta_{\text{диф}} > \vartheta_0$, то часть дифрагированных лучей будет выходить из цилиндра — пучок будет *расширяться.* При обратном соотношении $\vartheta_{\text{диф}} < \vartheta_0$ все дифрагированные лучи будут испытывать полное отражение от боковой поверхности цилиндра. А так как в реальных условиях интенсивность света и показатель преломления возрастают к оси пучка, то из-за искривления лучей пучок начнет *сжиматься и может стянуться в тонкий шнур.* Это и есть *самофокусировка.*

В промежуточном случае, когда $\vartheta_{\text{диф}} \approx \vartheta_0$, пучок будет проходить через нелинейную среду практически без изменения поперечных размеров. Он создает для себя как бы волновод, в котором и распространяется без рассеяния в стороны. Такой режим распространения называется *самоканализацией светового пучка.* Таким образом, самоканализация имеет место при усло-

вии $\vartheta_0 \approx \vartheta_{\text{диф}}$. Подставив сюда значения углов ϑ_0 и $\vartheta_{\text{диф}}$, а также выражение амплитуды A через мощность пучка

$$P = \frac{cn_0A^2}{8\pi}\,\frac{\pi D^2}{4} = \frac{cn_0D^2}{32}A^2,$$

получим так называемую *пороговую мощность*, выше которой начинается сжатие пучка. Она определяется соотношением

$$P_{\text{порог}} \approx c\frac{(0{,}61\lambda)^2}{16n_2}. \tag{125.3}$$

Расстояние от края среды, на котором фокусируются крайние лучи пучка, легко оценить из следующих соображений. В пучке угловое расхождение лучей из-за дифракции равно $2\vartheta_{\text{диф}}$. При критической мощности в результате отражения от боковой поверхности пучка крайние лучи делаются параллельными. Это произойдет на расстоянии

$$f_{\text{эф}} = \frac{D}{2\vartheta_{\text{диф}}} \approx \frac{n_0D^2}{2{,}44\lambda}. \tag{125.4}$$

Оно играет при самофокусировке роль *эффективного фокусного расстояния для крайних лучей пучка.* Если вместо расходящихся лучей взять лучи, параллельные оси пучка, то они сфокусируются на том же расстоянии $f_{\text{эф}}$.

Для сероуглерода CS_2, обладающего сравнительно большим значением $n_2 = 2\cdot 10^{-11}$ СГСЭ, при освещении рубиновым лазером ($\lambda = 694{,}3$ нм) пороговая мощность, вычисленная по формуле (125.3), равна $P_{\text{порог}} \approx 17$ кВт. Если диаметр пучка $D = 1$ мм, то формула (125.4) в этом случае дает $f_{\text{эф}} \approx 96$ см ($n_0 = 1{,}62$). В некоторых сортах оптического стекла $P_{\text{порог}} \sim 1$ Вт. В этих случаях явление самофокусировки можно наблюдать не только в мощных пучках импульсных лазеров, но и в малоинтенсивных пучках лазеров непрерывного действия.

§ 126. Параметрическая генерация света

1. Нелинейные оптические явления в кристаллах позволяют преобразовывать излучение лазера не только в излучения гармоник, суммарных и разностных частот, но и в *излучения с плавно перестраиваемой частотой.* Принцип такого преобразования был указан в 1962 г. С.А. Ахмановым и Р.В. Хохловым (1926–1977). Он заключается в следующем. Пусть на среду, нелинейная поляризация которой с точностью до квадратичных членов определяется выражением $\mathbf{P}_{\text{нл}} = \alpha_2 E\mathbf{E}$, падает мощная «волна накачки» $\mathbf{E}_{\text{н}} = \mathbf{A}_{\text{н}}\cos(\omega_{\text{н}}t - \mathbf{k}_{\text{н}}\mathbf{r})$ и две слабые волны $\mathbf{E}_1 = \mathbf{A}_1\cos(\omega_1 t - \mathbf{k}_1\mathbf{r})$ и $\mathbf{E}_2 = \mathbf{A}_2\cos(\omega_2 t - \mathbf{k}_2\mathbf{r})$, частоты кото-

рых связаны соотношением

$$\omega_{\text{н}} = \omega_1 + \omega_2. \tag{126.1}$$

Считая для простоты, что направления амплитуд всех волн совпадают, перейдем к скалярной форме записи. В первом приближении нелинейная поляризация среды будет равна $\alpha_2(E_{\text{н}} + E_1 + E_2)^2$. Возведя в квадрат, рассмотрим член $2\alpha_2 E_1 E_{\text{н}}$, представляющий собой произведение двух косинусов. Преобразуем его в сумму двух косинусов и возьмем слагаемое с разностной частотой $(\omega_{\text{н}} - \omega_1)$, которая, ввиду (126.1), равна ω_2. Так же поступим с произведением $2\alpha_2 E_2 E_{\text{н}}$. В результате из нелинейной поляризации $P_{\text{нл}}$ выделятся два члена с частотами ω_1 и ω_2:

$$\begin{aligned} P_{\text{нл}}(\omega_1) &= \alpha_2 A_2 A_{\text{н}} \cos[\omega_1 t - (\mathbf{k}_{\text{н}} - \mathbf{k}_2)\mathbf{r}], \\ P_{\text{нл}}(\omega_2) &= \alpha_2 A_1 A_{\text{н}} \cos[\omega_2 t - (\mathbf{k}_{\text{н}} - \mathbf{k}_1)\mathbf{r}]. \end{aligned} \tag{126.2}$$

Следовательно, возникнет переизлучение волн с теми же частотами ω_1 и ω_2. Это может привести к *усилению волн таких частот* за счет энергии волны накачки. Такое явление называется *параметрическим усилением света*, так как его можно рассматривать как результат *модуляции параметров среды* (показателя преломления) при ее взаимодействии с волной накачки. Оно было открыто в 1965 г. Ахмановым и Хохловым с сотрудниками в СССР, а также Джордмейном и Миллером в США и использовано ими для создания *когерентных генераторов света с плавно перестраиваемой частотой*. Взаимодействие с волной накачки будет особенно сильным, когда фазы волн (126.2) длительно совпадают с фазами обеих волн E_1 и E_2, т.е. когда соблюдается условие

$$\mathbf{k}_1(\omega_1) + \mathbf{k}_2(\omega_2) = \mathbf{k}_{\text{н}}(\omega_{\text{н}}). \tag{126.3}$$

Это условие называется условием *фазового синхронизма* между волной накачки и обеими волнами с частотами ω_1 и ω_2. Полученное ранее условие (124.7) является частным случаем условия (126.3). Чтобы в этом убедиться, достаточно записать (124.7) в виде

$$\mathbf{k}(\omega) + \mathbf{k}(\omega) = \mathbf{k}_2(2\omega)$$

и применить его к процессу образования волны с частотой ω из ее второй гармоники.

Если условие синхронизма выполнено, то энергия от волны накачки будет в нелинейной среде передаваться волнам с частотами ω_1 и ω_2. Для эффективного усиления этих волн надо волну накачки заставить многократно проходить через нелинейную среду (кристалл). Для этого последнюю, как в лазерах, помещают в *оптический резонатор между двумя зеркалами*. Оба зеркала должны иметь достаточно высокие коэффициенты отражения для волн обеих частот ω_1 и ω_2 и в то же время одно

из них, через которое входит волна накачки, должно быть в достаточной степени прозрачным для этой волны. При достаточно высоких коэффициентах отражения зеркал и большой мощности волны накачки возникает генерация на частотах ω_1 и ω_2, удовлетворяющих условиям (126.1) и (126.3).

Нет необходимости специально посылать в резонатор волны с частотами ω_1 и ω_2. Они сами возникают либо из-за всегда имеющихся *шумов*, либо из-за *тепловых флуктуаций*. Происходит *самовозбуждение* генератора с последующим усилением генерируемых волн при нелинейном взаимодействии их с волной накачки. В качестве волны накачки обычно используется вторая (или третья) гармоника рубинового или неодимового лазера.

2. *В изотропных средах в области нормальной дисперсии нельзя удовлетворить одновременно обоим условиям* (126.1) *и* (126.3). Действительно, допустим сначала, что все три волны с частотами $\omega_н$, ω_1, ω_2 распространяются в одном направлении. В этом случае условие (126.3) можно записать в виде $\omega_1 n_1 + \omega_2 n_2 = \omega_н n_н$, где n_1, n_2, $n_н$ — показатели преломления для соответствующих частот. С учетом условия (126.1) отсюда получаем $(n_н - n_1)\omega_1 + (n_н - n_2)\omega_2 = 0$, а это невозможно, так как $(n_н - n_1) > 0$ и $(n_н - n_2) > 0$. Из приведенного рассуждения следует, что волновое число $k_н$ всегда больше суммы волновых чисел k_1 и k_2, *независимо от направления волн*. Поэтому условию (126.3) нельзя удовлетворить и при различных направлениях векторов $\mathbf{k}_1$, $\mathbf{k}_2$, $\mathbf{k}_н$. Иначе получился бы векторный треугольник, одна сторона которого длиннее суммы длин двух других сторон.

Однако *синхронизм можно получить в некоторых кристаллах между обыкновенной и необыкновенной волнами*. Только теперь, когда частоты ω_1 и ω_2 могут не совпадать, для осуществления синхронизма имеется больше возможностей, чем в аналогичном случае при генерации второй или третьей гармоник (см. § 124). В принципе синхронизм мог бы осуществляться в четырех случаях:

$$\begin{aligned} &1)\ \mathbf{k}^e_н = \mathbf{k}^o_1 + \mathbf{k}^e_2, \quad 2)\ \mathbf{k}^e_н = \mathbf{k}^o_1 + \mathbf{k}^o_2, \\ &3)\ \mathbf{k}^o_н = \mathbf{k}^o_1 + \mathbf{k}^e_2, \quad 4)\ \mathbf{k}^o_н = \mathbf{k}^e_1 + \mathbf{k}^e_2, \end{aligned} \tag{126.4}$$

где индексы o и e относятся к обыкновенной и необыкновенной волнам. Разумеется, не всем этим условиям, и даже хотя бы одному из них, можно удовлетворить в реальных кристаллах. Так, в случае одноосного кристалла дифосфата калия KH_2PO_4 можно удовлетворить первым двум условиям. В первом случае в «направлении синхронизма» необыкновенная волна накачки будет генерировать в кристалле обыкновенную и необыкновенную волны с частотами ω_1 и ω_2, а во втором случае обе генерируемые

волны будут обыкновенными. Для одноосного кристалла ниобата лития $LiNbO_3$, обладающего очень большой нелинейностью, можно удовлетворить только второму из условий (126.4).

Поворачивая кристалл (или изменяя его температуру, а также накладывая постоянное электрическое поле), можно изменять частоты ω_1 и ω_2, для которых направление, перпендикулярное к зеркалам, является направлением синхронизма. Именно так действуют *параметрические генераторы когерентного света, позволяющие плавно перестраивать частоту.* КПД таких генераторов, определяемый как отношение мощностей параметрически генерируемых волн к мощности волны накачки, достигает нескольких процентов при выходной мощности в несколько десятков и сотен киловатт. Ясно, что генерируемые частоты ω_1 и ω_2 *всегда меньше* частоты $\omega_н$ волны накачки. Плавно перестраивая параметрический генератор света, можно пройти весь диапазон видимого света от красного до фиолетового, а также далеко проникнуть и в инфракрасную область спектра.

3. Если (126.1) умножить на постоянную Планка $\hbar$, то получится

$$\hbar\omega_н = \hbar\omega_1 + \hbar\omega_2. \qquad (126.5)$$

Это соотношение в квантовой физике интерпретируется как процесс *распада фотона* $\hbar\omega_н$ *на два фотона* $\hbar\omega_1$ и $\hbar\omega_2$, причем уравнение (126.5) выражает *закон сохранения энергии* для этого элементарного акта. Аналогично, генерация второй гармоники с квантовой точки зрения есть процесс *взаимодействия двух фотонов* с энергией $\hbar\omega$ каждый, в результате которого *рождается фотон* $\hbar\omega_2$ *с удвоенной частотой* $\omega_2 = 2\omega$. Точно так же можно интерпретировать генерацию третьей и высших гармоник, а также генерацию волн с суммарной и разностной частотами. Однако сейчас мы не будем входить в обсуждение всех этих вопросов, так как квантовые явления предполагается рассмотреть в пятом томе нашего курса.

ИМЕННОЙ УКАЗАТЕЛЬ

Аббе 115, 118, 122–124, 126, 127, 129, 166, 175–177, 386, 388, 393–395
Абрагам 590
Авогадро 553, 742
Айвс 268, 691, 692
Амичи 126, 127
Ангерер 289
Андерсон 334
Араго 28, 203, 288, 289, 508, 513, 547, 605
Аренс 493, 494
Аркадьев 304
Аскарьян 775
Ахманов 774, 776, 777

Бабине 296, 297, 327, 371, 502–504
Бальмер 603
Бартолинус 487
Басов 751
Беккерель 615
Беннет 763
Бессель 95, 314
Бийе 213, 215, 220, 221
Био 28, 605, 607
Бозе 742, 743
Больцман 150, 563, 589, 723–726, 728, 730, 731, 738, 740–742, 744, 745
Бонч–Бруевич 666
Бор 32, 743, 745
Борн 555, 590
Брадлей 692, 694
Браун 407
Бреге 29
Бриллюэн 66, 643, 645, 646, 648–650, 766
Бродхун 161
Бройль де 32, 45, 387
Брэгг 413–416, 418, 644, 645
Брюстер 28, 432, 433, 435, 446–448, 455, 456, 459, 460, 463–465, 476, 518, 759
Бугер 582, 583

Вавилов 151, 271–276, 511, 582, 583
Ван–Циттер 240, 242
Вебер 30, 40
Ведам 625
Вейерштрасс 124, 126
Верде 612, 615
Вильямс 337
Вин 578–580, 616, 617, 726, 727, 729, 730, 732, 735, 737, 738, 740, 742, 745, 747
Винер 267–269
Владимирский 648
Волластон 495, 497
Вольфке 363
Вуд 287, 334, 357, 476, 616
Вульф 413–416, 418, 644, 645

Габор 363–365
Гаген 477
Гайдингер 611
Гайзенберг 229
Галилей 184, 656, 657, 659, 676, 703
Гамильтон 129, 541, 543, 736
Гартнак 493, 494
Гаусс 78
Гельмгольц 77, 83, 85, 124, 132, 147, 388, 446
Герке 265, 266, 337–339, 342, 343
Герц 30, 314, 450, 455, 530, 633, 643, 768
Гершель 131, 132, 186–187
Гинзбург 273, 276
Глазебрук 493, 494, 497
Глан 493, 494
Горелик 209
Грегори 187, 190
Гримальди 20, 25, 211, 222
Грин 305
Гросс 527, 625, 646
Гук 20, 25
Гульстранд 143, 144
Гюйгенс 20, 21, 23–26, 47–49, 178-181, 183, 238, 242, 277–279, 281, 282, 291, 292, 294, 304, 305, 307, 309, 325, 335, 354, 374, 380, 393, 404, 487, 537

Дебай 417, 555, 646, 732
Декарт 14, 15, 20, 25
Делиль 288
Денисюк 370, 373, 374
Джаван 663, 763
Джасея 663
Джермер 32
Джинс 731, 735–740, 745, 747, 762

Джордмейн 771, 777
Доплер 426, 453, 582, 688–692, 695, 696, 698, 725, 726
Друде 268, 461, 463, 464, 473, 475, 477, 607
Дэвиссон 32

Жамен 249, 251, 252, 254, 256, 565, 584

Зеeман 31, 210, 596, 597, 601, 602, 604, 613, 616, 706
Зельмейер 551, 569, 575, 614
Зоммерфельд 66, 314

Ингл 353

Капица 599, 601
Каплянский 527, 625
Каратеодори 134, 136
Карно 725
Кассегрен 187, 190
Кельпер 180
Кеннеди 664
Кеплер 183, 184, 559, 666
Керр 31, 200, 583–586, 588–594, 596, 759, 774
Кеттелер 471, 474
Кирхгоф 277, 279, 281, 292, 304, 307, 717, 719 723, 735
Китинг 686
Клаузиус 46, 554, 721, 722, 735
Книппинг 407
Кольрауш 31, 40
Комптон 32
Коперник 692
Корбино 615
Корню 298–303
Коттон 31, 593, 594, 611
Кришнан 650
Куп 562
Кундт 564, 565, 612, 613
Курльбаум 737

Лагранж 77, 83, 85, 124, 132
Ладенбург 563
Ламберт 159, 160, 162, 163, 165, 398, 720, 721
Ламе 530
Ландау 646, 647
Ландсберг 642, 646, 649, 651
Ланжевен 586, 589, 593
Лаплас 44
Лармор 655, 675
Лауэ 407, 410 412, 414, 416
Лебедев 559, 724
Левин 573, 575
Лейт 363
Лемуан 590
Леонтович 646
Леру 564
Липпман 269–271, 373
Ллойд 213, 220–222, 234, 243, 541, 543
Ломоносов 28, 169, 186
Лоренц 31, 527, 548, 554, 555, 580, 596, 598, 602, 613, 655, 659, 675–678, 680, 681, 689, 703, 705–708, 710
Лоренц 554, 555
Лошмидт 639
Луммер 161, 265, 266, 337–339, 342, 343, 737
Лэнглей 737

Майкельсон 30, 223, 240, 256–258, 333, 334, 337, 341, 362, 404–407, 659–664, 669, 681, 705
Макалюзо 615
Максвелл 29, 30, 38, 39–41, 44, 45, 137, 139, 292, 314, 424, 425, 453, 461, 466, 478, 481, 485, 519, 548, 551, 581, 609, 622, 632, 633, 637, 659, 665, 707, 724, 767
Максутов 188–190
Малюс 28, 56, 57, 422, 433
Мандельштам 395, 579, 584, 639, 643, 645, 646, 648–651, 766
Маральди 288
Марци 20
Мах 256, 275
Мейман 754, 764
Меслин 215, 220
Миллер 777
Минковский 677–680, 707, 708
Михельсон 731
Молчанов 666
Морзе 315
Морли 257, 662
Мосотти 554
Муррей 663
Мутон 31, 593, 594

Нейхаусс 271
Нернст 268
Николь 491–495, 497
Ньютон 20, 21, 25, 27–29, 33, 84, 172, 186, 190, 247, 248, 437, 548, 549, 564, 656, 707

Орнштейн 641

Пастернак 625
Паули 665
Пашен 353, 737

Перо 263–265, 267, 337–339, 350, 751–753
Пилипецкий 775
Пифагор 20
Планк 10, 31, 387, 577, 715, 736–740, 742, 743, 745, 747, 779
Плачек 646, 647
Поггендорф 541
Пойнтинг 434, 436, 482, 501–503, 530, 532, 576, 634
Поккельс 594, 596, 759
Поль 214, 215, 220, 221, 289, 290, 398
Празмовский 493, 494
Прево 718, 719
Принсхейм 737
Прохоров 751
Пуанкаре 655, 675
Пуассон 288, 289, 725
Пуркинье 148
Пуччианти 565

Раман 650
Рамсден 95, 178, 180, 181, 183
Резерфорд 333
Рейхе 562
Рёмер 20, 694
Риттенгауз 333
Риттер 493, 494
Ритц 665
Рождественский 256, 388, 395, 565, 566, 569
Роулэнд 334, 350, 352, 353
Рошон 495
Рубенс 476, 477, 737
Рубинштейн 315
Рустамов 775
Рытов 325
Рэлей 60, 211, 278, 287, 314, 332, 336, 338, 341, 354, 357–359, 363, 366, 379–381, 384, 386, 390, 395, 451, 463, 528, 551, 576, 617, 632, 636, 639, 641, 731, 735–740, 745, 747, 762

Сенармон 495
Сивухин 45
Ситтер де 666
Смолуховский 637, 639, 641
Снеллиус 14, 15, 22, 33, 50, 53, 275, 486, 496, 544
Солейль 503, 504
Стефан 150, 723–726, 728, 730, 740–742
Стилуэлл 691, 692
Столетов 433, 448, 449

Тамм 273, 274
Таунс 663, 751
Твисс 407
Терхьюн 771
Тиндаль 632, 636
Томас 562
Томашек 666
Томпсон 493, 494
Торндайк 664

Уайтфорд 303
Угаров 680
Ульянин 721
Умов 532
Упатинекс 363

Фабелинский 325
Фабри 263–265, 267, 337–339, 350, 751–753
Фабрикант 750
Фарадей 29, 31, 553, 583, 600, 605, 611, 612, 614, 742
Ферма 49–54, 56–57, 87, 88, 340
Физо 29, 32, 249, 403, 404, 705
Фицжеральд 681
Форрестор 210
Фохт 541, 542, 603, 615
Франк И.М. 273, 274, 276
Франк (поляризационная призма) 493, 494
Франкен 771
Франклин 28
Фраунгофер 291, 293–296, 309, 315, 317, 320, 333, 378
Френель 28, 29, 203, 211–214, 216, 220, 221, 225, 269, 277–289, 291–294, 296, 299, 304, 307–311, 314, 315, 325, 354, 374, 393, 430–432, 434–436, 439, 440, 443, 446, 447, 449, 451, 454–460, 462, 463, 472, 473, 478, 508, 523, 525, 526, 533, 571, 608–611, 613, 705
Френкель 27
Фридрих 407
Фуко 29, 32, 492–494, 697, 699
Фурье 41, 58, 225–227, 230, 234, 239, 344–346, 355, 357–362, 370, 371, 390, 391, 435, 578, 644, 645, 732

Харрис 774
Хафель 686
Хелл 417
Хохлов 776, 777
Христиансен 564
Хьюиш 572

Цейсс 123, 175, 438
Цендер 256

Цернике 240, 242, 400, 641
Циолковский 711

Черенков 271–276

Шерер 417
Шмидт 188–190
Шредингер 560
Штарк 581, 602–604
Шустер 298, 299, 302

Эйлер 28, 29
Эйнштейн 30, 31, 637, 641, 655, 656, 658, 660, 665, 669, 670, 672, 676, 682, 699, 702, 709, 742–744, 749
Эйри 378–383, 386, 392, 518, 695
Эренфест 60, 61, 736
Эрриот 763

Юнг 28, 29, 147, 210, 211, 222, 248, 403

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

Аббе способ испытания объективов 127
— число 115
Аберрации оптических систем 101, 106, 113
Аберрационная кривая 107
Аберрация света 688, 692
Абсолютно матовое тело 721
— черное тело 719
Абсолютный оптический инструмент 136
Адаптация 151
Адиабатические инварианты 725, 728, 729
Аккомодация 141
—, область 141
Активная (инверсно заселенная) среда 750
Анаберрационная поверхность 55
Анализатор 421
Апертура 96
— полной поляризации 492
— числовая 166
Апертурный угол 97
Апохромат 118, 175
Астигматизм 101, 110
Астигматическая разность 102
Астрономическая рефракция 34
Атомные плоскости 410
— прямые 410
Ахромат 118, 176
Ахроматизм положения 176
— увеличения 177

Бабине компенсатор 502
— теорема 297
Бабине Солейля компенсатор 503
Баллистическая гипотеза 665
Беккереля формула 615
Бийе билинза 213
Бинокль 184
Бора правило частот 745
Брэгга Вульфа условие 413
Брюстера закон 432, 456, 459, 465
— угол 432
Бугера закон 582

Вектор импульса–энергии 710
Венцы 319
Верде постоянная 612
Видимая величина площадки 153
Видности кривая 148
Вина закон смещения 729, 730
— парадокс 616
— постоянная 742
— теорема 726
Виньетирование 98
Водянистая влага 139
Волна де Бройля 387
— дипольная 455
— монохроматическая 201
— необыкновенная 485
— неоднородная 43
— нормальная 610
— обыкновенная 484
— однородная 43
— поверхностная 43
— поляризации 455
— стоячая световая 267
Волновое сопротивление 274
— уравнение 45
— число спектроскопическое 229
Время восстановления 59, 63
— жизни 578
— затягивания 346
— установления 200
Вторичные волны Гюйгенса 24
Вторичный спектр 113
Входной зрачок 96
— люк 98
Выходной зрачок 97
— люк 98

Галилея преобразование 656
— принцип относительности 657
Гало 319
Гамма-излучение 9, 10
Генерация волн суммарной и разностной частот 773
— гармоник 769, 773
Геометрические законы отражения и преломления 426, 543
— — — на границе металла 469
Гигантские импульсы 760
Главная оптическая ось 73
Главное зеркало 185
— сечение кристалла 484
Главные нормальные сечения 102

Главные плоскости 81
— показатели преломления кристалла 524
— радиусы кривизны 102
— скорости света в кристалле 524
Главный азимут 474
— угол падения 474
Глаз близорукий 142
— дальнозоркий 142
— схематический 143
—, камера задняя 139
—, — передняя 139
—, область аккомодации 141
—, оптическая система 139
—, острота зрения 146
Глазная линза 178
Глубина проникновения 429, 470
— резкости 99
Голограмма 363
Голография 363
Гравитационное смещение спектральных линий 702
Группа волн 61
Гюйгенса принцип 24, 537
— — в формулировке Кирхгофа 304, 307
— — — — Френеля 278

Двойное лучепреломление 486
— — круговое 608
— — при механических деформациях 518
— — — течении жидкостей 519
Дебаеграмма 417
Дебая–Шерера–Холла метод 417
Демодуляция света 209
Денисюка метод в голографии 373
Диафрагма апертурная 96
— поля зрения 98
Диоптрия 91
Дисперсионная область 330
Дисперсия аномальная 552
— временная 619
— диэлектрических осей 522
— нормальная 552
— относительная 116
— отрицательная 563
— плазмы 569
— положительная 563
— пространственная 520, 526, 527, 619
— света 40
— средняя 116
— угловая 330
— коэффициент 115
Дисторсия 112
Дисторсия, условие отсутствия 120
Дифракционная решетка 320, 325
амплитудная 325
Дифракционная решетка вогнутая отражательная 350
— — двумерная 407
Рэлея 357
— — трехмерная (пространственная) 407
— —, второстепенные максимумы 321
— —, главные максимумы 321
— —, дисперсионная область 330
— —, метод скользящего падения 326
— —, период 320, 410
— —, порядок спектра 321
— —, пропускаемость (амплитудная прозрачность) 354
— —, разрешаемое расстояние 331
— —, разрешающая способность 331, 339–349, 379
— —, — — в рентгеновской области 418, 419
Дифракция 13, 277
— рентгеновских лучей 407
— Фраунгофера 293
— — на двумерной решетке 407
— — — круглом отверстии 316
— — — отверстиях 315
— — — трехмерной решетке 407
— Френеля 293
— — от края Луны 303
— — — — экрана 302
— —, метод подобия 304
Дихроизм 495
— круговой 611
Диэлектрическая проницаемость динамическая 41
— — комплексная 468
— — металла 467
— — поперечная 483
— — продольная 483
— — статическая 41
— — эффективная 478
Диэлектрические оси 521
Диэлектрический тензор 481, 520
Длина тубуса механическая 174
— — оптическая 174
Дополнительные экраны 297
Доплера эффект 688, 690, 691, 695
Друде метод 473
Дуализм волн и частиц 32

Естественная ширина спектральных линий 577
Естественное затухание 577

Желтое пятно 140
Жесткая фокусировка 92

Закон дисперсии волн 62
— независимости световых пучков 12, 13
— обратных квадратов 158
— отражения света 12, 14
— преломления света 12, 14
— прямолинейного распространения света 12, 24
— сохранения импульса–энергии 710
— фотометрического сложения 237
Замедление хода движущихся часов 682
— — часов гравитационное 701
Зеемана эффект 596
— — простой (нормальный) 602
— — сложный (аномальный) 602
— —, классическая теория 598
— —, пи-компоненты 597
— —, сигма-компоненты 597
Зеленый луч 36
Зельмейера формула 551
Зенитное расстояние 35
Зонная пластинка 286
Зрачок глаза 97, 140
Зрение периферическое 146
— фовеальное 146
Зрительные трубы и телескопы 182

Избирательное отражение 475
— поглощение 475
Излучательная способность 717
Излучение Вавилова–Черенкова 271
— видимое 9, 10
— вынужденное (индуцированное) 743
— инфракрасное 9, 10
— переходное 276
— равновесное (черное) 713, 719
— рентгеновское 9, 10
— спонтанное 743
— температурное 717
— ультрафиолетовое 9, 10
—, спектральная плотность 230
Измерение угловых диаметров звезд 303, 402
Изображение вторичное 393
— действительное 67
— мнимое 67
— обратное 81
— оптическое 67, 103
— первичное 393
— прямое 81
— стигматическое 67
Изохроматические линии (изохроматы) 516
— поверхности 516
Иммерсия 387
Инвариантность 706
Инверсная заселенность 563
Индикатриса 545
Интенсивность излучения 716
— — удельная 716
— колебаний (света) 155, 200
Интервал 677
— времениподобный 678
— пространственноподобный 678
— световой 678
Интерференционное отражение 266
— рассеяние 453
Интерференционные полосы 204
— — равного наклона 247
— — равной толщины 244
— —, видность 223, 237
— —, локализация 245, 247
— —, полуширина 261
— —, способ наблюдения объективный 214
— —, — — субъективный 214
— —, ширина 204
— —, условие контрастности 218, 219
Интерференционный член 201
Интерференция 14, 201
— в пленках 242
— многолучевая 258
— поляризованных лучей 508, 515
—, порядок 231
—, угол 219
Интерферометр Жамена 249
— Майкельсона 256
— — звездный 404
— Маха–Цендера 256
— Рождественского 256
— Фабри–Перо 263, 339
Ионосфера 571
Искривление плоскости изображения 110
Источники света мнимые 67

Камера-обскура 13, 394
Кандела 156
Каратеодори теорема 136
Кардинальные точки 81
Картезианский овал 55
Каустика 103
Квант света 737
Керра постоянная 585
— электрооптический эффект 583
— — —, теория Борна 590

Керра электрооптический эффект, теория Ланжевена 586
Кеттелера инварианты 471
Кирхгофа закон 719
Кирхгофа–Клаузиуса формула 722
Ковариантность 706
Когерентность 201, 237
— временная 233, 238
— полная 237
— пространственная 220, 222, 238
— частичная 237
—, время 233, 238
—, длина 233, 238
—, объем 238
—, площадь 223
—, степень 236
—, — комплексная 236
Колбочки 143
Колебание квазимонохроматическое 234
— модулированное 234
Коллектив (полевая линза) 178
Коллинеарное соответствие 79
Кома 109
Компенсатор 499
Конденсор 182
Коническая рефракция 541, 543
— —, конус 538, 539
Корню спираль 299, 302
Коррекционная оправа 177
— пластинка 189
Косинусов теорема 130
Коттона–Мутона постоянная 594
— — эффект 593
Коэффициент затухания (поглощения) 582
— отражения 433
— преломления 15
— пропускания 434
— рассеяния 636
— эллиптичности 463
Кристаллическая решетка 410
Кристалл-рефрактометр 489
Кристаллы двуосные 485, 527, 536
— одноосные 483, 485, 527, 536
— отрицательные 489
— положительные 489
Критическая опалесценция 641
— температура смешения 642
Кундта постоянная 613

Лагранжа–Гельмгольца инвариант 77
— — теорема 77, 124
Лазер 748
Лазер гелий-неоновый 762
— рубиновый 754
—, трехуровневая схема 760
—, четырехуровневая схема 760
Ламберта закон 159
Ландау и Плачека формула 647
Лауэ условия 410
Лауэграмма 411
Линза магнитная 191
— одиночная 197
— с непрерывным показателем преломления 195
— толстая 93
— тонкая 77
—, фокусное расстояние 78, 198
—, формула 78, 83
Линии кривизны 102
Линия прямого зрения 141
Ллойда зеркало 213
Локальная связь 617
Лоренца–Лоренца формула 554
Лоренца преобразование 675
— сокращение длины 681
Луммера–Бродхуна фотометр 161
Луммера–Гэрке пластинка 266
Лупа 172
Луч 14, 47, 528
— главный 97, 102
— мнимый 70
— необыкновенный 486
— обыкновенный 486
— параксиальный 75
Лучи остаточные 476
Лучистый поток 153
— —, интенсивность (лучистость) 153
Люкс 158
Люмен 156

Магнитная проницаемость динамическая 41
— — статическая 41
Мазер 749
Майкельсона опыт 659
— эшелон 334
Макалюзо–Корбино эффект 615
Максвелла закон 39
— эффект 519
Малюса закон 422
— теорема 56
Малых ступеней метод 148
Мандельштама Бриллюэна явление 643
— — — — вынужденное 648
Масса покоя 709

Масса релятивистская 709
Маха волны 275
Менисковые системы Максутова 188
Меслина опыт 220
Метод крюков Рождественского 566
Механика релятивистская 706
Механический эквивалент света 157
Микроволны 9
Микроскоп 174, 182, 387
Мировая линия 677
— точка 677
Многослойные покрытия 267
Модуляция амплитудная 234
— фазовая 234
Молекулы полностью анизотропные 586
— полярные 556
Монохроматическое векторное поле 423
Монохроматы 148
Мосотти–Клаузиуса формула 554
Мутная среда 632

Направления синхронизма 772
Некогерентность 201, 237
Нелокальность временная 617
— пространственная 617
Ньютона кольца 247

Обратная связь 751
Объектив иммерсионный 125
Объекты абсорбционные 399
— рефракционные 400
Одновременность 667
Окуляр Гюйгенса 178
— Кельнера 180
— компенсационный 176
— Рамсдена 178, 180
Оптика 9
— геометрическая 12, 44
— ионная 191
— линейная 42
— нелинейная 42, 748
— параксиальная (гауссова) 78
— физическая 9
— электронная 69, 191
Оптическая накачка 755
— область спектра 10
— сила 91
Оптические оси 483
— — второго рода (бинормали) 527
— — первого рода (бирадиали) 535
Оптический интервал 89
Оптическое детектирование 768
Ортотомная система 56
Освещенность 157
Освещенность естественная 166
Освещенность энергетическая (облученность) 157
Относительное отверстие 165
Отражение диффузное 17
избирательное 267
интерференционное 266
— правильное 16
Очки 143

Палочки 143
Парадокс близнецов 687
Параметрическая генерация света 776
Параметрическое усиление света 777
Перспектива 100
Плазменная (ленгмюровская) частота 570
Планка постоянная 10, 739
— формула 739
Пластинка в четверть волны 499
— — — —, главные направления 499
Плеохроизм 495
Плоскость поляризации (колебаний) 38, 420, 433
— —, вращение естественное 605, 620
— —, — магнитное (эффект Фарадея) 605, 611
— установки (наводки) 99
Плотность лучистой энергии 153, 715
— — — спектральная 715
Поверхностная прозрачность 434
Поверхность волновых векторов 545
— лучевая (волновая) 529
— нормалей 525
— показателей преломления 545
— равных амплитуд 43
— — фаз 43
Поглощательная способность 717
Показатель затухания 470
— — главный 468
— преломления 470
— — главный 468
— — комплексный 468
— — межпланетного и межзвездного пространств 453
— — необыкновенный 486
— — обыкновенный 486
— — рентгеновских лучей 453
Поккельса линейный электрооптический эффект 594
Покровное стекло 177
Полное отражение 16, 429, 436, 546
— —, предельный угол 16, 429

Полуширина спектральной линии 578
Поля опыт 214
Поляризатор 421
Поляризация вакуума 30
Поляризация света круговая 423
— — линейная 39, 420
— — эллиптическая 422, 473
— — — при полном отражении 438
Поляризуемость 766
Поляроиды 421, 495
Правило сумм 562
Прево правило 719
Призма дисперсионная 119
— — прямого зрения 119
— — сложная 119
— — разрешающая способность 339, 349
— поляризационная Аренса 493
— — Волластона 495
— — Глазебрука 493
— — Глана 493
— — Глана–Томпсона 493
— — Николя (николь) 491
— — Рошона 495
— — Сенармона 495
— — Франка–Риттера 493
— — Фуко 493
Призмы поляризационные двухлучевые 491
— — однолучевые 491
Принцип детального равновесия 713
— неопределенности 229
— суперпозиции 199
— таутохронизма 70
Просветление оптики 444
Просветляющие фильтры 759
Пространство изображений 69
— предметов 69
Пуркинье явление 149
Пучок лучей астигматический 103
— — гомоцентрический 67
— — опорный 364
— — предметный 364

Радиоволны 9
Радужная оболочка 139
Разность хода 205
— — оптическая 205
Разрешаемое расстояние 146, 379, 383, 386, 391, 393
Разрешающая способность 146
— — глаза 384
— — голограммы 393
— — микроскопа 386, 391, 393
— — объектива 378
Разрешающая способность спектральная 230, 263, 348
— — телескопа 383
— —, критерий Рэлея 332
Рассеяние света 13, 277, 632
— — молекулярное (рэлеевское) 637
— — —, деполяризация 639
— — —, тонкая структура 641
— — — комбинационное 650
— — — — вынужденное 653
— — — поверхностное 642
— — тиндалевское 632
Расстояние ясного зрения 142
Рефлектор 185
Рефрактометр интерференционный 251
Рефрактор 185
Рефракция 34
Рефракция атомная 555
— молекулярная 555
— удельная 554
Роговица 139
Родопсин 151
Рождение пар 30
Роулэнда круг 352
«Рыбий глаз» 139
Рэлея–Джинса формула 735
Рэлея закон 636
— формула 639

Самоканализация 775
Самофокусировка 775
Сателлиты 650
— красные (стоксовы) 651
— фиолетовые (антистоксовы) 651
Свет естественный 420
— квазимонохроматический 208, 234
— поляризованный 420
— — частично 420
Светимость 160
— энергетическая (излучаемость) 160
Световая эффективность 156
Световое давление 558
— поле вблизи фокуса 374
Световой вектор 269
— конус 680
— поток 155
Световые (фотометрические) величины 155
Светосила 165
Свеча международная 156
Связь массы с энергией 709
Сетчатка (ретина) 140, 143

Сила осциллятора 561
— света источника 155
— — — средняя сферическая 156
Система афокальная (телескопическая) 81, 84
— диоптрическая 83
— катоптрическая 83
Скачок фазы при отражении 234, 243, 432
— — — — — полном 438
Скин-эффект аномальный 479
Скин-эффект нормальный 478
Склера 139
Скорость групповая 57, 60, 63, 528
— движения энергии 64
— лучевая 528
— нормальная 55, 528
— передового фронта 65
— света в вакууме 667
— фазовая 58
Слепое пятно 144
Сложение скоростей релятивистское 703
Собственная длина 681
Собственное время 683
События абсолютно будущие 679
— — прошедшие 679
— — удаленные 679
Сопряженные точки 67
Сопутствующая система отсчета 683
Сопутствующие часы 683
Сосудистая оболочка 139
Спектр нормальный 330
— призматический 330
Статистически независимые величины 626
Стационарные состояния 89
Стекловидное тело 139
Степень поляризации 446
Стереографическая проекция 137
Стереотруба 184
Стефана–Больцмана закон 724
Стефана–Больцмана постоянная 742
Стильб 159
Столетова стопа 433
Сферическая аберрация 107
— — поперечная 108
— — продольная 108

Тахионы 680
Телескоп Гершеля–Ломоносова 186
— Грегори 187
— Кассегрена 187
— Ньютона 186
Телефон световой 437
Тени 13
Теорема Ван-Циттера–Цернике 242
— обращения 532
Теория света волновая 23
Теория света корпускулярная 21
— — электромагнитная 29
Ток поляризации 467
— проводимости 467
Точка ясного видения ближняя 141
— — — дальняя 141
Точки апланатические 122
— главные 81
— — обратные 85
— закругления 102
— ортоскопические 121
— узловые 84
— — обратные 85
— фокальные 80, 102
Третичный спектр 118
Труба Галилея 184
— Кеплера 183
Турмалин 495

Увеличение зрительной трубы 87
— лупы 173, 388
Увеличение микроскопа 170, 172, 388
— нормальное 167
— осевое (продольное) 85
— поперечное 81
— угловое 84
Угол аберрации 693
— зрения 141
— преломления вещественный 470
— проекции 97
— раскрытия 97
— рассеяния 634
— синхронизма 772
Уголковый отражатель 18
Ультрамикроскоп 387
Ультрафиолетовая катастрофа 736
Уравнение адиабаты для излучения 725
Ускоритель на встречных пучках 711
Условие Гершеля 131
— синусов 122, 131
— тангенсов (ортоскопии) 121
Уширение спектральных линий доплеровское 581
— — — естественное 577
— — — ударное 580
— — — штарковское 581

Фазовая пластинка 402
Фазовый контраст 399
— — негативный 402

Фазовый контраст позитивный 402
— синхронизм 771, 777, 778
Ферма принцип 49
Флуктуация 626
— относительная 626
— среднеквадратичная 626
Фокальные отрезки 103
— плоскости 80
Фокус 67
— главный 80
— параксиальный 106
Формула зеркала 75, 83
Фотон 32
Френеля бипризма 212
— закон лучевых скоростей света в кристалле 533
— — нормальных скоростей света в кристалле 523
— зеркала 211
— коэффициенты 431
— параллелепипед 439
— формулы 431, 472
Функция автокорреляционная 236
— корреляционная 236
— — взаимная 236
Фурье интеграл 226
Фурье ряд 225

Характеристическая функция 129
Хроматическая поляризация 513
Хрусталик 139, 141

Цветная фотография 270
— Липпмана 269
Цветного зрения теория 147
Центр интерференционной картины 230
Центральная ямка 141
Центрированная оптическая система 73

Часы идеальные 683
Черенковские счетчики 275
Четырехмерный вектор 707

Шмидта камера 188
Штарка эффект 602
Шустера зоны 298

Эйконал 46
точечный 129
, уравнение 46
Эйнштейна коэффициенты 744
— принцип относительности 658
— формула 641
Эйри кружок 378
— спирали 518
Электроны оптические 548
Энантиоморфизм 606
Энергетическая сила источника 155
Энергетические величины 155
Энергия кинетическая релятивистская 709
покоя 709
— полная 709
Эрмитов тензор 520
Эфир световой (мировой) 23
Эффективная температура излучения 762

Юнга опыт 210

Яркость 158
— энергетическая 158

Учебное издание

СИВУХИН Дмитрий Васильевич

ОБЩИЙ КУРС ФИЗИКИ

Том IV

ОПТИКА

Редактор *Д. А. Миртова*
Оригинал-макет: *О. Б. Широкова*
Оформление переплета: *А. А. Логунов*

ЛР № 071930 от 06.07.99
Подписано в печать 14.01.05. Формат 60×90/16
Бумага офсетная № 1. Печать офсетная
Усл. печ. л. 49,5. Уч.-изд. л. 51,17. Заказ №

Издательская фирма «Физико-математическая литература»
МАИК «Наука/Интерпериодика»
117864 Москва, Профсоюзная, 90
E-mail: fizmat@maik.ru; http://www.fml.ru

Отпечатано с диапозитивов
в РГУП «Чебоксарская типография № 1»
428019 Чебоксары, пр. И. Яковлева, 15

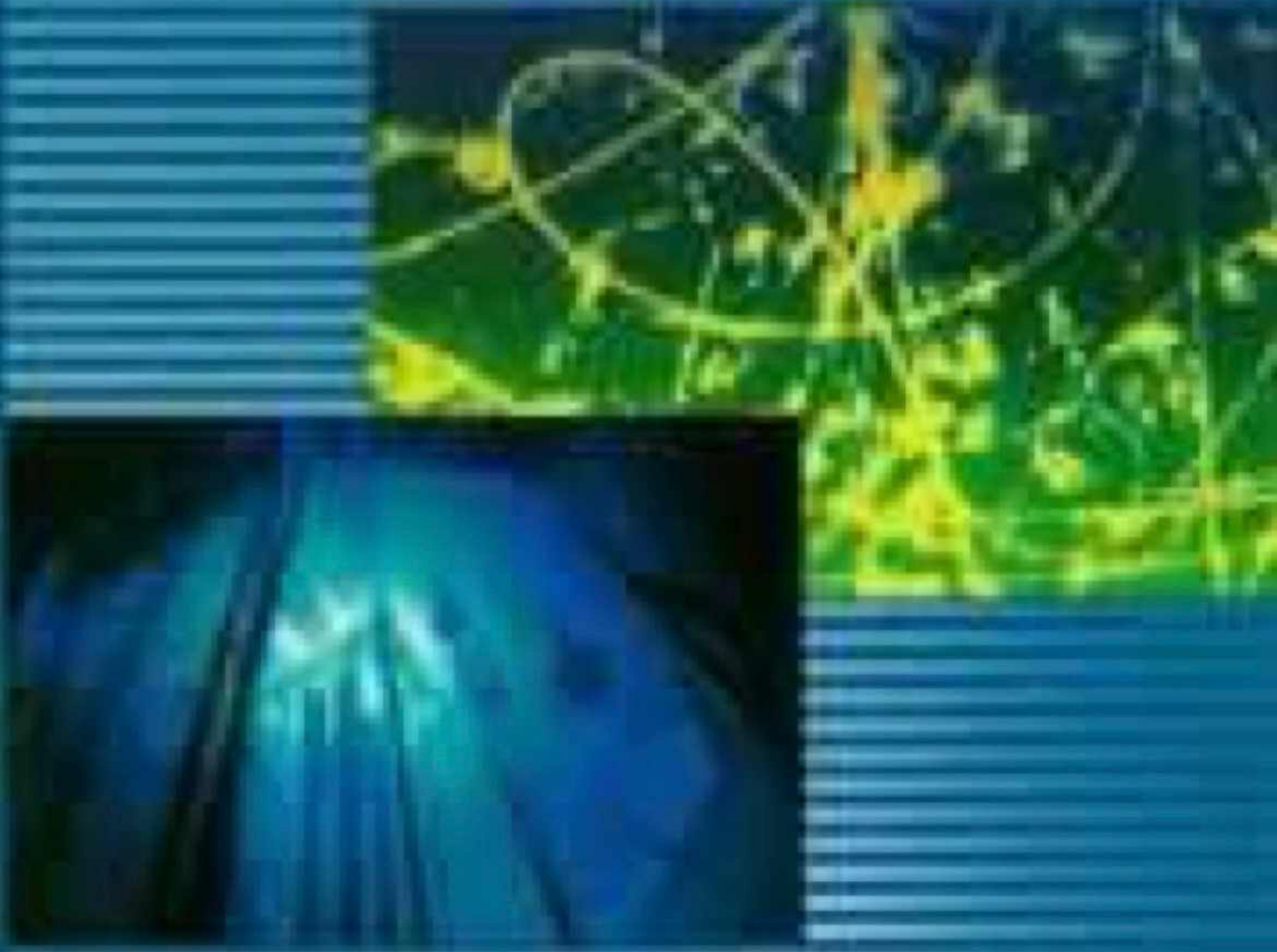
ISBN 5-9221-0228-1

ОБЩИЙ КУРС ФИЗИКИ

Д.В. Сивухин

АТОМНАЯ И ЯДЕРНАЯ ФИЗИКА

ТОМ V



УДК 530.1(075.8)
С34
ББК 22.38

Сивухин Д. В. **Общий курс физики.** Учеб. пособие: Для вузов. В 5 т. Т. V. **Атомная и ядерная физика.** — 2-е изд., стереот. — М.: ФИЗМАТЛИТ; Изд-во МФТИ, 2002. — 784 с. — ISBN 5-9221-0230-3; 5-89155-088-1.

Пятый том курса физики, широко известного у нас в стране и за рубежом. Книга написана на основе лекций, которые в течение ряда лет читались автором студентам Московского физико-технического института. Основное внимание уделено выяснению физического смысла и содержания основных законов и понятий атомной и ядерной физики, установлению границ применимости этих законов, развитию у студентов навыков физического мышления и умения ставить и решать конкретные задачи.

Первое издание пятого тома вышло в двух частях (в 1986 г. — первая часть, в 1989 г. — вторая).

Для студентов физических и математических факультетов университетов, физико-технических и инженерно-физических институтов, а также вузов, где физика является основной дисциплиной.

ISBN 5-9221-0230-3 (Т. V)
ISBN 5-9221-0229-X
ISBN 5-89155-088-1 (Т. V)
ISBN 5-89155-077-6

© ФИЗМАТЛИТ, 1989, 2002
© Издательство МФТИ, 2002

ОГЛАВЛЕНИЕ

Предисловие 7

Глава I. **Кванты света**

1. Энергия и импульс светового кванта 9
2. Фотоэлектрический эффект 14
3. Эффект Комптона 26
4. Эффект Доплера при движении источника света в вакууме с фотонной точки зрения 34
5. Отражение и преломление света в фотонной теории. Фотоны в среде 37
6. Излучение Вавилова–Черенкова. Эффект Доплера при движении источника света в среде 40
7. Фотоны в гравитационном поле 44
8. Некоторые опыты по обнаружению корпускулярных свойств света 46

Глава II. **Строение, энергетические уровни и спектры атома**

9. Ядерная модель атома и опыты Резерфорда 50
10. Определение заряда ядра из рассеяния рентгеновских лучей . . . 58
11. Спектральные закономерности 61
12. Постулаты Бора 64
13. Спектр водорода 67
14. Экспериментальное подтверждение постулатов Бора 79
15. Резонансное свечение и люминесценция 86
16. Принципиальные недостатки теории Бора 89

Глава III. **Волновые свойства частиц вещества**

17. Гипотеза де Бройля 92
18. Экспериментальные подтверждения гипотезы де Бройля 99
19. Статистическая интерпретация волн де Бройля и волновой функции 109
20. Соотношение неопределенностей 117

Глава IV. **Уравнение Шредингера. Квантование**

21. Уравнение Шредингера 128
22. Уравнение Шредингера и квантование 133
23. Гармонический осциллятор 138
24. Одномерные прямоугольные потенциальные ямы 142
25. Квантование в случае сферически симметричного силового поля 147
26. Система двух взаимодействующих частиц 149

27. Квантование водородоподобного атома в сферически симметричном случае . . . 153
28. Потенциальные барьеры . . . 157
29. К объяснению контактной разности потенциалов. Холодная эмиссия электронов из металлов . . . 167

Глава V. **Дальнейшее построение квантовой механики и спектры**

30. Операторный метод . . . 172
31. Момент импульса частицы . . . 181
32. Сложение угловых моментов . . . 190
33. Квантование водородного атома в общем случае . . . 195
34. Энергетические уровни и спектральные серии щелочных металлов . . . 199
35. Магнетизм атомов . . . 207
36. Опыты Штерна и Герлаха. Спин электрона . . . 211
37. Эффект Садовского и спин фотона . . . 217
38. Четыре квантовых числа электрона и тонкая структура спектральных термов . . . 226
39. Правила отбора при излучении и поглощении света . . . 234
40. Тонкая структура спектральных линий водорода и щелочных металлов . . . 238
41. Простой и сложный эффект Зеемана . . . 242
42. Магнитный резонанс . . . 250
43. Эффект Штарка . . . 259
44. Лэмбовский сдвиг уровней атомных электронов . . . 263
45. Физический вакуум и объяснение лэмбовского сдвига . . . 266

Глава VI. **Атомные системы со многими электронами**

46. Принцип тождественности одинаковых частиц. Принцип Паули . 270
47. Объяснение периодической системы химических элементов Д. И. Менделеева . . . 276
48. Рентгеновские лучи . . . 285
49. Атом гелия . . . 298
50. Химическая связь. Молекула водорода . . . 307
51. Параводород и ортоводород . . . 315
52. Молекулярные силы . . . 317

Глава VII. **Некоторые макроскопические квантовые явления**

53. Возможные состояния частицы в ограниченном объеме . . . 322
54. Теория Дебая теплоемкости твердых тел . . . 324
55. Типы связей атомов в твердых телах . . . 331
56. Колебания атомов в одномерной прямолинейной цепочке . . . 333
57. Фононы и квазичастицы . . . 340
58. Энергетические зоны в твердых телах . . . 348
59. Зонная структура и волны Блоха . . . 354
60. Сверхтекучесть. Опытные факты . . . 365
61. Понятие о теории сверхтекучести . . . 373
62. Понятие о теории сверхпроводимости . . . 381

Глава VIII. **Статические свойства атомного ядра**
63. Введение 390
64. Энергия связи ядра 400
65. Размеры ядра 410
66. Спин ядра и сверхтонкая структура спектральных линий 416
67. Влияние спина ядра на эффект Зеемана 427
68. Измерения спинов и магнитных моментов ядер методом магнитного резонанса. Опытные данные о спинах и магнитных моментах ядер 429
69. Четность. Закон сохранения четности 431
70. Электрические свойства и форма ядра 437

Глава IX. **Радиоактивность**
71. Введение 442
72. Законы радиоактивного распада 450
73. Альфа-распад 455
74. Бета-распад 467
75. Гамма-излучение ядер и внутренняя конверсия электронов 483
76. Эффект Мёссбауэра 487

Глава X. **Краткие сведения о ядерных моделях**
77. Общие сведения 495
78. Оболочечная модель ядра 498

Глава XI. **Прохождение заряженных частиц и гамма-квантов через вещество**
79. Введение 510
80. Прохождение тяжелых заряженных частиц через вещество 511
81. Прохождение легких заряженных частиц через вещество 519
82. Прохождение гамма-квантов через вещество 524
83. Другие проявления взаимодействия ядерных частиц с веществом 530

Глава XII. **Источники и методы регистрации ядерных частиц**
84. Ускорители 534
85. Источники нейтронов и других нейтральных частиц 555
86. Детекторы частиц 560

Глава XIII. **Ядерные реакции**
87. Терминология и определения 575
88. Законы сохранения в ядерных реакциях 579
89. Составное ядро 587
90. Ядерные реакции, идущие через составное ядро 590
91. Дополнительные сведения о ядерных реакциях 594

Глава XIV. **Нейтроны и деление атомных ядер**
92. История открытия нейтрона 602
93. Деление атомных ядер 606
94. Трансурановые элементы 617
95. Цепная реакция и ядерные реакторы 636

96. Природный ядерный реактор в Окло 649
97. Использование антинейтрино для контроля ядерного реактора . . 651
98. Термоядерная проблема 654
99. Нейтронная оптика 669

Глава XV. **Некоторые вопросы астрофизики**
100. Источники энергии звезд 683
101. Некоторые сведения из астрономии 695
102. Краткие сведения об эволюции звезд 699
103. Космические лучи 716

Глава XVI. **Элементарные частицы**
104. Что такое элементарные частицы 733
105. Классификация элементарных частиц 736
106. Античастицы 739
107. Законы сохранения энергии и импульса и их приложения 742
108. Законы сохранения электрического, лептонных и барионного зарядов 749
109. Другие законы сохранения и квантовые числа 753
110. Кварковая модель адронов 758

Таблицы 766
Именной указатель 769
Предметный указатель 773

Предисловие

Понимание явлений атомной физики, как и всех явлений микромира, невозможно без квантовых представлений. Поэтому в настоящем томе затрагиваются основные понятия и принципы квантовой механики. Делается это на основе экспериментальных фактов, которые одни только с полной убедительностью и вынуждают нас принять квантовые представления. Разумеется, в общей физике количественному рассмотрению доступны только простейшие задачи. Большинство вопросов рассматривается качественно на основе общих принципов.

В книге рассматриваются не только явления, происходящие в электронных оболочках атомов и молекул. Следуя историческому ходу развития, вначале вводится представление о фотонах и рассматриваются оптические явления, связанные с таким представлением. Затем идет основной материал, относящийся уже к атомной физике в узком смысле — физике электронных оболочек атомов. Наконец, коротко говорится о макроскопических квантовых явлениях — сверхтекучести, сверхпроводимости и пр.

Ядерная физика излагается в этой книге как раздел общего курса физики. Книга не ставит целью подробное изложение всех проблем этой науки. Почти все главы и параграфы книги при подробном изложении могли бы составить содержание специальных монографий. К ним и следует обращаться всем специализирующимся в различных областях ядерной физики. Здесь же приводится, да и то часто в общих чертах, только такой материал, в котором должен ориентироваться студент-физик, независимо от того, какую узкую специальность он изберет в дальнейшем.

В ядерной физике приходится в значительно большой степени апеллировать к опытным фактам, а не к теории, как в других разделах общей физики. Причина этого двоякая. Во-первых, еще не существует законченной теории ядерных сил, а следовательно, и основанной на ней будущей фундаментальной и последовательной теории ядра. Приходится довольствоваться основанными на опыте ядерными моделями, каждая из которых имеет ограниченную область применения. Во-вторых, по педагогическим соображениям нецелесообразно уводить студента, изучающего общую физику, в сторону сложных теорий, которые ему еще не доступны.

Эксперименты в области ядерной физики (и в особенности физики элементарных частиц) отличаются высокой степенью автоматизации с широким использованием ЭВМ. Поэтому обучение эксперименту

в этих областях физики должно проводиться не в лекционных аудиториях и лабораториях физического практикума, а в специальных институтах и лабораториях, изучающих ядерную физику и физику элементарных частиц. В общем курсе физики можно остановиться только в самых общих чертах на принципиальной стороне этих вопросов.

В главе XV кратко затронуты некоторые вопросы астрофизики. Источники энергии и эволюция звезд — это вопросы, с которыми должен быть знаком каждый физик. Более того, с ними, по крайней мере качественно, должен быть знаком и всякий образованный человек, независимо от его специальности, ибо их мировоззренческое значение несомненно.

В главе XVI приводятся основные сведения об элементарных частицах. Изложение здесь еще в большей степени не претендует на полноту, чем в остальных главах. Приводятся в основном только те сведения, которые необходимы в ядерной физике в узком смысле слова, т. е. в физике самого атомного ядра.

Физике элементарных частиц в современном ее состоянии должна быть посвящена отдельная книга и даже ряд книг.

Как и предыдущие тома, эта книга составлена на основе лекций, которые автор многократно читал в Московском физико-техническом институте.

Как и в остальных томах «Общего курса физики», в этой книге в качестве основной применяется система единиц СГС в ее гауссовой форме. Используется также оправдавшая себя удобная внесистемная единица энергии электронвольт (эВ) с ее кратными единицами (кэВ, МэВ, ГэВ). Международная система (СИ) совсем не применяется.

По мнению автора, в физике эта система неудовлетворительна. О ее принципиальных недостатках было сказано в предисловиях к I и III томам, а более подробно в § 85 тома III. Автор с удовлетворением прочитал, что в популярной книге по физике элементарных частиц известный физик-теоретик лауреат Нобелевской премии С. Вайнберг (р. 1933) отметил, что, по его мнению, введение СИ в физику было ошибкой. Автор может только присоединиться к этому мнению.

Рукопись этой книги рецензировалась членом-корреспондентом АН СССР профессором С. С. Герштейном, а также на кафедре экспериментальной физики Киевского государственного университета им. Т. Г. Шевченко, возглавляемой членом-корреспондентом АН УССР профессором И. С. Горбанем. Она была внимательно просмотрена заведующим кафедрой теории ядра и элементарных частиц Киевского государственного университета профессором В. К. Тартаковским. Много критических замечаний было сделано Г. И. Новиковой при редактировании книги. Все эти критические замечания способствовали улучшению книги и были учтены при окончательном ее редактировании. Всем перечисленным лицам автор выражает глубокую благодарность.

Глава I

КВАНТЫ СВЕТА

§ 1. Энергия и импульс светового кванта

1. К середине XIX века волновая природа света считалась доказанной окончательно. Ее подтверждали явления интерференции и дифракции света. А опыт Фуко (1819–1868), казалось, исключал всякую возможность корпускулярных представлений о свете (см. т. IV, § 3). Это действительно было бы так, если бы имелась в виду корпускулярная теория в ньютоновской форме (см. § 5).

Однако и волновая теория света, даже в ее электромагнитной форме, оказалась недостаточной для истолкования всей совокупности оптических явлений. Впервые это было осознано при рассмотрении проблемы *равновесного* (*черного*) *излучения.* Настойчивые попытки решить эту проблему в рамках волновых представлений на основе классических электродинамики и статистики закончились неудачей (см. т. IV, § 117). Формула, согласующаяся с опытом во всем диапазоне длин волн, была угадана Планком (1858–1947) в октябре 1900 г., сначала эмпирически. Немного позже Планк нашел и теоретический вывод своей формулы, доложенный им 14 декабря 1900 г. на заседании Немецкого физического общества. Это было исходным пунктом возникновения принципиально новых — *квантовых* — представлений. Сначала они касались только природы света, но затем постепенно проникли во все разделы физики.

Оказалось, что понятия и принципы классической физики, возникшие на основе изучения макроскопических объектов, неприменимы или ограниченно применимы в области атомных и субатомных масштабов. В этой области потребовались новые представления и законы, которые в конце концов и были найдены. Они составили основу новой, так называемой *квантовой* физики. Однако излагать квантовую физику систематически и дедуктивно, полностью отвлекаясь от истории ее развития, как это делается в теоретической физике, в общей физике было бы нецелесообразно. Сначала надо ознакомить начинающего с основными опытными фактами, которые одни только и могут убедить его в недостаточности и ограниченной применимости классических представлений. Они же, и это главное, позволяют наметить пути для введения новых представлений. Именно такой метод изложения принят в настоящей книге, дающей элементарное введение в квантовую физику.

2. Вернемся, однако, к истокам квантовых идей. При выводе своей формулы для равновесного излучения Планк ввел чуждую классической физике гипотезу, что *излучение и поглощение света веществом происходит не непрерывно, а конечными порциями,* или *квантами.* Так

как свойства равновесного излучения в полости не зависят от вещества стенок полости, вещество без ограничения общности он рассматривал как совокупность гармонических осцилляторов. А чтобы согласовать свою гипотезу с законами термодинамики и электродинамики, Планк принял, что энергия кванта $\mathscr{E}$, излучаемая или поглощаемая гармоническим осциллятором частоты ν, определяется выражением

$$\mathscr{E} = h\nu, \tag{1.1}$$

где h — универсальная постоянная, получившая название *постоянной Планка* (см. т. IV, § 118). Ее значение было вычислено самим Планком из экспериментальных результатов, полученных при изучении распределения энергии в спектре излучения *абсолютно черного тела.* Однако постоянная Планка, как и всякая фундаментальная постоянная, входит в множество других физических явлений. Все они дают независимые способы определения этой постоянной и в пределах ошибок измерений приводят к согласующимся результатам. По современным данным

$$h = 6{,}626176(36) \cdot 10^{-27} \text{ эрг} \cdot \text{с}\ ^{1)}. \tag{1.2}$$

В некоторых вопросах, в особенности в теоретической физике, более удобна постоянная

$$\hbar = 1{,}0545887(57) \cdot 10^{-27} \text{ эрг} \cdot \text{с}, \tag{1.3}$$

введенная Дираком (1902–1984). Ее также называют *постоянной Планка — перечеркнутой* или *дираковой.* Через эту постоянную энергия кванта излучения выражается формулой

$$\mathscr{E} = \hbar\omega, \tag{1.4}$$

где $\omega = 2\pi\nu$ — *циклическая частота излучения.* Мы будем пользоваться как выражением (1.1), так и выражением (1.4).

3. Сам Планк, как видно из изложенного, полагал, что квантовые свойства света проявляются *только в актах излучения и поглощения,* т. е. *при взаимодействии света с веществом.* Распространение же света в пространстве, по его воззрениям, происходит *непрерывно* и описывается классическими уравнениями Максвелла (1831–1879). Более радикальная и законченная форма была придана квантовой теории света Эйнштейном (1879–1955) в 1905 г. Руководствуясь некоторыми теоретическими соображениями и экспериментальными фактами, Эйнштейн пришел к представлению, что и *при распространении в пространстве свет ведет себя подобно совокупности каких-то частиц,* причем энергия каждой частицы определяется формулой Планка (1.1) или (1.4). Такие частицы позднее получили название *квантов света* или *фотонов.*

---

[1]) Здесь, как и всюду в дальнейшем, в скобках приведено стандартное отклонение последних двух цифр от их среднего значения.

Это не было простым возвратом к ньютоновской корпускулярной теории света. Нельзя смотреть на фотоны как на обычные частицы света, аналогичные материальным точкам классической механики и движущиеся по определенным траекториям в пространстве. Это видно уже из того, что фотонам свойственна *интерференция* и *дифракция.* Они обладают не только корпускулярными, но и *волновыми свойствами.* Такая особенность фотонов называется *корпускулярно-волновым дуализмом.* Было бы безнадежно пытаться истолковать корпускулярно-волновой дуализм в духе представлений классической физики. Человеческое воображение не в состоянии создать образ, обладающий одновременно и свойствами корпускулы, и свойствами волны. Однако природа богаче воображения человека. При ее изучении надо руководствоваться не тем, что доступно воображению человека, а тем, что дают наблюдения и опыт. Отметим уже сейчас, что обычные корпускулы — электроны, нейтроны, атомы, молекулы и пр. — также обладают волновыми свойствами. Опыты, заставляющие принять это заключение, будут рассмотрены в § 18. Поэтому обсуждение вопроса, как современная физика истолковывает корпускулярно-волновой дуализм, мы отложим до § 19, после того как будут изучены волновые свойства вещества.

4. Если фотон обладает энергией, то он должен обладать и импульсом, как этого требует теория относительности. Импульс фотона проявляется, например, в давлении света. Связь между энергией $\mathscr{E}$ и импульсом $\mathbf{p}$ при движении частицы в теории относительности выражается формулой

$$(\mathscr{E}/c)^2 - \mathbf{p}^2 = (m_0 c)^2 \tag{1.5}$$

(см. т. IV, § 111). (При этом предполагается, что во время движения внутреннее состояние частицы, а с ним и ее масса покоя m_0 остаются без изменения.) Фотон движется в вакууме со скоростью света c, т. е. является *релятивистской частицей.* Если бы масса покоя фотона m_0 была отлична от нуля, то его релятивистская масса

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}}$$

была бы бесконечно велика. Поэтому надо допустить, что для фотона $m_0 = 0$. В результате предыдущее соотношение запишется в виде

$$\mathscr{E} = pc. \tag{1.6}$$

(Знак минус при извлечении квадратного корня следует опустить, принимая, что импульс фотона $\mathbf{p}$ направлен в сторону распространения света.)

Неклассический характер соотношения (1.1) или (1.4) проявляется, между прочим, в том, что по классическим представлениям энергия должна быть связана не с частотой, а с *амплитудой* колебания. По этим представлениям корпускулярно-волнового дуализма не должно быть. Но если такой дуализм все же существует, то связь между корпускулярными и волновыми свойствами излучения не может ограничиваться соотношением (1.1) или (1.4). Корпускулярные свойства излучения

характеризуются *энергией* $\mathscr{E}$ и *импульсом* $\mathbf{p}$, волновые — *частотой* ω и волновым вектором $\mathbf{k}$. В теории относительности величины $\mathscr{E}$ и $\mathbf{p}$ объединяются в один *четырехмерный вектор энергии-импульса* $(\mathscr{E}, c\mathbf{p})$. Фаза волны $\omega t - \mathbf{kr}$, как показано в т. IV, § 107, *инвариантна относительно преобразования Лорентца.* А так как $(t, \mathbf{r}/c)$ — четырехмерный вектор, то отсюда следует, что частота ω и волновой вектор $\mathbf{k}$ также объединяются в *четырехмерный вектор* $(\omega, c\mathbf{k})$. Временные компоненты $\mathscr{E}$ и ω четырехмерных векторов $(\mathscr{E}, c\mathbf{p})$ и $(\omega, c\mathbf{k})$ одинаково преобразуются при преобразованиях Лорентца. Поэтому соотношение (1.1) (или (1.4)) удовлетворяет необходимому требованию *релятивистской инвариантности.* Но релятивистски инвариантное соотношение не может ограничиться связью только между временными компонентами четырехмерных векторов $(\mathscr{E}, c\mathbf{p})$ и $(\omega, c\mathbf{k})$. Связь должна существовать между самими *четырехмерными векторами.* Отсюда следует, что если гипотеза Планка $\mathscr{E} = \hbar\omega$ верна, то из нее и из требования релятивистской инвариантности с неизбежностью вытекает, что $\mathbf{p} = \hbar\mathbf{k}$. Поэтому мы в качестве гипотезы принимаем, что

$$\mathscr{E} = \hbar\omega, \quad \mathbf{p} = \hbar\mathbf{k}. \tag{1.7}$$

При этом для фотона $\mathscr{E}^2 - (\mathbf{p}c)^2 = (m_0c^2)^2 = 0$, т. е. обращение в нуль массы покоя фотона m_0 эквивалентно утверждению, что для фотона четырехмерный вектор энергии-импульса $(\mathscr{E}, c\mathbf{p})$ является световым вектором.

5. При взаимодействии с веществом фотоны могут испускаться, поглощаться и рассеиваться. Сохранение числа фотонов не имеет места. Зато должны выполняться *законы сохранения энергии и импульса.*

Свободный электрон может только рассеять, но не испустить или поглотить фотон. Чтобы показать это простейшим способом, воспользуемся системой отсчета, в которой электрон первоначально покоился. Пусть электрон испустил фотон с импульсом $\mathbf{p}_\text{ф}$ и энергией $\mathscr{E}_\text{ф}$. Обозначим импульс электрона после испускания через $\mathbf{p}_\text{э}$, а энергию — через $\mathscr{E}_\text{э}$. Из законов сохранения импульса и энергии следует

$$\mathbf{p}_\text{э} + \mathbf{p}_\text{ф} = 0, \quad \mathscr{E}_\text{э} + \mathscr{E}_\text{ф} = m_0c^2,$$

где m_0 — масса покоя электрона. Отсюда

$$(c\mathbf{p}_\text{э})^2 = (c\mathbf{p}_\text{э})^2, \quad \mathscr{E}_\text{э}^2 = \mathscr{E}_\text{ф}^2 - 2\mathscr{E}_\text{ф}m_0c^2 + (m_0c^2)^2.$$

Вычтем первое равенство из второго. Тогда с учетом соотношения (1.6) для фотона и соотношения (1.5) для электрона получим

$$\mathscr{E}_\text{ф}m_0c^2 = 0.$$

Отсюда следует $\mathscr{E}_\text{ф} = 0$, т. е. испускание невозможно. Таким же рассуждением убеждаемся, что невозможно и поглощение.

Полученный результат в известном смысле тривиален. Доказательство молчаливо предполагало, что масса покоя электрона до испускания равна его массе покоя после испускания. Это значит, что *внутреннее состояние электрона в результате испускания не изменилось.* В таком

случае полная энергия электрона может только возрастать за счет кинетической энергии, получаемой электроном при отдаче во время испускания. Испущенный фотон в свою очередь несет положительную энергию. Если бы испускание было возможно, то оно сопровождалось бы нарушением закона сохранения энергии.

В заключение несколько слов об обозначениях. В физике элементарных частиц под массой частицы принято понимать *массу покоя* и обозначать ее через m (опуская нуль в индексе). Однако, поскольку наряду с массой покоя нам придется пользоваться и релятивистской массой, мы сохраним для этих величин стандартные обозначения m_0 и m. Это в особенности будет делаться тогда, когда конкретный вид частицы не играет роли. Там же, где речь идет об электронах, протонах, нейтронах и других элементарных частицах, под $m_\mathrm{e}, m_\mathrm{p}, m_\mathrm{n}, \ldots$ мы будем понимать их массы покоя. Часто в общих рассуждениях массу покоя удобно обозначать через m_0, чтобы отличить ее от релятивистской массы m. В окончательных же результатах m_0 целесообразно заменить на $m_\mathrm{e}, m_\mathrm{p}, m_\mathrm{n}, \ldots$, чтобы явно указать, о массе покоя какой частицы идет речь.

ЗАДАЧИ

1. Определить длину волны λ_K, при которой энергия светового кванта равна энергии покоя электрона. Такая длина волны называется *комптоновской длиной* для электрона (см. § 3).

Ответ. $\lambda_\mathrm{K} = h/m_0c \approx 2{,}43 \cdot 10^{-10}$ см, где m_0 — масса покоя электрона.

2. Определить релятивистскую массу m светового кванта с длиной волны λ.

Ответ. $m = (\lambda_\mathrm{K}/\lambda)m_0$.

3. Если бы фотон обладал массой покоя $m_\mathrm{ф}$, то скорость света в вакууме должна была бы зависеть от длины волны. Исследуя экспериментально эту зависимость, можно было бы оценить нижний предел для массы фотона. Найти выражения для фазовой и групповой скорости света в вакууме в предположении, что $m_\mathrm{ф} \neq 0$.

Решение. Энергия фотона $\hbar\omega$ и его импульс $\hbar k$ связаны соотношением

$$(\hbar\omega)^2 - (c\hbar k)^2 = (m_\mathrm{ф}c^2)^2, \tag{1.8}$$

где, разумеется, под c следует понимать уже не скорость света в вакууме, а некоторую фундаментальную скорость, входящую в теорию относительности. Из уравнения (1.8) и определится фазовая скорость света в вакууме $v = \omega/k$:

$$\frac{1}{v} = \frac{1}{c}\sqrt{1 - \left(\frac{m_\mathrm{ф}c^2}{\hbar\omega}\right)^2} \approx \frac{1}{c}\left[1 - \frac{1}{2}\left(\frac{m_\mathrm{ф}c^2}{\hbar\omega}\right)^2\right]. \tag{1.9}$$

На опыте измеряется не фазовая, а групповая скорость света. Для нее из формулы (1.8) получается

$$u = \frac{d\omega}{dk} = c^2\,\frac{k}{\omega} = \frac{c^2}{v}, \tag{1.10}$$

или

$$u = c\left[1 - \frac{1}{2}\left(\frac{m_\Phi c^2}{\hbar\omega}\right)^2\right] = c\left[1 - \frac{1}{2}\left(\frac{c\lambda}{h}m_\Phi\right)^2\right]. \qquad (1.11)$$

Современные радиолокационные методы измерения скорости света при различных частотах приводят к результату $m_\Phi < 4 \cdot 10^{-21} m_e$, где m_e — масса электрона.

§ 2. Фотоэлектрический эффект

1. Одним из явлений, подтверждающих гипотезу фотонов, является *фотоэлектрический эффект*, к рассмотрению которого мы и перейдем.

В 1887 г. Генрих Герц (1857–1894) обнаружил, что освещение ультрафиолетовым светом *отрицательного электрода* искрового промежутка, находящегося под напряжением, облегчает проскакивание искры между его электродами. Занятый в то время исследованиями электромагнитных волн, предсказанных Максвеллом, Герц не обратил на это явление серьезного внимания. Первые исследования явления принадлежат Хальваксу (1859–1922), Риги (1850–1921) и в особенности А. Г. Столетову (1839–1896).

Сущность явления, обнаруженного Герцем, состоит в том, что *при освещении ультрафиолетовыми лучами отрицательно заряженного металлического тела оно теряет отрицательный заряд.* При освещении такими же лучами положительно заряженного тела потери заряда не наблюдается. Более того, если тело не было заряжено, то при освещении оно заряжается *положительно* до потенциала в несколько вольт. После открытия электрона в 1897 г. Дж. Дж. Томсоном (1856–1940) опытами самого Томсона, а также Ленарда (1862–1947) вскоре был найден удельный заряд e/m для частиц, теряемых телами при освещении. Он оказался таким же, как и для частиц катодных лучей. Тем самым было доказано, что при освещении тела теряют *электроны.*

Явление вырывания электронов из вещества при освещении его светом получило название *фотоэлектрического эффекта* или, короче, *фотоэффекта.* Различают *внешний* и *внутренний* фотоэффект. При внешнем фотоэффекте электроны освобождаются светом из поверхностного слоя вещества и переходят в другую среду, в частности в вакуум. При внутреннем фотоэффекте оптически возбужденные электроны остаются внутри освещаемого тела, не нарушая электрическую нейтральность последнего. Для обоснования гипотезы фотонов основное значение имеет *внешний фотоэффект*, который преимущественно и рассматривается в этом параграфе. О внутреннем фотоэффекте и о его применениях будет сказано несколько слов в конце этого же параграфа.

Электроны, вырванные под действием света, называются *фотоэлектронами.* Фотоэлектрическими свойствами обладают как металлы, так и диэлектрики, а также полупроводники и электролиты, причем необходимым (но недостаточным) условием фотоэффекта является заметное поглощение используемого света в поверхностном слое освещаемого

тела. Фотоэлектрический эффект вызывается не только ультрафиолетовыми лучами. Щелочные металлы — литий, натрий, калий, рубидий, цезий — весьма чувствительны к фотоэлектрическому действию и в видимой области спектра. А специальная обработка поверхностей этих и других металлов делает их способными испускать фотоэлектроны даже под действием инфракрасных лучей.

2. На рис. 1 показана принципиальная схема экспериментальной установки для исследования фотоэффекта. Фотоэлектроны, вырванные при освещении из катода, увлекаются приложенным напряжением к аноду и замыкают цепь. По скорости зарядки электрометра (вместо электрометра можно взять чувствительный гальванометр) можно определить силу электрического тока в цепи, а с ней и количество фотоэлектронов, достигающих анода в единицу времени. Опыты подобного рода в ранних исследованиях производились в газах. Но их лучше производить в вакууме, так как газ только осложняет явления, происходящие в поверхностном слое металла.

Фотоэффект с данного вещества сильно зависит от состояния его поверхности. Для получения однозначных результатов исследование этого явления необходимо производить в хорошем вакууме, предварительно тщательно очистив поверхность исследуемого тела. Перед помещением электродов в прибор материал, служащий катодом, подвергают перегонке в вакууме, а затем наносят тонким слоем на какую-либо подложку.

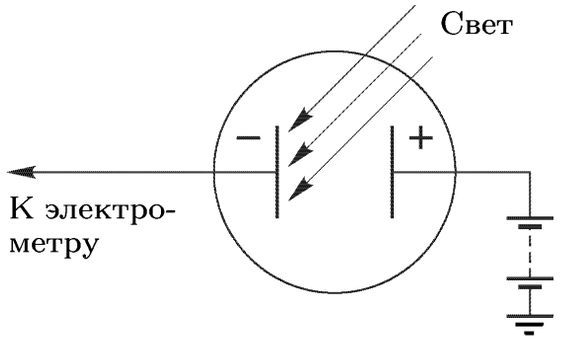


Рис. 1

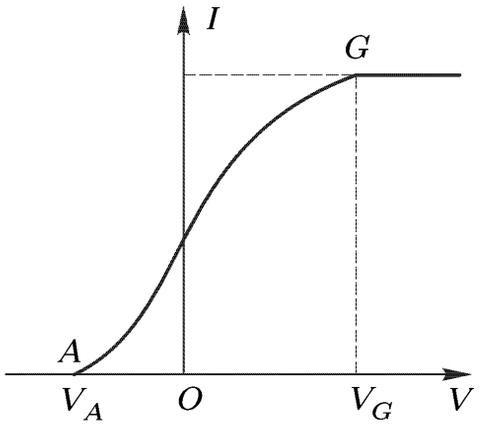


Рис. 2

Если при постоянных интенсивности и частоте падающего света менять напряжение V между анодом и катодом, то зависимость фототока I от V представится кривой, схематически изображенной на рис. 2. Она называется *характеристикой фотоэлемента*, т. е. того прибора, в котором наблюдается фотоэффект. При увеличении напряжения характеристика переходит в горизонтальную прямую, которой соответствует максимальный ток. Он называется *током насыщения*. Насыщение достигается тогда, когда все электроны, вырванные светом с поверхности катода, попадают на анод. Дальнейшее повышение напряжения не меняет силу фототока: *она определяется только количеством ежесекундно вырываемых электронов.*

Существование тока насыщения экспериментально было установлено А.Г. Столетовым. Им же было доказано на опыте, что *ток насыщения строго пропорционален интенсивности падающего света*, если только частота его остается постоянной. Точнее, ток насыщения пропорционален интенсивности *поглощаемого света*. Однако последняя в свою очередь пропорциональна интенсивности падающего света, а потому в формулировку закона и входит интенсивность того же света. Кроме того, закон пропорциональности Столетова строго соблюдается лишь при условии, когда ток насыщения образован только электронами, освобожденными светом со *светочувствительной поверхности тела*. Для этого светочувствительная поверхность должна помещаться в вакууме. В газонаполненных приборах могут наблюдаться заметные отступления от простой пропорциональности. Такие приборы обычно чувствительнее вакуумных, так как в них к току электронов, освобожденных светом со светочувствительной поверхности тела, добавляется еще *ток ионизации газа*, наполняющего прибор. Поэтому газонаполненными приборами, если они применяются для измерений, надо пользоваться с осмотрительностью.

3. Можно было бы попытаться дать качественное объяснение фотоэффекта с волновой точки зрения. Свободные электроны, в изобилии имеющиеся в металле, удерживаются в нем задерживающим полем, существующим на границе металла. Работа выхода электрона из металла составляет несколько электрон-вольт. В электрическом поле световой волны свободный электрон приходит в колебания. Когда энергия колебаний сделается достаточно большой, электрон может преодолеть задерживающее поле и уйти за пределы металла. Если же электрон связанный, то все будет происходить аналогично, но зависимость энергии колебаний от частоты будет носить более сложный, резонансный характер.

Однако такое объяснение не согласуется с опытом. Кинетическая энергия колеблющегося электрона заимствуется от световой волны. Казалось бы, что энергия вылетевшего электрона должна быть тем больше, чем больше интенсивность падающего света. Но опыты Ленарда и многочисленные исследования других ученых показали, что *максимальная скорость, с которой электроны вылетают из тела, от интенсивности падающего света совсем не зависит, а определяется только его частотой*. (Предполагается, что материал тела и состояние освещаемой поверхности остаются неизменными.)

Другое резкое расхождение с опытом получится, если на основе приведенного объяснения оценить *время возникновения фотоэффекта*. Действительно, возьмем, например, электрическую лампочку мощностью $P = 100$ Вт. Для простоты будем считать ее точечным и изотропно излучающим источником света. По классической волновой теории поток лучистой энергии распространяется от источника света *непрерывно* во все стороны. Пусть свет падает нормально на плоский фотокатод из цинка, расположенный на расстоянии r от лампочки. Энергетическая освещенность, создаваемая лампочкой на фотокатоде, будет $P/(4\pi r^2)$. Работа выхода электрона A из цинка

составляет около 3,74 эВ. Ясно, что для выхода электрона из металла энергия, накопленная им при вынужденных колебаниях, должна быть не меньше A. Максимальная энергия, которую может получить атом от излучения за время t, составляет $\mathscr{E}_{\text{макс}} = (P/4\pi r^2)\sigma t$, где σ — поперечное сечение атома. Энергия, переданная электрону за то же время, меньше $\mathscr{E}_{\text{макс}}$. Но она должна быть не меньше A, чтобы электрон мог выйти из металла, т. е. $\mathscr{E}_{\text{макс}} > A$, а потому должно быть $t > (4\pi r^2/\sigma P)A$. Среднее расстояние между атомами цинка найдется по формуле $d = (M/\delta N_{\text{A}})^{1/3}$, где N_{A} — постоянная Авогадро, M — молярная масса, δ — плотность цинка. Для цинка $M = 65$ г/моль, $\delta = 7$ г/см^3, а следовательно, $d = 2{,}49 \cdot 10^{-8}$ см. Сечение σ можно оценить по формуле $\sigma \approx d^2 \approx 6 \cdot 10^{-16}$ см^2. Если положить еще $r = 1$ м, то получится $t \gtrsim 1{,}25$ с. Таким образом, *по классической волновой теории фотоэффект должен протекать с запаздыванием.* Если учесть, что фотоэлементы реагируют на значительно более слабые световые потоки, чем в нашем примере, то и запаздывание фотоэффекта может получиться много больше того, которое мы вычислили. Между тем *опыт показывает, что фотоэффект протекает безынерционно, т. е. фототок появляется мгновенно — одновременно с освещением.* Именно на такой безынерционности основаны практически все научно-технические применения фотоэффекта.

4. Указанные трудности отпадают, если фотоэффект рассматривать с фотонной точки зрения. Взаимодействуя с электроном металла, фотон может обмениваться с ним энергией и импульсом. Такой процесс взаимодействия напоминает удар шаров, а потому его образно называют *столкновением.* Фотоэффект возникает при *неупругом столкновении* фотона с электроном. При таком столкновении фотон поглощается, а его энергия передается электрону. Таким образом, электрон приобретает кинетическую энергию не постепенно, а *сразу — в результате единичного акта столкновения.* Этим и объясняется безынерционность фотоэффекта.

Энергия поглощаемого фотона может затрачиваться на отрыв электрона от атома внутри металла. Оторванный электрон может взаимодействовать с атомом внутри металла, растрачивая энергию на тепло. Максимальной энергией вылетевший электрон будет обладать тогда, когда внутри металла он был *свободен*, т. е. не связан с атомом, а при вылете наружу не расходовал энергию на тепло. В этом случае кинетическая энергия электрона тратится только на преодоление задерживающих сил, действующих в поверхностном слое металла, т. е. на *работу выхода.* Предположим, что электрон получил кинетическую энергию при столкновении *только с одним фотоном.* Многофотонные процессы, которые будут рассмотрены в п. 9, возможны, но при слабых интенсивностях света (линейная оптика) маловероятны. Тогда *максимальная кинетическая энергия, которой будет обладать вылетевший электрон, определится формулой*

$$\frac{1}{2}\, m_{\text{e}} v_{\text{макс}}^2 = h\nu - A, \tag{2.1}$$

где A — работа выхода, а m_e — масса покоя электрона. Эта формула впервые была получена Эйнштейном и носит его имя.

Прежде чем анализировать формулу Эйнштейна, необходимо выяснить, как может «свободный электрон в металле» поглотить фотон. Не противоречит ли это утверждению, доказанному в конце предыдущего параграфа, согласно которому поглощение фотона свободным электроном несовместимо с законами сохранения энергии и импульса? На самом деле противоречия нет. Противоречие возникает из-за неудачной терминологии. «Свободный электрон в металле» в действительности *не свободен.* Он как бы заперт в ящике, вблизи стенок которого действует задерживающее поле. Фотон взаимодействует не только с электроном, но происходит взаимодействие обеих этих частиц с металлом в целом. При взаимодействии же трех тел законы сохранения энергии и импульса могут выполняться одновременно. Импульс фотона воспринимается как электроном, так и металлом, энергия же передается только электрону, так как масса металла может считаться бесконечно большой.

5. Из формулы Эйнштейна (2.1) вытекают два следствия, находящиеся в строгом согласии с опытом.

1) *Максимальная кинетическая энергия вырванных электронов линейно зависит от частоты освещающего света и не зависит от его интенсивности. Интенсивность оказывает влияние только на количество вырванных электронов, но совсем не влияет на их максимальную кинетическую энергию.* Любопытно, что тангенс угла наклона прямой (2.1) к оси частот ν равен постоянной Планка h (по оси абсцисс отложена частота ν, по оси ординат — максимальная кинетическая энергия электрона $m_e v_{\text{макс}}^2/2$). Это дает новый метод измерения постоянной Планка.

2) *Существует низкочастотная граница (порог) фотоэффекта, т. е. такая частота ν_0, ниже которой фотоэффект не наблюдается.* Она зависит от состава облучаемого тела и состояния его поверхности.

В самом деле, представим работу выхода в виде $A = h\nu_0$, где ν_0 — положительная постоянная, и запишем формулу (2.1) так:

$$\frac{1}{2} m_e v_{\text{макс}}^2 = h(\nu - \nu_0). \tag{2.2}$$

При $\nu < \nu_0$ правая часть отрицательна. А это невозможно, так как левая часть существенно положительна. Следовательно, при $\nu < \nu_0$ фотоэффект невозможен. Частота ν_0 и есть *низкочастотная граница фотоэффекта.* Существование такой границы совершенно непонятно с волновой точки зрения.

6. Экспериментальная проверка формулы Эйнштейна и следствий из нее имеет громадное значение не только для теории света, но и для всей физики. Она производится путем снятия характеристик фотоэлемента, в котором падающий свет возбуждает электрический ток. Принципиальная схема установки приведена на рис. 1. Существенно, что фототок должен измеряться не только при положительных значениях разности потенциалов между анодом и катодом (ускоряющее поле), *но и при отрицательных* (замедляющее поле). Для повышения

точности измерений прибор конструируют так, чтобы практически все фотоэлектроны, вырванные светом из катода, попадали на анод. С этой целью вместо плоского применяют *сферический конденсатор*, внешней обкладкой которого служит большая металлическая сфера, а внутренней — маленький шарик из исследуемого материала.

Фотоэлектрон, освобожденный светом, может претерпеть столкновение с атомом внутри поверхностного слоя металла. Из-за этого он может замедлиться и даже не выйти наружу. Приложенное электрическое поле способствует ускорению замедлившихся электронов и выходу их из металла. В этом причина, почему сила фототока возрастает с напряжением между катодом и анодом. Для проверки формулы Эйнштейна (2.1) надо измерить максимальную кинетическую энергию фотоэлектрона, которой он обладает по выходе из металла. Такие измерения затруднены *контактной разностью потенциалов*, весьма чувствительной к поверхностной обработке металла, а потому трудно контролируемой. Для исключения влияния контактной разности потенциалов на вольт-амперной характеристике фотоэлемента отмечают две точки: одну A, в которой ток обращается в нуль задерживающим полем; другую G, с которой начинается ток насыщения (см. рис. 2). Показания вольтметра, соответствующие этим точкам, обозначим через V_A и V_G. Это есть именно *показания вольтметра*, а не разности потенциалов между анодом и катодом. К последним должна быть добавлена контактная разность потенциалов V_c между теми же электродами. Полные разности потенциалов равны $V_A + V_c$ и $V_G + V_c$ соответственно.

При показаниях вольтметра меньше V_A все электроны задерживаются и не достигают анода. Когда вольтметр показывает V_A, анода начинают достигать электроны с максимальной скоростью $v_{\text{макс}}$. Приравнивая энергии электрона в точках A и G характеристики, получим

$$\frac{1}{2}\, m_{\text{e}} v_{\text{макс}}^2 = -e(V_A + V_c),$$

где e — заряд электрона по абсолютной величине. (Тепловыми скоростями мы пренебрегаем, так как средняя энергия теплового движения kT пренебрежимо мала по сравнению с энергией фотона $h\nu$.) Когда же вольтметр показывает V_G, наоборот, даже электроны нулевой скорости начинают достигать анода, поэтому

$$0 = -e(V_G + V_c).$$

Из последнего уравнения находится контактная разность потенциалов: $V_c = -V_G$. Первое же уравнение определяет искомую максимальную кинетическую энергию фотоэлектрона:

$$\frac{1}{2}\, m_{\text{e}} v_{\text{макс}}^2 = e(V_G - V_A). \tag{2.3}$$

Из уравнения $V_G = -V_c$ следует, что положение точки V_G на вольт-амперной характеристике фотоэлемента зависит только от контактной разности потенциалов, но не зависит от частоты ν падающего света. Напротив, положение точки V_A меняется с частотой ν, так как от

частоты зависит максимальная энергия $m_e v^2_{\text{макс}}/2$, которой определяется величина $V_A + V_c$. При увеличении частоты характеристики будут смещаться влево, но абсцисса V_G будет оставаться неизменной. Если характеристики строить в таком масштабе, чтобы на них ток насыщения всегда изображался ординатами одинаковой величины, то все они будут проходить через одну и ту же точку G.

В начале п. 6 было отмечено преимущество сферического конденсатора перед плоским. Теперь, после изложения идеи метода определения максимальной кинетической энергии фотоэлектрона, можно указать на другое преимущество. В случае плоского конденсатора для попадания фотоэлектронов на анод имеют значение не полные скорости их, а только составляющие, перпендикулярные к поверхности анода. Касательные составляющие приводят к тому, что характеристика фотоэлемента в точках A и G не пересекает горизонтальные прямые, а подходит к ним *асимптотически*. Это затрудняет определение положения самих точек A и G. В сферическом же конденсаторе с малым внутренним шариком этого не получается, так как все скорости фотоэлектронов практически радиальные, т. е. нормальны к поверхности анода.

Точная экспериментальная проверка формулы Эйнштейна была впервые осуществлена Ричардсоном (1879–1959) и Карлом Комптоном (1887–1954) в 1912 г., а еще более тщательно Милликеном (1868–1953) в 1916 г. Обе работы подтвердили формулу Эйнштейна, хотя и были не совсем свободны от возражений. В наиболее чистых условиях проверка была произведена П. И. Лукирским (1894–1954) и С. С. Прилежаевым в 1926 г. Схема их установки приведена на рис. 3. Прибор представлял собой сферический конденсатор, в котором создавался вакуум, с наружной обкладкой в виде слоя серебра, нанесенного на внутреннюю поверхность сферы S (с диаметром 11 см). Внутренней обкладкой служил шарик N (с диаметром 1,5 см) из металла, с поверхности которого изучался фотоэффект. Чтобы шарик можно было легко заменять, он навинчивался на стержень, связанный со шлифом. Кварцевая изоляция устраняла возможность влияния ползучих зарядов со стекла. Шарик N можно было освещать через отверстие A, закрытое кварцевой пластинкой. Свет поступал от ртутной дуги, предварительно пройдя через монохроматор M. Был изучен фотоэффект с металлов: Al, Zn, Sn, Ni, Ag, Cd, Pb, Cu, Pt. На рис. 4 приведены характеристики фотоэлемента для цинка (λ = 230,2; 253,7; 313,0 нм для кривых *1*, *2*, *3* соответственно). Для других металлов характеристики имели аналогичный вид. Все они построены в таком масштабе, в котором ток насыщения условно принят равным 100. При этом

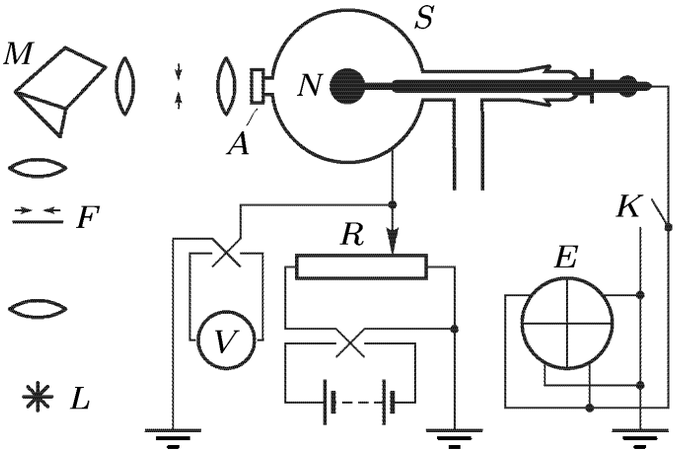


Рис. 3

условии все кривые для одного и того же металла, но при различных длинах волн сверху пересекались в одной и той же точке, в которой они переходили в горизонтальные участки. Это находится в согласии с тем, что было сказано выше. Авторы получили для постоянной Планка значение $h = 6{,}58 \cdot 10^{-27}$ эрг · с (если пересчитать их результат на современное значение заряда электрона).

Рис. 4

7. Для рентгеновских лучей энергия светового кванта $h\nu$ велика по сравнению с работой выхода A. Иными словами, частота ν велика по сравнению с частотой ν_0 низкочастотной границы фотоэффекта. В этом случае в уравнении (2.1) величиной A можно пренебречь и написать

$$\frac{1}{2} m_e v_{\text{макс}}^2 = h\nu, \tag{2.4}$$

или

$$eV = h\nu, \tag{2.5}$$

где eV — энергия электрона, выраженная через ускоряющее напряжение. Эта формула справедлива *не только для прямого фотоэффекта*, в котором энергия световых квантов переходит в кинетическую энергию электронов. Она относится и к *обратному фотоэффекту*, в котором рентгеновские кванты получаются за счет кинетической энергии электронов, бомбардирующих металл. Именно такой процесс происходит в рентгеновских трубках. В этом случае формула (2.5) определяет *максимальную частоту*, с которой могут испускаться рентгеновские лучи антикатодом при заданном напряжении на трубке. Существование максимальной граничной частоты подтверждается опытом. Это также говорит в пользу квантовой теории света.

При бомбардировке антикатода электроны тормозятся, из-за чего возникает так называемое *тормозное рентгеновское излучение*. Спектр этого излучения при разложении по длинам волн оказывается *сплошным*, как и спектр видимого белого света. По этой причине сплошное рентгеновское излучение называется *белым*. Его спектральная интенсивность при различных напряжениях на трубке представлена кривыми на рис. 5 (для вольфрамового антикатода). В сторону длинных волн кривая интенсивности спадает полого, асимптотически приближаясь к нулю с увеличением длины волны. Напротив, со стороны коротких волн кривая интенсивности резко обрывается при некотором значении длины волны, называемой *коротковолновой границей сплошного рентгеновского излучения*. Эта граница определяется формулой (2.5), из которой следует

$$\lambda_{\text{мин}} = \frac{hc}{eV} = \frac{12{,}40}{V} \text{ Å}, \tag{2.6}$$

где напряжение на трубке V измеряется в киловольтах. Коротковолновая граница не зависит от материала антикатода, а определяется *только напряжением на трубке*. Если увеличивать напряжение выше определенного предела, зависящего от материала антикатода, то на сплошное излучение накладываются узкие спектральные линии, составляющие так называемое *характеристическое излучение* антикатода. Но и в этом случае коротковолновая граница рентгеновского спектра существует и определяется прежней формулой (2.6). Существование такой границы является одним из наиболее ярких проявлений корпускулярных свойств рентгеновского излучения.

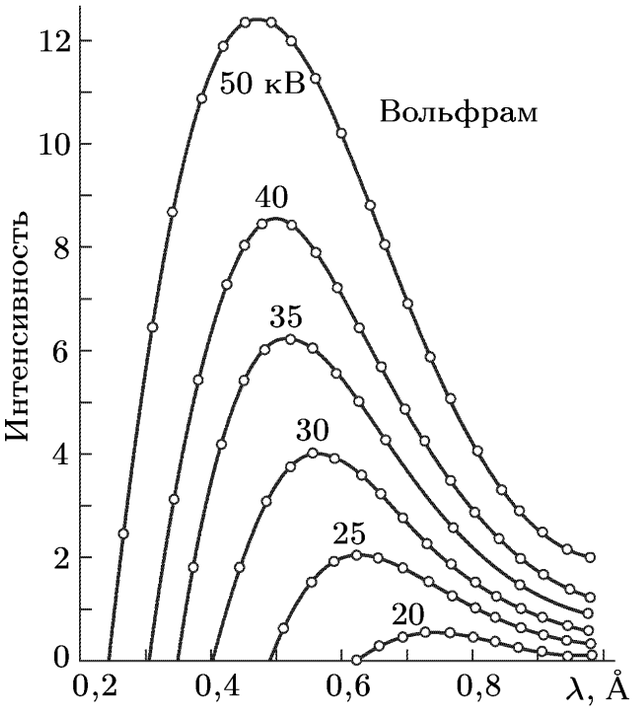


Рис. 5

Измерение коротковолновой границы рентгеновского излучения дает один из методов точного определения постоянной Планка h. Для этого служит формула (2.6). Та же формула применяется для измерения длин волн очень жесткого рентгеновского и гамма-излучения.

8. До сих пор при рассмотрении фотоэффекта мы акцентировали внимание на *корпускулярных* свойствах света. Но фотонам свойственны также и *волновые* свойства света. Проявляются ли они в явлениях фотоэффекта? На этот вопрос следует ответить утвердительно. Волновые свойства проявляются в закономерностях так называемого *селективного* (т. е. избирательного) *фотоэффекта*. Обозначим через I_λ силу фототока насыщения, отнесенную к единичному интервалу длин волн и к единице поглощенной лучистой энергии. Для фотоэффекта, называемого *нормальным*, величина I_λ монотонно убывает с длиной волны, стремясь к нулю на длинноволновой границе $\lambda = \lambda_0$ (рис. 6). Но если электрический вектор падающей волны *не перпендикулярен*

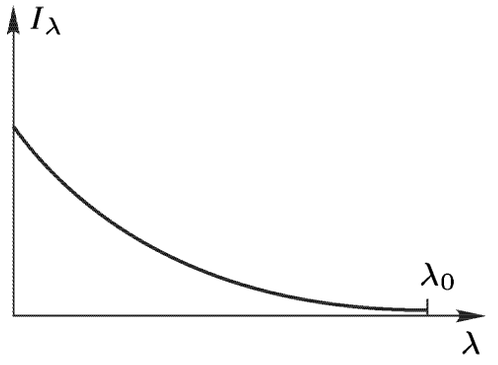


Рис. 6

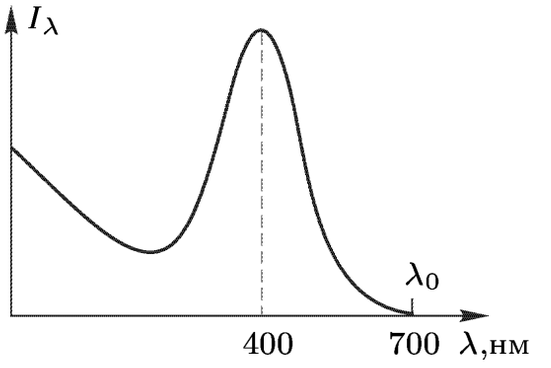


Рис. 7

к плоскости падения, то у ряда металлов, особенно щелочных, на кривой $I_\lambda(\lambda)$ наблюдается *резко выраженный максимум* в определенной области спектра. У натрия, например, максимум лежит при $\lambda = 340$ нм, у калия при $\lambda = 435$ нм, у алюминия при $\lambda = 215$ нм и т. д. (см. кривую на рис. 7, относящуюся к сплаву калия и натрия). Вероятно, аналогичные максимумы существуют и для других металлов, но лежат далеко в коротковолновой области спектра, а потому труднодоступны для наблюдения. Фотоэффект, характеризующийся подобными максимумами на кривой $I_\lambda(\lambda)$, и называется *селективным*.

Максимум на кривой $I_\lambda(\lambda)$ для селективного фотоэффекта напоминает резонансный максимум при вынужденных колебаниях гармонического осциллятора. Поэтому можно сказать, что электроны в металле при селективном фотоэффекте ведут себя так, как если бы они обладали *собственными частотами*, в окрестности которых и наблюдаются максимумы величины $I_\lambda(\lambda)$.

Другая особенность селективного фотоэффекта, в которой также проявляются волновые свойства света, состоит в том, что *интенсивность фототока сильно зависит от поляризации падающего света и от угла падения. Селективный фотоэффект не наблюдается, когда электрический вектор падающей волны перпендикулярен к плоскости падения, а также при нормальном падении света.* В обоих случаях в падающем свете есть составляющая электрического поля, касательная к границе металла, но нет нормальной составляющей. Явление происходит так, как если бы селективность фотоэффекта была обусловлена составляющей электрического вектора, *нормальной к поверхности металла.* Нормальная составляющая более эффективна для вырывания электрона из металла, чем касательная составляющая. В частности, селективность фотоэффекта выражена наиболее резко при скользящем падении поляризованного света, электрический вектор которого лежит в плоскости падения. Для иллюстрации высказанных утверждений на рис. 8 приведена зависимость $I_\lambda(\lambda)$ от λ для различных поляризаций падающего света, а на рис. 9 — зависимость от угла падения. Рисунки относятся к тому же сплаву калия и натрия, что и рис. 7.

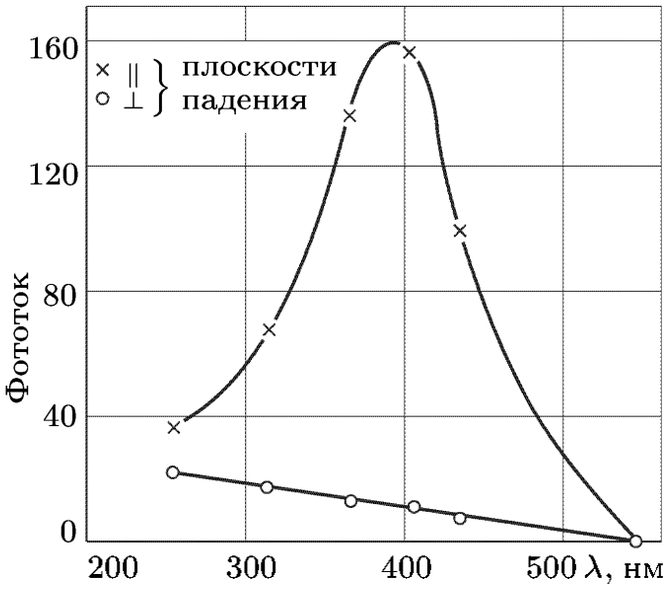


Рис. 8

9. С изобретением лазеров появилась возможность экспериментировать с мощными пучками света и наблюдать *нелинейные* особенности фотоэффекта. Если интенсивность света достаточно велика, то электрон, прежде чем покинуть катод, может претерпеть столкновение не с одним, а с *несколькими фотонами* (*многофотонный процесс*). В этом

случае вместо уравнения (2.1) следует написать

$$\frac{1}{2} m_e v_{\text{макс}}^2 = Nh\nu - A, \tag{2.7}$$

где N — число столкновений электрона с фотонами. Длинноволновая граница фотоэффекта существует по-прежнему, но граничная частота определяется выражением $\nu_0 = A/(hN)$, т. е. уменьшается в N раз по сравнению с однофотонным эффектом. Наблюдению такого многофотонного нелинейного эффекта длительное время препятствовало нагревание металла при лазерном освещении. Оно сопровождается *термоэлектронной эмиссией*, для которой, разумеется, длинноволновая граница не существует. Маскирующее влияние термоэлектронной эмиссии удалось устранить почти полностью применением *сверхкоротких импульсов лазерного излучения* длительностью 10^{-11}–10^{-12} с. Аналогично действует *скользящее освещение фотокатода* (угол падения $\approx 85°$). Таким путем длинноволновая граница многофотонного фотоэффекта надежно зафиксирована при $N = 2, 3, 4, 5$ для различных металлов (Na, Ag, Au и пр.), а также полупроводников при изменении интенсивности света от 0,1 до 10^3 МВт. А исследование влияния поляризации света и угла падения на ток насыщения позволило установить, что многофотонный фотоэффект вызывается исключительно составляющей электрического поля, *нормальной к поверхности катода*.

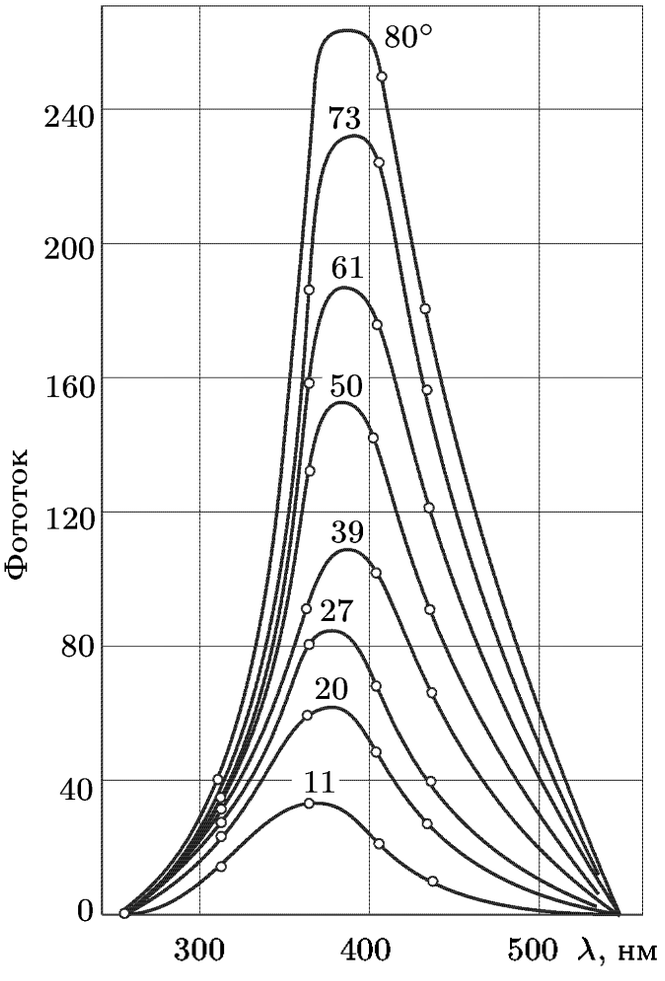


Рис. 9

10. В заключение кратко остановимся на *внутреннем фотоэффекте* и на применениях фотоэффекта. Внутренний фотоэффект может происходить в полупроводниках и диэлектриках. Под действием света часть электронов из валентной энергетической зоны переходит в зону проводимости (см. т. III, § 100). Концентрация носителей тока внутри тела увеличивается — возникает *фотопроводимость*, т. е. повышение электрической проводимости тела под действием света. Перераспределение электронов по различным энергетическим состояниям может привести также к изменению *внутреннего электрического поля* в кристалле. Это ведет к появлению *электродвижущей силы* (фото-ЭДС) на границах двух различных полупроводников или полупроводника

и металла при их освещении. Около границы образуется *переходный слой*, пропускающий ток только в одном направлении, т. е. обладающий *вентильными свойствами* (см. т. III, § 100).

Фотоэффект (как внешний, так и внутренний) используется в *фотоэлектронных приборах*, получивших разнообразные применения в науке и технике (в телевидении, космической технике и т. д.). Нашли широкое применение *фотоэлементы* с внешним фотоэффектом, т. е. двухэлектродные приборы, в которых падающая на поверхность катода лучистая энергия при внешнем приложенном напряжении между электродами превращается в энергию электрического тока. Электрическое сопротивление полупроводников падает при освещении; это используется для устройства *фотосопротивлений.* Возникновение фото-ЭДС при освещении приконтактной области двух различных соприкасающихся полупроводников используется в *фотодиодах* для непосредственного превращения лучистой энергии в электрическую. *Фотоэлектронные умножители* (см. т. III, § 103), усиливающие первоначальный фототок во много раз, позволяют регистрировать очень слабое излучение, вплоть до отдельных квантов.

ЗАДАЧИ

1. Определить максимальную скорость фотоэлектронов, вылетающих из никелевого электрода, освещаемого ультрафиолетовым светом с длиной волны $\lambda = 220$ нм. Работа выхода электрона из никеля $A = 4{,}84$ эВ.

Ответ. $v_{\text{макс}} = c\sqrt{\dfrac{2}{m_e c^2}\left(\dfrac{hc}{\lambda} - A\right)} = 510$ км/с.

2. Найти число электронов N, вырываемых светом в одну секунду из катода вакуумного фотоэлемента, если ток насыщения, протекающий через него при освещении, равен $I = 5 \cdot 10^{-10}$ А.

Ответ. $N = I/e = 3{,}1 \cdot 10^{9}$ с$^{-1}$.

3. Вычислить длину волны λ для длинноволновой границы фотоэффекта на цинке, если работа выхода электрона из цинка $A = 3{,}74$ эВ.

Ответ. $\lambda = hc/A = 330$ нм.

4. Максимальная кинетическая энергия фотоэлектронов при освещении цезиевого электрода монохроматическим светом оказалась равной $\mathscr{E}_{\text{макс}} = 0{,}15$ эВ. Вычислить длину волны λ света, применявшегося при освещении, если работа выхода электрона из цезия $A = 1{,}89$ эВ.

Ответ. $\lambda = hc/(\mathscr{E}_{\text{макс}} + A) = 600$ нм.

5. Уединенный медный шарик облучается ультрафиолетовым светом с длиной волны $\lambda = 200$ нм. До какого максимального потенциала зарядится шарик, если работа выхода электрона из меди $A = 4{,}47$ эВ?

Ответ. $V = \dfrac{hc/\lambda - A}{e} = 1{,}04$ В.

6. При каких длинах волн λ облучающего света шарик в условиях предыдущей задачи заряжаться не будет?

Ответ. $\lambda \geqslant hc/\lambda = 274$ нм.

7. Вакуумный фотоэлемент с катодом из молибдена освещается монохроматическим светом с длиной волны $\lambda = 250$ нм. При наложении задерживающей разности потенциалов фототок уменьшается и обращается в нуль, когда она достигает значения $V = 1{,}8$ В. Определить внешнюю контактную

разность потенциалов между молибденом и материалом анода, если работа выхода электрона из молибдена $A = 4{,}27$ эВ.

Ответ. $V_c = \dfrac{hc/\lambda - A}{e} - V = -1{,}40$ В.

Отрицательный знак означает, что при контакте молибдена с материалом анода потенциал молибдена окажется ниже.

8. Определить наибольшую скорость электронов v на аноде рентгеновской трубки, если минимальная длина волны сплошного рентгеновского излучения составляет $\lambda = 0{,}1$ нм.

Ответ. $v = h/m_e\lambda = 7{,}3 \cdot 10^3$ км/с.

§ 3. Эффект Комптона

1. В 1922 г. Артур Комптон (1892–1962) открыл явление, которое, как и фотоэффект, подтверждает гипотезу фотонов. Комптон изучал рассеяние *жесткого* рентгеновского излучения на телах, состоящих из *легких атомов* (графит, парафин и пр.). Схема его установки показана на рис. 10. Источником рентгеновского излучения служила рентгеновская трубка с молибденовым антикатодом. Узкий пучок монохроматического рентгеновского излучения выделялся диафрагмами D_1 и D_2 и рассеивался на исследуемом теле R. Для исследования спектрального состава рассеянного излучения оно после прохождения ряда диафрагм попадало на кристалл K рентгеновского спектрографа, а затем в ионизационную камеру или на фотопластинку P. Оказалось, что в рассеянном излучении, наряду с исходной длиной волны λ, появляется *смещенная линия* с длиной волны $\lambda' > \lambda$. Изменение длины волны $\lambda' - \lambda$ в длинноволновую сторону спектра при рассеянии излучения получило название *комптоновского смещения*, а само явление — *эффекта Комптона*. Опыт показал, что *комптоновское смещение* $\lambda' - \lambda$ *для исследованных веществ не зависит от состава рассеивающего тела и длины падающей волны* λ. Оно пропорционально квадрату синуса половины угла рассеяния θ.

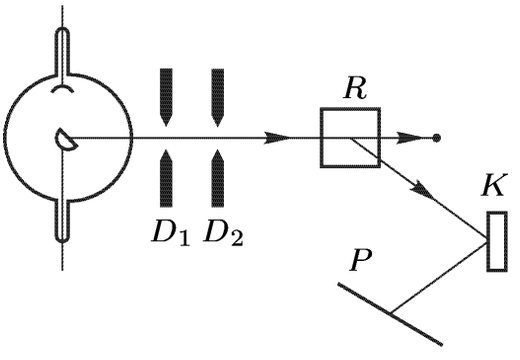


Рис. 10

На рис. 11 представлены результаты измерений на графите при различных углах рассеяния для K-линии молибдена ($\lambda = 0{,}0712605$ нм). Сверху показана форма линии исходного излучения (т. е. угловое распределение интенсивности в линии). Ниже сделано то же самое для рассеянного излучения при различных значениях угла рассеяния. Ясно видно, что первоначально одиночная линия в результате рассеяния становится двойной. Уширение обеих компонент рассеянного излучения обусловлено движением электронов и атомов, на которых происходит рассеяние.

2. Классическая теория не смогла объяснить закономерности комптоновского рассеяния (см. п. 8). Они были поняты только на основе *квантовой теории*, предложенной независимо самим Комптоном

и Дебаем (1884–1966). По их представлениям рассеяние рентгеновского кванта с изменением длины волны есть результат *одиночного акта* столкновения его с электроном.

В легких веществах, с которыми производились опыты Комптона, энергия связи электрона с атомом мала по сравнению с энергией, передаваемой ему рентгеновским квантом при столкновении. Энергия, передаваемая атому квантом при столкновении, тем больше, чем больше угол рассеяния. Поэтому указанное условие выполняется тем лучше, чем больше угол рассеяния. В легких атомах энергией связи электрона внутри атома можно пренебречь при всех углах рассеяния, т. е. все электроны считать свободными. Это и делается в теории Комптона–Дебая. Тогда одинаковость комптоновского смещения $\lambda' - \lambda$ для всех веществ становится понятной само собой. Действительно, в теории с самого начала предполагается, что рассеивающее вещество в сущности состоит только из свободных *электронов*, т. е. индивидуальные особенности вещества никак не учитываются. Но это допустимо только для легких атомов. Для внутренних электронов тяжелых атомов такая идеализация не годится. Учет энергии связи ведет к отступлениям от простой формулы (3.3), выводимой ниже. Такие отступления действительно были обнаружены.

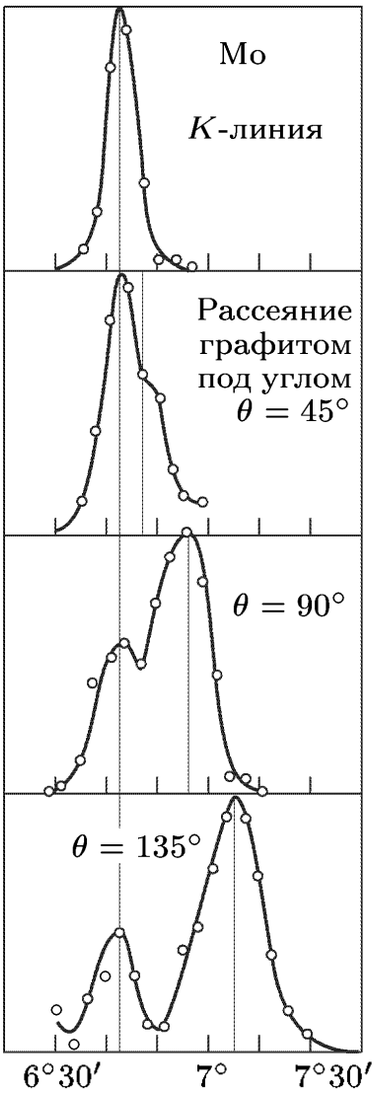


Рис. 11

3. Рассмотрим теперь столкновение фотона со свободным электроном. При взаимодействии этих двух частиц должны соблюдаться *законы сохранения энергии и импульса.* Поскольку при столкновении с фотоном электрон может получить релятивистские скорости, столкновение должно рассматриваться на основе *релятивистской механики.* Обозначим через $\mathscr{E}_\text{ф}$ и $\mathbf{p}_\text{ф}$ энергию и импульс фотона до рассеяния, а через $\mathscr{E}'_\text{ф}$ и $\mathbf{p}'_\text{ф}$ — после рассеяния. Для электрона полная энергия и импульс до рассеяния будут соответственно $\mathscr{E}_0 = m_0c^2$ и 0 (электрон до рассеяния покоился), а после рассеяния $\mathscr{E}_\text{e}$ и $\mathbf{p}_\text{e}$. Тогда законы сохранения энергии и импульса дают

$$\mathscr{E}_\text{ф} + \mathscr{E}_0 = \mathscr{E}'_\text{ф} + \mathscr{E}_\text{e}, \quad \mathbf{p}_\text{ф} = \mathbf{p}'_\text{ф} + \mathbf{p}_\text{e},$$

или

$$\mathscr{E}_\text{e} = \mathscr{E}_\text{ф} - \mathscr{E}'_\text{ф} + \mathscr{E}_0, \quad \mathbf{p}_\text{e} = \mathbf{p}_\text{ф} - \mathbf{p}'_\text{ф}.$$

Отсюда

$$\left(\frac{\mathscr{E}_\text{e}}{c}\right)^2 - \mathbf{p}_\text{e}^2 = \frac{(\mathscr{E}_\text{ф} - \mathscr{E}'_\text{ф} + \mathscr{E}_0)^2}{c^2} - (\mathbf{p}_\text{ф} - \mathbf{p}'_\text{ф})^2. \qquad (3.1)$$

Для каждой частицы квадрат четырехмерного вектора в пространстве Минковского есть инвариант, т. е.

$$\left(\frac{\mathscr{E}}{c}\right)^2 - \mathbf{p}^2 = \mathrm{Inv}, \tag{3.2}$$

причем для фотона этот инвариант равен нулю. Это дает

$$\left(\frac{\mathscr{E}_\mathrm{e}}{c}\right)^2 - \mathbf{p}_\mathrm{e}^2 = \left(\frac{\mathscr{E}_0}{c}\right)^2, \quad \left(\frac{\mathscr{E}_\Phi}{c}\right)^2 = \mathbf{p}_\Phi^2, \quad \left(\frac{\mathscr{E}'_\Phi}{c}\right)^2 = \mathbf{p}'^2_\Phi.$$

С учетом этого из (3.1) получаем

$$\frac{\mathscr{E}_\Phi \mathscr{E}'_\Phi}{c^2} + \frac{\mathscr{E}'_\Phi \mathscr{E}_0}{c^2} - \frac{\mathscr{E}_\Phi \mathscr{E}_0}{c^2} - \mathbf{p}_\Phi \mathbf{p}'_\Phi = 0,$$

или

$$1 - \cos\theta = m_0 c\left(\frac{1}{p'_\Phi} - \frac{1}{p_\Phi}\right),$$

где θ — угол рассеяния, т. е. угол между векторами $\mathbf{p}_\Phi$ и $\mathbf{p}'_\Phi$.

До сих пор квантовая природа света, во всяком случае в ее специальной форме (1.7), не использовалась. Мы оперировали лишь понятиями энергии и импульса излучения. Наше рассмотрение справедливо *и в классической физике*, поскольку в ней излучению также присущи определенные энергия и импульс, подчиняющиеся законам сохранения. На основе этих законов сохранения мы и рассмотрели взаимодействие электрона с определенной порцией излучения. Теперь мы учтем, что такой порцией излучения является фотон с начальным импульсом $p_\Phi = h/\lambda$ и конечным $p'_\Phi = h/\lambda'$, где λ и λ' — длины волн падающего и рассеянного излучения. В результате получим

$$\lambda' - \lambda = \lambda_\mathrm{K}(1 - \cos\theta) = 2\lambda_\mathrm{K}\sin^2\frac{\theta}{2}, \tag{3.3}$$

где

$$\lambda_\mathrm{K} = \frac{h}{m_\mathrm{e} c} = 2{,}4263096(15)\cdot 10^{-10}\ \text{см}, \tag{3.4}$$

причем в последней формуле массу покоя электрона мы обозначили через m_e (см. конец § 1).

Из (3.3) следует независимость комптоновского смещения $\lambda' - \lambda$ от вещества рассеивающего тела (для легких атомов). Качественно этот результат уже был получен выше. Формула (3.3) приводит также к независимости комптоновского смещения от первоначальной длины волны λ. Это специфично лишь для *релятивистского* рассмотрения, если дополнительно учесть, что масса покоя одного из сталкивающихся тел (фотона) равна нулю.

Универсальная постоянная λ_K, определяемая формулой (3.4), является одной из важнейших атомных постоянных. Она называется *комптоновской длиной для электрона.* Комптоновская длина представляет собой изменение длины волны фотона при его рассеянии на

угол $\theta = \pi/2$ на свободном неподвижном электроне. Другой физический смысл комптоновской длины будет указан в § 17. Существует комптоновская длина для протона, нейтрона и других элементарных частиц. Она также определяется выражением (3.4), если в нем массу электрона заменить на массу соответствующей частицы. В теоретических исследованиях предпочитают пользоваться другой универсальной постоянной:

$$\lambda\!\!\!^{-}_{\mathrm{K}} = \frac{\hbar}{m_e c} = \frac{\lambda_{\mathrm{K}}}{2\pi} = 3{,}861592(4) \cdot 10^{-11} \text{ см}. \qquad (3.5)$$

Ее также называют *комптоновской длиной для электрона* (*перечеркнутой* или *приведенной*).

Следует особо подчеркнуть, что электрон, на котором рассеивается фотон, в рассматриваемой теории предполагается неподвижным. Если же электрон движется, то при столкновении он может передать свою кинетическую энергию фотону, а сам остановиться. Этот процесс сопровождается уменьшением длины волны фотона и называется *обратным эффектом Комптона.*

4. Формула (3.3) показывает, что рассеяние фотонов на неподвижных свободных электронах всегда должно сопровождаться комптоновским увеличением длины волны. Каково же происхождение несмещенной линии (см. рис. 11)? Она вызывается связанными электронами. В этом случае рассеяние происходит на *атомах*, массы которых могут считаться бесконечно большими. Для бесконечно тяжелого атома комптоновская длина, а с ней и смещение $\lambda' - \lambda$, согласно формулам (3.3) и (3.4), обращаются в нуль. Атом получает импульс от падающего фотона, но его энергия остается без изменения. С ростом атомного номера относительное число связанных электронов увеличивается. Поэтому происходит и возрастание интенсивности несмещенной компоненты по сравнению с интенсивностью смещенной.

Как же меняется соотношение между интенсивностями смещенной и несмещенной компонент при изменении угла рассеяния? Выше было отмечено, что условие, при котором электрон может считаться свободным, выполняется тем лучше, чем больше угол рассеяния. Отсюда следует, что при увеличении угла рассеяния увеличивается относительная доля свободных электронов, а с ней и отношение интенсивностей смещенной и несмещенной компонент.

Чем больше энергия фотона, тем в меньшей степени проявляется связь электрона с атомом. Вот почему для наблюдения эффекта Комптона должны применяться *жесткие рентгеновские лучи.* Однако когда энергия фотона начинает превышать удвоенную энергию покоя электрона $2m_e c^2$, при взаимодействии его с веществом начинается процесс *образования пар*, т. е. превращение фотона в пару электрон–позитрон. С увеличением энергии фотона этот процесс идет относительно все быстрее и быстрее и при энергиях, значительно превышающих $2m_e c^2$, подавляет комптоновское рассеяние. Для видимого света энергия связи электронов в атоме превышает энергию фотона. Вот почему в видимой области эффект Комптона не наблюдается.

Отметим еще, что рассеяние фотонов на атомах *когерентно*, а на свободных электронах *некогерентно*. Последнее очевидно, поскольку свободные электроны и их движения независимы, а потому независимы и рассеяния на них. Колебания же связанных электронов, вызванные падающей волной, согласованы. Поэтому волны, рассеянные на связанных электронах, могут интерферировать между собой и с падающей волной. Именно такая интерференция проявляется при прохождении рентгеновских лучей через кристаллы и определяется известными условиями Лауэ и Вульфа–Брэгга (см. т. IV, § 61).

5. При рассеянии рентгеновского фотона на свободном электроне электрон получает отдачу, как это изображено параллелограммом на рис. 12. Из этого параллелограмма получаем

$$p'_{\text{е}} \cos\varphi + p'_{\text{ф}} \cos\theta = p_{\text{ф}}, \quad p'_{\text{е}} \sin\varphi = p'_{\text{ф}} \sin\theta.$$

Исключая отсюда $p'_{\text{е}}$ и учитывая соотношение $p_{\text{ф}}/p'_{\text{ф}} = \lambda'/\lambda$, получим

$$\operatorname{tg}\varphi = \frac{\sin\theta}{\lambda'/\lambda - \cos\theta}.$$

Отношение λ'/λ найдем с помощью формулы (3.3). После несложных преобразований получаем

$$\operatorname{tg}\varphi = \frac{\operatorname{ctg}(\theta/2)}{1 + \lambda_{\text{К}}/\lambda} = \frac{\operatorname{ctg}(\theta/2)}{1 + h\nu/(m_{\text{е}}c^2)}. \tag{3.6}$$

Этой формулой определяется направление вылета электрона отдачи. Аналогично нетрудно получить его энергию и импульс (см. задачу 2 к этому параграфу).

6. Процесс, изображенный на рис. 12, наблюдался в камере Вильсона (1869–1959). Напомним, что в камере Вильсона быстрым адиабатическим расширением создается пересыщенный водяной пар (см. т. II, § 119). Заряженная частица, пролетая через камеру, ионизует окружающий воздух. На ионах оседают капельки тумана, делающие видимым след заряженной частицы. Такой след называется *треком*. Направление движения электрона отдачи (OA на рис. 12) можно, следовательно, фиксировать в камере Вильсона. Энергия же и величина импульса электрона отдачи определятся длиной его пробега, т. е. длиной трека. Импульс $p'_{\text{е}}$ электрона отдачи может, таким образом, считаться известным.

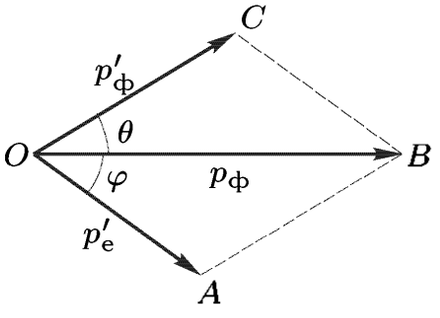


Рис. 12

Но как определить отрезки OB и OC, изображающие импульсы падающего и рассеянного фотонов? Фотоны не заряжены и не оставляют следов в камере Вильсона. Однако начало отрезка OB отмечается точкой, в которой начинается трек электрона отдачи, а длина этого отрезка определяется величиной импульса падающего фотона. Этими

данными отрезок OB определен полностью. Остается определить только отрезок OC, для чего достаточно знать положение точки C. Но она будет определена, если рассеянный фотон сам претерпит комптоновское рассеяние, в результате чего будет отмечен новый электрон отдачи. В этом случае все три отрезка OA, OB и OC будут известны, и можно убедиться, что OB действительно является диагональю параллелограмма, построенного на OA и OC, как того требует квантовая теория эффекта Комптона. Можно, в частности, убедиться, что между углами рассеяния фотона и электрона отдачи действительно выполняется соотношение (3.6). Конечно, все это нельзя получить с полной достоверностью в результате только одного опыта. На основании одного опыта нельзя утверждать, что новый трек оставлен не каким-то случайным электроном, не имеющим отношения к рассматриваемому явлению, а именно новым электроном отдачи, о котором говорилось выше. Это можно утверждать только с той или иной *степенью вероятности*, на основе статистической обработки результатов очень большого числа опытов. Такая статистическая обработка и подтвердила правильность исходных положений фотонной теории эффекта Комптона. Она показала, в частности, что *законы сохранения энергии и импульса выполняются в элементарных актах рассеяния фотонов на электронах.*

7. Д. В. Скобельцыным (р. 1892) наблюдались электроны отдачи в камере Вильсона, помещенной в сильное магнитное поле, при облучении ее γ-лучами радиоактивного препарата. Их траекториями были окружности радиуса

$$R = \frac{cp'_e}{eB}, \tag{3.7}$$

где B — напряженность магнитного поля. Измерив этот радиус, можно с помощью этой формулы вычислить импульс, а затем по формуле

$$\mathscr{E}^2 = (p'_e c)^2 + (m_0 c^2)^2 \tag{3.8}$$

и энергию электрона. Оказалось, что это действительно *релятивистские электроны*, а потому пользоваться теорией относительности для них обязательно.

8. В связи с последним результатом затронем вопрос о невозможности классического истолкования эффекта Комптона. Поскольку электроны в легких атомах могут считаться свободными, никакими собственными частотами они не обладают. Рассмотрим с классической точки зрения их поведение в поле монохроматической электромагнитной волны. Пренебрежем сначала действием магнитного поля. В этом приближении, если пренебречь также затуханием, электрон приходит в колебания, частота которых совпадает с частотой электромагнитной волны. Колеблющийся электрон переизлучает электромагнитные волны, которые и представляют собой рассеянное излучение. Поэтому с классической точки зрения частота рассеянного излучения должна совпадать с частотой падающего.

Этот вывод, однако, был получен в пренебрежении действием магнитного поля. Если учесть магнитное поле $\mathbf{B}$ и снова пренебречь затуханием, то движение свободного электрона будет описываться уравнением

$$\dot{\mathbf{p}} = -e\left(\mathbf{E} + \frac{1}{c}[\mathbf{vB}]\right).$$

В нерелятивистском приближении это — линейное дифференциальное уравнение с переменными коэффициентами. Благодаря этому при вынужденных колебаниях электронов могут возбудиться частоты, отличные от частоты падающей волны. Конечно, такая возможность может проявиться только тогда, когда скорость электрона v не пренебрежимо мала по сравнению со скоростью света. Особо существенной она может быть при релятивистских скоростях, когда движение электрона к тому же осложнено нелинейными эффектами. Но в эффекте Комптона появляются именно релятивистские электроны отдачи, как это следует из опытов с камерой Вильсона. Казалось бы, появляется надежда, что при рассеянии на свободных электронах может возникать изменение частоты электромагнитной волны.

Однако если бы даже это изменение частоты подчинялось наблюдаемой зависимости от угла рассеяния, такую надежду следует оставить. Достаточно поставить вопрос, откуда в непрерывной электромагнитной волне могут появиться релятивистские электроны отдачи? (А опыт показывает, что, как и в случае фотоэффекта, они при освещении гамма-лучами появляются без запаздывания!) Невообразимо, как может непрерывная электромагнитная волна сжаться в плотный сгусток энергии и сосредоточить свое действие на отдельном электроне! Поэтому следует признать, что опыт вынуждает нас принять квантовую точку зрения, согласно которой комптоновское рассеяние излучения и электроны отдачи появляются в результате *одиночного действия* фотона на отдельный электрон.

ЗАДАЧИ

1. В результате комптоновского рассеяния на свободном покоящемся электроне длина волны фотона с энергией $\mathscr{E}_\text{ф}$ увеличилась в α раз. Найти кинетическую энергию $\mathscr{E}_\text{e}$ электрона отдачи.

Ответ. $\mathscr{E}_\text{e} = \mathscr{E}_\text{ф}(\alpha - 1)/\alpha$.

2. Фотон рентгеновского излучения с энергией $\mathscr{E}_\text{ф}$ в результате комптоновского рассеяния на свободном покоящемся электроне отклонился от первоначального направления на угол θ. Определить кинетическую энергию $\mathscr{E}_\text{кин}$ и импульс p'_e электрона отдачи. Показать на основе геометрических соображений, что импульс электрона по абсолютной величине окажется больше импульса падающего фотона, если фотон отклонится от первоначального направления на угол $\theta = \pi/2$.

Ответ.

$$\mathscr{E}_\text{кин} = \frac{hc}{\lambda}\left(1 - \frac{\lambda}{\lambda + 2\lambda_\text{K}\sin^2(\theta/2)}\right), \quad (p'_\text{e}c)^2 = \mathscr{E}^2_\text{кин} + 2m_\text{e}c^2\mathscr{E}_\text{кин}.$$

3. Фотон с длиной волны λ рассеялся на движущемся свободном электроне. В результате электрон остановился, а фотон отклонился от первоначального направления на угол θ. Найти изменение длины волны фотона $\lambda' - \lambda$ в таком процессе. Свести эту задачу к задаче о рассеянии фотона на неподвижном электроне.

Ответ. $\lambda' - \lambda = -(2h/m_e c)\sin^2(\theta/2)$. В результате рассеяния длина волны фотона уменьшается.

4. Гамма-квант с энергией $h\nu$ рассеивается на неподвижном электроне. Найти направления рассеянного кванта, чтобы при взаимодействии с веществом он мог породить электрон–позитронную пару. Найти также предельное значение $\theta_{\text{пред}}$ угла рассеяния, при котором этот процесс возможен для гамма-квантов очень высоких энергий.

Ответ. Рассеянный гамма-квант может породить электрон-позитронную пару, если рассеяние происходит внутрь конуса

$$\cos\theta > \frac{1}{2} + \frac{m_e c^2}{h\nu}.$$

Если же направление рассеяния лежит вне этого конуса, то рождение электрон–позитронных пар невозможно. Для квантов очень высоких энергий ($h\nu \to \infty$) пары могут образовываться, если угол рассеяния не превосходит $\theta_{\text{пред}} = 60^\circ$.

5. Показать, что при взаимодействии с веществом фотон сколь угодно высокой энергии, испытавший комптоновское рассеяние на неподвижном электроне строго назад ($\theta = \pi$), не может породить электрон–позитронную пару.

6. Пусть ν — частота фотона до комптоновского рассеяния на электроне в системе отсчета, где электрон покоится после соударения с фотоном, а ν' — частота фотона после соударения в системе отсчета, где электрон покоился до соударения (т. е. в лабораторной системе). Показать, что $\nu = \nu'$ (теорема взаимности).

Решение. Обозначим рукописными буквами $\mathscr{P}$ и $\mathscr{H}$ четырехмерные импульсы электрона и фотона $(\mathbf{p}, \mathscr{E}/c)$ до соударения. Те же величины после соударения обозначим такими же буквами, но со штрихами. При упругом столкновении четырехмерный импульс системы фотон–электрон сохраняется:

$$\mathscr{P} + \mathscr{H} = \mathscr{P}' + \mathscr{H}'.$$

Переписав это соотношение в виде $\mathscr{P} - \mathscr{H}' = \mathscr{P}' - \mathscr{H}$ и возведя его в квадрат, получим

$$\mathscr{P}^2 + \mathscr{H}'^2 - 2\mathscr{P}\mathscr{H}' = \mathscr{P}'^2 + \mathscr{H}^2 - 2\mathscr{P}'\mathscr{H}.$$

Квадрат 4-импульса частицы в пространстве Минковского [1]) есть инвариант. Для электрона он равен

$$\mathscr{P}^2 = \mathbf{p}^2 - \frac{\mathscr{E}_e^2}{c^2} = \mathbf{p}'^2 - \frac{\mathscr{E}_e'^2}{c^2} = -m_0^2 c^2,$$

где через m_0 обозначена масса покоя электрона. Для фотона такой же инвариант равен нулю: $\mathscr{H}^2 = 0$ (так как масса покоя фотона равна нулю).

[1]) Напомним, что в таком пространстве квадрат 4-вектора равен разности квадратов пространственной и временной составляющих. В скалярное произведение 4-векторов произведение временных составляющих также входит со знаком минус.

Таким образом,

$$\mathscr{P}\mathscr{H}' = \mathscr{P}'\mathscr{H}.$$

В системе отсчета, где электрон до столкновения покоился,

$$\mathscr{P} = (0, m_0 c), \quad \mathscr{H}' = (\mathbf{p}'_\phi, h\nu'/c), \quad \text{так что} \quad \mathscr{P}\mathscr{H}' = -m_0 h\nu'.$$

В системе же отсчета, где электрон покоится после соударения,

$$\mathscr{P}' = (0, m_0 c), \quad \mathscr{H} = (\mathbf{p}_\phi, h\nu/c), \quad \mathscr{P}'\mathscr{H} = -m_0 h\nu.$$

Отсюда $\nu = \nu'$, что и требовалось доказать.

7. Показать, что частота фотона после комптоновского рассеяния в системе отсчета, в которой электрон покоится после соударения, равна частоте падающего фотона в лабораторной системе.

8. Показать, что при комптоновском рассеянии в системе отсчета, все время связанной с электроном, частоты падающего и рассеянного фотонов одинаковы, а импульсы равны по величине, но противоположно направлены.

§ 4. Эффект Доплера при движении источника света в вакууме с фотонной точки зрения

1. Фотоэлектрический эффект и эффект Комптона — типично *квантовые явления*, не допускающие классической трактовки. В этом и следующих трех параграфах рассматриваются явления, которые допускают классическое и квантовое объяснения, согласующиеся между собой. Рассмотрим сначала эффект Доплера (1803–1853) в нерелятивистском приближении.

Возьмем какую-либо инерциальную систему отсчета, в которой источник света массы M движется в вакууме со скоростью v. Энергия источника слагается из кинетической энергии $Mv^2/2$ и внутренней энергии $\mathscr{E}$ возбужденных атомов. При испускании света внутренняя энергия источника изменяется. Начальное значение ее обозначим через $\mathscr{E}$, конечное — через $\mathscr{E}'$. Кроме того, из-за давления излучения источник испытывает отдачу — его скорость получает приращение $(v' - v)$. По законам сохранения энергии и импульса

$$\frac{1}{2}Mv^2 + \mathscr{E} = \frac{1}{2}Mv'^2 + \mathscr{E}' + \mathscr{E}_{\text{изл}}, \tag{4.1}$$

$$Mv = Mv' + \mathbf{p}_{\text{изл}}, \tag{4.2}$$

где $\mathscr{E}_{\text{изл}}$ и $\mathbf{p}_{\text{изл}}$ — энергия и импульс излучения в рассматриваемой инерциальной системе отсчета.

Возведем второе уравнение в квадрат, разделим полученное соотношение на $2M$ и вычтем его из первого. Тогда получим

$$\mathscr{E} - \mathscr{E}' = \mathscr{E}_{\text{изл}} - \mathbf{v}'\mathbf{p}_{\text{изл}} - \frac{\mathbf{p}^2_{\text{изл}}}{2M},$$

или на основании (4.2)

$$\mathscr{E} - \mathscr{E}' = \mathscr{E}_{\text{изл}} - \mathbf{v}\mathbf{p}_{\text{изл}} + \frac{\mathbf{p}^2_{\text{изл}}}{2M}.$$

Если масса источника M велика, то последним членом в этом уравнении можно пренебречь. В этом приближении

$$\mathscr{E} - \mathscr{E}' = \mathscr{E}_{\text{изл}} - \mathbf{v}\mathbf{p}_{\text{изл}}. \tag{4.3}$$

В этом соотношении можно еще исключить импульс излучения. Излучение — существенно *релятивистский объект*. Для него импульс выражается через энергию соотношением $p_{\text{изл}} = \mathscr{E}_{\text{изл}}/c$. Подстановка этого выражения в предыдущее уравнение дает

$$\mathscr{E} - \mathscr{E}' = \mathscr{E}_{\text{изл}}\left(1 - \frac{v}{c}\cos\theta\right), \tag{4.4}$$

где θ — угол между направлением движения излучающего тела и направлением излучения (т. е. угол между векторами $\mathbf{v}$ и $\mathbf{p}$).

Соотношение (4.4) было получено *без использования каких бы то ни было квантовых представлений*. При его выводе использовались только законы сохранения энергии и импульса, справедливые как в классической, так и в квантовой физике. Примем теперь, что излучение происходит в форме фотонов. Предположим, что испущен один фотон. Если частоту фотона в системе отсчета, в которой излучатель движется (т. е. в рассматриваемой нами инерциальной системе отсчета), обозначить через ν', то $\mathscr{E}_{\text{изл}} = h\nu'$. Частоту же излучения в системе отсчета, где излучатель покоится, обозначим через ν. Тогда, рассматривая процесс испускания в такой системе, можно написать $\mathscr{E} - \mathscr{E}' = h\nu$, так как в этой системе изменение внутренней энергии излучателя происходит только за счет испускания светового кванта. Подставляя эти выражения в соотношение (4.4) и сокращая на h, получим

$$\nu' = \frac{\nu}{1 - \dfrac{v}{c}\cos\theta}. \tag{4.5}$$

Эта формула верна только в *первом порядке*, т. е. с точностью до членов первой степени относительно v/c, так как она была получена на основе нерелятивистской механики. Поэтому вместо формулы (4.5) с той же точностью можно пользоваться формулой

$$\nu' = \nu\left(1 + \frac{v}{c}\cos\theta\right). \tag{4.6}$$

2. Получим теперь для эффекта Доплера *релятивистскую формулу*, которой можно пользоваться уже при любых значениях скорости v. В этом случае нет смысла разделять полную (релятивистскую) энергию тела на кинетическую и внутреннюю. Под $\mathscr{E}$ и $\mathscr{E}'$ мы будем теперь понимать *полные энергии* тела в рассматриваемой инерциальной системе отсчета до и после испускания света соответственно. Соответствующие энергии покоя тела до и после испускания обозначим через $\mathscr{E}_0$ и $\mathscr{E}'_0$. Снова будем оперировать с энергиями и импульсами, вводя квантовые представления лишь на заключительном этапе вычислений.

Законы сохранения энергии и импульса при испускании света удобно записать в форме

$$\mathscr{E}' = \mathscr{E} - \mathscr{E}_{\text{изл}}, \quad \mathbf{p}' = \mathbf{p} - \mathbf{p}_{\text{изл}}. \tag{4.7}$$

Возведем оба равенства в квадрат, умножив предварительно второе из них на c, а затем вычтем почленно. С учетом соотношений

$$\mathscr{E}^2 - (\mathbf{p}c)^2 = \mathscr{E}_0^2, \quad \mathscr{E}'^2 - (\mathbf{p}'c)^2 = \mathscr{E}_0'^2, \quad \mathscr{E}_{\text{изл}} = cp_{\text{изл}}$$

получим

$$\mathscr{E}_0'^2 = \mathscr{E}_0^2 - 2\mathscr{E}\mathscr{E}_{\text{изл}} + 2c^2\mathbf{p}'\mathbf{p}_{\text{изл}}.$$

Учтем теперь, что

$$p = mv = \frac{\mathscr{E}}{c^2}\,v = \frac{\mathscr{E}\beta}{c},$$

где $\beta = v/c$. Тогда

$$\mathscr{E}_0^2 - \mathscr{E}_0'^2 = 2\mathscr{E}\mathscr{E}_{\text{изл}}(1 - \beta\cos\theta), \tag{4.8}$$

или

$$\mathscr{E}_0 - \mathscr{E}_0' = \frac{2\mathscr{E}}{\mathscr{E}_0 + \mathscr{E}_0'}\,\mathscr{E}_{\text{изл}}(1 - \beta\cos\theta).$$

Здесь, как и прежде, θ означает угол между направлением движения излучающего тела и направлением распространения излучения. Если масса тела велика, то в знаменателе последнего соотношения можно положить $\mathscr{E}_0' = \mathscr{E}_0$. Учтя еще, что $\mathscr{E} = \mathscr{E}_0/\sqrt{1-\beta^2}$, придем к соотношению

$$(\mathscr{E}_0 - \mathscr{E}_0')\sqrt{1-\beta^2} = \mathscr{E}_{\text{изл}}(1 - \beta\cos\theta). \tag{4.9}$$

Теперь воспользуемся квантовыми соотношениями $\mathscr{E}_0 - \mathscr{E}_0' = h\nu$ и $\mathscr{E}_{\text{изл}} = h\nu'$, совершенно такими же, как и в нерелятивистском случае. В результате и получим *релятивистскую формулу* для эффекта Доплера:

$$\nu' = \nu\,\frac{\sqrt{1-\beta^2}}{1 - \beta\cos\theta}. \tag{4.10}$$

3. Обе формулы фотонной теории для эффекта Доплера, нерелятивистская (4.5) и релятивистская (4.10), совпадают с соответствующими формулами, которые дает классическая волновая теория света (см. т. IV, § 107, 108). Глубокая причина такого совпадения состоит в том, что в фотонной теории обе формулы получаются из соотношений (4.3) и (4.9). Эти же соотношения, поскольку они выведены из одних только законов сохранения энергии и импульса при испускании света, остаются справедливыми независимо от того, представляет ли свет непрерывные волны или состоит из отдельных фотонов. Единственное предположение квантового характера, которое было сделано при выводе формул для частоты испускаемого света, состоит в том, что энергия фотона однозначно определяется его частотой. Более того, связь между этими величинами должна быть специального вида:

энергия фотона пропорциональна частоте света. Значение же коэффициента пропорциональности, т. е. постоянная Планка h, выпадает из окончательных формул для частоты колебаний. А то обстоятельство, что связь между $\mathscr{E}$ и ν должна быть именно такой, видно из того, что при преобразовании Лорентца величины $\mathscr{E}$ и ν преобразуются одинаково. Только благодаря этому, как уже отмечалось выше, соотношение $\mathscr{E} = h\nu$ релятивистски инвариантно.

Мы вернемся к эффекту Доплера в § 6, где предполагается, что источник света движется не в вакууме, а в среде с показателем преломления n.

§ 5. Отражение и преломление света в фотонной теории. Фотоны в среде

1. В XIX веке считалось, что опытом Фуко окончательно опровергнута корпускулярная теория света. Этот опыт доказал, что скорость света в воде меньше, чем в вакууме. Однако по волновой теории такое соотношение между скоростями относится к *фазовой скорости* света. В опыте же Фуко, если бы свет действительно представлял собой волны, то измерялась бы *групповая скорость* (см. т. IV, § 108). Поэтому с чисто логической точки зрения аргументация физиков XIX века должна быть дополнена указанием на то, что в опыте Фуко скорость света измерялась в практически недиспергирующих средах. В этом случае фазовая скорость совпадает с групповой. Заключение о несостоятельности корпускулярной теории остается верным, с той существенной оговоркой, что оно относится не к корпускулярной теории вообще, а к корпускулярной теории *в форме Ньютона*. Как же обстоит дело в фотонной теории Эйнштейна?

Отражение и преломление света в корпускулярной теории Ньютона рассматривалось на основе классической ньютоновской механики. К фотонам классическая механика неприменима. Природа фотонов — *двойственная*. При распространении они ведут себя как *волны* и только при взаимодействии с веществом проявляют *свойства частиц*. А поскольку отражение и преломление являются специальными случаями распространения света, фотонная теория при рассмотрении этих явлений должна приводить к тем же результатам, что и классическая волновая теория.

2. Этот вопрос заслуживает более подробного рассмотрения, которое и приводится ниже. В среде, как и в вакууме, можно ввести представление о распространении света в виде фотонов. При этом среду мы будем считать *неподвижной*. Иными словами, будем рассматривать явления в системе отсчета, относительно которой среда неподвижна. Такая система является *выделенной*, или *привилегированной*. Вопрос заключается в том, каковы энергия и импульс фотона в среде? Заметим, что *энергия и импульс фотона слагаются из энергии и импульса электромагнитного поля и среды*. Но в нашем рассмотрении идет речь о сумме этих величин, т. е. о *полных* энергии и импульсе, связанных

с распространением фотона в среде. Среду мы будем считать изотропной и характеризовать ее свойства показателем преломления $n(\omega)$.

По фотонным представлениям при отражении и преломлении *общее число фотонов не меняется. Если граница неподвижна, то не меняются и энергии отраженного и прошедшего фотонов.* Докажем это утверждение сначала для отражения, предполагая, что фотон падает на среду из вакуума. Энергия падающего фотона $\mathscr{E}_e = \hbar\omega$, т. е. однозначно определяется частотой ω. Такое же соотношение между энергией и частотой должно существовать и для отраженного фотона, поскольку он распространяется в таком же вакууме, как и падающий фотон. Но при отражении от неподвижной границы частота ω не меняется. В противном случае происходило бы изменение цвета отраженного луча, чего на самом деле не происходит. Таким образом, энергия отраженного фотона $\mathscr{E}_r$ также равна $\hbar\omega$, т. е. $\mathscr{E}_r = \mathscr{E}_e$.

3. Для прошедшего фотона указанное утверждение не столь очевидно. Чтобы его доказать, рассмотрим плоскопараллельный пучок, состоящий из N_e монохроматических фотонов, падающий из вакуума под любым углом на неподвижную границу среды. Пусть число отраженных фотонов будет N_r, прошедших N_d, а соответствующие им энергии $\mathscr{E}_r$ и $\mathscr{E}_d$. По закону сохранения энергии

$$N_e\mathscr{E}_e = N_r\mathscr{E}_r + N_d\mathscr{E}_d. \tag{5.1}$$

Условие сохранения общего числа фотонов:

$$N_e = N_r + N_d. \tag{5.2}$$

Наконец, по доказанному $\mathscr{E}_r = \mathscr{E}_e$. Отсюда получается $\mathscr{E}_d = \mathscr{E}_e$, что и требовалось доказать. Таким образом, и для фотонов в среде справедливо соотношение $\mathscr{E} = \hbar\omega$.

Рассмотрим теперь импульс фотона в среде. Как и в случае энергии, можно написать

$$N_e p_e = N_r p_r + N_d p_d, \quad N_e = N_r + N_d, \tag{5.3}$$

где индексы e, r, d относятся к падающему, отраженному и прошедшему фотонам. Рассмотрим сначала поведение касательных составляющих импульсов. Направим ось X вдоль границы раздела, а ось Z — перпендикулярно к ней. Любая теория должна приводить к закону отражения, из которого следует, что $p_{ex} = p_{rx}$. Но тогда для p_{ex}, p_{rx}, p_{dx} получается в точности такая же система уравнений, как и для соответствующих энергий. Из нее находим

$$p_{ex} = p_{rx} = p_{dx}, \tag{5.4}$$

т. е. касательные составляющие импульсов всех трех фотонов одинаковы.

Совсем иначе ведут себя нормальные составляющие. В этом случае по-прежнему

$$N_e p_{ez} = N_r p_{rz} + N_d p_{dz}.$$

Однако если падающий фотон приближается к среде, то отраженный удаляется от нее, так что $p_{rz} = -p_{ez}$. В результате получится

$$(N_e + N_r)p_{ez} = N_d p_{dz},$$

или

$$(N_e + N_r)p_{ez} = (N_e - N_r)p_{dz},$$

или, наконец,

$$p_d = \frac{N_e + N_r}{N_e - N_r} \frac{\cos\varphi}{\cos\psi} p_e,$$

где φ — угол падения, а ψ — угол преломления.

Конечно, связь между p_d и p_e определяется свойствами среды и не может зависеть от характера поляризации падающего света. Это позволяет упростить вычисления. Предположим, что падающий свет поляризован перпендикулярно к плоскости падения. Тогда электрические векторы всех фотонов будут коллинеарны, так что на основании формулы Френеля

$$\frac{N_r}{N_e} = \frac{n\cos\psi - \cos\varphi}{n\cos\psi + \cos\varphi}$$

(см. т. IV, § 65). Действительно, в монохроматической волне все фотоны когерентны, и поэтому N_r/N_e дает отношение *амплитуд*, а не интенсивностей отраженной и падающей волн. Выбор знака в последней формуле определен из тех соображений, что числа фотонов N_r и N_e существенно положительны. В результате получается

$$p_d = np_e. \tag{5.5}$$

4. Резюмируя изложенное, приходим к заключению, что, как и в вакууме, энергия и импульс фотона в веществе определяются формулами

$$\mathscr{E} = \hbar\omega, \quad \mathbf{p} = \hbar\mathbf{k}. \tag{5.6}$$

Однако в веществе волновое число

$$k = n\omega/c, \tag{5.7}$$

т. е. в n раз больше, чем в вакууме. Если ввести фазовую скорость $v = c/n$ и длину волны $\lambda = 2\pi v/\omega$ в рассматриваемой среде, то получатся формулы

$$k = \omega/v = 2\pi/\lambda, \tag{5.8}$$

одинаково применимые к вакууму и веществу, так как в вакууме $v = c$. Связь между энергией и импульсом фотона в среде принимает вид

$$\mathscr{E} = pv = pc/n. \tag{5.9}$$

Разумеется, это соотношение справедливо только в привилегированной системе отсчета, в которой среда покоится.

Из изложенного видно, что фотонная теория преломления света аналогична ньютоновской. Обе теории определяют угол преломления светового луча из закона сохранения импульса, точнее — его тангенциальной составляющей. Только ньютоновская теория связывает этот

закон с направлением силы, действующей на световую корпускулу нормально к границе раздела, что в случае фотона лишено смысла. Количественное же различие между результатами обеих теорий связано с тем, что *импульс ньютоновской корпускулы пропорционален ее скорости v, а импульс фотона в среде обратно пропорционален.* По этой причине фотонная теория приводит к такому же выражению для показателя преломления, как и классическая волновая теория.

Представление о фотонах в среде применимо не только к преломлению света, но и ко многим другим явлениям, о которых говорится в последующих параграфах.

5. В заключение — одно принципиальное замечание, затрагивающее уже *физическое толкование* корпускулярно-волнового дуализма. Как уже было отмечено выше, общее число фотонов при отражении и преломлении не меняется. Если на границу раздела падает всего один фотон, то после падения он будет обнаружен либо в виде отраженного, либо в виде прошедшего фотона. Импульс у отраженного фотона — такой же по величине, как и у падающего, но имеет иное направление. Импульс прошедшего фотона меняет не только направление, но и величину. Не является ли это нарушением закона сохранения импульса в *элементарном процессе*, т. е. для единичного фотона? Не является ли этот закон лишь статистическим законом, как это фактически требовалось в пп. 2 и 3? Такое заключение отвергается современной квантовой теорией.

Явления интерференции фотонов приводят к заключению, что при падении фотона на границу раздела возникает *новое состояние*, не поддающееся описанию на языке классической теории. Оно характеризуется тем, что *фотон отчасти находится о состоянии отражения, отчасти в состоянии преломления. Закон сохранения импульса соблюдается и в элементарном процессе.* Если же произвести опыт по обнаружению фотона (в квантовой механике его называют *измерением*), то будет обнаружен либо отраженный, либо прошедший фотон. Но измерение меняет состояние системы, нарушения закона сохранения импульса не происходит. В каком состоянии будет обнаружен фотон, — в состоянии отражения или в состоянии преломления, — это заранее предсказать с достоверностью невозможно. Можно указать лишь *вероятности* того и другого состояний.

§ 6. Излучение Вавилова–Черенкова. Эффект Доплера при движении источника света в среде

1. Пусть источник света движется в изотропной среде с показателем преломления $n(\omega)$. Законы сохранения энергии и импульса при испускании света можно записать совершенно так же, как и в случае вакуума, т. е. в виде (4.7), сохранив при этом прежние обозначения. Разумеется, мы пользуемся привилегированной системой отсчета, в которой рассматриваемая среда *неподвижна.* С помощью одних только законов сохранения, конечно, нельзя решить вопрос, произойдет или не

произойдет испускание света. Для этого необходимо привлечь более детальные электродинамические уравнения. Но *если законы сохранения не выполняются, то испускание невозможно.* Законы сохранения не раскрывают механизма явления. Они могут дать только *необходимое*, но не достаточное условие испускания.

С уравнениями (4.7) мы поступим в точности так же, как в § 4, с той только разницей, что связь энергии с импульсом излучения запишем в виде $p_{\text{изл}} = n\mathscr{E}_{\text{изл}}/c$. Тогда получим

$$\mathscr{E}_0^2 - \mathscr{E}_0'^2 = \mathscr{E}_{\text{изл}}^2(n^2-1) - 2\mathscr{E}\mathscr{E}_{\text{изл}}(\beta n\cos\theta - 1). \tag{6.1}$$

При $n = 1$, как и должно быть, это соотношение переходит в (4.8). Однако, в отличие от вакуума, уравнение (6.1) учитывает то обстоятельство, что в явлении участвуют не только излучающий атом и само излучение, но и *среда*, в которой движется атом. Это учитывается феноменологически — посредством показателя преломления среды.

2. Применим уравнение (6.1) к движению частицы, *внутреннее состояние которой при движении не меняется*, т. е. частица все время находится в одном и том же квантовом состоянии. В случае устойчивой частицы это будет всегда, когда энергии взаимодействия ее с окружающей средой недостаточно, чтобы перевести частицу в более высокое энергетическое состояние. Так будет обстоять дело, например, при движении электрона, протона и другой, даже неустойчивой, элементарной частицы, если только во время полета она не распадется или не испытает какого-либо другого превращения. Во всех этих случаях от внутренних степеней свободы частицы можно отвлечься, считая ее как бы бесструктурной. Масса m_0, а с ней и энергия покоя $\mathscr{E} = m_0c^2$ частицы во время движения остаются постоянными. В уравнении (6.1) следует положить $\mathscr{E}_0' = \mathscr{E}_0$, что дает

$$\mathscr{E}_{\text{изл}}(n^2-1) - 2\mathscr{E}(\beta n\cos\theta - 1) = 0,$$

если только испускание света действительно происходит ($\mathscr{E}_{\text{изл}} \neq 0$). С учетом соотношения $\mathscr{E} = \mathscr{E}_0/\sqrt{1-\beta^2}$ отсюда получаем

$$\mathscr{E}_{\text{изл}} = \frac{2\mathscr{E}_0}{\sqrt{1-\beta^2}(n^2-1)}(\beta n\cos\theta - 1), \tag{6.2}$$

а потому

$$\cos\theta = \frac{1}{\beta n} + \frac{\sqrt{1-\beta^2}(n^2-1)}{2\beta n}\left(\frac{\mathscr{E}_{\text{изл}}}{\mathscr{E}_0}\right). \tag{6.3}$$

Если частица испустила один квант, то $\mathscr{E}_{\text{изл}} = \hbar\omega$, а потому

$$\cos\theta = \frac{1}{\beta n(\omega)} + \frac{\sqrt{1-\beta^2}(n^2(\omega)-1)}{2\beta n(\omega)}\,\frac{\hbar\omega}{\mathscr{E}_0}. \tag{6.4}$$

Второе слагаемое в этой формуле учитывает отдачу, которую получает излучающая частица при испускании кванта $\hbar\omega$. Если пренебречь этим

слагаемым (что допустимо, когда оно по модулю очень мало по сравнению с единицей), то получится классическая формула

$$\cos\theta = \frac{1}{\beta n(\omega)}. \tag{6.5}$$

Из формулы (6.5) следует, что $\cos\theta > 0$, т. е. испускание света может происходить *только вперед.* Для испускания света частоты ω необходимо, чтобы было $\beta n(\omega) > 1$. То же получается и из общей формулы (6.4), так как оба слагаемых в правой части этой формулы существенно положительны. Если ввести фазовую скорость света в среде по формуле $n(\omega) = c/v_{\text{фаз}}$, то получится

$$v > v_{\text{фаз}}, \tag{6.6}$$

т. е. *для испускания света необходимо, чтобы скорость частицы превосходила фазовую скорость света в среде.* Отсюда следует, что испускание невозможно при движении частицы в вакууме, так как в этом случае $v_{\text{фаз}} = c$. Энергия излучения черпается за счет кинетической энергии частицы, так как при испускании света внутреннее состояние частицы остается без изменения. По этой причине движение частицы замедляется. Однако ускорение само по себе на механизм явления не влияет. В этом смысле говорят, что испускание света возможно и при *равномерном движении* заряженной частицы в среде, если только выполнено условие (6.6).

Рассмотренное явление было открыто в 1934 г. и экспериментально исследовано П. А. Черенковым (р. 1904) в лаборатории и под руководством С. И. Вавилова (1891–1951). Его называют *излучением Вавилова–Черенкова.* Классическая теория явления была создана И. Е. Таммом (1895–1971) и И. М. Франком (р. 1908), а квантовая — В. Л. Гинзбургом (р. 1916) (см. т. IV, § 38).

3. Рассмотрим теперь случай, когда *при движении внутреннее состояние излучающей частицы изменяется.* Примером может служить движущийся атом или ион. В этом случае испускание света происходит за счет внутренней и кинетической энергии атома. Как и в § 4, частоту излучаемого света неподвижного атома обозначим через ω, а движущегося — через ω'. (Эти обозначения отличаются от обозначений предыдущего пункта.) Кроме того,

$$\mathscr{E}_0^2 - \mathscr{E}_0'^2 = (\mathscr{E}_0 + \mathscr{E}_0')(\mathscr{E}_0 - \mathscr{E}_0') = [2\mathscr{E}_0 - (\mathscr{E}_0 - \mathscr{E}_0')](\mathscr{E}_0 - \mathscr{E}_0').$$

При подстановке этого выражения в (6.1), ради упрощения, пренебрежем квадратами малых величин $\mathscr{E}_{\text{изл}}$ и $(\mathscr{E}_0 - \mathscr{E}_0')$. Обе эти величины, как видно из дальнейшего, пропорциональны $\hbar^2$. В указанном приближении, следовательно, учитываются члены, линейные по $\hbar$ и отбрасываются члены, пропорциональные $\hbar^2$. Благодаря этому из окончательного выражения для частоты излучаемого света выпадает постоянная Планка $\hbar$, и для эффекта Доплера получается классический результат. В нем, с точностью до малых квантовых поправок, содержатся все существенные выводы. Выполнив указанную подстановку и учтя при

этом, что $\mathscr{E} = \mathscr{E}_0/\sqrt{1-\beta^2}$, получим классическую формулу

$$\mathscr{E}_{\text{изл}} = \frac{\sqrt{1-\beta^2}(\mathscr{E}_0 - \mathscr{E}_0')}{1-\beta n\cos\theta}. \tag{6.7}$$

Возможны три случая.

1) $1-\beta n > 0$, т. е. скорость источника меньше фазовой скорости света в среде. В этом случае тем более $1-\beta n\cos\theta > 0$, т. е. знаменатель в (6.7) положителен. Так как при испускании всегда $\mathscr{E}_{\text{изл}} > 0$, то испускание возможно лишь при условии $\mathscr{E}_0 - \mathscr{E}_0' > 0$, т. е. атом в результате испускания должен переходить с более высокого на более низкий энергетический уровень. В частности, невозбужденный атом в рассматриваемом случае излучать не может. Чтобы от энергии перейти к частоте, заметим, что $\mathscr{E}_{\text{изл}} = \hbar\omega'$, а $\mathscr{E}_0 - \mathscr{E}_0' = \hbar\omega$, так как в случае неподвижного атома вся его внутренняя энергия при переходе с верхнего уровня $\mathscr{E}_0$ на нижний $\mathscr{E}_0'$ превращается в излучение. В результате из (6.7) получается формула

$$\omega' = \frac{\omega\sqrt{1-\beta^2}}{1-\beta n\cos\theta} = \frac{\omega\sqrt{1-\beta^2}}{1-(v/v_{\text{фаз}})\cos\theta}. \tag{6.8}$$

Она отличается от соответствующей формулы для вакуума (4.10) только тем, что вместо скорости c в ней стоит фазовая скорость света в среде. На поперечном эффекте Доплера ($\cos\theta = 0$) влияние среды никак не сказывается. Он определяется исключительно релятивистским замедлением времени (см. т. IV, § 107).

2) $1-\beta n < 0$, т. е. скорость источника больше фазовой скорости света в среде. В этом случае, на который обратили внимание В. Л. Гинзбург и И. М. Франк, говорят о *сверхсветовом* или *аномальном эффекте Доплера*. Здесь в свою очередь есть две возможности.

Во-первых, испускание света происходит под таким углом θ, что знаменатель в (6.7) положителен, т. е. $\cos\theta < 1/\beta n$. Иными словами, излучение направлено *вне черенковского конуса*, т. е. конуса, определяемого уравнением $\cos\theta = 1/\beta n$. Как и в случае (1), испускание возможно, если $\mathscr{E}_0 > \mathscr{E}_0'$, т. е. атом должен переходить с более высокого энергетического уровня на более низкий. Частота ω' определяется прежней формулой (6.8).

Во-вторых, испускание света происходит под таким углом θ, что знаменатель в (6.7) отрицателен, т. е. $\cos\theta > 1/\beta n$. Это значит, что излучение распространяется *внутри* черенковского конуса $\cos\theta = 1/\beta n$. Для возможности испускания необходимо, чтобы было $\mathscr{E}_0 < \mathscr{E}_0'$, — при испускании атом должен переходить с более низкого на более высокий уровень, т. е. возбуждаться. Испускание света и возбуждение атома происходят за счет кинетической энергии атома. В этом случае $\mathscr{E}_0' - \mathscr{E}_0 = \hbar\omega$, т. е. ω означает частоту, с которой испускался бы свет неподвижным атомом при переходе с верхнего уровня $\mathscr{E}_0'$ на нижний $\mathscr{E}_0$. Она равна той частоте света, которую способен поглощать неподвижный атом при обратном переходе с нижнего уровня $\mathscr{E}_0$ на верхний $\mathscr{E}_0'$.

Что касается $\mathscr{E}_{\text{изл}}$, то эта величина определяется прежним выражением $\mathscr{E}_{\text{изл}} = \hbar\omega'$. Поэтому получается формула

$$\omega' = \frac{\omega\sqrt{1-\beta^2}}{\beta n\cos\theta - 1} = \frac{\omega\sqrt{1-\beta^2}}{(v/v_{\text{фаз}})\cos\theta - 1}. \tag{6.9}$$

3) $1 - \beta n = 0$, т. е. атом движется со скоростью, равной фазовой скорости света в среде. В этом случае, если атом заряжен, появляется излучение Вавилова–Черенкова.

§ 7. Фотоны в гравитационном поле

1. Рассмотрим с квантовой точки зрения изменение частоты света и искривление светового луча в гравитационном поле. Первый эффект уже рассматривался классически в т. I (§ 72) и в т. IV (§ 109) на основе принципа эквивалентности поля тяготения и ускоренного движения. Полученные там результаты выводятся здесь из закона сохранения энергии с использованием связи между энергией и частотой фотона: $\mathscr{E} = \hbar\omega$.

Согласно теории относительности всякая энергия обладает массой, причем инертная и гравитационная массы равны между собой. Применим это положение к ограниченному пучку света с энергией $\mathscr{E}$, распространяющемуся в постоянном гравитационном поле. Гравитационный потенциал поля $\varphi(\mathbf{r})$ может меняться в пространстве. Поскольку свет обладает гравитационной массой $m = \mathscr{E}'/c^2$, гравитационное поле над ним совершает работу. Если свет переходит из точки с гравитационным потенциалом φ в точку с гравитационным потенциалом $\varphi + d\varphi$, то энергия света получает приращение

$$d\mathscr{E} = -Gm\,d\varphi = -G\frac{\mathscr{E}}{c^2}\,d\varphi,$$

где G — гравитационная постоянная. Интегрируя это уравнение между точками 1 и 2, получим

$$\ln\frac{\mathscr{E}_2}{\mathscr{E}_1} = \frac{G}{c^2}(\varphi_1 - \varphi_2). \tag{7.1}$$

Это соотношение имеет общий характер и не содержит еще никаких квантовых предположений. Оно в равной мере справедливо и в классической, и в квантовой физике. Но получить из него соотношение для частот возможно лишь с использованием зависимости между энергией и частотой, которая дается в квантовой теории. В самом деле, допустим, что световой пучок состоит всего из одного фотона частоты ω. В этом случае $\mathscr{E} = \hbar\omega$, и соотношение (7.1) переходит в

$$\ln\frac{\omega_2}{\omega_1} = \frac{G}{c^2}(\varphi_1 - \varphi_2). \tag{7.2}$$

Постоянная Планка выпала из окончательного результата. Результат не зависит от ее числового значения. Так и должно быть во всех случаях, если окончательный результат совпадает с классическим.

Первоначально соотношение (7.2) было проверено астрономически — по смещению спектральных линий в поле тяготения звезд. Открытие эффекта Мёссбауэра (р. 1929) позволило Паунду (р. 1919) и Ребке в 1960 г. проверить его в земных условиях. В их опытах было измерено изменение частоты света при прохождении в поле тяжести Земли всего 19,6 м по вертикали. Этот вопрос будет разобран в ядерной физике.

2. Искривление светового луча в поле тяготения также может быть разобрано на основе принципа эквивалентности без привлечения квантовых представлений. Впервые вопрос был поставлен и решен именно таким путем Эйнштейном в 1911 г., еще до создания общей теории относительности. Впрочем, *полное решение вопроса может быть дано только в рамках общей теории относительности.* Решение, приводимое ниже, дает правильный результат лишь с точностью до числового множителя. Оно по существу классическое, хотя по форме выглядит как квантовое.

Предположим, что фотон пролетает мимо Солнца или другого массивного небесного тела массы M. Если бы не было поля тяготения, то он двигался бы прямолинейно. Фотон обладает инертной массой, которую мы обозначим через m (разумеется, речь идет о релятивистской массе, так как масса покоя фотона равна нулю). По принципу эквивалентности инертная масса всегда равна массе гравитационной. Поэтому фотон будет подвергаться воздействию силы тяготения GMm/R^2, направленной к центру Солнца (R — расстояние от центра Солнца). Влияние касательной составляющей этой силы было выяснено в предыдущем пункте — она вызывает изменение частоты световой волны. Нормальная составляющая искривляет траекторию фотона, т. е. световой луч. Поэтому при прохождении мимо Солнца световой луч должен отклоняться к его центру.

Вычислим угол отклонения светового луча. В отсутствие поля тяготения луч был бы прямой линией AB (рис. 13). Будем считать, что в поле тяготения он мало отличается от AB. Задача сводится к вычислению импульса $\int F_n\,dt$ нормальной силы F_n, действующей на фотон, за все время движения. Интеграл должен быть вычислен вдоль *истинной траектории* фотона. Но в рассматриваемом случае можно применить *метод возмущений*, заменив при вычислении интеграла истинную траекторию невозмущенной прямолинейной траекторией AB фотона. Допустим, что невозмущенный луч касается края Солнца. Тогда, как видно из рис. 13,

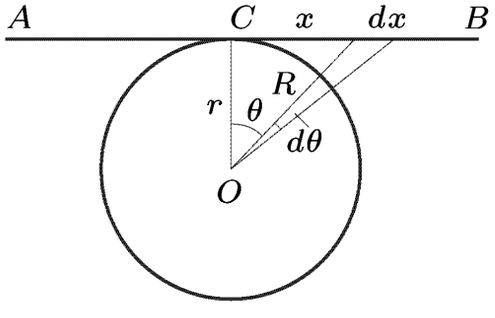


Рис. 13

$$F_n = G\frac{Mm}{R^2}\cos\theta = G\frac{Mm}{r^2}\cos^3\theta,$$

$$x = r\,\mathrm{tg}\,\theta, \quad dx = \frac{r}{\cos^2\theta}\,d\theta,$$

где r — радиус Солнца. Следовательно,

$$\int F_n\,dt = \int F_n \frac{dx}{c} = \frac{GMm}{cr}\int\limits_{-\pi/2}^{+\pi/2} \cos\theta\,d\theta = \frac{2GMm}{cr}.$$

Этот импульс нормальной силы должен быть равен изменению количества движения фотона. В рассматриваемом приближении количество движения фотона меняется только по направлению, но не по величине. Его изменение равно $mc\varphi$, где φ — угол поворота светового луча. Приравнивая оба выражения, получим

$$\varphi = \frac{2GM}{c^2 r}.$$

Общая теория относительности приводит к вдвое большему результату:

$$\Phi = \frac{4GM}{c^2 r}. \tag{7.3}$$

Для Солнца эта величина равна $\Phi = 1{,}75''$, что согласуется с экспериментом.

§ 8. Некоторые опыты по обнаружению корпускулярных свойств света

1. А.Ф. Иоффе (1880–1960) и Н.И. Добронравов (1891–1949) в 1925 г. экспериментально подтвердили квантовую картину фотоэффекта, о которой говорилось в § 2. Вместо видимых и ультрафиолетовых лучей они пользовались рентгеновскими лучами, кванты которых много крупнее. В их опытах микроскопическая пылинка, заряжавшаяся и перезаряжавшаяся при освещении рентгеновскими лучами, уравновешивалась в электрическом поле конденсатора, аналогичного тому, с помощью которого Милликен определял заряд электрона. Пока заряд пылинки оставался постоянным, она неподвижно висела в электрическом поле конденсатора и наблюдалась сбоку с помощью микроскопа. Но если пылинку освещали слабым потоком рентгеновских лучей, то происходил фотоэффект. Время от времени на пылинку попадал рентгеновский квант, освобождавший электрон. Заряд пылинки менялся на заряд одного электрона, и пылинка выходила из равновесия. На опыте оказалось, что именно такой квантовой картине фотоэффекта соответствовало поведение пылинки.

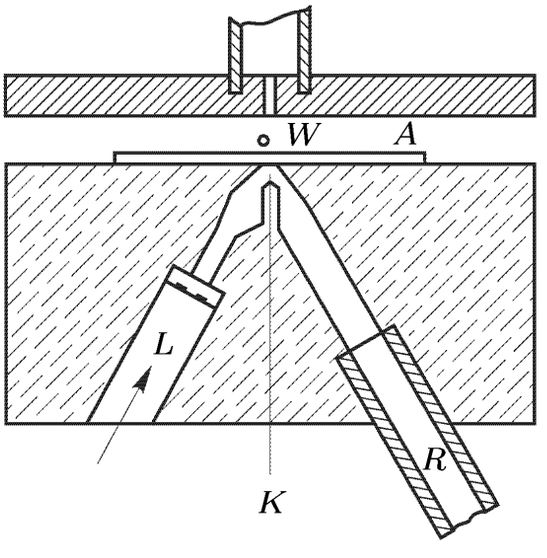


Рис. 14

На рис. 14 показана схема опыта Иоффе и Добронравова. В толстой эбонитовой пластинке просверлены отверстия L и R. Через отверстие R из образовавшейся полости откачивался воздух, чтобы полость стала прозрачной для ультрафиолета. Через отверстие L,

закрывавшееся кварцевым окошком, проходили ультрафиолетовые лучи, освещавшие конец алюминиевой проволочки K с диаметром 0,2 мм. Образовавшиеся фотоэлектроны ускорялись электрическим напряжением 12 000 В, приложенным между проволочкой и алюминиевой фольгой A, закрывающей полость сверху. Толщина фольги ($\sim 5 \cdot 10^{-3}$ мм) подбиралась так, что она практически не поглощала рентгеновские лучи, возбуждавшиеся в ней при торможении электронов. Освещение кончика проволочки K подбиралось настолько слабым, что число фотоэлектронов и связанных с ними рентгеновских импульсов составляло около 1000 в секунду. Алюминиевая фольга одновременно служила нижней обкладкой конденсатора. От нее на расстоянии примерно 0,02 см уравновешивалась висмутовая пылинка W размером около $6 \cdot 10^{-5}$ см.

Опыты показали, что в среднем раз в 30 минут пылинка выходила из равновесия, т. е. с такой средней частотой рентгеновские лучи вырывали из нее электрон. В течение указанного времени образовывалось около $N = 30 \cdot 60 \cdot 1000 = 1{,}8 \cdot 10^6$ ренгтгеновских импульсов. По классическим представлениям энергия каждого импульса должна распространяться во все стороны в виде сферической волны. Каждый из таких импульсов отдавал бы пылинке ничтожную часть своей энергии из-за малости телесного угла, под которым пылинка видна из ближайшего места фольги, где возбуждались рентгеновские лучи. Кроме того, эта энергия распределялась бы между множеством электронов пылинки. При таких условиях было бы совершенно невероятно, чтобы в течение 30 минут большая доля энергии электронов пылинки сосредоточилась только на одном электроне, который должен вылететь из пылинки. Ясно, что с точки зрения классической волновой теории результаты опытов Иоффе и Добронравова непонятны. Напротив, в квантовой теории они вполне естественны.

Следующее элементарное вычисление показывает, что квантовая интерпретация явления выдерживает и количественную проверку. Если считать, что для рентгеновского фотона все направления движения после вылета равновероятны, то вероятность попадания фотона в пылинку будет равна $\Omega/4\pi$, где Ω — телесный угол, под которым пылинка видна из точки, откуда исходят рентгеновские лучи. Если ежесекундно источник испускает в среднем n фотонов, то за время t он испустит $N = nt$ фотонов. Из них на пылинку попадет в среднем $N\Omega/4\pi = nt\Omega/4\pi$ фотонов. Среднее время t в течение которого на пылинку попадает один фотон, определится из условия $nt\Omega/4\pi = 1$, откуда

$$t = \frac{4\pi}{n\Omega}.$$

В опытах Иоффе и Добронравова было $n = 1000$ с$^{-1}$, $\Omega = (\pi/4) \times \times (6 \cdot 10^{-5}/0{,}02)^2$, если пылинку считать шариком. Подстановка этих значений в предыдущее выражение дает $t \approx 1800$ с $= 30$ мин, что согласуется со временем, наблюдавшимся в опытах Иоффе и Добронравова.

2. По классическим представлениям лучистая энергия распространяется от источника в разные стороны одновременно. Но это не так, если распространение происходит квантами. Отдельные кванты испускаются независимо друг от друга. Этот эффект, если он действительно существует, должен быть выражен тем отчетливее, чем крупнее кванты. Для его обнаружения на опыте Боте (1891–1957) в 1924 г. пользовался рентгеновскими лучами.

В опытах Боте кванты излучения обнаруживались счетчиками Гейгера (1882–1945). Счетчик Гейгера представляет собой маленький цилиндрический конденсатор A, наполненный газом под давлением в несколько мм рт. ст. (рис. 15). Внутренней обкладкой конденсатора служит тонкая металлическая

нить, натянутая вдоль оси цилиндрика, или тщательно заточенное металлическое острие, укрепленное внутри цилиндрика на изоляторе. Внутренняя и наружная обкладки конденсатора соединены через батарею B и очень большое сопротивление R, нижний конец которого заземлен. Таким образом, между обкладками конденсатора поддерживается разность потенциалов около 1000 В. Вблизи нити или острия создается сильное резко неоднородное электрическое поле. Если в счетчик попадает какая-либо частица (в том числе и фотон), то она ионизует молекулы газа. Образовавшиеся электроны и ионы ускоряются этим полем и при столкновениях с молекулами газа сами ионизуют их. Новые ионы и электроны также ускоряются, и т. д. Возникает электронно-ионная лавина, через счетчик проходит импульс электрического тока заметной силы. Из-за этого приложенное напряжение перераспределяется. Основная часть напряжения будет приходиться на сопротивление R, а напряжение на конденсаторе счетчика упадет почти до нуля. Ток через счетчик прекратится, напряжение на конденсаторе возрастет, счетчик вернется в рабочее состояние. Импульс электрического тока, прошедший через счетчик, можно обнаружить чувствительным электрометром E, присоединенным к внутренней обкладке конденсатора счетчика. По этим импульсам можно регистрировать (считать) каждую отдельную частицу, вызвавшую первичную ионизацию газа, если только время зажигания разряда меньше времени между попаданиями в счетчик двух последовательных частиц. Современные счетчики снабжены автоматически действующими устройствами, производящими такую регистрацию. Счетчик Гейгера сыграл и продолжает играть большую роль при изучении ядерной физики, космических лучей, элементарных частиц и пр.

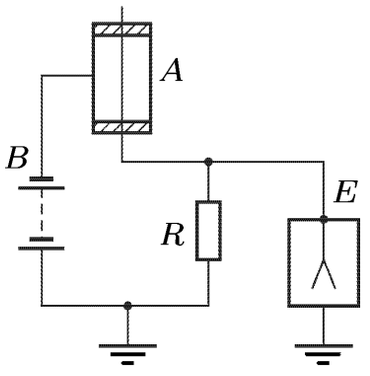


Рис. 15

В опыте Боте применялись два быстродействующих счетчика C_1 и C_2 (рис. 16), регистрировавших прохождение каждого рентгеновского кванта. Симметрично между счетчиками помещалась тонкая железная или медная фольга A. Фольга освещалась сбоку рентгеновскими лучами R достаточной жесткости и благодаря этому сама становилась источником характеристического рентгеновского излучения (рентгеновская флуоресценция). Попадание рентгеновского кванта в один из счетчиков вызывало немедленное (менее чем через 0,001 с) вздрагивание нити электрометра. Такие вздрагивания регистрировались автоматически на движущейся ленте. Если бы из источника A во все стороны распространялись волны, как того требует классическая теория, то срабатывания счетчиков происходили бы одновременно и согласованно (с малыми случайными вариациями). Если же источник A испускал бы рентгеновские кванты вправо и влево беспорядочно и независимо, то срабатывания счетчиков были бы также беспорядочными во времени и независимыми. Опыт показал, что все происходит так, как если бы рентгеновское излучение флуоресценции распространялось в виде квантов, случайно направленных то к правому, то к левому счетчику. Так и должно быть по квантовым представлениям.

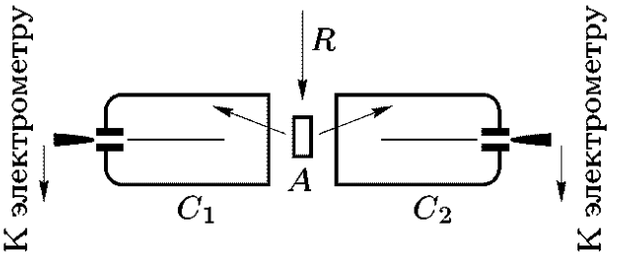


Рис. 16

3. В обоих описанных опытах, по существу, изучались *флуктуационные явления* в слабых потоках рентгеновского излучения. Наблюдать аналогичные явления с видимым светом затруднительно из-за малой величины световых квантов в этой области спектра. Чувствительность существующих объективных методов регистрации еще недостаточна, чтобы регистрировать отдельные кванты видимого света. Наблюдения слабых световых потоков проводил С. И. Вавилов. Он воспользовался тем, что периферические участки сетчатки человеческого глаза (палочки), достаточно долго пробывшего в темноте, обладают необходимой чувствительностью к слабому свету, с которым приходилось иметь дело на опыте. Вавилов воспользовался также существованием порога зрительного ощущения. Если число квантов, попадающих при кратковременной вспышке на периферический участок сетчатки, превышает некоторую минимальную величину n_0, то глаз видит вспышку. Если же оно меньше n_0, то вспышка не видна совсем. По оценке Вавилова n_0 составляет всего несколько десятков и, возможно, даже несколько квантов. Существование порога зрительного ощущения систематически использовалось Вавиловым при изучении люминесценции и вместе с Черенковым при открытии явления, носящего их имя. Используя указанные свойства глаза, Вавилов разработал чувствительный метод визуальных наблюдений флуктуации интенсивности видимого света в слабых световых потоках. В результате многолетних экспериментальных исследований таких флуктуации Вавилов пришел к заключению, что они происходят аналогично флуктуациям числа молекул газа, т. е. так, как если бы свет состоял из конечных порций, или квантов.

Глава II

СТРОЕНИЕ, ЭНЕРГЕТИЧЕСКИЕ УРОВНИ И СПЕКТРЫ АТОМА

§ 9. Ядерная модель атома и опыты Резерфорда

1. В первой четверти XX-го века было установлено, что атом состоит из положительно заряженного *ядра* и окружающей его *электронной оболочки.* Линейные размеры ядра порядка 10^{-13}–10^{-12} см. Размеры самого атома, определяемые электронной оболочкой, примерно в 10^5 раз больше. Однако почти вся масса атома (не менее 99,95 %) сосредоточена в ядре. Это связано с тем, что ядро состоит из «тяжелых» протонов и нейтронов, а электронная оболочка — из одних только «легких» электронов ($m_{\mathrm{p}} = 1836{,}15m_{\mathrm{e}}$, $m_{\mathrm{n}} = 1838{,}68m_{\mathrm{e}}$). Число электронов в оболочке нейтрального атома равно заряду ядра, если за единицу принять элементарный заряд (т. е. заряд электрона по абсолютной величине). Но электронная оболочка может терять или приобретать электроны. Тогда атом становится электрически заряженным, т. е. превращается в положительный или отрицательный *ион.*

Химические свойства атома определяются электронной оболочкой, точнее, ее наружными электронами. Такие электроны сравнительно слабо связаны с атомом и поэтому наиболее подвержены электрическим воздействиям со стороны наружных электронов соседних атомов. То же относится к силам притяжения или отталкивания между нейтральными атомами и молекулами (к молекулярным силам). Напротив, протоны и нейтроны прочно связаны внутри ядра. Чтобы воздействовать на ядро, нужны силы, в миллионы раз превосходящие те силы, которые достаточны для отрыва наружных электронов атома. Однако строение и свойства электронной оболочки определяются в конце концов электрическим полем ядра атома.

2. Если изложенная модель атома соответствует действительности, то атом должен быть в высокой степени прозрачным для пронизывающих его частиц. Для пучка электронов это было установлено еще Ленардом. Однако окончательное экспериментальное доказательство этой модели атома было дано Резерфордом (1871–1937) в 1911 г. Поэтому она по справедливости называется *моделью Резерфорда.* По предложению и под руководством Резерфорда его ученики Гейгер и Марсден (1889–1970) исследовали количественно рассеяние α-частиц, испускаемых радиоактивными веществами. В их опытах параллельный пучок α-частиц направлялся в вакууме на тонкую металлическую фольгу и рассеивался ею. Применялся визуальный метод регистрации рассеянных α-частиц. При ударе о флуоресцирующий экран из сернистого цинка α-частица оставляла на нем вспышку (сцинтилляцию). Отдель-

ные сцинтилляции можно было наблюдать в темноте через лупу или микроскоп. И экспериментаторы производили подсчет таких сцинтилляций.

Оказалось, что подавляющее число α-частиц рассеивалось на небольшие углы порядка 1–3°. Угловое распределение таких частиц хорошо описывалось кривой случайных ошибок Гаусса (1777–1855). Однако наблюдались также отдельные α-частицы, отклоняющиеся на большие углы, доходившие до 150°. Относительное число таких частиц было ничтожно. Например, при прохождении через платиновую фольгу пучка α-частиц от RaC [1]) из 8000 падающих частиц в среднем только одна частица отклонялась на угол, превышавший 90°. Но и этого было бы слишком много, если бы большие отклонения возникали в результате накопления множества случайных отклонений.

Резерфорд сделал вывод, что каждое большое отклонение появляется в результате *единичного акта* взаимодействия какого-то практически точечного силового центра с близко пролетающей α-частицей. Таким силовым центром и является положительно заряженное *ядро атома.* Сама α-частица есть также атомное ядро, именно ядро атома гелия. Это подтверждается тем, что α-частица может быть получена в результате двукратной ионизации атома гелия, как это было уже ранее установлено тем же Резерфордом. Электростатическое взаимодействие между этими двумя ядрами и вызывает рассеяние α-частиц на большие углы.

Изложенное подтверждается снимками треков α-частиц в камере Вильсона. Обычно конец трека α-частицы никакими особенностями не отличается. Но изредка наблюдаются треки, заканчивающиеся изломами и «вилками». Такой случай воспроизведен на рис. 17. Здесь зафиксировано столкновение α-частицы с ядром. В результате столкновения направление движения α-частицы резко изменилось, а пришедшее в движение ядро оставило новый трек, который вместе с треком самой α-частицы образовал «вилку».

3. Резерфорд разработал и количественную теорию рассеяния α-частиц. В этой теории к взаимодействию α-частицы с ядром приме-

[1]) На самом деле исследовавшиеся α-частицы испускались RaC′, являющимся короткоживущим продуктом RaC. (Период полураспада RaC′ — около 10^{-6} с.)

няется *закон Кулона*. Это, конечно, гипотеза, так как α-частица может подходить к ядру на расстояния порядка 10^{-12} см (см. задачу 1 к этому параграфу), а на таких расстояниях закон Кулона не был проверен экспериментально. Разумеется, движение α-частицы в поле ядра Резерфордом рассматривалось *классически*. Применимость классической механики в рассматриваемом случае можно обосновать с помощью квантовой механики, возникшей несколько позднее (см. § 20). Наконец, масса ядра предполагается большой по сравнению с массой α-частицы, так что ядро может считаться неподвижным. От последнего предположения легко освободиться, заменив массу α-частицы *приведенной массой*.

В опытах Резерфорда применялись очень тонкие металлические фольги с толщиной порядка 10^{-5}–10^{-4} см. В таких случаях при рассеянии на большие углы можно было не учитывать многократные столкновения α-частицы с атомными ядрами. Вероятность двукратных, а тем более многократных столкновений с большими отклонениями ничтожна. Ничтожна вероятность рассеяния на большие углы и на электронах ввиду малости их масс. Многократные столкновения с ядрами и с электронами атомных оболочек играют роль только при очень малых углах рассеяния. Такие углы мы исключим из рассмотрения. Тогда, учитывая взаимодействие α-частицы только с одним ядром, к которому α-частица подходит наиболее близко, мы придем к *задаче двух тел*. От всех остальных ядер α-частица проходит много дальше, и поэтому взаимодействием с ними пренебрегается. Таким образом, теория Резерфорда применима для больших отклонений, когда отклонение вызывается только электрическим полем *одного ядра*, так что по сравнению с этим отклонением все прочие отклонения, вместе взятые, пренебрежимо малы. Соответствующее рассеяние называется *резерфордовским*. Оно является *упругим* в том смысле, что кинетическая энергия α-частицы в результате рассеяния не изменяется, т. е. не растрачивается на возбуждение атомов, а тем более атомных ядер.

Сформулированная задача формально аналогична *задаче Кеплера* (1571–1630) о движении планеты вокруг Солнца. И тут и там сила взаимодействия тел — центральная и меняется обратно пропорционально квадрату расстояния между ними. В случае планеты это сила притяжения, в случае α-частицы — сила отталкивания. Это проявляется в том, что планета (в зависимости от ее полной энергии) может двигаться и по эллипсу, и по гиперболе, а α-частица — только по гиперболе. Но в математических вычислениях это не имеет значения. Угол рассеяния α-частицы θ равен углу между асимптотами ее гиперболической траектории (рис. 18). Для него была получена формула

$$\operatorname{ctg}\frac{\theta}{2} = \frac{mbv^2}{2Ze^2} \tag{9.1}$$

(см. т. I, § 58). Здесь m — масса α-частицы, v — ее скорость в «бесконечности», т. е. вдали от ядра, Ze — заряд ядра, $2e$ — заряд α-частицы,

равный удвоенному элементарному заряду e. (Число Z называется *зарядовым числом* ядра. Ради краткости его часто называют просто *зарядом ядра*, подразумевая, что за единицу принят элементарный заряд e.) Через b обозначено *прицельное расстояние*, т. е. длина перпендикуляра, опущенного из ядра на невозмущенную прямолинейную траекторию α-частицы (или, что то же самое, на касательную к реальной траектории, когда α-частица находилась бесконечно далеко от ядра).

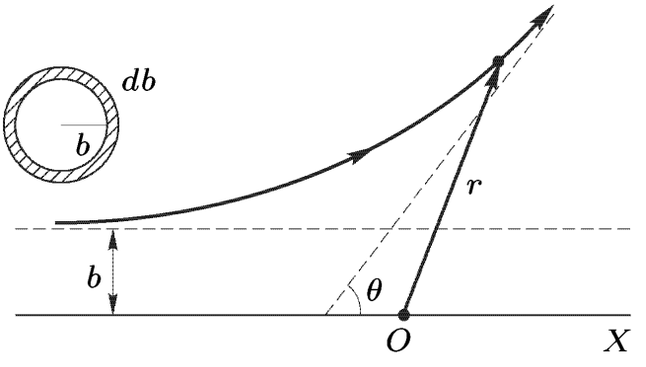


Рис. 18

4. Экспериментальной проверке в области атомных явлений, разумеется, доступна не сама формула (9.1), а *статистические следствия* из нее. Введем так называемое *дифференциальное эффективное сечение рассеяния*. Обозначим через I интенсивность плоскопараллельного пучка α-частиц, налетающих на ядро, т. е. число α-частиц пучка, проходящих в единицу времени через единичную площадку, перпендикулярную к потоку. Из этого числа через элементарную площадку $d\sigma$, также перпендикулярную к потоку, проходит $dN_1 = I\,d\sigma$ α-частиц. После рассеяния эти частицы попадают в элементарный телесный угол $d\Omega$. Конечно, величина телесного угла $d\Omega$ и направление его оси определяются величиной и положением площадки $d\sigma$. Поэтому dN_1 имеет также смысл числа α-частиц, рассеиваемых ядром в единицу времени в телесный угол $d\Omega$. Отношение dN_1 к I равно $d\sigma$ и имеет размерность площади. Оно и называется *дифференциальным эффективным сечением ядра для рассеяния α-частиц в телесный угол $d\Omega$*. Это понятие применяется к рассеянию не только α-частиц, но и *любых частиц*, а также к *другим процессам*, происходящим с частицами. Таким образом, по определению

$$d\sigma = \frac{dN_1}{I}, \tag{9.2}$$

т. е. *дифференциальное эффективное сечение рассеяния есть отношение числа частиц, рассеянных атомом в единицу времени в телесный угол $d\Omega$, к интенсивности I падающих частиц.*

Определим теперь дифференциальное сечение для рассеяния α-частиц на отдельном ядре атома. Задача сводится к определению величины площадки $d\sigma$, пройдя через которую α-частица после рассеяния попадает внутрь заданного телесного угла $d\Omega$. Возьмем за ось X прямолинейную траекторию той α-частицы, которой соответствует прицельное расстояние $b = 0$ (такая частица испытала бы с ядром лобовое столкновение). Используя цилиндрическую симметрию, для упрощения заменим $d\sigma$ на кольцевую площадку $d\sigma = 2\pi b\,db$, перпендикулярную к потоку. Внутренний радиус такой площадки равен b, наружный $b + db$, а центр расположен на оси X (рис. 18, слева вверху). Интервалу $b, b + db$ соответствует интервал углов рассеяния θ, $\theta + d\theta$,

причем по формуле (9.1)

$$db = \frac{Ze^2}{mv^2}\,\frac{d\theta}{\sin^2(\theta/2)}.$$

Введя телесный угол $d\Omega = 2\pi \sin\theta\, d\theta$, в который рассеиваются α-частицы, прошедшие через кольцевую площадку, нетрудно получить

$$d\sigma = \left(\frac{Ze^2}{mv^2}\right)^2 \frac{d\Omega}{\sin^4(\theta/2)}. \tag{9.3}$$

В таком виде формула (9.3) справедлива для *любой* элементарной площадки $d\sigma$, а не только для кольцевой. Она и называется *формулой Резерфорда* (см. задачу 3 к § 20).

5. Прежде чем пойти дальше, введем понятие *полного сечения рассеяния* или какого-либо другого процесса. Оно определяется как *отношение полного числа частиц, претерпевших рассматриваемый процесс в единицу времени, к интенсивности падающего пучка частиц.* Полное сечение σ может быть получено из дифференциального сечения $d\sigma$ путем интегрирования его по всем возможным значениям $d\Omega$. В случае рассеяния α-частиц в формуле (9.3) следует сначала положить $d\Omega = 2\pi \sin\theta\, d\theta$, а затем выполнить интегрирование в пределах от $\theta = 0$ до $\theta = \pi$ (см. задачу 2 к этому параграфу). Это дает $\sigma = \infty$. Результат этот понятен. Чем дальше площадка $d\sigma$ удалена от оси X, тем меньше угол рассеяния θ. Частицы, проходящие через удаленные площадки, практически не отклоняются, т. е. проходят в окрестности угла рассеяния $\theta = 0$. Суммарная площадь таких площадок, а с ней и полное число рассеянных частиц бесконечно велики. Бесконечно велико и полное поперечное сечение рассеяния. Впрочем, этот вывод имеет формальный характер, так как при малых углах рассеяния формула Резерфорда (9.3) неприменима.

6. Приведем теперь формулу (9.3) к виду, доступному для экспериментальной проверки. Акты рассеяния α-частиц различными атомами *независимы.* Отсюда следует, что если n — число ядер (атомов) в единице объема, то число α-частиц, рассеиваемых объемом V в единицу времени в телесный угол $d\Omega$, определяется выражением

$$dN = VnI\, d\sigma = VnI \left(\frac{Ze^2}{mv^2}\right)^2 \frac{d\Omega}{\sin^4(\theta/2)}. \tag{9.4}$$

В таком виде формула Резерфорда и была подтверждена на опыте. В частности, на опыте было показано, что при постоянстве $d\Omega$ величина $dN \sin^4(\theta/2)$ постоянна, т. е. не зависит от угла рассеяния θ, как это и должно быть согласно формуле (9.4).

Подтверждение формулы Резерфорда на опыте может рассматриваться как *косвенное доказательство закона Кулона* на таких малых расстояниях, на какие могут сближаться центры α-частицы и взаимодействующего с ней ядра. Другим доказательством могут служить опыты Блэкетта (1897–1974) по рассеянию α-частиц в газах. Фотографировалось большое количество треков α-частиц в камере Вильсона,

измерялись угловые отклонения их и подсчитывалось, как часто встречаются определенные углы рассеяния. Эти опыты также подтвердили формулу Резерфорда. Но главной целью их была проверка закона Кулона. Оказалось, что при расстояниях между центрами α-частицы и взаимодействующего ядра в случае воздуха от $3 \cdot 10^{-12}$ до $5 \cdot 10^{-10}$ см, а в случае аргона от $7 \cdot 10^{-12}$ до 10^{-9} см закон Кулона подтверждается экспериментально. Отсюда не следует, что этот закон справедлив на любых расстояниях между центрами взаимодействующих ядер. Опыты по упругому рассеянию легких ядер, ускоренных ускорителями, также на легких, но неподвижных ядрах показали, что наблюдаются резкие отступления от закона Кулона, когда указанное расстояние уменьшается до 10^{-12} см и меньше. На таких расстояниях проявляют свое действие *ядерные силы* притяжения, перекрывающие кулоновские силы отталкивания ядер.

7. Формулу (9.4) можно применить для измерения заряда ядра. Для этого надо измерить dN и I. После этого можно вычислить Z, так как все прочие величины в формуле (9.4) могут считаться известными. Основная трудность состоит в том, что величины dN и I очень сильно отличаются друг от друга. В первых опытах они измерялись на различных установках, т. е. в разных условиях, что вносило значительные ошибки. В опытах Чэдвика (1891–1974) этот недостаток был устранен. Рассеивающая фольга имела форму кольца AA' (рис. 19), радиоактивный препарат R (источник α-частиц) и флуоресцирующий экран S из ZnS устанавливались на оси кольца на одинаковых расстояниях от него.

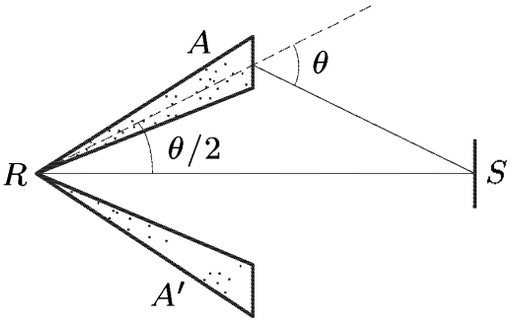


Рис. 19

Для подсчета сцинтилляций от α-частиц, рассеянных фольгой, отверстие кольца AA' закрывалось экраном, непрозрачным для α-частиц. Наоборот, для измерения I производился подсчет сцинтилляций, когда отверстие было свободно, а кольцо AA' закрыто. Так как в этом случае число сцинтилляций было очень велико, то для его уменьшения перед экраном S устанавливался вращающийся диск с узким вырезом. Зная ширину выреза и сосчитав число сцинтилляций, можно вычислить I. Чэдвик нашел для платины $Z = 77{,}4$, серебра $Z = 46{,}3$, меди $Z = 29{,}3$. Атомные или порядковые номера этих элементов в периодической системе Менделеева равны соответственно 78, 47, 29. Тем самым был подтвержден уже известный результат, впервые установленный Мозли (1887–1915), что заряд ядра Z совпадает с атомным номером элемента (см. § 48).

8. Вернемся снова к модели атома, обоснованной опытами Резерфорда. Могут ли атомное ядро и окружающая его электронная оболочка образовать устойчивую систему, какой, несомненно, является атом? Если бы это было возможно, то эти частицы не могли бы находиться в покое. В противном случае получилась бы электростатическая система (практически) точечных зарядов, между которыми действуют

кулоновские силы, а такая система, согласно теореме Ирншоу (см. т. III, § 9), неустойчива. Кулоновы силы меняются обратно пропорционально квадрату расстояния между взаимодействующими частицами. Но так же меняются гравитационные силы между телами планетной системы. Устойчивость планетной системы обеспечивается вращением планет вокруг Солнца. Поэтому Резерфорд естественно пришел к *планетарной модели атома*, в которой электроны вращаются вокруг ядра.

Однако, согласно классической электродинамике, при движении заряда меняется и электромагнитное поле, источником которого является заряд. В частности, *электрический заряд, движущийся ускоренно, излучает электромагнитные волны.* Вращающийся электрон имеет ускорение, а потому должен непрерывно излучать. Теряя энергию на излучение, электрон непрерывно приближался бы к ядру и в конце концов упал бы на него (см. задачу 4 к этому параграфу). Таким образом, и при наличии движения получается *неустойчивая* модель атома. Можно было бы предположить, что закон Кулона и прочие законы, определяющие электромагнитное поле в электродинамике, нарушаются в случае элементарных частиц и малых расстояний. Можно было бы учесть ядерные силы и ввести неизвестные нам гипотетические силы, обеспечивающие устойчивость атома. Но и это не спасает положения. Каковы бы ни были силы, согласно общим принципам классической механики спектр излучения атома должен состоять из нескольких основных частот и соответствующих им обертонов. Опыт приводит к совсем иной закономерности, выражаемой *комбинационным принципом Ритца* (1878–1909) (см. § 11). Приходится констатировать, что классическая механика и электродинамика оказались не в состоянии объяснить *существование атомов* как устойчивых систем атомных ядер и электронов. Решение этой проблемы было получено только в рамках квантовой механики,

ЗАДАЧИ

1. На какое расстояние r при лобовом столкновении могут сближаться центры α-частицы с энергией $\mathscr{E} = 6$ МэВ и неподвижного ядра золота? Заряд ядра золота $Z = 79$.

Ответ. $r = 2Ze^2/\mathscr{E} = 3{,}8 \cdot 10^{-12}$ см.

2. Определить сечение $\Delta\sigma$ резерфордовского рассеяния α-частицы на атомном ядре на угол рассеяния, превышающий θ.

Ответ.

$$\Delta\sigma = 4\pi \left(\frac{Ze^2}{mv^2}\right)^2 \operatorname{ctg}^2 \frac{\theta}{2}. \tag{9.5}$$

3. В опытах по рассеянию α-частиц применялась платиновая фольга толщиной $\delta = 8 \cdot 10^{-5}$ см. Пробег α-частицы от радия C′ в воздухе (при 15 °C и давлении 760 мм рт. ст.) оказался равным 6,96 см. Из графика, выражающего соотношение между пробегом и кинетической энергией $\mathscr{E}$ α-частицы, можно найти, что $\mathscr{E} \approx 5{,}9$ МэВ. Определить относительную долю α-частиц, рассеивавшихся на угол, превышающий $\theta = 90°$. Плотность платины $\rho = 21{,}5$ г/см^3, заряд ядра $Z = 78$, атомная масса $A = 195$.

Р е ш е н и е. Согласно формуле (9.5) поперечное сечение рассеяния в углы $\theta > \pi/2$ определяется выражением

$$\Delta\sigma = 2\pi \left(\frac{Ze^2}{\mathscr{E}}\right)^2,$$

так как кинетическая энергия α-частицы $\mathscr{E} = mv^2/2$. Число рассеивающих ядер в фольге $N = S\delta n$, где S — полная площадь фольги, а n — число ядер в единице объема. На одно ядро приходится площадь $s = S/N = 1/\delta n$. В углы $\theta > \pi/2$ попадают α-частицы, рассеивающиеся только в результате единичных актов столкновения с ядрами. Такие столкновения *независимы*. Поэтому относительная доля частиц, рассеиваемых в такие углы, будет

$$\frac{\Delta\sigma}{s} = 2\pi \left(\frac{Ze^2}{\mathscr{E}}\right)^2 \delta n.$$

Подставляя сюда $n = \rho N_\text{A}/A$, где N_A — постоянная Авогадро, получим

$$\frac{\Delta\sigma}{s} = 2\pi \left(\frac{Ze^2}{\mathscr{E}}\right)^2 \frac{\rho}{A}\,\delta N_\text{A} = \frac{1}{8200}.$$

4. Через какой промежуток времени t электрон, вращающийся вокруг протона по окружности радиуса $a_0 = 0{,}53 \cdot 10^{-8}$ см, упал бы на ядро вследствие потерь энергии на излучение, если бы к нему были применимы классические механика и электродинамика?

Р е ш е н и е. Несмотря на наличие излучения, при приближенном вычислении ускорения электрона можно пользоваться формулой $|v| = v^2/a$, где a — переменное расстояние электрона от ядра. При вращении по окружности $mv^2/a = e^2/a^2$, откуда

$$\frac{mv^2}{2} = \frac{e^2}{2a}.$$

Полная энергия электрона

$$\mathscr{E} = \frac{mv^2}{2} - \frac{e^2}{a} = -\frac{e^2}{2a}.$$

Потеря энергии на излучение в единицу времени

$$-\frac{d\mathscr{E}}{dt} = \frac{2}{3}\frac{e^2}{c^3}v^2 = \frac{2}{3}\frac{e^2}{c^3}\frac{v^4}{a^2} = \frac{2}{3}\frac{e^2}{m^2c^3a^4}.$$

Подставив сюда выражение для $\mathscr{E}$ получим

$$\frac{d}{dt}\left(\frac{1}{a}\right) = \frac{4}{3}\frac{e^4}{m^2c^3a^4},$$

откуда

$$a_0^3 - a^3 = 4\frac{e^4}{m^2c^3}\,t.$$

Полагая здесь $a = 0$, находим искомое время падения:

$$t = \frac{m^2c^3a_0^3}{4e^4} = 1{,}6 \cdot 10^{-11}\text{с}.$$

§ 10. Определение заряда ядра из рассеяния рентгеновских лучей

1. В § 3 при рассмотрении рассеяния рентгеновских лучей на свободных электронах мы интересовались только изменением длины волны в зависимости от угла рассеяния. Для решения этого вопроса была достаточна простая квантовая теория, основанная только на законах сохранения энергии и импульса. Если же требуется определить интенсивность и поляризацию рассеянного излучения в различных направлениях, то необходимо уже пользоваться полной системой уравнений квантовых электродинамики и релятивистской механики. Именно так такая задача была решена О. Клейном (1894–1977) и Нишиной (1890–1951) в 1929 г. и более строго И. Е. Таммом в 1930 г. Рассмотрение этого вопроса далеко выходит за рамки настоящего руководства. Однако и простая классическая теория приводит к правильному результату в предельном случае, когда энергия падающего кванта $h\nu$ мала по сравнению с собственной энергией электрона $m_e c^2$ (или, что то же самое, когда длина волны λ велика по сравнению с комптоновской длиной λ_K для электрона). Такой случай представляет определенный интерес, так как для легких элементов он дает независимый метод определения заряда ядра Z. Рассмотрим этот вопрос при указанном условии, что $h\nu \ll m_e c^2$. Однако рентгеновские кванты будем предполагать все же настолько жесткими, что их энергия велика по сравнению с энергией связи электронов, так что электроны могут считаться свободными. Удовлетворить обоим условиям можно только для *легких элементов.*

2. Свободный электрон в монохроматическом электрическом поле $E = E_0 \cos \omega t$ получает ускорение

$$\ddot{x} = -\frac{e}{m} E = -\frac{e}{m} E_0 \cos \omega t$$

(если пренебречь действием магнитного поля). Согласно классической электродинамике такой электрон излучает, теряя энергию в виде рассеянного излучения. Энергия, рассеиваемая электроном в единицу времени, дается выражением

$$-\frac{d\mathscr{E}}{dt} = \frac{2}{3}\frac{e^2}{c^3}\ddot{x}^2 = \frac{2}{3}\frac{e^4}{m^2 c^3} E_0^2 \cos^2 \omega t$$

(см. т. III, § 141). Среднее по времени значение этой величины равно

$$-\overline{\frac{d\mathscr{E}}{dt}} = \frac{1}{3}\frac{e^4}{m^2 c^3} E_0^2.$$

Рассеянием на тяжелых атомных ядрах можно полностью пренебречь, так как в этом случае в знаменатель последней формулы войдет большая величина — масса заряженной частицы в квадрате.

Если падающая волна плоская, то плотность потока электромагнитной энергии численно равна вектору Пойнтинга

$$S = \frac{c}{4\pi} EH = \frac{c}{4\pi} E_0^2 \cos^2 \omega t.$$

Ее среднее значение по времени $\overline{S} = (c/8\pi)E_0^2$. Разделив среднюю рассеиваемую энергию на $\overline{S}$, получим *полное поперечное сечение рассеяния* на свободном электроне:

$$\sigma_{\text{Т}} = \frac{8\pi}{3}\frac{e^4}{m_{\text{e}}^2 c^4} = 0{,}6652448(33)\cdot 10^{-24}\ \text{см}^2. \tag{10.1}$$

Эта формула была получена еще Томсоном на заре электронной теории. Величина $\sigma_{\text{Т}}$ называется *томсоновским поперечным сечением рассеяния для электрона*. Ее можно представить в виде

$$\sigma_{\text{Т}} = \frac{8}{3}\pi r_{\text{e}}^2, \tag{10.2}$$

где r_{e} — так называемый *классический радиус электрона*:

$$r_{\text{e}} = \frac{e^2}{m_{\text{e}}c^2} = 2{,}8179380(70)\cdot 10^{-13}\ \text{см}. \tag{10.3}$$

Согласно (10.1) интенсивность рассеяния в рентгеновской области спектра совсем не зависит от частоты падающего излучения. Напротив, в оптической области интенсивность света, рассеянного атомами и молекулами, а также любыми малыми неоднородностями среды, пропорциональна четвертой степени частоты (см. т. IV, § 98). Это различие связано с тем, что в рентгеновской области спектра ускорение электрона (а от него зависит рассеяние) определяется самим электрическим полем. В оптической же области размеры атомов и молекул малы по сравнению с длиной световой волны. В пренебрежении резонансными эффектами здесь напряженностью электрического поля определяется дипольный момент частицы $p = \beta E$. Рассеяние же пропорционально квадрату его второй производной по времени, т. е. четвертой степени частоты ω. То же относится и к неоднородностям среды, если только их линейные размеры малы по сравнению с λ.

Точная формула Клейна–Нишины для полного сечения комптоновского рассеяния на неподвижном свободном электроне имеет вид

$$\frac{\sigma}{\sigma_{\text{Т}}} = \frac{3}{4}\left\{\frac{1+\gamma}{\gamma^3}\left[\frac{2\gamma(1+\gamma)}{1+2\gamma} - \ln(1+2\gamma)\right] + \frac{1}{2\gamma}\ln(1+2\gamma) - \frac{1+3\gamma}{(1+2\gamma)^2}\right\}, \tag{10.4}$$

где $\gamma = h\nu/m_{\text{e}}c^2$ — отношение энергии падающего кванта к энергии покоя электрона. Формула эта хорошо согласуется с опытными данными. По сравнению с формулой Томсона (10.2) она дает монотонное убывание σ с возрастанием энергии падающего кванта. Это убывание иллюстрируется табл. 1. Максимального значения σ достигает при $\gamma \to 0$. Оно равно $\sigma_{\text{Т}}$.

3. Перейдем теперь к рассеянию падающего пучка рентгеновского излучения всем атомом. В случае достаточно жесткого излучения электроны атомной оболочки будут рассеивать это излучение *независимо*, т. е. *некогерентно*. Тем более это будет справедливо для электронов различных атомов. Заметив это, рассмотрим параллельный пучок

рентгеновских лучей с единичной площадью поперечного сечения, распространяющийся в направлении оси X. Между двумя поперечными сечениями этого пучка с координатами x и $x + dx$ находится $n\,dx$ атомов, где n — число атомов в единице объема. Если атом нейтрален,

Таблица 1

Поперечное сечение комптоновского рассеяния при различных энергиях первичных фотонов

γ	$\sigma/\sigma_{\text{т}}$	γ	$\sigma/\sigma_{\text{т}}$
0,05	0,913	5	0,191
0,1	0,84	10	0,123
0,2	0,737	20	0,0754
0,33	0,637	50	0,0376
0,5	0,563	100	0,0215
1	0,431	200	0,0122
2	0,314	500	0,00556
3	0,254	1000	0,00304

то в нем содержится Z электронов. В единицу времени они рассеивают энергию $Z\sigma I$, а электроны всех $n\,dx$ атомов — энергию $Z\sigma I n\,dx$, где I — интенсивность падающего пучка. Из-за такого рассеяния интенсивность пучка уменьшается на $I\varkappa\,dx$, где $\varkappa$ называется *коэффициентом рассеяния* и определяется выражением

$$\varkappa = \sigma n Z. \tag{10.5}$$

Ослабление пучка происходит не только из-за рассеяния, но и из-за *поглощения* рентгеновских лучей. Поглощение сопровождается выделением тепла внутри тела, так что оно принципиально может быть отделено от рассеяния. Сам коэффициент рассеяния $\varkappa$ пропорционален n, т. е. плотности ρ тела. Поэтому на опыте целесообразно измерять отношение $\varkappa/\rho$. Очевидно, $\rho = nAm_{\text{н}}$, где $m_{\text{н}}$ — масса атома водорода, A — относительная атомная масса рассеивающего вещества. Используя для σ томсоновское значение (10.1), нетрудно получить

$$\frac{\varkappa}{\rho} = \frac{\sigma_{\text{т}}}{m_{\text{н}}}\,\frac{Z}{A} \approx 0{,}40\,\frac{Z}{A}\ \text{см}^2/\text{г}. \tag{10.6}$$

Найденная на опыте величина $\varkappa/\rho$ оказалась для легких элементов не зависящей от длины волны и равной приблизительно 0,20 $\text{см}^2/\text{г}$. Используя это значение, из (10.6) получаем $Z/A \approx 1/2$, т. е у легких элементов (за исключением водорода) зарядовое число Z (численно совпадающее с порядковым номером элемента) равно приблизительно половине массового числа A. Это действительно приближенно оправдывается в начале периодической системы элементов. Физическая причина такой закономерности и отступлений от нее будет выяснена при изучении атомного ядра.

§ 11. Спектральные закономерности

1. Накаленные твердые тела испускают *сплошные спектры.* У газов (наряду со сплошной областью) наблюдаются *линейчатые* и *полосатые спектры.* Линейчатый спектр состоит из ряда закономерно расположенных более или менее узких спектральных линий. В полосатом спектре полосы кажутся сплошными при наблюдении в спектроскоп малой разрешающей силы. При применении спектрального аппарата высокой разрешающей силы они распадаются на множество тесно расположенных спектральных линий.

К началу XX-го века было выяснено, что *линейчатые спектры газов испускаются атомами и ионами, а полосатые — молекулами.* Поэтому их называют также *атомными* и *молекулярными спектрами.* Атомный спектр водорода удается наблюдать при электрическом разряде в вакуумной водородной трубке только тогда, когда большая часть молекул водорода диссоциировала на атомы. Но в парах йода полосы молекулярного спектра в основном исчезают уже в процессе диссоциации молекул I_2 на атомы I.

Наличие многих спектральных линий атома указывает на сложность его внутренней структуры. Неудивительно, что богатейший материал, накопленный эмпирически при изучении спектров, послужил в XX-м веке основным фундаментом, на котором развилась теория строения атома.

Положение спектральной линии в спектре характеризуется *длиной волны* λ или *частотой* $\nu = c/\lambda$. Частота более удобна для выражения спектральных закономерностей. Но для ее вычисления надо знать скорость света c, а она была измерена с недостаточной точностью, во всяком случае до самого последнего времени. Длина же волны λ измеряется спектральным аппаратом с высокой точностью — до седьмого десятичного знака и выше. Поэтому вместо ν спектроскописты употребляют так называемое *спектроскопическое волновое число* $\overline{\nu}$. Это есть число волн, укладывающихся в вакууме на 1 см длины:

$$\overline{\nu} = 1/\lambda \ ^{1)}. \tag{11.1}$$

В спектроскопии частоту и спектроскопическое волновое число принято обозначать одной и той же буквой ν. Во избежание недоразумений

---

[1]) В настоящее время приведенные соображения утратили свое значение. Методами нелинейной оптики удалось на опыте измерить частоту световых колебаний и притом с точностью, превышающей измерение длины волны в спектроскопии. Скорость же света в вакууме связана с частотой ν и длиной волны λ соотношением $c = \lambda\nu$. Поэтому в октябре 1983 г. Генеральная конференция по мерам и весам приняла новое определение метра. По этому определению скорость света в вакууме принимается равной

$$c = 2{,}99792458 \cdot 10^8 \text{ м/с} \quad \textit{точно}.$$

Метр же определяется через расстояние, проходимое светом в вакууме в течение одной секунды. Поэтому в принципе было бы безразлично, пользоваться ли в спектроскопии величиной $1/\lambda$ или величиной c/λ.

мы этого делать не будем. Частоту будем обозначать через ν, а спектроскопическое волновое число — через $\overline{\nu}$, т. е. в последнем случае ставить над ν черточку. Впрочем, вместо $\overline{\nu}$ предпочтительнее пользоваться обозначением $1/\lambda$.

2. Основным законом спектроскопии, установленным эмпирически в 1908 г., является *комбинационный принцип Ритца.* Он состоит в том, что все многообразие спектральных линий рассматриваемого атома может быть получено путем попарных комбинаций гораздо меньшего числа величин, называемых *спектральными термами* или сокращенно просто *термами.* Частота (волновое число) каждой спектральной линии выражается разностью двух термов:

$$\overline{\nu} = \frac{1}{\lambda} = T_{n_1} - T_{n_2}. \tag{11.2}$$

Термы принято считать существенно положительными и нумеровать их так, чтобы с возрастанием номера терма его величина уменьшалась. В приведенной формуле, например, должно быть $n_1 < n_2$, $T_{n_1} > T_{n_2}$. Если фиксировать n_1 и придавать n_2 всевозможные возрастающие значения, начиная с $n_2 = n_1 + 1$, то получится система линий, называемая *спектральной серией.* Совокупность спектральных серий составляет *спектр* рассматриваемого элемента (атома).

Рассмотрим две спектральные линии одной и той же серии:

$$\overline{\nu}_{12} = T_{n_1} - T_{n_2}, \quad \overline{\nu}_{13} = T_{n_1} - T_{n_3},$$

предполагая, что $\overline{\nu}_{12} > \overline{\nu}_{13}$ (а следовательно, $n_2 > n_3$). Вычитая из первого равенства второе, получим

$$\overline{\nu}_{12} - \overline{\nu}_{13} = T_{n_3} - T_{n_2}.$$

Но это есть волновое число какой-то спектральной линии того же элемента, принадлежащей к серии с начальным термом T_{n_3}. Таким образом, из комбинационного принципа следует, что *разность частот (волновых чисел) двух спектральных линий одной и той же серии атома дает частоту (волновое число) спектральной линии какой-то другой серии того же атома.* Впрочем, такой линии может не оказаться в спектре, так как на комбинации термов друг с другом накладываются некоторые ограничения, называемые *правилами отбора.*

3. Аналитические выражения для термов подавляющего большинства элементов неизвестны. В лучшем случае они представляются приближенными эмпирическими или полуэмпирическими формулами. Для атомов щелочных металлов такие формулы будут приведены в § 34. Исключение составляет простейший атом — атом водорода, состоящий из одного протона и одного электрона. Для атома водорода терм с высокой степенью точности имеет вид

$$T_n = \frac{R_{\text{H}}}{n^2} \quad (n = 1, 2, 3, \ldots), \tag{11.3}$$

где R_{H} — постоянная, называемая *постоянной Ридберга* (1854–1919) для водорода. Ее числовое значение равно

$$R_{\text{H}} = 109678{,}76(1)\ \text{см}^{-1}. \tag{11.4}$$

Такое же выражение, но с другим числовым значением постоянной Ридберга справедливо для всех изотопов водорода и всех одноэлектронных ионов (см. § 27).

Из выражения (11.3) путем комбинаций получаются следующие спектральные серии:

Серия Лаймана:

$$\overline{\nu} = \frac{1}{\lambda} = R_{\text{H}}\left(1 - \frac{1}{n^2}\right), \quad n = 2, 3, 4, \ldots \tag{11.5}$$

Эта серия была открыта Лайманом (1874–1954) в 1916 г. в ультрафиолетовой области спектра.

Серия Бальмера:

$$\overline{\nu} = \frac{1}{\lambda} = R_{\text{H}}\left(\frac{1}{2^2} - \frac{1}{n^2}\right), \quad n = 3, 4, 5, \ldots \tag{11.6}$$

Четыре первые линии этой серии лежат в видимой области спектра и обозначаются через H_α, H_β, H_γ, H_δ, остальные — в области ультрафиолета. На этих четырех линиях Бальмером (1825–1898) в 1885 г. и была выявлена закономерность, выражаемая формулой (11.6). С этого началось систематическое исследование спектральных серий. Мы приводим схематическое изображение серии Бальмера (рис. 20). Таблица 2, в которой приведены вычисленные и наблюденные значения длин

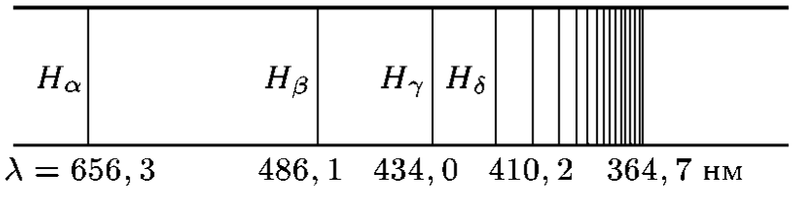


Рис. 20

волн для первых девяти линий серии Бальмера, показывает, с какой высокой точностью формула (11.6) согласуется с опытом. Длины волн отнесены к воздуху, где они были экспериментально измерены. Поэтому вычисленные по формуле (11.6) длины волн были пересчитаны для воздуха.

Серия Пашена:

$$\overline{\nu} = \frac{1}{\lambda} = R_{\text{H}}\left(\frac{1}{3^2} - \frac{1}{n^2}\right), \quad n = 4, 5, 6, \ldots \tag{11.7}$$

Эту серию предсказал Ритц в 1908 г. на основе комбинационного принципа. Все линии этой серии получаются путем комбинаций из линий серии Бальмера. Рассматриваемая серия в том же году была обнаружена Пашеном (1865–1947) в инфракрасной области спектра.

Таблица 2

Длины волн бальмеровской серии водорода

n	Обозначение	λ, нм		n	Обозначение	λ, нм	
		вычислено	наблюдено			вычислено	наблюдено
3	H_α	656,279	656,285	8	H_ζ	388,905	388,906
4	H_β	486,133	486,132	9	H_η	383,539	383,540
5	H_γ	434,047	434,046	10	H_θ	379,790	379,791
6	H_δ	410,174	410,173	11	H_ι	377,067	377,065
7	H_ε	397,008	397,007				

Серия Брэккета:

$$\overline{\nu} = \frac{1}{\lambda} = R_{\text{H}} \left(\frac{1}{4^2} - \frac{1}{n^2} \right), \quad n = 5, 6, 7, \ldots \tag{11.8}$$

Серия Пфунда:

$$\overline{\nu} = \frac{1}{\lambda} = R_{\text{H}} \left(\frac{1}{5^2} - \frac{1}{n^2} \right), \quad n = 6, 7, 8, \ldots \tag{11.9}$$

Эти серии лежат в далекой инфракрасной области спектра. Они были открыты в 1922 г. и 1924 г. соответственно. Конечно, серия Брэккета получается путем комбинаций из линий серии Пашена, а серия Пфунда (1879–1949) — из линий серии Брэккета.

Максимальная длина волны для серии Лаймана получается при $n = 2$. Она равна $\lambda = 4/3R_{\text{H}} = 121{,}56713$ нм. Соответствующая линия называется *резонансной линией водорода*. Максимальная частота (волновое число) получится по формулам (11.5)–(11.9) при $n = \infty$. Эта частота называется *границей серии*. Для серии Бальмера, например, граница серии равна

$$\overline{\nu}_\infty = \frac{R_{\text{H}}}{4} = 27\,419{,}69 \text{ см}^{-1} \quad \text{или} \quad \lambda_\infty = \frac{4}{R_{\text{H}}} = 364{,}70142 \text{ нм}.$$

При приближении к границе серии спектральные линии сгущаются — разность длин волн между ними асимптотически стремится к нулю; стремятся к нулю и интенсивности линий. За границей серии спектр не обрывается, а становится сплошным. Эта закономерность проявляется в спектральных сериях не только водорода, но и других элементов. Здесь также существуют границы серий, за которыми следует сплошной спектр.

§ 12. Постулаты Бора

1. Законы классической физики, по существу, приспособлены для описания *непрерывных процессов*. Между тем резкость спектральных линий, испускаемых атомами химических элементов, указывает на то, что процессам внутри атомов свойственна известная *прерывность* или

дискретность. Наряду с непрерывностью дискретность должна быть отражена в основных физических законах. Это ясно понял Нильс Бор (1885–1962), сформулировавший в 1913 г. два постулата.

1) *Атом (и всякая атомная система) может находиться не во всех состояниях, допускаемых классической механикой, а только в некоторых избранных (квантовых) состояниях, характеризующихся определенными прерывными, дискретными значениями энергии* $\mathscr{E}_1$, $\mathscr{E}_2$, $\mathscr{E}_3, \ldots$ *В этих состояниях, вопреки классической электродинамике, атом не излучает. Поэтому они называются стационарными состояниями.*

Квантовая механика, пришедшая на смену теории Бора, автоматически приводит к стационарным состояниям с дискретными уровнями энергии. Правда, она допускает и стационарные состояния с непрерывными уровнями энергии. Однако в последнем случае электроны и ядра не образуют связанную систему. По крайней мере некоторые из них движутся инфинитно. В атомах же и молекулах составляющие их частицы связаны, т. е. движутся финитно. В этих случаях энергетические уровни дискретны, как того и требует первый постулат Бора (см. § 13, п. 3, а также § 22). Однако, как уже было упомянуто выше, постулат Бора в квантовой механике не является независимым утверждением. Он является следствием основных принципов квантовой механики.

Сам Бор первоначально принял, что движение электронов в стационарных состояниях атомов характеризуется теми же понятиями и величинами, что и движение частиц в классической механике. Иными словами, оно происходит по определенным траекториям и в каждый момент времени характеризуется определенными значениями координат и скоростей электронов. Такая точка зрения несовместима с основными принципами квантовой механики. Квантовая механика отвергает представление о траекториях, по которым якобы движутся электроны и другие микрочастицы (см. § 20). Поэтому мы, принимая первый постулат Бора, будем характеризовать стационарные состояния только значениями энергии, но не будем вводить никаких наглядных представлений о движении электронов в атомах.

2) *При переходе из стационарного состояния с большей энергией* $\mathscr{E}_{n_2}$ *в стационарное состояние с меньшей энергией* $\mathscr{E}_{n_1}$, *энергия атома изменяется на* $\mathscr{E}_{n_2} - \mathscr{E}_{n_1}$. *Если такое изменение происходит из-за излучения, то при этом испускается один фотон с энергией*

$$h\nu \equiv \hbar\omega = \mathscr{E}_{n_2} - \mathscr{E}_{n_1}. \tag{12.1}$$

Такое же соотношение справедливо и для случая поглощения, когда падающий фотон переводит атом с низшего энергетического уровня на более высокий, а сам исчезает. Соотношение (12.1) называется *правилом частот Бора.*

2. Таким образом, атомная система переходит из одного стационарного состояния в другое *скачками.* Такие скачки называются *квантовыми.* Что происходит с системой во время скачка, т. е. между двумя последовательными стационарными состояниями, — на этот вопрос теория Бора не давала ответа. Некоторые ведущие физики считали

подобные вопросы принципиально не имеющими смысла, так как (по их мнению) они не допускают экспериментальной проверки. Они выдвигали в качестве теоретико-познавательного принципа требование, чтобы наука не пыталась описывать явления природы способами, не допускающими экспериментальной проверки. Наука должна оперировать только тем, что принципиально доступно наблюдению. Но в такой общей формулировке это требование совершенно бессодержательно, из него нельзя получить ничего нового. Оно получило бы содержание, если бы было указано, что принципиально доступно наблюдению, а что принципиально недоступно. Но никакого общего критерия для решения этого вопроса заранее указать нельзя. На этот вопрос может ответить *только теория*, а правильность последней контролируется опытом. Различные теории отличаются одна от другой, между прочим, тем, как они решают поставленный вопрос. Отказ теории Бора от ответа на вопрос, что происходит во время квантового скачка, указывает на ее недостаток, на ее неполноту. Однако если бы теория отказалась от рассмотрения конкретных явлений до решения этого вопроса, то она вне всяких сомнений оказалась бы бесплодной. Здесь дело обстоит так же, как с законом всемирного тяготения. Ньютон потому достиг успеха, что он сосредоточил внимание на выводе конкретных следствий из указанного закона, совершенно оставив в стороне вопрос о физической природе тяготения.

3. Возможны и *многофотонные процессы*, когда переход атома с одного уровня на другой сопровождается испусканием или поглощением не одного, а сразу нескольких фотонов. Но такие процессы *нелинейны*. Они могут происходить только в сильных полях, например в таких, которые возникают в гигантских импульсах лазерного излучения. Их мы рассматривать не будем.

Наконец, возможны и *безызлучательные переходы*, когда энергия выделяется или поглощается, например, не в виде света, а в виде тепла. Атом может совершить переход с одного уровня на другой не обязательно в результате излучения или поглощения фотона, а, скажем, в результате столкновения с другой частицей.

4. Правило частот Бора (12.1) объясняет комбинационный принцип Ритца. Из него для спектроскопического волнового числа $\overline{\nu} = \nu/c$ получается

$$\overline{\nu} = \frac{1}{\lambda} = \frac{\mathscr{E}_{n_2}}{ch} - \frac{\mathscr{E}_{n_1}}{ch}. \tag{12.2}$$

Сравнивая это соотношение с (11.2), находим

$$T_n = -\frac{\mathscr{E}_n}{ch}. \tag{12.3}$$

Тем самым раскрывается физический смысл спектральных термов. Они определяются энергетическими уровнями атома.

Становится понятным и другой экспериментальный результат. Если в спектре излучения атомарного водорода наблюдается серия Лаймана, то наблюдаются и все прочие спектральные серии: Бальмера, Пашена и пр. Напротив, в спектре поглощения несветящегося атомарного

водорода наблюдается только серия Лаймана, а все прочие серии не наблюдаются. Дело в том, что для появления в спектре излучения серии Лаймана необходима энергия, достаточная для удаления электрона с самого низкого энергетического уровня. При этом могут возбуждаться и вышележащие энергетические уровни, т. е. удаляться электроны с них, так как для этого требуются меньшие энергии. В результате станут возможными переходы между всевозможными энергетическими уровнями, т. е. в спектре излучения появятся всевозможные спектральные серии. Иначе происходит со спектром поглощения. Если водород не излучает, то его атомы находятся в нормальном состоянии, т. е. на самом низком энергетическом уровне. На других уровнях атомов нет. Поэтому при поглощении света возможны только переходы с самого низкого энергетического уровня на более высокие. При таких переходах в спектре поглощения появится только серия Лаймана. Чтобы при поглощении возникли и другие серии, необходимо, чтобы уже до освещения были возбуждены и другие энергетические уровни. Но тогда появится и спектр испускания рассматриваемого газа. Линии поглощения таких серий могут наблюдаться только на фоне спектра испускания газа. Поэтому необходимо, чтобы последние были менее яркими, чем соответствующие линии освещающего излучения. Аналогично ведут себя и спектры других атомов.

5. Совокупность значений энергии стационарных состояний атома $\mathscr{E}_1, \mathscr{E}_2, \mathscr{E}_3, \ldots$ образует *энергетический спектр атома*. По классической теории излучаемые частоты атома (в простейшем случае одномерного движения) также можно представить в виде ряда $\nu_1, \nu_2, \nu_3, \ldots$, перенумеровав их одним индексом. В этом некоторые авторы видят какое-то противоречие между классической теорией и опытом. Действительно, говорят они, опыт приводит к комбинационному принципу Ритца. Согласно этому принципу и теории Бора каждая частота получается путем комбинации двух термов и должна нумероваться *двумя индексами*, например $\nu_{12}, \nu_{13}, \ldots$ Первый индекс указывает номер начального, второй — конечного термов. Поэтому частоты должны образовывать не одномерный ряд, а прямоугольную таблицу, или *матрицу*. Такое возражение является недоразумением. Мы указываем на него только потому, что оно действительно встречается и может быть воспринято читателем как действительное, а не фиктивное возражение. На самом деле, как хорошо известно из математики, элементы прямоугольной матрицы образуют либо конечное, либо счетное множество. Их всегда можно перенумеровать с помощью *одного индекса*, даже в том случае, когда число элементов в каждой строке и каждом столбце матрицы бесконечно велико.

§ 13. Спектр водорода

1. Определение значений энергии атома $\mathscr{E}_1, \mathscr{E}_2, \mathscr{E}_3, \ldots$ в стационарных состояниях называется *квантованием*, точнее — *квантованием энергии* атома. Бор предложил правило квантования для водородного атома, приводящее к правильным результатам. Проблема квантования

в общем виде была сформулирована в квантовой механике, и притом не только для водородного атома, но и для *любых* атомных систем (см. § 22). Она довольно сложна. Правило квантования Бора представляет только исторический интерес. Тем не менее полезно привести простое решение задачи о квантовании для атома водорода или водородоподобного атома, близко примыкающее к идеям Бора. В основе такого решения лежит аналогия с классической механикой и эмпирически установленное выражение для спектральных термов атома водорода.

Примем, что спектральные термы и соответствующие им уровни энергии атома водорода имеют *бальмеровский вид*:

$$T_n = \frac{Z^2 R}{n^2}, \tag{13.1}$$

$$\mathscr{E}_n = -chT_n = -ch\frac{Z^2 R}{n^2}, \tag{13.2}$$

где R — *постоянная Ридберга*, а зарядовое число Z ядра введено ради удобства. Целое число n называется *главным квантовым числом*. С возрастанием n соседние уровни энергии атома сближаются, и при $n \to \infty$ расстояние между ними стремится к нулю. Дискретность энергетического спектра становится все менее и менее заметной. Поэтому можно ожидать, что в таком предельном случае квантовая система будет вести себя, как классическая. Это положение было выдвинуто Бором и названо им *принципом соответствия*.

Принцип соответствия позволяет выразить постоянную Ридберга R через фундаментальные постоянные, характеризующие атом. Для общности будем рассматривать *водородоподобный атом*. Так называется ион с зарядом ядра $+Ze$, вокруг которого вращается один электрон. При $Z = 1$ он переходит в обычный нейтральный атом водорода H, при $Z = 2$ — в однократно ионизованный атом гелия He^+, при $Z = 3$ — в дважды ионизованный атом лития Li^{++} и т. д. Разумеется, наше рассуждение будет предполагать справедливость выражения (13.2) и для водородоподобного атома. Одинакова ли постоянная Ридберга для различных водородоподобных атомов — это будет выяснено в дальнейшем.

2. Для простоты Бор принял, что электрон вращается вокруг ядра по окружности. Позднее Зоммерфельд (1868–1951) обобщил рассуждения Бора на случай эллиптических орбит. Однако с появлением квантовой механики это обобщение потеряло значение, и мы его рассматривать не будем. Ограничимся более простым случаем круговых орбит. По классическим представлениям частота излучаемого света равна частоте обращения электрона по круговой орбите. Для низких частот это безусловно верно, как показывает сравнение классической теории с опытом в области радиодиапазона. К таким частотам и относятся приводимые ниже вычисления. Здесь частоты, вычисленные по квантовой и классической теориям, должны совпадать, как этого требует принцип соответствия. Ядро мы будем считать бесконечно тяжелым, а потому неподвижным. При вращении по окружности радиуса r с циклической

частотой ω

$$m\omega^2 r = \frac{Ze^2}{r^2}, \tag{13.3}$$

откуда $\omega = Ze^2/(Lr)$, где $L = mr^2\omega$ — момент количества движения электрона. Полная энергия электрона слагается из кинетической и потенциальной энергий и равна

$$\mathscr{E} = \frac{1}{2}\, mr^2\omega^2 - \frac{Ze^2}{r} = -\frac{Ze^2}{2r}.$$

Следовательно, по классической теории должно быть

$$\omega = -\frac{2\mathscr{E}}{L}. \tag{13.4}$$

С другой стороны, уровни энергии водородоподобного атома должны иметь бальмеровский вид (13.2). Отсюда следует, что при переходах атома с одного уровня на другой величина $\mathscr{E}_n n^2$ должна сохраняться: $\mathscr{E}_n n^2 = \text{const}$. Поэтому при больших квантовых числах n и малых их изменениях Δn должно выполняться соотношение

$$\frac{\Delta\mathscr{E}}{\mathscr{E}_n} + 2\frac{\Delta n}{n} = 0.$$

Отсюда с учетом правила частот Бора $\Delta\mathscr{E} = \hbar\omega$ получается

$$\omega = -\frac{2\mathscr{E}}{\hbar n}\,\Delta n, \tag{13.5}$$

причем у $\mathscr{E}$ мы опустили индекс n и считаем $\Delta n > 0$, чтобы не вводить отрицательных частот. Наименьшая частота соответствует переходу $\Delta n = 1$. Это — *основная частота*. Значениям $\Delta n = 2, 3, \ldots$ соответствуют ее *гармоники*, или *обертоны*. По принципу соответствия основная частота в формуле (13.5) должна совпадать с классической частотой (13.4). Это возможно только тогда, когда

$$L = n\hbar. \tag{13.6}$$

Значит, по теории Бора момент количества движения, по крайней мере при больших квантовых числах n, *квантуется*, т. е. может принимать только избранные значения (13.6).

Из формулы (13.3) теперь получаем

$$(mr^2\omega)^2 = Ze^2 rm = (n\hbar)^2.$$

Отсюда

$$r = \frac{n^2\hbar^2}{Ze^2 m}, \tag{13.7}$$

а следовательно,

$$\mathscr{E}_n = -\frac{Ze^2}{2r} = -\frac{(Ze^2)^2 m}{2\hbar^2 n^2}. \tag{13.8}$$

Сравнением этой формулы с (13.2) находим

$$R_\infty = \frac{me^4}{4\pi c\hbar^3} = \frac{2\pi^2 me^4}{ch^3} = 109\,737{,}309 \pm 0{,}12 \text{ см}^{-1}. \qquad (13.9)$$

Мы снабдили R индексом ∞, чтобы указать, что при получении формулы (13.9) масса ядра M считалась бесконечной, а само ядро неподвижным. В этом приближении постоянная Ридберга одинакова для всех водородоподобных атомов.

Теоретическое значение постоянной Ридберга (13.9), хотя и очень близко к экспериментальному значению для атомов водорода $R_\text{H} = 109\,677{,}576$ см$^{-1}$, но при спектроскопической точности измерений их различие все же очень велико. Оно связано с тем, что при выводе формулы (13.9) не учитывалась конечность массы ядра M. Чтобы учесть это, надо массу электрона m заменить на приведенную массу $Mm/(M+m)$. Тогда получится

$$R = \frac{R_\infty}{1 + m/M}. \qquad (13.10)$$

В этом приближении постоянная Ридберга зависит от массы ядра, а потому ее значения для различных водородоподобных атомов отличаются друг от друга, хотя и очень мало. Для атома водорода формула (13.10) дает $R = 109\,677{,}6$ см$^{-1}$, что хорошо согласуется с экспериментом.

Формула (13.10) может служить для вычисления постоянной Ридберга R_∞ для бесконечно тяжелого ядра. Для этого достаточно воспользоваться спектроскопическим значением R, например, для водорода, а также значением m/M из масс-спектроскопических измерений.

3. Формула (13.8) получена для больших значений квантовых чисел n. Но она остается справедливой *при любых значениях* n, так как при ее выводе был постулирован бальмеровский вид термов (13.1), в котором на значения n не наложено никаких ограничений. Наша задача состояла лишь в том, чтобы на основе принципа соответствия получить теоретические формулы (13.9) и (13.10) для постоянных Ридберга R_∞ и R. Но эти постоянные, очевидно, не зависят от n, а потому их значения и можно было получить, проводя вычисления при больших n.

В спектроскопии спектральные термы и уровни энергии принято изображать горизонтальными линиями, а переходы между ними — стрелками. Стрелкам, направленным от высших уровней энергии к низшим, соответствуют линии излучения; стрелкам, проведенным в обратных направлениях, — линии поглощения. В качестве примера на рис. 21 таким путем изображен спектр водорода. Уровни энергии здесь нумеруются квантовым числом n. За нуль принята энергия уровня с $n = \infty$. Этот уровень изображен верхней горизонтальной штриховой прямой. Все энергетические уровни, расположенные ниже, дискретны. Им соответствуют отрицательные значения полной энергии атома. Выше штриховой линии энергия не квантуется, т. е. энергетический спектр непрерывен. Но при $\mathscr{E} < 0$ движение электрона финитно, а при $\mathscr{E} > 0$ — инфинитно. Это непосредственно следует из соответствующей теоремы классической механики (см. т. I, § 25, 57), поскольку при больших n

ее можно применять. Соответствующая теорема может быть доказана и в последовательной квантовой механике (см. § 22), т. е. совершенно строго.

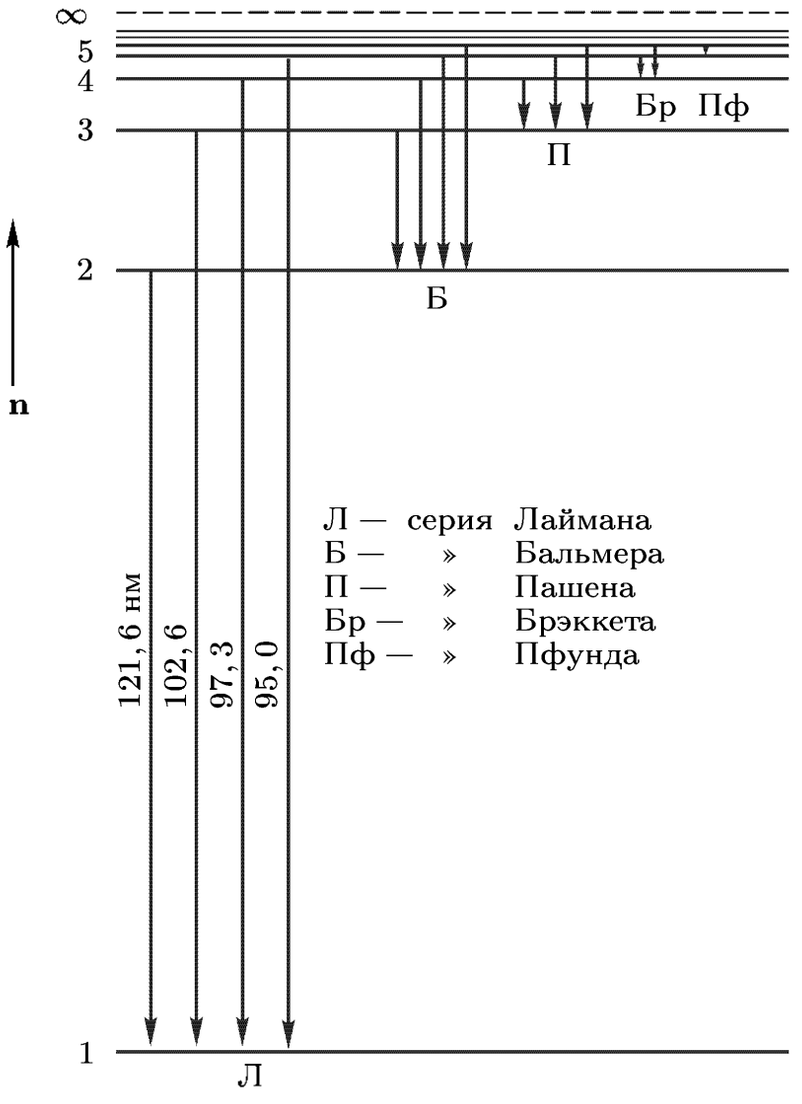


Рис. 21

Таким образом, *ядро и электрон образуют связанную систему — атом — только в случае дискретного энергетического спектра.* При непрерывном энергетическом спектре электрон может как угодно далеко удаляться от ядра. В этом случае пару частиц — ядро и электрон — можно лишь условно называть атомом. Если понимать под атомом только связанные состояния, то можно сказать, что уровни энергии атома всегда дискретны, как это и постулировал Бор.

Наличие несвязанных электронов делает, однако, возможными квантовые переходы между состояниями непрерывного энергетического спектра, а также между такими состояниями и состояниями дискретного спектра энергии. Это проявляется в виде сплошного спектра испускания или поглощения, накладывающегося на линейчатый спектр атома. Вот почему, в частности, спектр атома не обрывается на границе серии, а продолжается за нее в сторону более коротких волн, где он становится сплошным. Квантовые переходы из состояний непрерывного энергетического спектра, т. е. из состояний, в которых

атом ионизован, в состояния дискретного спектра сопровождаются рекомбинацией электронов с соответствующими положительными ионами. Излучение, возникающее при таких переходах, называется *рекомбинационным.*

Переход атома из нормального состояния на более высокий энергетический уровень дискретного спектра есть *возбуждение атома.* Переход же атома с одного из уровней дискретного спектра в область сплошного спектра превращает атом в несвязанную систему. Это есть процесс *ионизации атома.* Если вначале атом находился в нормальном состоянии, то, очевидно, минимальная энергия ионизации атома определяется выражением $\mathscr{E}_{\text{ион}} = \mathscr{E}_\infty - \mathscr{E}_1 = -\mathscr{E}_1$, т. е. для водородоподобного атома

$$\mathscr{E}_{\text{ион}} = (Ze^2)^2 \frac{m}{2\hbar^2} = 2\pi c\hbar Z^2 R_\infty = chZ^2 R_\infty, \tag{13.11}$$

или

$$\mathscr{E}_{\text{ион}} = 21{,}8 \cdot 10^{-12} Z^2 \text{ эрг} = 13{,}6 Z^2 \text{ эВ}. \tag{13.12}$$

4. До 1913 г., т. е. до появления теории Бора, водороду приписывались еще две серии:

$$\frac{1}{\lambda} = R_{\text{H}} \left(\frac{1}{1{,}5^2} - \frac{1}{\overline{n}^2} \right), \quad \overline{n} = 2, 3, 4, \ldots, \tag{13.13}$$

$$\frac{1}{\lambda} = R_{\text{H}} \left(\frac{1}{2^2} - \frac{1}{(\overline{n} + 1/2)^2} \right), \quad \overline{n} = 2, 3, 4, \ldots \tag{13.14}$$

Серия (13.13) была открыта Фаулером (1889–1944) в смеси H и He, а серия (13.14) наблюдалась Пикерингом в спектре планетарной туманности (ζ Кормы). Однако, согласно теории Бора, линии этих серий принадлежат не водороду H, а однократно ионизованному гелию He^+. Они содержатся в спектральных сериях последнего:

$$\frac{1}{\lambda} = 4R_{\text{He}} \left(\frac{1}{3^2} - \frac{1}{n^2} \right), \quad n = 4, 5, 6, \ldots, \tag{13.13а}$$

$$\frac{1}{\lambda} = 4R_{\text{He}} \left(\frac{1}{4^2} - \frac{1}{n^2} \right), \quad n = 5, 6, 7, \ldots, \tag{13.14а}$$

если пренебречь различием постоянных Ридберга для водорода и гелия. В действительности эти постоянные немного отличаются одна от другой, как видно из формулы

$$\frac{R_{\text{He}}}{R_{\text{H}}} = \frac{1 + m_{\text{e}}/M_{\text{H}}}{1 + m_{\text{e}}/M_{\text{He}}} = 1{,}00041. \tag{13.15}$$

Если, однако, не учитывать различие между R_{He} и R_{H}, то в этом приближении линии серии (13.13) совпадут с линиями серии (13.13а) при четных n, а линии серии (13.14) — с линиями серии (13.14а) при нечетных n. И действительно, серии (13.13а) и (13.14а) получались экспериментально в чистом гелии. Конечно нет никаких оснований

делить единые серии (13.13а) и (13.14а) на две половины, выделяя из них серии Фаулера (13.13) и Пикеринга (13.14). В действительности R_{He} немного больше R_{H}. Благодаря этому спектральные линии однократно ионизованного гелия He^+ оказываются немного смещенными в коротковолновую сторону спектра относительно соответствующих бальмеровских линий водорода. Такой эффект называется *изотопическим смещением спектральных линий.*

Разумеется, этот эффект существует и для других химических элементов и их ионов, хотя для многоэлектронных атомов его и нельзя трактовать столь же просто, как в случае атомов водородоподобных. Действительно, различные изотопы одного и того же химического элемента или иона отличаются только массами ядер. Но они имеют одинаковые заряды ядер, а потому и одинаковые электронные оболочки. Процессы же излучения света происходят как раз *в электронных оболочках.*

Говорить об изотопическом смещении линий гелия относительно линий водорода на первый взгляд кажется нелогичным, поскольку гелий не является изотопом водорода. Однако в эффекте смещения спектральных линий речь идет не об атомах гелия, а о его *однократно ионизованных* атомах. Последние же в рассматриваемом вопросе ведут себя как изотопы водорода. Так же обстоит дело и в случае других элементов.

Заметим, что изотопическое смещение спектральных линий обусловлено не только различием масс изотопов, но и *различием размеров их атомных ядер.* Забегая вперед, это нетрудно понять, если проблему квантования рассматривать на основе уравнения Шредингера (см. § 22). В самом деле, для различных изотопов кулоновские и ядерные силовые поля внутри ядра несколько отличаются друг от друга. Различны также размеры областей, занимаемых этими полями. Это ведет к небольшому различию волновых функций и соответствующих им собственных значений энергии. Влияние размеров ядра на изотопическое смещение спектральных линий особенно существенно для тяжелых ядер. Здесь изотопическое смещение, вызванное различием размеров ядер, того же порядка, что и изотопическое смещение, вызванное различием их масс.

5. Известно, что наряду с обычным водородом существуют два изотопа его, называемые *дейтерием* и *тритием.* Ядро атома обычного водорода состоит всего из одного протона. Такой водород называется *протием.* Ядро атома дейтерия называется *дейтроном*, а ядро атома трития — *тритоном.* Дейтрон состоит из протона и нейтрона, тритон — из протона и двух нейтронов. Эти ядра тяжелее ядра протия соответственно в 2 и 3 раза. Атом дейтерия обозначается через D или $^2\mathrm{H}$, а атом трития — через T или $^3\mathrm{H}$. Дейтерий — стабильный изотоп, а тритий — радиоактивный. Тритий образуется искусственно при бомбардировке ускоренными дейтронами литиевой или бериллиевой мишеней. В ничтожных количествах он образуется также в атмосфере под действием космических лучей. В природных соединениях воды на каждые 6800 атомов протия приходится в среднем один атом дейтерия. Благодаря

относительно большой разности масс D и H различие их физических свойств выражено значительно сильнее, чем для других изотопов. Так, плотность тяжелой воды D_2O примерно на 10% больше плотности обыкновенной воды H_2O, а температура плавления (при нормальном атмосферном давлении) составляет 3,8 °C. Но изотопическое смещение спектральных линий для D и H невелико. Действительно, например,

$$\frac{R_D}{R_H} = \frac{1 + m_e/M_H}{1 + m_e/M_D} = 1{,}000272.$$

Отсюда следует, что для изотопического сдвига частот или длин волн дейтерия относительно водорода получается

$$\frac{\Delta\nu}{\nu} = -\frac{\Delta\lambda}{\lambda} = \frac{R_D - R_H}{R_H} = m_e\left(\frac{1}{M_H} - \frac{1}{M_D}\right) = 0{,}000272.$$

Так, бальмеровская линия D_α дейтерия смещена относительно соответствующей линии H_α водорода в коротковолновую сторону спектра всего на $|\Delta\lambda| = 0{,}179$ нм.

6. Расчеты, приведенные выше, обоснованы только при больших квантовых числах n. Однако Бор первоначально принял, что электроны в стационарных состояниях движутся по определенным орбитам, как частицы в классической механике. С этой точки зрения и при малых n должна быть справедлива не только формула (13.8) для энергии стационарного состояния, но и формула (13.7), определяющая радиус орбиты электрона. Квантовая механика от орбит отказалась. Однако формула (13.7) сохраняет смысл и в квантовой механике. Она по порядку величины определяет размеры атома в соответствующих стационарных состояниях. Точный смысл этого выражения будет указан в § 27. Сейчас же заметим, что радиус орбиты электрона в нормальном состоянии атома обычного водорода (т. е. при $n = 1$) называется *боровским радиусом* и обозначается через $r_Б$. Согласно (13.7) он равен

$$r_Б = \frac{\hbar^2}{me^2} = 0{,}52917 \cdot 10^{-8} \text{ см}. \tag{13.16}$$

По порядку эта величина совпадает с размерами атомов, к которым ранее приводила кинетическая теория вещества. Напряженность электрического поля ядра на первой боровской орбите атома водорода

$$E = \frac{e}{r_Б^2} = 17{,}2 \cdot 10^6 \text{ СГСЭ} = 51{,}5 \cdot 10^8 \text{ В/см}.$$

Вообще, величина $E \sim 10^8$ В/см является характерным масштабом для напряженностей внутриатомных электрических полей. Во внешних полях с напряженностью такого порядка атомы быстро ионизуются.

Легко видеть, что скорость движения электрона по стационарной круговой орбите определяется выражением

$$v = \frac{Z}{n}\alpha c, \tag{13.17}$$

где α — безразмерная постоянная:

$$\alpha = \frac{e^2}{\hbar c} = \frac{1}{137{,}03604(11)} \approx \frac{1}{137}, \tag{13.18}$$

называемая *постоянной тонкой структуры.* Для движения по первой боровской орбите атома водорода

$$v = \alpha c = \frac{c}{137} = 0{,}00730c. \tag{13.19}$$

Если постоянную α ввести в формулу (13.8), то получится

$$\mathscr{E}_n = -\frac{Z^2}{2n^2}\,\alpha^2 mc^2, \tag{13.20}$$

т. е. $\mathscr{E}$ выразится через собственную энергию электрона mc^2.

ЗАДАЧИ

1. Найти границы спектральной области, в пределах которой расположены линии серии Бальмера атомарного водорода.

Ответ. $\lambda_{\text{Б}\infty} \leqslant \lambda < \lambda_{\text{Бк}}$, где $\lambda_{\text{Б}\infty} = 4/R_{\text{H}} = 364{,}705$ нм, $\lambda_{\text{Бк}} = 36/(5R_{\text{H}}) = 656{,}468$ нм (красная линия).

2. Используя данные, полученные при решении предыдущей задачи, определить длины волн коротковолновых границ водородных серий Лаймана и Пашена.

Ответ. Для серии Лаймана $\lambda_{\text{Л}\infty} = \lambda_{\text{Б}\infty}/4 = 91{,}1762$ нм. Для серии Пашена $\lambda_{\text{Пш}\infty} = (9/4)\lambda_{\text{Б}\infty} = 820{,}586$ нм.

3. Вычислить энергию, которую надо сообщить атому водорода, чтобы его серия Бальмера содержала только одну спектральную линию.

Ответ. $\mathscr{E}_1 \leqslant \mathscr{E} \leqslant \mathscr{E}_2$, где $\mathscr{E}_1 = (8/9)hcR_{\text{H}} = 12{,}09$ эВ, $\mathscr{E}_2 = (15/16)hcR_{\text{H}} = 12{,}75$ эВ.

4. Какие спектральные линии появятся в спектре атомарного водорода при облучении его ультрафиолетовым светом с длиной волны 100 нм?

Ответ. $\lambda_1 = (4/3)R_{\text{H}} = 121{,}5682$ нм, $\lambda_2 = (9/8)R_{\text{H}} = 102{,}57317$ нм, $\lambda_3 = (36/5)R_{\text{H}} = 656{,}46828$ нм.

5. Будет ли атом водорода поглощать излучение с волновым числом $1/\lambda = 3R$?

Ответ. Такое поглощение происходит и сопровождается ионизацией атома.

6. Первоначально неподвижный атом водорода испустил фотон с частотой, соответствующей головной линии серии Лаймана. Определить относительное изменение частоты фотона $\Delta\nu/\nu_0$ из-за отдачи. Какую скорость v приобрел атом после излучения фотона?

Ответ.

$$\frac{\nu_0 - \nu}{\nu_0} = \frac{h\nu_0}{2m_{\text{ат}}c^2} = \frac{3}{16}\,\frac{m_{\text{e}}}{m_{\text{ат}}}\,\alpha^2 = 5{,}44 \cdot 10^{-9},$$

$$v = \frac{3}{8}\,\frac{m_{\text{e}}}{m_{\text{ат}}}\,\alpha^2 c = 326 \text{ см/с},$$

где $\alpha = e^2/\hbar c = 1/137$ — постоянная тонкой структуры.

7. Найти потенциалы ионизации ионов He^+ и Li^{++}.

Ответ. $V_{\text{He}} = 4V_{\text{H}} = 54$ В; $V_{\text{Li}} = 9V_{\text{H}} = 122$ В.

8. Найти длину волны резонансной линии He^+.

Ответ. $\lambda = 30{,}375$ нм.

9. Определить наименьшую энергию, которую надо сообщить в основном состоянии трижды ионизованному атому бериллия, чтобы возбудить полный спектр этого атома.

Ответ. $\mathscr{E}_{\text{Be}} = hcZ^2R_{\text{H}} = 217{,}5747$ эВ.

10. Фотон головной серии Лаймана иона гелия He^+ поглощается водородным атомом в основном состоянии и ионизует его. Определить кинетическую энергию $\mathscr{E}_{\text{кин}}$, которую получит электрон при такой ионизации.

Ответ. $\mathscr{E}_{\text{кин}} = 2\mathscr{E}_{\text{H, ион}} = 27{,}2$ эВ, где $\mathscr{E}_{\text{H, ион}}$ — энергия ионизации атома водорода.

11. Оценить по теории Бора число спектральных линий, которые можно получить в различных сериях спектра испускания атомарного водорода в газоразрядной трубке при давлении $P = 5$ мм рт. ст. и температуре $T = 300$ К.

Решение. Водородные серии Лаймана, Бальмера, Пашена и т. д. излучаются изолированными атомами, т. е. практически тогда, когда диаметр боровской орбиты $2r = 2n^2r_{\text{Б}}$ не превосходит среднего расстояния между атомами $l = (kT/P)^{1/3}$. Пользуясь этим, максимально допустимый номер орбиты n можно оценить по формуле $n \approx \sqrt{l/(2r_{\text{Б}})}$. Максимально возможное число линий, которые можно наблюдать в трубке при давлении и температуре, указанных в условии задачи, будет равно:

в серии Лаймана $n - 1 \approx 12$,

в серии Бальмера $n - 2 \approx 11$,

в серии Пашена $n - 3 \approx 10$.

12. В спектрах некоторых звезд наблюдается $m \sim 40$ линий водородной серии Бальмера. При каком наименьшем числе N штрихов дифракционной решетки можно разрешить эти линии в спектре первого порядка?

Ответ. $N \approx m^3/8 = 8000$ штрихов.

13. Вблизи спектральной линии водорода $\lambda_1 = 486{,}1320$ нм Юри (1893–1981) в 1932 г. обнаружил близкую линию $\lambda_2 = 485{,}9975$ нм. Предполагая, что эта линия обусловлена небольшой примесью к обычному водороду его изотопа, определить относительную атомную массу m_x/m_{H} этого изотопа.

Решение. Из формулы (13.10) нетрудно получить

$$\frac{\lambda_1}{\lambda_2} \approx \frac{1 - m_{\text{e}}/m_{\text{H}}}{1 - m_{\text{e}}/m_x},$$

где m_x — атомная масса неизвестного изотопа. Отсюда

$$\frac{\lambda_1 - \lambda_2}{\lambda} \approx \frac{m_{\text{e}}}{m_x} - \frac{m_{\text{e}}}{m_{\text{H}}},$$

если пренебречь величиной m_{e}/m_x по сравнению с единицей. (В пределах точности расчета под λ можно понимать любое из значений λ_1, λ_2.) Следовательно,

$$\frac{m_{\text{e}}}{m_x} = \frac{m_{\text{e}}}{m_{\text{H}}} - \frac{\lambda_1 - \lambda_2}{\lambda}.$$

Так как $m_{\text{e}}/m_{\text{H}} = 1/1835$, то отсюда получаем $m_{\text{e}}/m_x = 1/3727$. Следовательно, $m_x/m_{\text{H}} \approx 2$. Линия λ_2 принадлежит дейтерию.

14. Определить разности длин волн бальмеровской серии водорода и дейтерия в видимой области спектра.

Ответ. $\lambda_{H_\alpha} - \lambda_{D_\alpha} = 0{,}179$ нм; $\lambda_{H_\beta} - \lambda_{D_\beta} = 0{,}132$ нм; $\lambda_{H_\gamma} - \lambda_{D_\gamma} = 0{,}118$ нм; $\lambda_{H_\delta} - \lambda_{D_\delta} = 0{,}116$ нм.

15. Определить разрешающую способность R (не путать с постоянной Ридберга) спектрального аппарата, необходимую для наблюдения изотопического смещения спектральных линий дейтерия относительно линий водорода. Какова должна быть ширина b основания призмы из тяжелого флинта с дисперсией $dn/d\lambda = 1000$ см$^{-1}$ (в диапазоне красного света) в призменном спектрографе, если его применять для обнаружения изотопического смещения головной (длинноволновой) линии серии Бальмера?

Ответ.

$$R > 2\frac{m_{\mathrm{H}}}{m_{\mathrm{e}}} \approx 3700; \quad b > \frac{R}{dn/d\lambda} \approx 3{,}7 \text{ см}.$$

16. Серия Лаймана наблюдается в смеси атомарных водорода и трития. Определить разрешающую способность R спектрального прибора, достаточную для разрешения изотопической структуры спектральных линий этой серии. Как меняется требуемая разрешающая способность при переходе к другим сериям (Бальмера, Пашена) той же смеси? Можно ли разрешить изотопическую структуру спектральных линий рассматриваемой смеси в видимой области спектра с помощью стеклянной призмы с основанием $b = 1$ см и дисперсией показателя преломления $dn/d\lambda = 1000$ см$^{-1}$? Каково должно быть эффективное число отражений $N_{\text{эф}}$ и порядок m наблюдаемого спектра, чтобы разрешить и исследовать ту же структуру с помощью интерферометра Фабри–Перо?

Ответ. Разрешающая способность должна быть не меньше

$$\frac{3}{2}\frac{m_{\mathrm{p}}}{m_{\mathrm{e}}} \approx 2800.$$

Она одинакова для всех линий спектральных серий смеси. Разрешающая способность призмы $b\,dn/d\lambda = 1000$, т. е. недостаточна для разрешения

$$m < \frac{3}{2}\frac{m_{\mathrm{p}}}{m_{\mathrm{e}}} < N_{\text{эф}}m.$$

17. Зная постоянные Ридберга для водорода и гелия ($R_{\mathrm{He}} = 109\,722{,}267$ см$^{-1}$), а также R_∞, вычислить отношение $m_{\mathrm{He}}/m_{\mathrm{H}}$ массы атома гелия к массе атома водорода.

Ответ.

$$\frac{m_{\mathrm{He}}}{m_{\mathrm{H}}} \approx \frac{R_\infty - R_{\mathrm{H}}}{R_\infty - R_{\mathrm{He}}} = 3{,}971.$$

18. Удельный заряд электрона e/m можно вычислить спектроскопически. Произвести это вычисление, пользуясь спектроскопическими данными

$$R_{\mathrm{H}} = 109\,677{,}576 \text{ см}^{-1}, \quad R_{\mathrm{D}} = 109\,707{,}419 \text{ см}^{-1},$$

молярными массами атомарных водорода и дейтерия

$$\mathscr{M}_{\mathrm{H}} = 1{,}008142 \text{ г/моль}, \quad \mathscr{M}_{\mathrm{D}} = 2{,}014755 \text{ г/моль}$$

и постоянной Фарадея

$$F = N_{\mathrm{A}}e = 2{,}89366 \cdot 10^{14} \text{ СГСЭ/моль} = 9632{,}19 \text{ СГСМ/моль}.$$

Решение. Постоянные Ридберга для водорода и дейтерия:

$$R_{\mathrm{H}} = \frac{R_\infty}{1 + m_{\mathrm{e}}/M_{\mathrm{H}}}, \quad R_{\mathrm{D}} = \frac{R_\infty}{1 + m_{\mathrm{e}}/M_{\mathrm{D}}},$$

где M_{H} и M_{D} — массы ядер водорода и дейтерия. Отсюда получаем

$$\frac{1}{m_{\text{e}}} = \frac{1}{R_{\text{D}} - R_{\text{H}}}\left(\frac{R_{\text{H}}}{M_{\text{H}}} - \frac{R_{\text{D}}}{M_{\text{D}}}\right).$$

Умножая обе части этого соотношения на e и принимая во внимание, что $(M_{\text{H}}+m_{\text{e}})N_{\text{A}} = \mathscr{M}_{\text{H}}$, $(M_{\text{D}}+m_{\text{e}})N_{\text{A}} = \mathscr{M}_{\text{D}}$ (здесь m_{e}, M_{H} и M_{D} — в граммах), $N_{\text{A}}e = F$, найдем

$$\frac{e}{m_{\text{e}}} = \frac{F}{R_{\text{D}} - R_{\text{H}}}\left(\frac{R_{\text{H}}}{\mathscr{M}_{\text{H}} - N_{\text{A}}m_{\text{e}}} - \frac{R_{\text{D}}}{\mathscr{M}_{\text{D}} - N_{\text{A}}m_{\text{e}}}\right).$$

Так как масса электрона мала по сравнению с массой атома, то нет необходимости вычислять величину $N_{\text{A}}m_{\text{e}}$ с высокой точностью. Достаточно принять

$$N_{\text{A}}m_{\text{e}} = 6{,}022 \cdot 10^{23} \cdot 9{,}11 \cdot 10^{-28} = 5{,}49 \cdot 10^{-4} \text{ г/моль};$$

получаем

$$\mathscr{M}_{\text{H}} - N_{\text{A}}m_{\text{e}} = 1{,}007593 \text{ г/моль}; \quad \mathscr{M}_{\text{D}} - N_{\text{A}}m_{\text{e}} = 2{,}014206 \text{ г/моль}.$$

Подставляя эти значения в предыдущую формулу, находим

$$\frac{e}{m_{\text{e}}} = 5{,}2732 \cdot 10^{17} \text{ СГСЭ/г} = 1{,}7590 \cdot 10^{7} \text{ СГСМ/г}.$$

Точность этих вычислений теряется при определении разности $R_{\text{D}} - R_{\text{H}}$. Поэтому сами величины R_{D} и R_{H} надо знать с исключительно высокой точностью. Существенно, что постоянная Ридберга R_∞ в вычисления не входит.

19. Позитроний (химический символ Ps) представляет собой связанную систему из электрона и позитрона, вращающихся вокруг центра масс этой системы. Позитроний образуется при столкновении медленных позитронов с атомами вещества и захвате позитроном атомного электрона. Это — нестабильная система, так как электрон и позитрон очень быстро аннигилируют в два (за время $\sim 10^{-10}$ с) или три (за время $\sim 10^{-7}$ с) гамма-кванта. Найти уровни энергии, энергию ионизации и длину волны резонансной линии для позитрония.

Ответ.

$$\mathscr{E}_{\text{Ps}} = -\frac{hc}{2n^2}R_\infty; \quad \mathscr{E}_{\text{ион}} = \frac{hc}{2}R_\infty = 6{,}80 \text{ эВ}, \quad \lambda_{\text{рез}} = 243 \text{ нм}.$$

20. При торможении в веществе отрицательно заряженных элементарных частиц (мюонов μ^-, π^--мезонов, K^--мезонов и пр.) они могут захватываться кулоновским полем ядра, замещая электроны электронной оболочки. Наибольшее значение имеет захват отрицательных мюонов. Практически может замещаться лишь один электрон. Получающиеся в результате такой замены системы называются *мезоатомами*. Масса мюона $m_\mu = 207m_{\text{e}}$. Вычислить по теории Бора радиус первой круговой орбиты (K-орбиты) мюона в мезоатоме. Рассчитать энергетические уровни мезоатома. Какое излучение будет наблюдаться при переходе на K-орбиту мюона с более высоких орбит? Почему исследование такого излучения применяется для выяснения структуры тяжелых атомных ядер вблизи их поверхностей? Массой мюона по сравнению с массой ядра пренебречь.

Решение. Электронная оболочка атома практически не оказывает влияния на движение мюона в мезоатоме, так как последний находится либо

очень близко от ядра, либо внутри ядра. Поэтому из-за сферической симметрии электронного облака создаваемое им электрическое поле в месте нахождения мюона может считаться равным нулю. Если орбита мюона проходит вне ядра, то для радиуса K-орбиты мюона получается

$$r_1 = \frac{\hbar^2}{4\pi^2 e^2 m_\mu Z} = \frac{0{,}0026}{Z} \cdot 10^{-8} \text{ см},$$

что примерно в 200 раз меньше соответствующего значения r_1 для водородоподобного атома с тем же значением зарядового числа Z. В том же предположении для уровней энергии получаем

$$\mathscr{E}_n = -hcR_\infty Z^2 \frac{m_\mu}{m_e} = -2820 Z^2 \text{ эВ}.$$

Отсюда видно, что излучение, возникающее при переходе мюона на K-орбиту с более высоких орбит, будет расположено в рентгеновской области спектра, а при больших Z — в области γ-лучей. При больших Z K-орбита мюона проходит внутри ядра атома. В этом случае приведенные выше формулы становятся неприменимыми. Результаты сильно зависят от распределения электрического заряда в ядре, с чем и связана возможность использования мезоатомов для изучения структуры атомного ядра.

§ 14. Экспериментальное подтверждение постулатов Бора

1. Результаты, изложенные в предыдущем параграфе, являются убедительным экспериментальным подтверждением постулатов Бора. Постулаты Бора лежат в основе физических представлений, которыми руководствовались при создании и совершенствовании оптических квантовых генераторов (лазеров) (см. т. IV, гл. XI), открывших принципиально новые пути не только в физике, но и в технике и других отраслях науки. То обстоятельство, что эти приборы работают в соответствии с предсказаниями теории, свидетельствует о правильности последней, а следовательно, и о правильности самих постулатов Бора.

В этом параграфе приводятся другие экспериментальные факты, подтверждающие постулаты Бора. Сюда относятся прежде всего опыты Джеймса Франка (1882–1964) и Густава Герца (1887–1975), начатые незадолго до появления теории Бора. Первоначальная цель их состояла в измерении *потенциалов ионизации атомов*. Но эти опыты принесли экспериментальное подтверждение постулатов Бора, т. е. решили более важную задачу. В опытах Франка и Герца через исследуемый газ проходили ускоренные электроны. При столкновении с атомами газа последние могли переходить в *возбужденные состояния*, каждое из которых характеризуется определенным значением энергии. Если энергетические уровни атома дискретны, как утверждает первый постулат Бора, то кинетическая энергия электронов должна быть не меньше некоторой минимальной величины, чтобы они начали возбуждать атомы газа. Такой метод возбуждения атомов особенно пригоден для инертных газов и паров металлов, так как атомы этих веществ

не обладают электронным сродством, т. е. не проявляют склонности к захвату электронов и образованию отрицательных ионов. Франк и Герц в своих первых опытах пользовались парами ртути. Разумеется, ртутный пар должен быть сравнительно плотным, чтобы электроны сталкивались с атомами достаточно часто. Опыты производились при различных плотностях ртутного пара. С этой целью лампа, в которую было введено несколько капель ртути, помещалась в печь, температуру которой можно было менять и поддерживать постоянной во время опыта.

2. Схема опыта Франка и Герца приведена на рис. 22. Ускоряющая разность потенциалов V приложена между накаливаемым катодом K

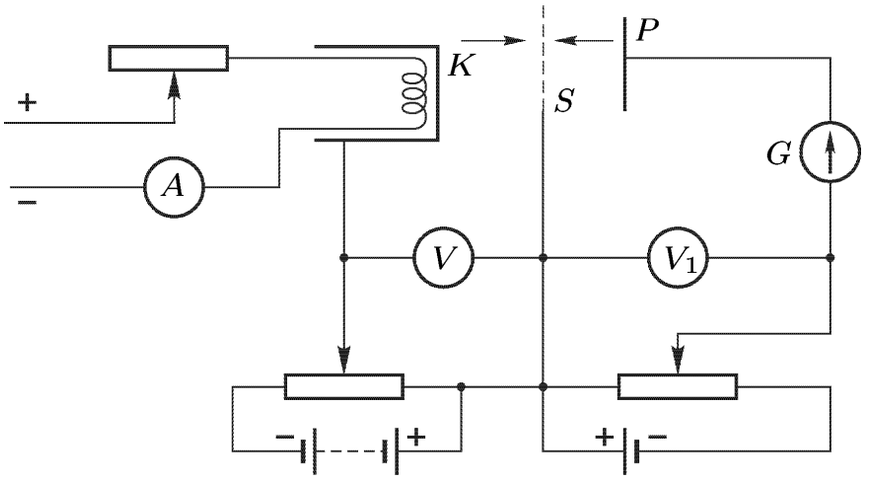


Рис. 22

и сеткой S. Выйдя из K электроны, пройдя через пары ртути, направляются к сетке S. Если их кинетическая энергия еще достаточно велика, то после прохождения через сетку они могут преодолеть небольшой задерживающий потенциал V_1 порядка 0,5 В и попасть на собирающий электрод (коллектор) P. В противном случае они не смогут преодолеть задерживающий потенциал V_1 и на коллектор не попадут. Электронный ток I на коллектор измеряется гальванометром. Амперметр A служит для контроля постоянства тока накала. Стеклянный баллон, в который заключены накаливаемый катод, сетка и коллектор, на рис. 22 не изображен.

Если бы в сосуде поддерживался вакуум, то примерная зависимость термоэлектронного тока I от ускоряющего напряжения V представлялась бы кривой, приведенной на рис. 23. Из-за пространственного заряда эта кривая в самом начальном участке следовала бы закону трех вторых. При больших напряжениях получался бы ток насыщения, не зависящий от V (см. т. III, § 101). Опыт показал, что при наличии в сосуде ртутных паров или других газов кривая $I = I(V)$ имеет качественно совсем другой вид. На ней появляются резко выраженные максимумы и минимумы тока. Для ртути (рис. 24) расстояния между двумя соседними максимумами или минимумами равны 4,9 В. Такие максимумы и минимумы свидетельствуют о дискретном характере энергетических уровней атома, причем для атома ртути ближайший энергетический

уровень лежит выше нормального (наинизшего) уровня примерно на 4,9 эВ.

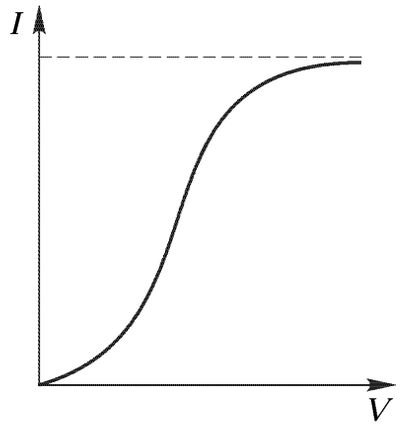


Рис. 23

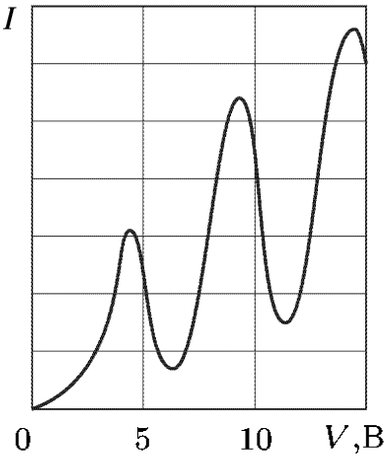


Рис. 24

3. Действительно, столкновения электрона с атомами ртути могут быть *упругими* и *неупругими*. При упругих столкновениях внутреннее состояние атома ртути не меняется, а потому кинетическая энергия электрона не тратится на его возбуждение. Она может переходить только в кинетическую энергию атома как целого. Но с этим процессом можно совсем не считаться, так как масса атома ртути почти в 400 тысяч раз превышает массу электрона. При столкновении с атомом электрон отражается от него, как от неподвижной стенки, меняя только направление своего движения. При этом можно не учитывать теплового движения самого атома, так как средняя энергия такого движения $(3/2)kT$ очень мала по сравнению с энергией, приобретаемой электроном при ускорении в электрическом поле. Пока кинетическая энергия электрона меньше $\mathscr{E}_1 \approx 4{,}9$ эВ, он не может переводить атом ртути в возбужденное состояние — все столкновения будут упругими. Понятно, что среди множества упругих столкновений электрона с атомами ртути имеются не только столкновения, в которых электрон теряет направление первоначального движения, но и столкновения, при которых это направление почти восстанавливается. Если электрон был в состоянии преодолеть задерживающий потенциал до столкновений, то он сможет сделать то же самое и после нескольких упругих столкновений при условии, что направление его движения осталось прежним. Поэтому, пока столкновения упругие, электронный ток I растет с ростом напряжения V.

Положение меняется, когда энергия электрона достигает значения $\mathscr{E}_1 = 4{,}9$ эВ или больше. Тогда электрон сможет затратить свою кинетическую энергию или часть ее *на возбуждение атома ртути*. В этом случае столкновения называются *неупругими*. Если при этом оставшейся кинетической энергии электрона окажется недостаточно, чтобы преодолеть задерживающий потенциал V_1, то электрон на коллектор P не попадет, и с ростом напряжения V электронный ток I начнет убывать. При достижении энергии $\mathscr{E}_1$ ток I не падает скачкообразно до нуля, так как не все скорости электронов одинаковы, хотя бы

потому, что при вылете из нити получается тепловой разброс скоростей. Кроме того, для преодоления задерживающего поля имеет значение не полная скорость электрона, а только ее *продольная составляющая*, т. е. составляющая вдоль поля; поперечная составляющая роли не играет. А эта составляющая меняется при упругих столкновениях — сохраняется только полная скорость. Все это ведет к сглаживанию кривой $I = I(V)$. Этим же объясняется, почему максимумы на кривой тока получаются не точно при энергии $\mathscr{E}_1$, достаточной для возбуждения атома, а при энергии несколько меньшей.

При дальнейшем возрастании ускоряющего напряжения электрон, оставшийся после упругого столкновения в ускоряющем поле, может вторично набрать энергию, достаточную для возбуждения первого энергетического уровня атома $\mathscr{E}_1$. Если он снова претерпит неупругое столкновение и потеряет энергию, то при достаточно высоком потенциале V он может в третий раз ускориться до энергии, необходимой для возбуждения того же первого энергетического уровня атома, и т. д. В результате таких многократных возбуждений уровня $\mathscr{E}_1$ на кривой $I = I(V)$ и появляются максимумы вблизи значений энергии $\mathscr{E}_1$, $2\mathscr{E}_1$, $3\mathscr{E}_1$ и т. д. Наличие *контактной разности потенциалов* между катодом и сеткой искажает показания вольтметра V, смещая всю кривую $I = I(V)$ вправо или влево. Однако контактная разность потенциалов исключается, если величину $\mathscr{E}_1$ определять по расстоянию между соседними максимумами тока $I = I(V)$, которое от нее не зависит.

4. Неупругое столкновение с возбуждением первого возбужденного уровня $\mathscr{E}_1$ может и не произойти. Электрон может продолжать увеличивать свою энергию в электрическом поле, пока она не достигнет значения, достаточного для возбуждения второго энергетического уровня $\mathscr{E}_2$, третьего уровня $\mathscr{E}_3$ и т. д. Ускоряющие напряжения, соответствующие этим уровням энергии, называются *потенциалами возбуждения* или *критическими потенциалами*. Таким образом, в принципе на кривой $I = I(V)$ вблизи критических потенциалов могут появиться новые максимумы. Однако в плотных газах этого фактически не происходит. Так было и в первых опытах Франка и Герца (при температуре, например, 210 °С давление ртутных паров составляет 24 мм рт. ст.). Дело в том, что для накопления нужной энергии электрон без неупругих столкновений должен пройти в ускоряющем электрическом поле расстояние, большое по сравнению с длиной свободного пробега между двумя последовательными неупругими столкновениями. Это условие в случае плотных газов не выполняется. Возбуждение высших энергетических уровней $\mathscr{E}_2$, $\mathscr{E}_3$, ..., возможно при более низком давлении газа и более совершенной аппаратуре (см. п. 6).

5. Однако, незначительно изменив схему рис. 22, можно приспособить ее для измерения *потенциалов ионизации атомов. Потенциал ионизации — это наименьшее ускоряющее напряжение, при котором атомы начинают ионизоваться и образовываются положительные ионы.* Понятно, что для возможности ионизации электроны должны быть ускорены до энергии, не меньшей *энергии ионизации атома.* В случае ртути для этого температура печи должна быть невысокой,

чтобы давление ртутных паров составляло примерно 0,5 мм рт. ст. Схема рис. 22 сохраняется. Только теперь положительный полюс вольтметра V_1 надо соединить с отрицательным, а не с положительным полюсом вольтметра V. Тогда всегда потенциал коллектора будет примерно на 0,5 В ниже потенциала катода, каково бы ни было значение ускоряющего сеточного напряжения. Благодаря этому электроны, испускаемые катодом, не смогут попадать на коллектор. Ток через гальванометр G не пойдет, пока нет положительных ионов, т. е. пока ускоряющий сеточный потенциал ниже потенциала ионизации. Как только ускоряющее сеточное напряжение достигнет потенциала ионизации, начнут образовываться положительные ионы — через гальванометр пойдет ток. Показание вольтметра V в этот момент и даст потенциал ионизации ртути. Конечно, в это показание должна быть введена поправка на контактную разность потенциалов между катодом и сеткой, которая находится из предыдущих измерений первого потенциала возбуждения ртути.

6. Чтобы повысить разрешающую способность прибора и наблюдать максимумы, соответствующие высшим энергетическим уровням ртути, можно воспользоваться прежней схемой, изображенной на рис. 22, но вместо одной взять *две сетки*, между которыми накладывается слабое ускоряющее напряжение порядка 0,1 В. Основное ускоряющее напряжение создается между катодом и первой сеткой, а тормозящее напряжение — между второй сеткой и коллектором. Этим достигается то преимущество, что создается большое пространство между сетками, *почти свободное от поля*. В таком пространстве электроны, ускоренные до потенциала возбуждения или выше, главным образом и претерпевают неупругие столкновения. Напротив, расстояние между катодом и ускоряющей сеткой должно быть невелико, чтобы между ними было относительно мало столкновений. С этим и связано повышение разрешающей способности прибора. Понятно, что давление газа в трубке должно быть невысоким (для ртутных паров порядка 5–6 мм рт. ст.). Таким путем у атома ртути обнаруживается второй потенциал возбуждения $V = 6{,}7$ В. Третий потенциал возбуждения $V = 10{,}4$ В является уже ионизационным.

Более совершенный и чувствительный метод измерения критических потенциалов был разработан Герцем. В этом методе электроны, вылетевшие из катода K (рис. 25) и ускоренные потенциалом V, попадают в свободное от поля пространство, ограниченное цилиндрической сеткой, изображенной на рис. 25 штриховой линией (меридиональный разрез). Там происходят их столкновения с атомами газа. При упругих столкновениях электроны, не теряя энергии, отклоняются и могут вылетать через боковую поверхность сетки в слабое задерживающее поле, создаваемое потенциалом $V_1 \approx 0{,}1$ В, наложенным между сеткой и окружающим ее собирающим электродом P (коллектором). Энергия электронов достаточна, чтобы преодолеть V_1 и попасть на P. Если же при увеличении ускоряющего напряжения V появятся неупругие столкновения, то после столкновения энергия электрона может оказаться недостаточной, чтобы преодолеть задерживающее поле и попасть на коллектор P. Ток на коллектор P измеряется гальванометром G. Измерения производятся при двух близких значениях задерживающего потенциала: $V_1 = 0$ и, например, $V_1 = 0{,}1$ В.

Пока столкновения упругие, показания гальванометра в обоих случаях практически одинаковы. Если же ускоряющее напряжение V достигает значения одного из критических потенциалов, то появляются неупругие столкновения, и во втором случае (т. е. при $V_1 = 0{,}1$ В) ток в гальванометре G резко

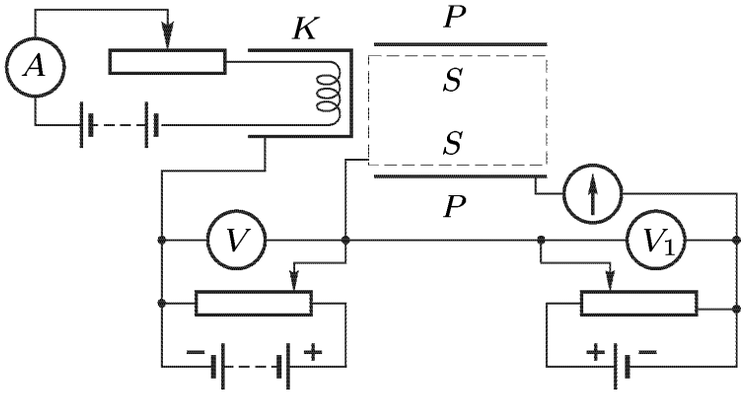


Рис. 25

падает. Разность показаний гальванометра при $V_1 = 0$ и $V_1 = 0{,}1$ В может служить мерой для числа электронов, вылетевших из сетки с энергией меньше 0,1 эВ, т. е. для числа электронов, испытавших неупругие столкновения в окрестности рассматриваемого критического потенциала. Так как электроны сталкиваются с атомами там, где электрического поля нет, то после столкновения они не могут вновь набрать энергию, достаточную для возбуждения атома. По этой причине на кривой $I = I(V)$ уже не появятся равноотстоящие максимумы, соответствующие одному и тому же критическому потенциалу. Кратного повторения таких максимумов, как было раньше, не произойдет. Непосредственное исключение контактной разности потенциалов окажется невозможным. Однако контактную разность потенциалов можно найти и затем учесть ее влияние, если известно значение хотя бы одного критического потенциала. Можно также воспользоваться смесью двух газов, для одного из которых критические потенциалы известны.

7. Опыты Франка и Герца подтверждают также и второй постулат Бора, т. е. правило частот. Оказывается, что при ускоряющем напряжении меньше первого критического потенциала $\approx 4{,}9$ В пары ртути не светятся. В этом случае нет возбужденных атомов ртути, а следовательно, и их переходов в невозбужденное состояние. При повышении ускоряющего напряжения примерно до 4,9 В появляются первые возбужденные атомы. При переходах их в нормальное состояние должно появиться свечение. И действительно, в этом случае наблюдается свечение, состоящее из одной резонансной ультрафиолетовой линии $\lambda = 253{,}7$ нм.

Конечно, для наблюдения этого свечения стеклянный баллон, в котором находятся пары ртути и остальная аппаратура, не годится, так как стекло непрозрачно для ультрафиолетовых лучей. Не годится и стеклянная оптика для исследования спектрального состава излучения. Подходящими материалами могут быть кварц или флюорит, прозрачные соответственно приблизительно до 180 и 120 нм.

По длине волны линии $\lambda = 253{,}7$ нм можно вычислить и первый критический потенциал ртути, и притом значительно точнее, чем по максимумам на кривой $I = I(V)$, получаемой в опытах Франка и Герца.

Для этой цели можно воспользоваться формулой

$$\mathscr{E}_2 - \mathscr{E}_1 = eV = h\nu = \frac{hc}{\lambda}.$$

Подстановка в нее числовых значений дает

$$\lambda = \frac{1240}{V(\text{В})} \text{ нм}. \tag{14.1}$$

Полагая в этой формуле $\lambda = 253{,}7$ нм, найдем $V = 4{,}887$ В, что хорошо согласуется со значением 4,9 В, полученным в опытах Франка и Герца.

При более совершенной методике эксперимента у атомов ртути обнаруживается второй критический потенциал $V = 6{,}7$ В. При этом при переходе в основное состояние появляется излучение с длиной волны $\lambda = 184{,}9$ нм. Этой спектральной линии по формуле (14.1) соответствует $V = 6{,}705$ В, что также находится в превосходном согласии с опытом. Третий критический потенциал ртути 10,4 В является уже ионизационным. Следовательно, при $V > 10{,}4$ В должны происходить переходы из несвязанных состояний на все нижележащие энергетические уровни. И действительно, опыт показал, что при этом возбуждается полный атомный спектр ртути.

8. Для возбуждения и наблюдения свечения газа ударами электронов установку лучше несколько изменить, так как теперь ускоряющее поле уже не требуется. Принципиальная схема, пригодная для наблюдения свечения, приведена на рис. 26. Ускоряющее напряжение V создается между подогреваемым катодом K и окружающей его цилиндрической сеткой S. В свою очередь сетка S окружена цилиндрическим анодом A. Зазор между катодом K и сеткой S должен быть небольшим, чтобы в нем происходило мало столкновений. Напротив, пространство между сеткой S и анодом A должно быть относительно большим, чтобы почти все столкновения происходили именно в этом пространстве. Это пространство должно быть свободно от поля, для чего анод и сетка соединяются между собой. Таким образом, как и в установке Герца (см. рис. 25), ускоряющий промежуток отделен от пространства, где электроны сталкиваются с атомами газа и возбуждают их.

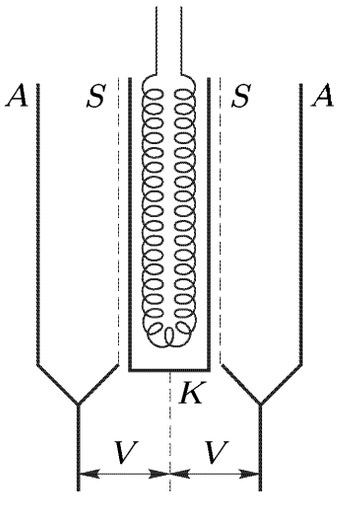


Рис. 26

Допустим, например, что лампа заполнена парами натрия. Опыт показывает, что пары натрия не светятся, если ускоряющее напряжение V меньше 2,1 В. При напряжении 2,1 В начинает возбуждаться только желтая D-линия натрия $\lambda = 589{,}6$ нм. Значит, она является резонансной линией, а $V = 2{,}1$ В — резонансным потенциалом. Его величину можно уточнить, подставив значение λ в формулу (14.1). Это дает $V = 2{,}103$ В. При дальнейшем увеличении ускоряющего напряжения появляются остальные линии спектра испускания натрия.

§ 15. Резонансное свечение и люминесценция

1. В добавление к предыдущему параграфу рассмотрим опыты Вуда (1868–1955) по возбуждению свечения в газах, начатые в 1904–1905 гг. В эвакуированный стеклянный баллон Вуд помещал кусочек металлического натрия. После нагревания баллон наполнялся парами этого металла. Конденсорная линза концентрировала на парах натрия свет от горелки, в пламя которой вводилась поваренная соль. Такой источник света обильно испускал две желтые линии D_1 и D_2 натрия с длинами волн $\lambda_{D_1} = 589{,}6$ нм и $\lambda_{D_2} = 589{,}0$ нм. На пути падающего пучка пары натрия начинали светиться ярким желтым светом, распространявшимся во все стороны, а не только в направлении падающего света. Спектроскопическое исследование показало, что свечение состоит из тех же двух желтых линий натрия D_1 и D_2. Оказалось также, что линии D_1 и D_2 сильно поглощаются парами натрия. С повышением температуры, а следовательно, и плотности паров поглощение увеличивается. Глубина проникновения падающего света в пар уменьшается. Вследствие этого продольные размеры светящейся области сокращаются. При достаточной плотности паров свечение сосредоточивается в тонком поверхностном слое в месте вхождения падающего пучка света. При этом обе линии D_1 и D_2 расширяются и в конце концов сливаются.

Аналогичное явление Вуд наблюдал и в парах ртути. Возбуждающим светом служило ультрафиолетовое излучение ртути с длиной волны $\lambda = 253{,}7$ нм. Конечно, во избежание сильного поглощения сосуд, содержащий ртутные пары, в этом случае должен быть изготовлен из кварца, а источником света может служить, например, ртутная кварцевая лампа. Оказалось, что ртутные пары сильно поглощают свет именно с длиной волны $\lambda = 253{,}7$ нм, а затем переизлучают его во все стороны с той же длиной волны. Явление удалось наблюдать, освещая ртутные пары и второй линией ртути $\lambda = 185{,}0$ нм, хотя в этом случае наблюдение значительно труднее, так как указанная линия поглощается гораздо сильнее. Впоследствии удалось распространить наблюдения и на другие элементы. Не вызывает сомнения, что описанное явление существует во всех веществах, хотя его воспроизведение практически осложняется сильным поглощением и необходимостью иметь подходящий источник света.

2. В то время, когда Вуд наблюдал описанные явления, их, конечно, истолковывали классически. Атому свойственны определенные собственные частоты, с которыми он и излучает спектральные линии. При освещении атома монохроматическим светом определенной частоты в нем возбуждаются вынужденные колебания, и атом начинает переизлучать свет с той же частотой. Этот эффект должен быть выражен особенно сильно, когда частота внешнего излучения совпадает с одной из собственных частот атома, т. е. при резонансе. Поэтому явление получило название *резонансного излучения* или *свечения*.

Конечно, приходящий в колебание атом может не только переизлучать полученную энергию, но, например, передавать ее без излучения

окружающим атомам. Эта часть энергии в конце концов выделяется в виде тепла. Такой эффект проявится в ослаблении или тушении резонансного свечения. Он будет выражен тем сильнее, чем сильнее взаимодействие между атомами рассматриваемого газа. В частности, это произойдет при увеличении плотности газа или при добавлении к нему атомов постороннего вещества. Так, в парах ртути при давлении 0,001 мм рт. ст. резонансное свечение хорошо выражено. Но при добавлении водорода под давлением 0,2 мм рт. ст. интенсивность свечения убывает вдвое, а при дальнейшем добавлении водорода ослабевает еще значительнее. Аналогично действует добавление и других газов.

3. Приведенное классическое толкование резонансного излучения на первый взгляд представляется весьма убедительным и единственно возможным. При критическом рассмотрении это оказывается не совсем так. Действительно, резонансное излучение не есть обособленное явление, а представляет частный случай более общего явления — *излучения спектральных линий.* В основе обоих этих явлений должен лежать *один и тот же* механизм. Происхождение спектральных линий и спектральных серий классическая физика объяснить не могла. Это удалось сделать только квантовой теории. Естественно думать, что и резонансное излучение должно иметь также квантовое истолкование. Последнее действительно существует и заключается в следующем.

Квант падающего света $h\nu$ должен возбудить атом, т. е. перевести его с нормального энергетического уровня $\mathscr{E}_1$ на вышележащий уровень. Пусть $\mathscr{E}_2$ — ближайший уровень, лежащий выше $\mathscr{E}_1$. Если $h\nu < \mathscr{E}_2 - \mathscr{E}_1$, то энергии светового кванта $h\nu$ недостаточно, чтобы перевести атом на уровень $\mathscr{E}_2$. В этом случае возбуждение атома световым квантом невозможно. Для возбуждения уровня $\mathscr{E}_2$ необходимо, чтобы $\mathscr{E}_2 \leqslant \mathscr{E}_1 + h\nu$. Максимальное возбуждение будет происходить при условии $\mathscr{E}_2 = \mathscr{E}_1 + h\nu$, когда вся энергия кванта $h\nu$ расходуется на возбуждение уровня $\mathscr{E}_2$. Через короткое время атом с уровня $\mathscr{E}_2$ возвращается на ближайший нижележащий уровень, т. е. $\mathscr{E}_1$. При этом излучается световой квант $h\nu' = \mathscr{E}_2 - \mathscr{E}_1$. Таким образом, $\nu' = \nu$, т. е. частоты возбуждающего и переизлучаемого света одинаковы. В этом и состоит резонансное излучение, если его рассматривать с квантовой точки зрения.

То обстоятельство, что в парах натрия наблюдается излучение не одиночной линии, а двух близких линий D_1 и D_2, объясняется тем, что второй возбужденный уровень $\mathscr{E}_2$ атома натрия в действительности состоит из двух близко расположенных уровней.

С квантовой точки зрения, как и с классической, понятно и явление *тушения свечения* при увеличении концентрации газа. Действительно, атом, находящийся на возбужденном уровне $\mathscr{E}_2$, может передать свою энергию без излучения окружающим атомам. Этот эффект усиливается с увеличением плотности газа или при добавлении к нему посторонних примесей.

Изложенное представляет только основную схему явления, но не охватывает всех его сторон. В частности, мы оставили в стороне все

вопросы, связанные с уширением энергетических уровней и спектральных линий. Но это не входит в нашу задачу.

4. Постулаты Бора составляют основу для понимания явлений *люминесценции*, которые до этого оставались совершенно непонятными. Ответим сначала на вопрос, что такое люминесценция?

Все тела испускают электромагнитные волны, интенсивность и степень поляризации которых в различных участках спектра определяются температурой тела и его поглощательной способностью в соответствии с законом Кирхгофа (1824–1887). Такое излучение называется *температурным* или *равновесным*. Помимо температурного излучения, многие тела в результате различных внешних возбудителей дают избыточное излучение, не определяющееся температурой тела. Таково, например, свечение экрана в телевизоре, свечение газа в газоразрядной трубке при прохождении через него электрического тока, свечение сахара при раскалывании, свечение некоторых живых организмов (светлячков), свечение гниющего дерева и т. д. Все это — примеры *холодного свечения*. Такое *избыточное над температурным излучение называется люминесценцией, если его длительность после прекращения внешнего воздействия (послесвечение) значительно превышает период световых колебаний.*

Первая часть этого определения и самый термин «люминесценция» были введены Е. Видеманом (1826–1899). Вторая часть — критерий длительности — был введен С. И. Вавиловым, чтобы отличить люминесценцию от более кратковременных явлений вторичного излучения — отражения и рассеяния света, тормозного излучения заряженных частиц, излучения Вавилова–Черенкова и пр. Конечно, этот критерий не определен вполне жестко и может служить только для *общей ориентировки*. В ряде случаев он не позволяет провести резкую грань между люминесценцией и нелюминесценцией. Раньше явления люминесценции делили на *флуоресценцию* и *фосфоресценцию*. Под флуоресценцией понимали излучение, прекращающееся мгновенно после прекращения действия вызывающего его внешнего возбудителя. Если же свечение продолжается некоторое время и после удаления возбудителя, то его называли фосфоресценцией. В свете более поздних исследований было выяснено, что никакой принципиальной разницы между флуоресценцией и фосфоресценцией нет. Различие между ними чисто количественное — по времени длительности послесвечения. Поэтому указанное деление потеряло смысл. Термин «флуоресценция» применяется и сейчас, но только для того, чтобы качественно подчеркнуть кратковременность свечения. Например, длительность резонансного свечения разреженных газов, как показали специальные измерения, порядка 10^{-8}–10^{-9} с.

К приведенному определению люминесценции следует добавить некоторые замечания. Поскольку температура — *макроскопическое понятие*, разделять полное излучение на температурное и люминесценцию имеет смысл не для отдельных молекул и атомов, а *только для их совокупностей* — тел, имеющих определенную температуру. Если состояние тела столь неравновесно, что говорить об его температуре не приходится, то в этом случае нельзя говорить и о температурном излучении, и о люминесценции. Так будет и при свечении отдельных атомов и молекул. По этой причине для свечения газов, исследованного Вудом, мы предпочли нейтральный термин «резонансное излучение» вместо часто употребляемого термина «резонансная флуоресценция». Действительно, такое свечение наблюдается и в разреженных газах. Термин же «резонансная флуоресценция» оправдан, когда речь идет о свечении сравнительно плотных газов.

Вещества, в которых возбуждается люминесценция, называются *люминофорами*. По методу возбуждения свечения различают:

1) *фотолюминесценцию* — люминесценцию, возникающую при освещении люминофора видимым или ультрафиолетовым светом;

2) *рентгенолюминесценцию* — свечение люминофора при облучении его рентгеновскими или γ-лучами;

3) *катодолюминесценцию* — свечение люминофора при бомбардировке его электронами (катодными лучами);

4) *радиолюминесценцию* — свечение под действием ядерных излучений: α-частиц, β-частиц, γ-лучей, протонов и т. д.;

5) *электролюминесценцию* — свечение люминофора, например газонаполненной вакуумной трубки, в электрическом поле;

6) *хемилюминесценцию* — свечение тел при химических реакциях;

7) *триболюминесценцию* — свечение, возникающее при растирании, раздавливании или раскалывании некоторых кристаллов (например, сахара);

8) *кандолюминесценцию* — свечение (но не температурное), возникающее при помещении некоторых веществ в пламя, например в пламя газовой горелки.

5. Оставляем в стороне вопрос о полноте и обоснованности этого списка. Вопросам люминесценции, и даже ее отдельным видам, посвящена обширная специальная литература, и мы не можем входить в их обсуждение. Существенно только отметить *квантовый характер люминесценции.* Как и всякое излучение, люминесценция возникает в результате квантовых переходов излучающей системы с каких-то возбужденных энергетических уровней на уровни более низкие. С этой точки зрения становится понятным один из характерных признаков люминесценции — *длительность свечения,* ибо многие тела могут находиться в возбужденных состояниях длительное время. Характер люминесценции определяется структурой энергетического спектра тела, средним временем его пребывания в возбужденных состояниях, правилами отбора при поглощении и излучении света и т. д.

Для фотолюминесценции Стоксом (1819–1903) в 1852 г. было установлено правило, согласно которому *длина волны λ' света люминесценции больше длины волны λ возбуждающего света.* Это правило указывает на квантовый характер фотолюминесценции. Действительно, если квант возбуждающего света переводит систему с нормального уровня $\mathscr{E}_1$ на возбужденный уровень $\mathscr{E}_2$, то $\mathscr{E}_2 \leqslant \mathscr{E}_1 + h\nu$. При обратном возвращении системы на прежний уровень $\mathscr{E}_1$ частота излучаемого света будет $\nu' = (\mathscr{E}_2 - \mathscr{E}_1)/h$, а потому $\nu' \leqslant \nu$. Система может вернуться с уровня $\mathscr{E}_2$ не на прежний уровень $\mathscr{E}_1$, но на уровень, лежащий несколько выше. Тогда тем более $\nu' < \nu$.

Однако правило Стокса допускает исключения. Допустим, что световой квант переводит систему с какого-то возбужденного уровня $\mathscr{E}'_1$ на вышележащий уровень $\mathscr{E}'_2$. Тогда, конечно, $\mathscr{E}'_2 \leqslant \mathscr{E}'_1 + h\nu$. Но так как $\mathscr{E}'_1 > \mathscr{E}_1$, то может случиться, что $\mathscr{E}'_2 > \mathscr{E}_1 + h\nu$. В этом случае при возвращении системы на невозмущенный уровень $\mathscr{E}_1$ будет испускаться квант с частотой $\nu' > \nu$.

§ 16. Принципиальные недостатки теории Бора

Теория Бора — важный этап в понимании внутриатомных явлений. Классическая физика, созданная при изучении макроскопических явлений, столкнулась с принципиальными трудностями, когда ее стали применять к атомам, молекулам и вообще к явлениям микромира. Теория Бора подготовила почву для уяснения того важнейшего

факта, что *для понимания явлений микромира классических понятий и классических законов недостаточно.* В области микромира нужны *принципиально новые понятия и законы.* Здесь первенствующая роль должна принадлежать кванту действия, открытому Планком. Теория Бора послужила мощным стимулом для постановки многих экспериментальных исследований, принесших важные результаты. Даже в тех случаях (а таких случаев было большинство), когда теория была не в состоянии дать количественного объяснения многих явлений, два постулата Бора служили руководящей нитью при классификации и качественной интерпретации этих явлений. На их основе, например, был классифицирован громадный эмпирический материал атомной и молекулярной спектроскопии.

Но двух постулатов Бора, конечно, недостаточно для построения полной теории. Они должны быть дополнены прежде всего *правилами квантования*, с помощью которых могут быть вычислены уровни энергии атома. Бор предложил правило квантования круговых орбит электронов в одноэлектронных атомах. Оно выражается формулой (13.6). При получении этой формулы Бор, как он сам признается, исходил из требования подогнать теоретическую формулу для уровней энергии под эмпирически установленную формулу Бальмера для спектральных термов водорода. Несколько позже Зоммерфельд обобщил правило квантования Бора на случай движения электрона по эллиптическим орбитам. Но и после этого правило квантования относилось только к атому с одним электроном. Не удалось распространить правила квантования на многоэлектронные атомы, даже на простейший после водорода атом гелия, состоящий из ядра и двух электронов. От теории, конечно, нельзя требовать, чтобы она давала аналитическое решение задачи трех тел. Такое решение может оказаться невозможным. Но теория должна указать *принципиальный метод* численного решения с точностью, достаточной для сравнения с экспериментом [1]. Этого теория Бора дать не могла. Она тем более не могла дать принципиального решения и более сложного вопроса — образования молекул, даже простейшей молекулы водорода, состоящей из двух протонов и двух электронов.

Теория Бора, даже в простейшем случае одноэлектронных атомов, позволяла вычислять только частоты спектральных линий, но не их

[1] Именно так обстоит дело в небесной механике. Аналитическое решение задачи трех тел получить не удалось. Но в небесной механике разработаны эффективные приближенные методы численного решения задачи, дающие необходимую точность. Правда, Зундман в 1912 г. нашел аналитическое решение в виде рядов. Но для получения числовых результатов эти ряды непригодны, так как они сходятся очень медленно. Например, чтобы с помощью рядов Зундмана достигнуть той же точности, которую дают современные астрономические ежегодники, надо просуммировать примерно $10^{8\,000\,000}$ членов. Это выходит далеко за пределы возможностей современных и будущих самых больших электронных счетных машин.

интенсивности и поляризацию. Правда, для нахождения интенсивностей и поляризации она пользовалась *принципом соответствия.* Но принцип соответствия мог быть оправдан только при больших квантовых числах, где вычисления могли быть произведены классически. Теория Бора распространяла результат и на малые квантовые числа, для чего не было никаких оснований. Таким образом, в конце концов интенсивность и поляризация определялись классически.

Но основной принципиальный недостаток теории Бора — в ее *непоследовательности.* Она принимала существование только стационарных состояний атома или, как говорил сам Бор, стационарных орбит электронов. Это совершенно непонятно с точки зрения классической механики. В то же время к движению электронов в стационарных состояниях она применяла законы классической механики, хотя и считала неприменимой классическую электродинамику (поскольку нет излучения). По шуточному замечанию Г. Брэгга (1862–1942), в теории Бора по понедельникам, средам и пятницам надо применять классические законы, а по вторникам, четвергам и субботам — квантовые. Два постулата Бора, если не пользоваться представлениями об орбитах электронов в атомах, как это делалось нами выше, *проверены экспериментально* и потому должны считаться *правильными.* Но сама теория Бора в целом является только промежуточным этапом на пути к более совершенной и последовательной теории. Это лучше других понимал сам Бор, которому принадлежит главная заслуга в осмысливании принципиальных положений квантовой механики, пришедшей на смену теории Бора.

Планк ввел представление о квантовом характере процессов излучения и поглощения света. Эйнштейн распространил квантование света и на его распространение в пространстве, введя представление о фотонах. А теория теплоемкостей Эйнштейна прямо указывала на то, что введенная Планком постоянная h проявляется не только в световых явлениях, но и в процессах, происходящих в веществе. Дальнейший шаг сделал Бор. Успех теории Бора при вычислении постоянной Ридберга и размеров атома выявил значение постоянной Планка как *универсальной фундаментальной величины* для описания всех видов материи, а не только для описания корпускулярно-волнового дуализма света. И действительно, постоянная Планка, имеющая размерность действия (т. е. размерность произведения импульса на координату или энергии на время), пронизывает содержание всей новейшей (квантовой) физики.

Глава III
ВОЛНОВЫЕ СВОЙСТВА ЧАСТИЦ ВЕЩЕСТВА

§ 17. Гипотеза де Бройля

1. Создание последовательной теории для описания явлений атомных и субатомных масштабов было начато и вчерне завершено в 1925–1926 гг. Такая теория получила название *квантовой механики.* Сначала возникло то направление в квантовой механике, которое получило название *матричной механики.* Ее основные идеи были заложены в основополагающей работе Гейзенберга (1901–1976) «О квантовомеханическом истолковании кинематических и механических соотношений». Систематическое построение матричной механики было дано Борном (1882–1970) и Иорданом (1902–1980), к которым в дальнейшем присоединился и сам Гейзенберг. К этому направлению примыкает и та форма квантовой механики, которая практически одновременно и независимо была разработана Дираком. Немного позже в работах Шредингера (1887–1961) появилось другое направление, названное *волновой механикой.* Вскоре было выяснено, что эти два направления, отличаясь по форме, тождественны по своему физическому содержанию.

В общем курсе нецелесообразно говорить о весьма абстрактной матричной механике. Ограничимся только изложением, далеко не полным, физических представлений волновой механики. Разумеется, мы не можем подробно излагать сложный математический аппарат, составляющий неотъемлемую и весьма важную часть квантовой механики. Это делается в курсах теоретической физики.

2. Построению волновой механики Шредингера предшествовали работы Луи де Бройля (р. 1892). В 1923–1924 гг. он выдвинул и развил идеи о волнах вещества. К тому времени в оптике уже сложилась парадоксальная, но подтверждаемая опытом ситуация: в одних явлениях (интерференции, дифракции,...) свет ведет себя как *волны*; другие явления (фотоэффект, эффект Комптона,...) показывают с неменьшей убедительностью, что он ведет себя и как *частицы.* Де Бройль поставил вопрос, не распространяется ли подобный корпускулярно-волновой дуализм и на обычное вещество? Если это действительно так, то каковы волновые свойства частиц вещества? Ответ, подтвержденный в дальнейшем опытами, оказался положительным.

Пусть частица движется в свободном пространстве с постоянной скоростью v. Де Бройль предположил, что с такой частицей связана какая-то плоская монохроматическая волна

$$\Psi = \Psi_0 e^{i(\mathbf{kr}-\omega t)}, \tag{17.1}$$

распространяющаяся в направлении скорости v [1]. О природе этой волны, т. е. о физическом смысле функции Ψ, де Бройль не мог сказать ничего определенного. Отвлечемся временно и мы от обсуждения этого вопроса. Волны типа (17.1) получили название *фазовых волн*, *волн вещества* или *волн де Бройля*.

Попытаемся установить рациональную связь между корпускулярными и волновыми характеристиками частицы, которая совсем не зависит от физической природы величины Ψ. Будем руководствоваться требованием, чтобы эта связь была релятивистски инвариантна. Корпускулярные свойства частицы характеризуются ее энергией $\mathscr{E}$ и импульсом $\mathbf{p}$, волновые — частотой ω и волновым вектором $\mathbf{k}$. Под $\mathscr{E}$ мы будем понимать *полную энергию* частицы в смысле теории относительности. Она определяется однозначно требованием, чтобы *энергия и импульс образовывали четырехмерный вектор* $(\mathscr{E}/c, \mathbf{p})$ (см. т. IV, § 111, п. 3). Частоту ω определим из требования, чтобы фаза волны $\omega t - \mathbf{kr}$ была релятивистски инвариантна (см. по этому поводу § 19, п. 9). Тогда ω и $\mathbf{k}$ будут образовывать четырехмерный вектор $(\omega/c, \mathbf{k})$. Если потребовать, чтобы временные и пространственные компоненты четырехмерных векторов $(\mathscr{E}/c, \mathbf{p})$ и $(\omega/c, \mathbf{k})$ были пропорциональны друг другу, то получатся релятивистски инвариантные соотношения

$$\mathscr{E} = \hbar\omega, \tag{17.2}$$

$$\mathbf{p} = \hbar\mathbf{k}. \tag{17.3}$$

Они будут совпадать с соответствующими соотношениями для фотонов, если для всех частиц $\hbar$ положить равной постоянной Планка, что мы и сделаем. Такой выбор $\hbar$ логически не необходим, а оправдывается последующими результатами. Соотношения (17.2) и (17.3) и были постулированы де Бройлем.

Во всякой инерциальной системе отсчета волновой вектор $\mathbf{k}$ определен абсолютно однозначно, поскольку соотношением (17.3) он однозначно выражается через импульс частицы $\mathbf{p} = mv$. Напротив, соотношение (17.2) такой абсолютной однозначностью не отличается. Здесь однозначность навязана искусственно — требованием, чтобы $\mathscr{E}$ и ω были временными компонентами четырехмерных векторов. В принципе же энергия определена всегда с точностью до произвольной аддитивной постоянной. Мы увидим далее (см. § 19), что и частоту ω можно переопределить так, чтобы она также содержала аддитивную постоянную.

---

[1] В оптике монохроматическая волна записывалась в виде $e^{i(\omega t - \mathbf{kr})}$. Сейчас мы употребляем комплексно сопряженное выражение $e^{i(\mathbf{kr} - \omega t)}$. Оба способа написания совершенно равноправны. Но в квантовой механике укоренилось написание волны именно в форме (17.1). Однако независимо от способа написания под фазой волны следует понимать выражение $\omega t - \mathbf{kr}$.

3. Рассмотрим некоторые свойства волн де Бройля, вытекающие из соотношений (17.2) и (17.3). Прежде всего из (17.3) получаем выражение для длины волны де Бройля:

$$\lambda = \frac{2\pi}{k} = \frac{2\pi\hbar}{p} = \frac{h}{p}. \tag{17.4}$$

Эта величина в каждой инерциальной системе отсчета определена однозначно.

Для фазовой скорости волн де Бройля формулы (17.2) и (17.3) дают

$$v_{\text{ф}} = \frac{\omega}{k} = \frac{\mathscr{E}}{p}. \tag{17.5}$$

В релятивистской теории $\mathscr{E} = mc^2$, $p = mv$, где v — скорость частицы, а m — релятивистская масса. В этом случае

$$v_{\text{ф}} = \frac{c^2}{v}. \tag{17.6}$$

Поскольку всегда $v \leqslant c$, отсюда следует, что $v_{\text{ф}} \geqslant c$. Для фотонов в вакууме $v = c$, а потому в этом случае $v_{\text{ф}} = c$. Полученный результат не должен нас смущать, поскольку на величину фазовой скорости не накладывается никаких ограничений. К тому же в дальнейшем будет показано, что, согласно современной физической интерпретации, фазовая скорость волн де Бройля имеет чисто символическое значение, так как эта интерпретация относит ее к числу принципиально ненаблюдаемых величин.

Принципиально наблюдаемой величиной является групповая скорость волн де Бройля

$$v_{\text{гр}} = \frac{d\omega}{dk} = \frac{d\mathscr{E}}{dp}. \tag{17.7}$$

Эта величина не содержит никакой неопределенности, поскольку не только dp, но и приращение энергии $d\mathscr{E}$ определены однозначно. При любой скорости движения частицы $d\mathscr{E} = v\,dp$, так что всегда

$$v_{\text{гр}} = v, \tag{17.8}$$

т. е. групповая скорость волн де Бройля равна скорости частицы. Заменив теперь в формуле (17.6) v на $v_{\text{гр}}$, получаем

$$v_{\text{ф}} v_{\text{гр}} = c^2. \tag{17.9}$$

Формально можно образовать величину, аналогичную длине волны де Бройля (17.4). Для этого заметим, что длина четырехмерного вектора энергии-импульса частицы в пространстве Минковского равна $\sqrt{(\mathscr{E}/c)^2 - p^2}$. Это есть инвариант, равный $m_0 c$, где m_0 — масса покоя частицы. Поделив на него постоянную Планка h, получим инвариантную величину

$$\lambda_{\text{К}} = \frac{h}{m_0 c}, \tag{17.10}$$

имеющую размерность длины. Она представляет собой комптоновскую длину частицы. Таким образом, формально $\lambda_{\text{К}}$ можно рассматривать

как длину волны де Бройля, которой соответствует величина импульса, равная инвариантной длине четырехмерного вектора энергии-импульса частицы в пространстве Минковского.

4. Де Бройль использовал представление о фазовых волнах для наглядного толкования таинственного правила квантования Бора (13.6) в случае одноэлектронного атома. Он рассмотрел фазовую волну, бегущую вокруг ядра по круговой орбите электрона. Если на орбите длина волны λ укладывается целое число раз (рис. 27), то волна при обходе вокруг ядра будет всякий раз возвращаться в исходную точку с той же фазой и амплитудой. В каждой точке орбиты установится неизменный колебательный режим во времени и не возникнет излучения. В этом случае орбита получится стационарной. Если же указанное условие не выполняется, то при обходе вокруг ядра фаза и амплитуда волны не возвратятся к своим исходным значениям — стационарного состояния не получится. Исходя из этих соображений, де Бройль записал условие стационарности орбиты, или правило квантования, в виде

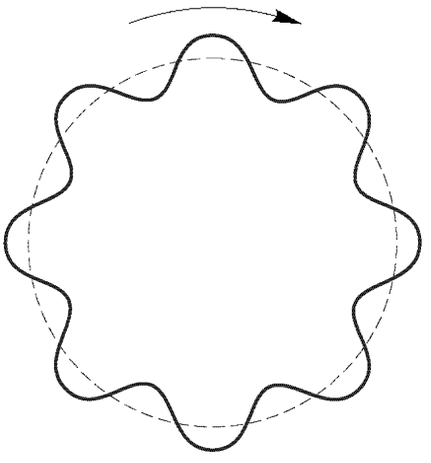


Рис. 27

$$\frac{2\pi R}{\lambda} = n, \tag{17.11}$$

где R — радиус круговой орбиты, а n — целое число (главное квантовое число). Полагая здесь $\lambda = h/p = 2\pi\hbar/p$ и замечая, что $L = Rp$ есть момент количества движения электрона, получим

$$L = n\hbar, \tag{17.12}$$

что совпадает с условием (13.6). В этом де Бройль видел успех своей концепции фазовых волн. В дальнейшем условие (17.11) удалось обобщить и на случай эллиптических орбит, когда длина волны λ меняется вдоль траектории электрона. Казалось, что это еще больше усиливало успех теории.

На самом деле этот успех призрачный. В рассуждении де Бройля предполагается, что волна распространяется не в пространстве, а вдоль линии — вдоль стационарной орбиты электрона. Такая идеализация соответствует приближению геометрической (лучевой) оптики. Этим приближением можно пользоваться в предельном случае, когда длина волны λ пренебрежимо мала по сравнению с радиусом орбиты электрона, т. е. при больших квантовых числах. А тогда проблема квантования несущественна. Чтобы действительно получить существенно новое, надо заменить геометрическую оптику волновой. Это и было сделано Шредингером.

5. К полученным результатам можно прийти и другим путем. Для этого введем *показатель преломления* μ волн де Бройля — важную

величину, имеющую и самостоятельное значение. Пространство, в котором распространяется волна де Бройля, условимся называть *средой*. Если в среде нет силового поля, то среда будет однородной. Показатель преломления среды может быть определен лишь с точностью до произвольного постоянного множителя, так как для преломления фазовых волн на границе раздела двух сред имеет значение только отношение показателей преломления этих сред. Во всякой волновой теории μ обратно пропорционален фазовой скорости волны. В случае волн де Бройля $\mu \sim 1/v_{\text{фаз}} = v/c^2$. Опуская постоянный множитель, можно принять

$$\mu = v. \tag{17.13}$$

Определяемый этой формулой показатель преломления условно будем называть *абсолютным*. Формула (17.13) сохраняет смысл и в том случае, когда скорость частицы v меняется от точки к точке, т. е. при наличии силовых полей. Скорость v, а с ней и μ в каждой точке однозначно определяются уравнением энергии $\mathscr{E} + U = \text{const}$, в котором предполагается, что потенциальная функция U зависит только от координат, но не зависит явно от времени.

В предельном случае коротких длин волн распространение последних происходит вдоль независимых линий или *лучей* (см. т. IV, § 6). Этот случай называется *геометрической* или *лучевой оптикой*. Распространение волнового возмущения вдоль лучей формально аналогично движению частицы классической механики по траекториям. Радиус кривизны R луча или траектории частицы определяется формулой

$$\frac{1}{R} = \frac{\partial}{\partial N}(\ln \mu) = \frac{\partial}{\partial N}(\ln v), \tag{17.14}$$

где дифференцирование производится в направлении главной нормали $\mathbf{N}$ к лучу или траектории (см. т. IV, § 4). Эта формула, конечно, может быть использована (вместо уравнений Ньютона) для определения формы луча или траектории частицы.

Вернемся теперь к выводу правила квантования, данному де Бройлем. Условие применимости геометрической оптики к движению электрона вокруг ядра атома выражается формулой

$$\left|\lambda \frac{d\mu}{dr}\right| \ll \mu, \quad \text{т. е.} \quad \left|\lambda \frac{dv}{dr}\right| \ll v.$$

Подставив сюда $\lambda = 2\pi\hbar/p$, $\mu = v$ и ограничиваясь нерелятивистским приближением, запишем это так:

$$\left|2\pi\hbar \frac{dv}{dr}\right| \ll pv = 2K, \tag{17.14а}$$

где K — кинетическая энергия электрона. Скорость v найдется из уравнения энергии

$$\frac{mv^2}{2} - \frac{Ze^2}{r} = \mathscr{E} = \text{const}.$$

Отсюда находим

$$mvr\frac{dv}{dr} = -\frac{Ze^2}{r} = U,$$

где U — потенциальная энергия. Если электрон движется по окружности, то $mvr = n\hbar$, а потому

$$n\hbar\frac{dv}{dr} = U = \mathscr{E} - K = -2K,$$

так как при движении по окружности $\mathscr{E} + K = 0$. Сопоставляя полученное соотношение с неравенством (17.14а), находим

$$n \gg 2\pi,$$

т. е. квантовое число n должно быть большим в согласии с тем, что было установлено выше.

6. Все изложенное представляет собой чисто умозрительное, гипотетическое построение, а потому не имеет доказательной силы. Истинное доказательство или опровержение полученных результатов может дать только опыт. В каких же явлениях природы могут проявиться волновые свойства вещества, если они действительно существуют? Независимо от физической природы волн сюда относятся явления интерференции и дифракции. Непосредственно наблюдаемой величиной в них является длина волны λ. Во всех случаях длины волн де Бройля определяются формулой (17.4). Применим ее к нерелятивистскому движению частиц.

Для электронов, ускоренных разностью потенциалов V, импульс определяется формулой $p = \sqrt{2m_e eV}$, так что в абсолютной системе единиц

$$\lambda_e = \frac{h}{\sqrt{2m_e eV}}. \tag{17.15}$$

Положим здесь $hc = 1{,}2399 \cdot 10^{-4}$ эВ $\cdot$ см, $m_e c^2 = 511\,003$ эВ. Тогда получится практическая формула

$$\lambda_e = \sqrt{\frac{150{,}42}{V_{(\text{В})}}} \cdot 10^{-8} \text{ см} = \frac{1{,}2264}{\sqrt{v_{(\text{В})}}} \text{ нм}. \tag{17.16}$$

Для протонов

$$\lambda_p = \frac{0{,}02862}{\sqrt{v_{(\text{В})}}} \text{ нм}. \tag{17.17}$$

Вычислим еще длину волны де Бройля для молекул неподвижного газа при абсолютной температуре T. Задача эта — не совсем определенная, поскольку молекулы движутся с тепловыми скоростями, распределенными по закону Максвелла. Не вдаваясь в обоснование (см. задачу к следующему параграфу), будем понимать под v среднюю квадратичную скорость молекулы. Тогда ее импульс будет $p = \sqrt{3mkT}$. Отсюда легко получить для атомов гелия ($m_{\text{He}} = 6{,}7 \cdot 10^{-24}$ г)

$$\lambda_{\text{He}} = \frac{1{,}26}{\sqrt{T}} \text{ нм}. \tag{17.18}$$

Для молекул водорода

$$\lambda_{H_2} = \frac{1{,}78}{\sqrt{T}} \text{ нм}. \tag{17.19}$$

а для тепловых нейтронов

$$\lambda_n = \frac{2{,}52}{\sqrt{T}} \text{ нм}. \tag{17.20}$$

Эти формулы показывают, что для электронов, ускоренных до потенциала 100–10 000 В, для атомов гелия и молекул водорода при комнатной температуре, а также для тепловых нейтронов и других «медленных» легких частиц длины волн де Бройля того же порядка, что и длины волн мягких рентгеновских лучей. Поэтому дифракцию таких частиц надо пытаться искать методами, аналогичными тем, которые применяются в случае рентгеновских лучей. Однако гипотеза де Бройля представлялась настолько фантастичной, что сравнительно долго никто из экспериментаторов не пытался подвергнуть ее экспериментальной проверке.

ЗАДАЧИ

1. Обобщить нерелятивистские формулы (17.15), (17.16) и (17.17) на случай релятивистских электронов и протонов. При каком значении ускоряющего потенциала V можно пользоваться нерелятивистскими формулами, чтобы ошибка не превосходила одного процента?

Ответ.

$$\lambda = \frac{h}{\sqrt{2m_0eV}}\left(1 + \frac{eV}{2m_0c^2}\right)^{-1/2}, \tag{17.15а}$$

где m_0 — масса покоя частицы. Для электронов

$$\lambda_e = \frac{1{,}2264}{\sqrt{v_{(В)}}}\left(1 + 0{,}978 \cdot 10^{-6} V_{(В)}\right)^{-1/2} \text{ нм}, \tag{17.16а}$$

для протонов

$$\lambda_p = \frac{0{,}02862}{\sqrt{v_{(В)}}}\left(1 + 0{,}533 \cdot 10^{-9} V_{(В)}\right)^{-1/2} \text{ нм}. \tag{17.17а}$$

Нерелятивистскими формулами при указанной точности расчета можно пользоваться для электронов при $V \lesssim 20$ кэВ, для протонов при $V \lesssim 40$ МэВ.

2. Найти приближенное выражение для длины волны де Бройля ультрарелятивистской частицы, т. е. такой частицы, кинетическая энергия $\mathscr{E}_{кин}$ которой много больше энергии покоя m_0c^2. При каких значениях $\mathscr{E}_{кин}$ полученная формула будет давать ошибку, не превосходящую 1 %? Найти λ для ультрарелятивистской частицы с кинетической энергией $\mathscr{E}_{кин} = 10$ ГэВ.

Ответ. $\lambda = hc/\mathscr{E}_{кин}$. При $\mathscr{E}_{кин} > 100\, m_0c^2$ ошибка не превосходит 1 %. При $\mathscr{E}_{кин} = 10$ ГэВ $\lambda = 1{,}25 \cdot 10^{-14}$ см.

3. При какой скорости частицы ее дебройлевская и комптоновская длины волн равны между собой?

Ответ. $v = c/\sqrt{2}$.

§ 18. Экспериментальные подтверждения гипотезы де Бройля

1. Интерференция электронов при отражении от кристаллов была обнаружена, но не понята еще до появления гипотезы де Бройля. Производя опыты по рассеянию электронов тонкими металлическими фольгами в 1921–1923 гг., Дэвиссон (1881–1958) и Кэнсман наблюдали определенно выраженную зависимость интенсивности рассеянного пучка от угла рассеяния. Положение и величина получающихся максимумов на кривой рассеяния существенно зависели от скорости электронов. В одном из опытов, в котором электроны рассеивались никелевой пластинкой, стеклянный прибор лопнул и пластинка окислилась. После длительного прокаливания пластинки в вакууме и атмосфере водорода произошла перекристаллизация с образованием некоторого количества крупных кристаллов. При повторении опыта по рассеянию электронов с этой пластинкой кривая рассеяния резко изменилась: количество максимумов сильно возросло, а сами максимумы сделались значительно более отчетливыми. На рис. 28 приведены полярные диаграммы рассеяния электронов до прокаливания пластинки ($а$) и после прока-

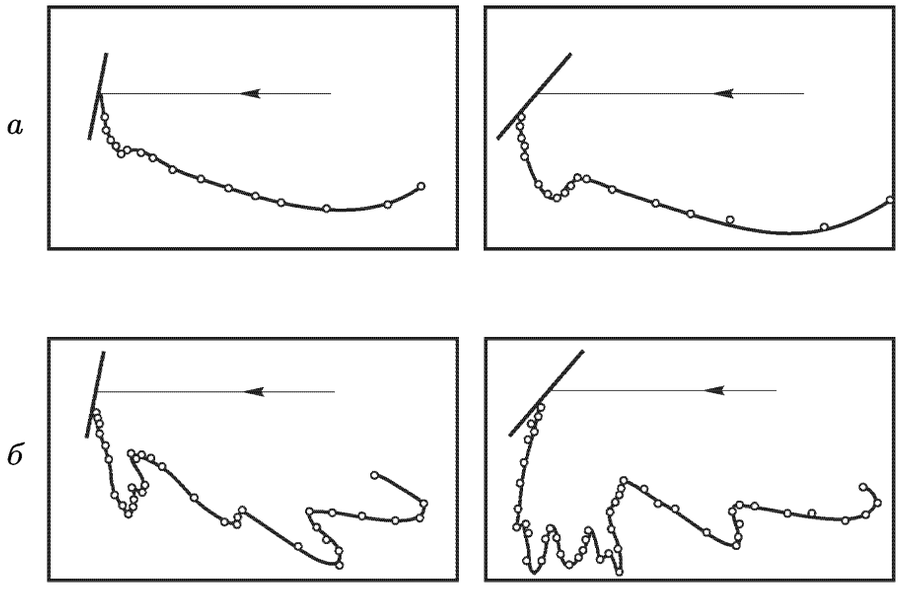


Рис. 28

ливания ($б$). На этих диаграммах вдоль радиусов-векторов отложены отрезки, пропорциональные интенсивности рассеяния электронов в соответствующих направлениях.

Происхождение максимумов и минимумов на кривых рассеяния долгое время оставалось непонятным, пока их не истолковали как результат интерференционного отражения волн де Бройля от соответствующих атомных плоскостей крупных кристаллов, образовавшихся в результате перекристаллизации. Это истолкование было подтверждено в 1927 г. опытами Дэвиссона и Джермера (1896–1971). На этих опытах была открыта дифракция электронов. С них началось систематическое исследование этого явления.

2. В опытах Дэвиссона и Джермера прежде всего использовался *метод Л. Брэгга* (1890–1971), но применительно не к рентгеновским лучам, а к волнам де Бройля. Параллельный пучок электронов одинаковой скорости, создаваемый «электронной пушкой» A (рис. 29), направлялся на монокристалл никеля. Рассеянные электроны улавлива-

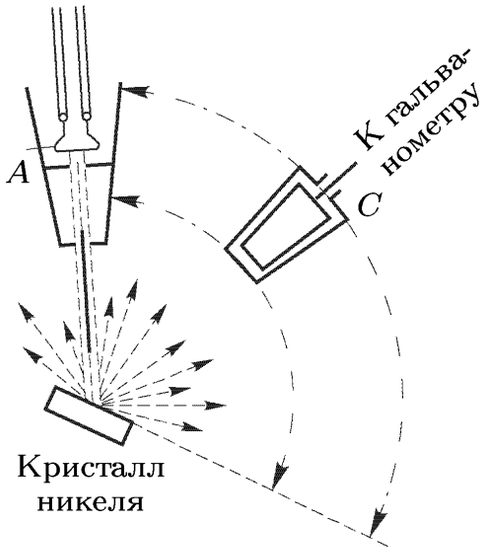


Рис. 29

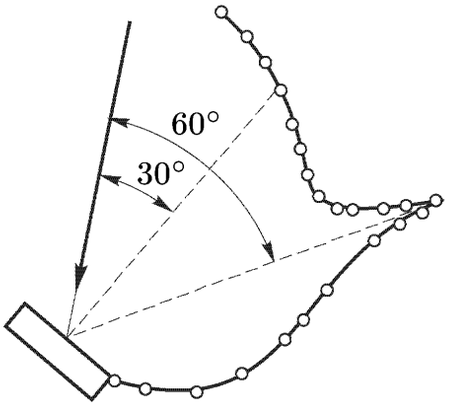


Рис. 30

лись коллектором C, соединенным с гальванометром. Коллектор можно было устанавливать под любым углом к направлению падающего пучка, вращая его все время в одной плоскости (плоскости рисунка). По показанию гальванометра можно было судить об интенсивности пучков электронов, рассеянных в различных направлениях. Типичная полярная диаграмма интенсивности рассеяния электронов представлена на рис. 30. На ней имеется резко выраженный максимум, соответствующий зеркальному отражению электронов, когда угол падения равен углу отражения. Тот же опыт, повторенный с поликристаллической пластинкой никеля, состоящей из множества мельчайших беспорядочно ориентированных кристалликов, не обнаружил никакого преимущественного направления при отражении электронов.

В опытах Дэвиссона и Джермера гипотеза де Бройля была подвергнута и количественной проверке. Как известно, отражение рентгеновских лучей от кристаллов носит *интерференционный характер* (см. т. IV, § 61). От различных параллельных атомных плоскостей кристалла исходят волны, как бы испытавшие зеркальное отражение на каждой из этих плоскостей. Если выполнено условие Брэгга–Вульфа

$$2d\sin\varphi = m\lambda, \tag{18.1}$$

где φ — угол скольжения, d — межплоскостное расстояние, $m = 1, 2, 3, \ldots$, то эти волны при интерференции усиливают друг друга. В результате и возникает отраженная волна. Необходимым условием такой трактовки отражения является выполнение неравенства $m\lambda/(2d) < 1$, т. е. малость длины волны. Такое условие выполняется и в случае волн де Бройля при ускоряющем напряжении в десятки и сотни вольт. Поэтому, даже не зная детально самого механизма отражения

волн де Бройля, можно ожидать, что оно также интерференционное, и по этой причине условие (18.1) должно выполняться и для волн де Бройля [1].

В случае монохроматических рентгеновских лучей длину волны λ во время опыта сохраняют постоянной. На опыте меняют угол скольжения φ и замечают, при каком значении φ наступает интерференционное отражение. В случае волн де Бройля значительно удобнее во время опыта угол φ сохранять неизменным, добиваясь интерференционного отражения путем изменения ускоряющего напряжения, т. е. длины дебройлевской волны λ. Так и поступили экспериментаторы.

По теории максимумы отражения должны появиться только при тех значениях λ, которые получаются по формуле (18.1) при целых значениях m. Подставляя в эту формулу значение λ из (17.16), получим для нерелятивистских электронов

$$\sqrt{V} = \frac{1{,}226}{2d \sin \varphi}. \tag{18.2}$$

Здесь V выражено в вольтах, а d — в нанометрах. На рис. 31 приведена кривая, полученная в опытах Дэвиссона и Джермера с монокристаллом никеля при $\varphi = 80°$, $d = 0{,}203$ нм. По оси абсцисс отложено

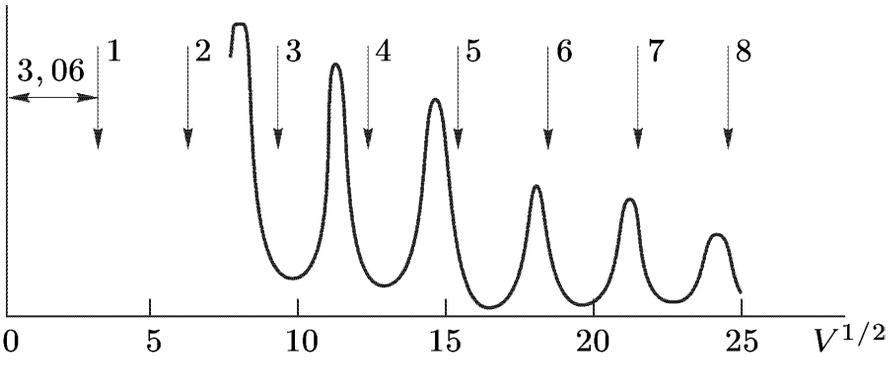


Рис. 31

значение $\sqrt{V}$, а по оси ординат — относительная интенсивность отражения. На том же рисунке стрелками показано положение максимумов, найденное по формуле (18.2). Ожидалось, что максимумы должны быть равноотстоящими, а расстояние между ними должно составлять 3,06 $\text{В}^{1/2}$. Все это подтвердилось на опыте, но только при больших значениях m ($m = 6, 7, 8$). При малых m получились систематические отступления от формулы (18.2).

3. Причина указанного расхождения была выяснена Бете (р. 1906), который показал, что электронным волнам де Бройля в кристалле надо

[1] В сущности, отражение и рассеяние длинных волн (длина волны которых велика по сравнению с межмолекулярными и межатомными расстояниями) также можно рассматривать как интерференционное. Роль атомных плоскостей в этих случаях играют какие-то параллельные и равноотстоящие слои. Но эти слои являются макроскопическими (см. т. IV, § 66, 99), а не моноатомными.

приписать показатель преломления, больший показателя преломления их в вакууме. В самом деле, положительно заряженные ионы кристаллической решетки металла и отрицательные электроны между ними пространственно не совпадают. Поэтому в металле существует электрическое поле, потенциал которого периодически меняется от точки к точке. При грубом рассмотрении его можно заменить постоянным потенциалом V_0, который получается из истинного потенциала путем усреднения его по пространству. Такой усредненный потенциал V_0 называется *внутренним потенциалом* металла. Если потенциал внешнего пространства принять равным нулю, то величина V_0 должна быть положительна, чтобы электроны могли удерживаться внутри металла. Действительно, в этом случае потенциальная энергия электрона внутри металла будет отрицательной — электрон как бы будет находиться в потенциальной яме постоянной глубины V_0, на стенках которой потенциал скачкообразно меняется от нуля снаружи до постоянного значения V_0. Наличием внутреннего потенциала металла и можно объяснить увеличение показателя преломления при переходе из вакуума в металл.

Действительно, пусть наружный электрон падает на металл. Если он прошел ускоряющий потенциал V, то его скорость будет $v_1 \sim \sim \sqrt{V}$. В металле скорость этого электрона возрастет до $v_2 \sim \sqrt{V+V_0}$. Поэтому при входе в металл траектория электрона и связанная с ним волна де Бройля должны испытать преломление. Согласно (17.13) относительный показатель преломления металла для этого процесса будет

$$\mu_{21} = \frac{v_2}{v_1} = \sqrt{1+\frac{V_0}{V}}. \tag{18.3}$$

Индекс 21 ради краткости опустим, т. е. будем пользоваться здесь обозначением $\mu \equiv \mu_{21}$ (не путать с «абсолютным» показателем преломления).

С учетом преломления электронных волн де Бройля условие Брэгга–Вульфа (18.1) следует писать в виде

$$2d\mu\cos\psi = m\lambda \tag{18.4}$$

(см. т. IV, § 33, 61; потеря полуволны при отражении здесь не возникает). Под φ здесь понимается угол преломления, тогда как φ означает угол скольжения (а не падения). Поэтому закон преломления следует писать в виде

$$\frac{\cos\varphi}{\sin\psi} = \mu.$$

Из него находим $\cos\psi = (1/\mu)\sqrt{\mu^2-\cos^2\varphi}$. Таким образом, условие Брэгга–Вульфа принимает вид

$$2d\sqrt{\mu^2-\cos^2\varphi} = m\lambda. \tag{18.5}$$

Правильность приведенного объяснения подтверждается расчетами. На опыте при фиксированном угле φ измеряются значения ускоряющего потенциала V, при которых получаются брэгговские максимумы различных порядков (m =3, 4, 5). По формуле (17.16) вычисляются соответствующие им длины волн де Бройля. Затем по формуле (18.5) находят показатели преломления μ. Наконец, по формуле (18.3) определяют внутренний потенциал V_0 металла. Оказалось, что в пределах ошибок измерений V_0 зависит только от природы металла, но не зависит от порядка отражения m (т. е. от V) и от угла φ. Так и должно быть, если только объяснение, данное выше, правильное. Для никеля, например, $V_0 \approx$ 15 В. Такого же порядка внутренние потенциалы и для других металлов. Впрочем, для объяснения ряда явлений заменять истинный потенциал в металле усредненным недостаточно. Надо принять во внимание, что истинный потенциал периодически меняется в пространстве. На этой основе построена так называемая *динамическая теория интерференции волн де Бройля*, созданная также Бете в 1928 г.

4. Метод, с помощью которого были получены приведенные результаты, мы назвали *методом Брэгга*. Исследование дифракции рентгеновских лучей производится также двумя другими главными методами: *методом Лауэ* (1879–1960) и *методом Дебая–Шерера–Хелла* (см. т. IV, § 61). Такие же методы можно применять и для исследования дифракции волн де Бройля. Дэвиссон и Джермер в 1927 г. применяли метод Лауэ. В методе Лауэ узкий пучок рентгеновских лучей пропускается через монокристалл. Дифрагированные лучи попадают на фотопластинку и фиксируются на ней в виде системы симметрично расположенных пятен. Спектр рентгеновских лучей, как было выяснено в т. IV, § 61, должен быть *обязательно сплошным*. При изучении дифракции электронов по методу Лауэ надо было бы экспериментировать с пучками электронов с непрерывным разбросом скоростей, что практически невозможно. Скорости электронов, выходящих из электронной пушки, одинаковы или, точнее, непрерывно распределены в узких пределах. Условия Лауэ (разумеется, исправленные с учетом показателя преломления электронных волн де Бройля) для пучка электронов определенной скорости, вообще говоря, не выполняются и интерференция не возникает. Поэтому для получения дифракционных максимумов приходится шаг за шагом *непрерывно менять скорость электронов* (путем изменения ускоряющего потенциала, наложенного на пушку) и одновременно положение коллектора.

5. В опыте Дэвиссона и Джермера использовался монокристалл никеля (кристалл кубической системы), сошлифованный вдоль плоскости с миллеровскими индексами (111) (см. т. II, § 135). Перпендикулярно к сошлифованной плоскости в вакууме падал пучок электронов, испускаемый электронной пушкой (рис. 32). Для улавливания рассеянных электронов применялся двойной цилиндр Фарадея с надежной кварцевой изоляцией между наружным и внутренним цилиндрами. Внутренний цилиндр был соединен с чувствительным гальванометром. Чтобы устранить неупруго рассеянные электроны, потерявшие при рассеянии

значительную часть своей энергии, на внешний цилиндр накладывался задерживающий (отрицательный) потенциал, величина которого была в десять раз меньше ускоряющего потенциала. В результате приемника могли достигать только электроны, отразившиеся упруго, а именно такие, у которых скорость при отражении уменьшилась не более чем на 10%. Исследовалось количество рассеянных электронов в зависимости от 1) угла рассеяния; 2) скорости электронов в пучке; 3) азимута, т. е. угла поворота кристалла вокруг оси, перпендикулярной к рассеивающей плоскости.

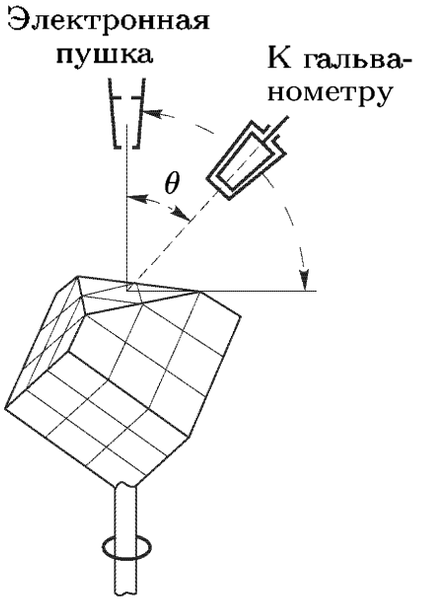


Рис. 32

На рис. 33 приведена полярная диаграмма, определяющая зависимость числа рассеянных электронов от угла рассеяния θ при том положении кристалла, которое соответствует рис. 32. В этом положении сошлифованная поверхность кристалла покрыта правильными рядами атомов, перпендикулярными к плоскости падения (плоскости рисунка), расстояние между которыми равно $d = 0{,}215$ нм. Коллектор мог поворачиваться, оставаясь все время в плоскости падения. Менялся также ускоряющий потенциал. При угле $\theta = 50°$ и различных ускоряющих напряжениях наблюдался максимум, но своего полного развития он достигал при ускоряющем напряжении 54 В ($\lambda = 0{,}167$ нм). Его можно истолковать как дифракционный максимум первого порядка от плоской дифракционной

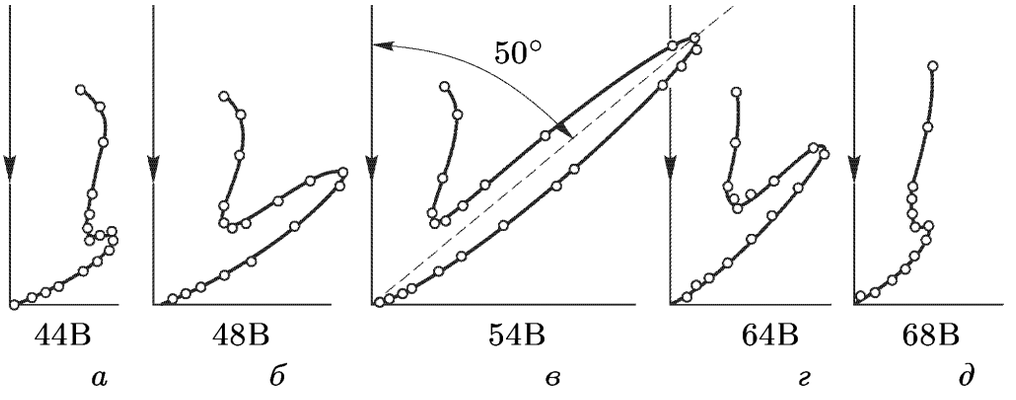


Рис. 33

решетки с периодом $d = 0{,}215$ нм. В самом деле, вычисление угла θ по формуле $d\sin\theta = \lambda$ приводит к результату $\theta = 51°$, что хорошо согласуется с экспериментом.

На опыте производились также измерения числа рассеянных электронов при неизменном положении коллектора и постоянстве угла θ, но при различных азимутах кристалла. Иначе говоря, кристалл мог поворачиваться вокруг оси, перпендикулярной к сошлифованной плоскости. Результаты таких измерений изображены графически на рис. 34

для углов $\theta = 44°$ и $\theta = 50°$ и соответствующих им ускоряющих напряжений. Большая часть максимумов на этих рисунках отвечает пространственной интерференции, т. е. взаимной интерференции волн, рассеянных атомами пространственной решетки кристалла. При повороте кристалла на 360° кривые трижды повторяются. Это связано с тем, что перпендикуляр к отражающей плоскости (111) является по-

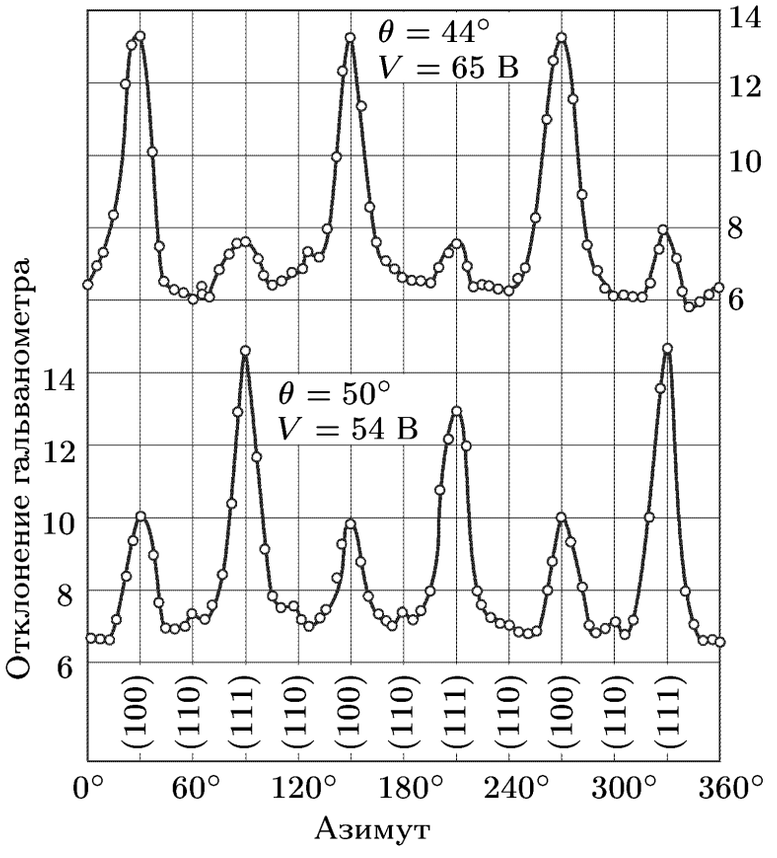


Рис. 34

воротной осью кристалла третьего порядка. Для полного количественного согласия с теорией необходимо было учесть показатель преломления кристалла. Небольшое число максимумов соответствует интерференционному усилению волн, рассеянных атомами, адсорбированными в поверхностном слое кристалла. Этим объясняется, почему на нижней половине рис. 34 из трех больших максимумов средний несколько ниже двух крайних.

6. Исследование дифракции электронов методом Дебая–Шерера–Хелла производилось, начиная с 1928 г., Д. П. Томсоном (1892–1975), П. С. Тартаковским (1895–1939) и некоторыми другими физиками. Томсон пропускал тонкий монохроматический пучок быстрых электронов (ускоряющий потенциал от 17,5 до 56,5 кВ, соответствующие длины волн, вычисленные по релятивистской формуле (17.16а), от 0,092 до 0,052 нм) сквозь поликристаллическую фольгу толщиной $\sim 10^{-5}$ см. (Использование быстрых электронов связано с тем, что более медленные электроны сильно поглощаются фольгой.) На фотопластинке, поставленной за фольгой, получилось центральное пятно, окруженное дифракционными кольцами. На рис. 35 воспроизведена дифракционная картина, полученная описанным методом с золотой фольгой.

Очень простым опытом Томсон показал, что дифракционная картина образовывалась самими рассеянными электронами, а не вторичными рентгеновскими лучами, возбуждаемыми ими: при включении маг-

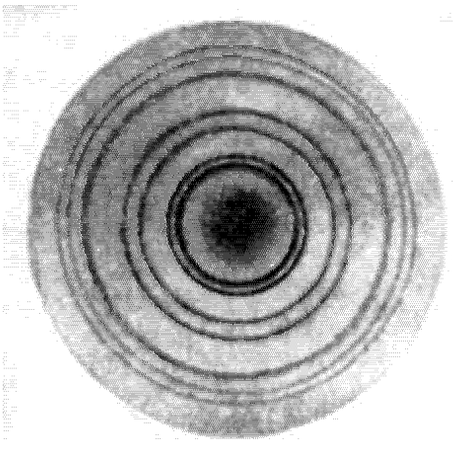


Рис. 35

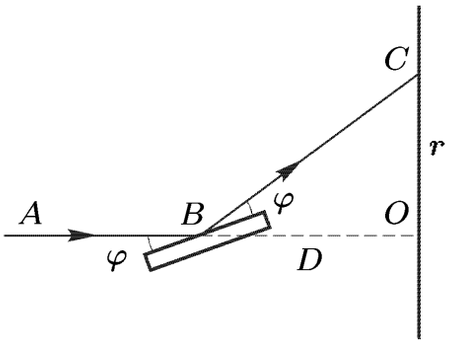


Рис. 36

нитного поля вся дифракционная картина смещалась и искажалась. Этого, конечно, не должно было бы быть, если бы картина возникала в результате дифракции рентгеновских лучей.

Происхождение дифракционных колец в случае электронной дифракции — такое же, как и в случае дифракции рентгеновских лучей (см. т. IV, § 61). Поликристаллическая фольга, пронизываемая электронным лучом, состоит из множества мельчайших ($\sim 10^{-6}$ см) беспорядочно ориентированных кристалликов. Как уже указывалось выше, электронный луч в методе Дебая–Шерера–Хелла должен быть монохроматическим. Но при фиксированной длине волны λ среди множества кристалликов найдутся такие, при отражении от которых выполняется условие Брэгга–Вульфа $2d\sin\varphi = m\lambda$ (рис. 36), где φ — угол скольжения, а d — межплоскостное расстояние. (Ради простоты мы отвлеклись от усложнения, связанного с преломлением электронных лучей. Это обстоятельство легко учесть.) Статистически совокупность таких кристалликов обладает симметрией вращения вокруг направления падающего луча ABO. Поэтому точки C на фотопластинке, куда попадают соответствующие лучи и вызывают их почернение, должны располагаться вдоль колец с центром в O. Найдем связь между радиусом кольца r и длиной волны λ. При малых углах скольжения

$$\sin\varphi \approx \varphi = \frac{m\lambda}{2d} = \frac{r}{2D},$$

где D — расстояние от фольги до фотопластинки. Таким образом, должно выполняться соотношение

$$\frac{r}{\lambda} = \text{const}. \tag{18.6}$$

Оно будет справедливо и при учете преломления. Если воспользоваться формулой (17.16) (для медленных электронов) или (17.16а) (для быст-

рых электронов), то (18.6) преобразуется в соотношения между r и V. Такие соотношения подтвердились на опыте.

Другое количественное доказательство дифракции электронов в кристаллах получится, если сравнить электронограмму кристалла с рентгенограммой того же кристалла. С их помощью можно вычислить постоянную кристаллической решетки. Оказалось, что эти различные методы приводят к одинаковым значениям.

Аналогичные исследования дифракции электронов по методу Дебая–Шерера–Хелла производились П. С. Тартаковским. Он пользовался менее быстрыми электронами (ускоряющий потенциал до 1700 В, $\lambda > 0{,}297$ нм) и алюминиевой фольгой. Исследования велись при постоянном угле рассеяния. При изменении скорости электронов наблюдался ряд максимумов, для которых выполнялось условие Брэгга–Вульфа.

7. Дифракция электронов наблюдалась также на отдельных атомах и молекулах. Не останавливаясь подробно на этом вопросе, ограничимся замечанием, что еще в 1920 г. Рамзауэр (1879–1955) открыл явление, названное его именем. Оно состоит в аномальном увеличении проницаемости некоторых газов для весьма медленных электронов. Иными словами, полное эффективное сечение атома (см. § 9, а также т. II, § 87) по отношению к проходящему через газ пучку электронов становится аномально малым при уменьшении их скорости. Явление особенно отчетливо выражено для благородных газов: аргона, криптона и ксенона. С классической точки зрения (см. § 9) эффективное сечение атома должно было бы монотонно падать с увеличением скорости электронов. Напротив, опыт показывает, что для аргона, криптона и ксенона оно имеет глубокий минимум при ускоряющем напряжении $\approx 0{,}6$ В. По обе стороны от указанного минимума эффективное сечение резко возрастает, достигая максимума при 6 (ксенон), 11 (криптон) и 13 В (аргон). Это объясняется интерференционными ослаблениями и усилениями волн де Бройля при прохождении их через электрическое поле внутри атома. Грубо говоря, влияние такого поля сводится к действию неоднородной среды, показатель преломления которой внутри атома непрерывно меняется от точки к точке.

8. До сих пор говорилось о дифракции электронов. Атомы и молекулы также претерпевают дифракцию. Как показал Штерн (1888–1969), для тяжелых атомов, когда длина волны де Бройля $\lambda = h/mv$ очень мала, дифракционная картина либо совсем не получается, либо получается весьма расплывчатой. Но для легких атомов и молекул (H_2, He) наблюдаются весьма четкие дифракционные картины, с большой точностью подчиняющиеся формуле $\lambda = h/mv$. Как видно из формул (17.18) и (17.19), для молекул водорода и атомов гелия средняя длина волны при комнатной температуре порядка 0,1 нм, т. е. того же порядка, что и постоянная кристаллической решетки. Однако эти частицы обладают тепловым разбросом скоростей. Для получения четкой дифракционной картины их предварительно необходимо монохроматизировать. Подходящий селектор скоростей молекул схематически изображен на рис. 37.

Это два диска, насаженные на общую ось, с одинаковыми, но повернутыми одна относительно другой щелями. Диски могут приводиться во вращение вокруг этой оси. На диски направляется пучок атомов или молекул параллельно оси вращения. Селектор в принципе работает так же, как зубчатое колесо в известном методе Физо (1819–1896) измерения скорости света. В отношении дифракционных картин, получаемых с атомарными и молекулярными пучками, заметим, что эти пучки не могут проникать в глубь кристаллов. Кристаллы действуют на них как плоские двумерные отражательные решетки. На рис. 38 приведена картина дифракции атомов гелия от кристалла фтористого лития, т. е. угловое распределение дифрагированных атомов.

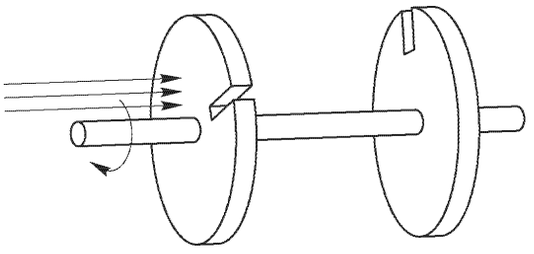


Рис. 37

9. Остановимся еще кратко на *дифракции нейтронов*. Нейтрон — нейтральная элементарная частица, масса которой почти точно равна массе протона. Благодаря отсутствию электрического заряда он взаимодействует только с ядрами атомов — посредством ядерных сил. На электроны атомных оболочек, а следовательно и на фотопластинку,

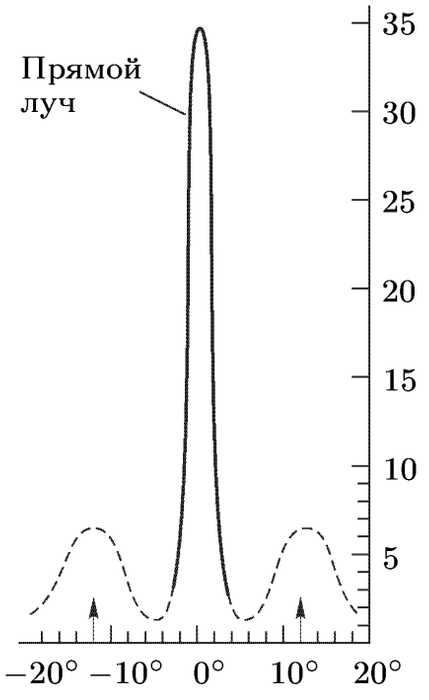


Рис. 38

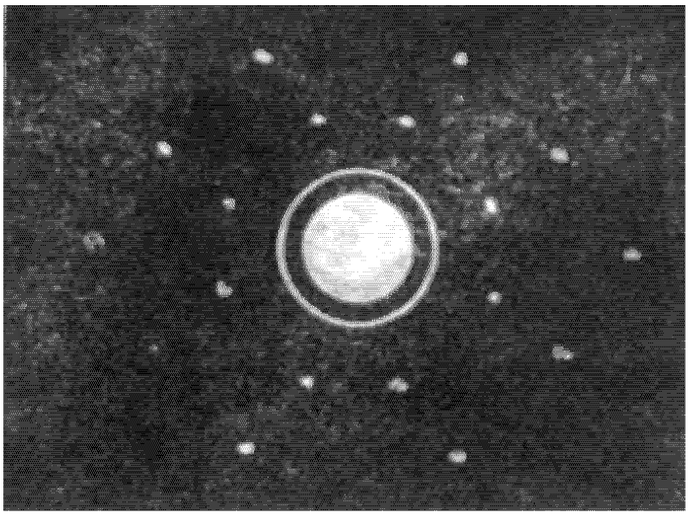


Рис. 39

нейтрон не действует. Чтобы тем не менее применить фотопластинку, фотослой покрывают индиевой фольгой, в которой в местах попадания нейтронов происходят ядерные реакции с выделением электронов и γ-квантов, действующих на фотоэмульсию. Для получения дифракционной картины (нейтронограммы) по методу Лауэ «белый» пучок тепловых нейтронов от ядерного реактора направляют на крупный монокристалл, в котором и дифрагируют нейтроны.

На рис. 39 приведена нейтронограмма, полученная при прохождении пучка нейтронов через монокристалл NaCl. Помимо центрального пятна, получилась система симметрично расположенных пятен, соответствующая поворотной оси четвертого порядка кристалла NaCl. Привлекает внимание изолированность пятен. Это, как было выяснено в т. IV, § 61, связано с тем, что фотопластинка поставлена не в фраунгоферовой области дифракции, а близко от кристалла — *в области применимости геометрической оптики*, — но все же достаточно далеко, чтобы отдельные пучки, на которые разделяется в кристалле падающая волна, уже успели разойтись.

Резюмируя, можно сказать, что волновые свойства частиц не только доказаны экспериментально, но и получили обширные научно-технические применения (электронография, нейтронография и пр.).

ЗАДАЧА

На селектор скоростей (рис. 37) из печи направлен пучок атомов или молекул, скорости которых в печи распределены по закону Максвелла. Селектор выделяет пучок частиц, абсолютные скорости которых заключены в интервале $v, v + dv$ постоянной ширины dv. При каком значении v (выделяемом скоростью вращения дисков) через селектор пройдет максимальное число частиц?

Ответ. $v^2 = (3/2)kT/m$, т. е. v должно быть средней квадратичной скоростью.

§ 19. Статистическая интерпретация волн де Бройля и волновой функции

1. Каков же физический смысл волн де Бройля и какова их связь с частицами вещества?

Одна из идей, которой некоторое время придерживался Шредингер, а затем быстро отказался от нее, состоит в следующем. Никакого дуализма волн и частиц в действительности не существует. Существуют только волны. Частицы же представляют собой *суперпозиции волн*. Дело в том, что в силу математической теоремы Фурье (1768–1830) из волн различных частот и направлений всегда можно составить *волновой пакет*, т. е. такое волновое образование, что при наложении в определенный момент времени волны будут усиливать друг друга в какой-то малой области пространства, а вне этой области произойдет их полное гашение. *Такой волновой пакет и есть частица.* Интенсивность волн де Бройля рассматривается как величина, пропорциональная плотности среды, из которой образуется частица. Казалось, что подтверждением такой точки зрения служит то обстоятельство, что центр волнового пакета, подобно центру группы волн, должен в вакууме распространяться с групповой скоростью. А согласно формуле (17.8) групповая скорость волн де Бройля равна скорости частицы.

Однако волновой пакет не может вести себя как частица сколь угодно длительное время. Причина этого в том, что даже в вакууме волны де Бройля обладают *дисперсией*. Действительно, связь между

импульсом частицы p и ее энергией $\mathscr{E}$ выражается формулой

$$\left(\frac{\mathscr{E}}{c}\right)^2 - p^2 = (m_0 c)^2. \tag{19.1}$$

Подставив в нее значения $\mathscr{E}$ и p из формул (17.2) и (17.3), получим закон дисперсии волн де Бройля

$$\left(\frac{\hbar\omega}{c}\right)^2 - (\hbar k)^2 = (m_0 c)^2. \tag{19.2}$$

Из него видно, что фазовая скорость волн де Бройля $v_\Phi = \omega/k$ зависит от частоты ω, — в этом и состоит дисперсия. Допустим теперь, что в какой-то момент времени, скажем $t = 0$, волны де Бройля усиливают друг друга в некоторой малой области пространства, а в остальном пространстве волновое поле обращается в нуль. Как будет эволюционировать такой волновой пакет во времени? Монохроматические волны разных частот, из которых образован пакет, будут расходиться с различными фазовыми скоростями. Это приведет к деформации, расплыванию и в конце концов к *распаду* первоначального волнового пакета. Таким образом, частица, если бы она представляла волновое образование, была бы *неустойчива* и быстро распадалась бы. Это ни в какой мере не соответствует действительности. Таким образом, *частица не может быть волновым пакетом, образованным из волн де Бройля.*

2. Нельзя принять и противоположную точку зрения: первичными являются *частицы*, а волны представляют их *образования*, т. е. возникают в среде, состоящей из частиц, подобно звуку, распространяющемуся в воздухе. Действительно, такая среда должна быть достаточно плотной, ибо о волнах в среде частиц имеет смысл говорить лишь тогда, когда среднее расстояние между частицами очень мало по сравнению с длиной волны. А в типичных случаях, как мы видели, для волн де Бройля это условие не выполняется.

Но если бы даже мы преодолели это затруднение, то все же указанная точка зрения должна быть отвергнута. В самом деле, она означает, что волновые свойства присущи системам многих частиц, а не отдельным частицам. Между тем волновые, интерференционные свойства частиц не исчезают и при малых интенсивностях падающих пучков. Это было доказано прямыми опытами Бибермана, Сушкина и Фабриканта (р. 1907) с электронами и Яноши (1912–1978) с фотонами.

В опытах Бибермана, Сушкина и Фабриканта применялись столь слабые пучки электронов, что средний промежуток времени между двумя последовательными прохождениями электрона через дифракционную систему был примерно в 30 000 раз больше времени, затрачиваемого одним электроном на прохождение всего прибора. При таких условиях взаимодействие между электронами, конечно, не играло никакой роли. Между тем при достаточно длительной экспозиции возникала дифракционная картина, ничем не отличающаяся от картины,

получаемой при короткой экспозиции с пучками электронов, интенсивность которых была примерно в 10^7 раз больше. Важно только, что в обоих случаях общее число электронов, попавших на фотопластинку, было одинаково. Это показывает, что и *отдельные частицы обладают волновыми свойствами.* Косвенным доказательством этого является также то, что волновые свойства проявляют электроны атомных оболочек, например единственный электрон атома водорода, когда о среде, образованной электронами, говорить не приходится.

3. Приведенные две попытки физического толкования волн де Бройля являются чисто гипотетическими. В свете экспериментальных фактов обе они оказались несостоятельными. Мы не намерены в дальнейшем идти по пути гипотез, а предпочитаем более надежный метод — *метод принципов.* Допустим, что пучок частиц (для определенности будем говорить об электронах) падает на какое-либо дифракционное устройство, например кристалл. Как выяснено выше, дифракция свойственна и волнам де Бройля, сопутствующим только одной частице. Поэтому можно предположить, что падающий пучок состоит всего из одного электрона. При прохождении соответствующей электронной волны де Бройля через кристалл она разбивается на несколько дифракционных пучков. Нельзя допустить, что в каждом из таких пучков находится какая-то доля электрона. Электрон действует всегда как целое и никогда не обнаруживается часть электрона, — в этом проявляется атомизм, свойственный микромиру. Допустим, что на пути одного из дифрагированных пучков поставлен счетчик для улавливания электронов. Если счетчик срабатывает, то он всегда обнаруживает *целый электрон,* а отнюдь не его часть. Из этого нельзя заключить, что до обнаружения электрон находился только в одном рассматриваемом пучке, а потому все остальные дифрагированные пучки никакой роли не играли — их просто не существовало. Такая точка зрения означала бы, что электрон проходит через экспериментальное устройство как материальная точка классической механики. *Это несовместимо с явлениями интерференции и дифракции электронов.* Если повторить тот же опыт с другим электроном, то электрон обнаружится также в одном из дифрагированных пучков, но, вообще говоря, не в том же самом. Подобные трудности вынудили Борна предложить *статистическую интерпретацию* волн де Бройля, позволяющую сочетать атомизм частиц с их волновыми свойствами.

4. Согласно статистической интерпретации волны де Бройля следует рассматривать как *волны вероятности.* Более определенно: *интенсивность волны де Бройля в каком-либо месте пространства пропорциональна вероятности обнаружить частицу в этом месте.* Но статистические или вероятностные свойства частиц могут быть установлены на опыте не с одной частицей, а лишь со *многими частицами* или только с одной частицей, если опыт при определенных условиях *повторен многократно.* Говорить о статистике и вероятности имеет смысл лишь по отношению к определенной совокупности элементов, к которым эти понятия относятся. Это может быть либо совокупность

многих одновременно наблюдаемых элементов, либо один элемент, многократно наблюдаемый в последовательные моменты времени. Такие совокупности элементов в квантовой механике называются *квантовыми ансамблями. Квантовый ансамбль*, и в этом состоит одно из основных положений квантовой механики, *реализуется путем задания каких-то макроскопических параметров*. Это, конечно, не означает, что волновые свойства присущи ансамблям частиц, а не самим частицам. Ансамбли необходимы только для *обнаружения* таких свойств.

Статистическая интерпретация явлений появилась впервые в статистической физике. Однако в основе классической статистической физики лежат динамические законы, допускающие в принципе возможность точного предсказания поведения во времени каждой индивидуальной микрочастицы системы. Для этого нужно только знать начальные условия для всех микрочастиц и силы взаимодействия между ними. Так, поведение газа в состоянии покоя макроскопически определяется его температурой, давлением и объемом. Но это поведение можно было бы в принципе описать положениями и скоростями каждой молекулы в каждый момент времени. По отношению к макроскопическому описанию положения и скорости молекул газа играют роль *скрытых параметров*, т. е. величин, не проявляющихся макроскопически. Можно было бы думать, что и в основе статистического описания, даваемого квантовой механикой, лежат также какие-то неизвестные динамические законы, управляющие поведением неизвестных скрытых параметров, более тонко и детально определяющих состояния систем, чем это делается на современном уровне. Такая точка зрения действительно существует. Но она пока что не привела ни к каким положительным результатам. А потому вопрос о скрытых параметрах мы оставим в стороне и ограничимся статистическим описанием, которое дает современная квантовая механика. Существенно заметить, что такое описание достаточно для предсказания и интерпретации экспериментальных результатов, по крайней мере на современном уровне.

5. Как со статистической точки зрения объясняется дифракция частиц, например электронов? Перед попаданием на дифракционное устройство электроны проходят определенную ускоряющую разность потенциалов, которой соответствует одно и то же значение длины волны де Бройля. Ускоряющий потенциал и является тем макроскопическим параметром, который выделяет квантовый ансамбль частиц. Пусть регистрация электронов производится фотопластинкой. В какое место фотопластинки попадет индивидуальный электрон, с достоверностью предсказать невозможно; это можно сделать только с той или иной степенью вероятности. Вероятность попадания электрона в то или иное место фотопластинки пропорциональна интенсивности волны де Бройля в этом месте. Отдельный электрон оставляет на фотопластинке (после ее проявления) пятнышко. Если электронов мало, то фотопластинка будет напоминать мишень, простреленную небольшим количеством пуль. В расположении пятнышек на фотопластинке не обнаружится никакой закономерности. Закономерность выявится статистически, когда на пластинку попадет очень много электронов.

В этом случае они преимущественно попадут в те места фотопластинки, где должны получиться дифракционные максимумы волн де Бройля. Совокупность соответствующих пятнышек и есть дифракционная картина, получающаяся на опыте. Пока электростатическое отталкивание между электронами несущественно, дифракционная картина получится одной и той же независимо от того, образуется ли она электронами, последовательно проходящими по одному через прибор, или сразу интенсивным пучком одинаково ускоренных электронов, в котором содержится то же число частиц.

6. Символический характер волн де Бройля проявляется, между прочим, в том, что функция (17.1), представляющая плоскую волну де Бройля $\Psi(\mathbf{r}, t)$, *существенно комплексна.* С комплексными выражениями такого типа мы имели дело и в классической физике, но тогда физический смысл мы связывали только с вещественными частями этих выражений. Последние можно было и не дополнять соответствующими мнимыми частями. Это делалось только в целях сокращения математических преобразований и удобства интерпретации окончательных результатов. Не так обстоит дело в квантовой механике. Здесь функция $\Psi(\mathbf{r}, t)$ *принципиально комплексна.* Все физические величины, имеющие реальный физический смысл, выражаются через *всю* комплексную функцию $\Psi(\mathbf{r}, t)$, а не только через ее вещественную часть. Конечно, можно было бы совсем избежать применения комплексных функций. Но тогда для изображения волны де Бройля потребовалась бы не одна, а *две* вещественные функции. Одна из них представляла бы вещественную часть, а другая — коэффициент при мнимой части комплексной функции Ψ. Но это только усложнило бы дело.

Теперь вероятность обнаружить частицу в каком-либо месте пространства можно представить квадратом модуля функции Ψ в том же месте, т. е.

$$|\Psi|^2 = \Psi^*\Psi.$$

В случае плоской волны де Бройля (17.1) этот квадрат модуля равен

$$\Psi^*(\mathbf{r}, t)\Psi(\mathbf{r}, t) = \Psi_0^*\Psi_0 = \text{const},$$

т. е. равновероятно обнаружить частицу в любом месте пространства. Всякий другой результат для равномерно движущейся частицы в течение бесконечного времени несовместим с однородностью пространства. Но он не получился бы, если бы вместо комплексного выражения (17.1) для волны де Бройля взять вещественное, например sin или cos. Это может служить одним из оправданий использования комплексных выражений вместо вещественных. Но будет ли наблюдаться интерференция волн, если каждая из них существенно комплексна? Легко видеть, что да. Действительно, пусть две волны де Бройля представляются выражениями

$$\Psi_1 = e^{i(\mathbf{kr}-\omega t)}, \quad \Psi_2 = e^{i(\mathbf{kr}-\omega t-\delta(r))},$$

т. е. между обеими волнами существует разность фаз $\delta(r)$. При их наложении получается волна $\Psi = \Psi_1 + \Psi_2$. Вероятность обнаружить

частицу в каком-либо месте пространства будет пропорциональна

$$\Psi^*\Psi = (\Psi_1^* + \Psi_2^*)(\Psi_1 + \Psi_2) = (\Psi_1^*\Psi_1 + \Psi_2^*\Psi_2) + \\ + (\Psi_2^*\Psi_1 + \Psi_1^*\Psi_2) = 2 + (e^{i\delta} + e^{-i\delta}) = 2(1 + \cos\delta).$$

Она содержит интерференционный член $2\cos\delta(\mathbf{r})$, меняющийся в пространстве от -2 до $+2$, в зависимости от разности фаз $\delta(\mathbf{r})$. Таким образом, интерференция будет наблюдаться.

7. До сих пор говорилось только о плоских волнах де Бройля, представляемых выражением (17.1). Такие волны сопровождают свободное равномерное движение частиц. Необходимо теперь обобщить полученные результаты на случай *произвольных движений частицы в произвольных силовых полях*. В этих случаях полное описание состояния частицы в квантовой механике дается не плоской волной де Бройля, а какой-то более сложной комплексной функцией $\Psi(\mathbf{r}, t)$, зависящей от координат и времени. Она называется *волновой функцией*. В частном случае свободного движения частицы волновая функция переходит в плоскую волну де Бройля (17.1). Сама по себе волновая функция вводится как некоторый *вспомогательный символ* и не относится к числу непосредственно наблюдаемых величин. Но ее знание позволяет *статистически предсказывать* значения величин, которые получаются экспериментально и потому имеют реальный физический смысл. Эта связь волновой функции с тем, что наблюдается в действительности, будет раскрываться постепенно по мере дальнейшего изложения. Сейчас же мы ограничимся только одной стороной этого вопроса.

Через волновую функцию определяется *относительная вероятность* обнаружения частицы в различных местах пространства. На этой стадии, когда говорится только об отношениях вероятностей, волновая функция принципиально определена с точностью до произвольного постоянного множителя. Если во всех точках пространства волновую функцию умножить на одно и то же постоянное (вообще говоря, комплексное) число, отличное от нуля, то получится новая волновая функция, описывающая в точности *то же состояние*. Не имеет смысла говорить, что Ψ равна нулю во всех точках пространства, ибо такая «волновая функция» никогда не позволяет заключить об относительной вероятности обнаружения частицы в различных местах пространства. Но неопределенность в определении Ψ можно значительно сузить, если *от относительной вероятности перейти к абсолютной*. Распорядимся неопределенным множителем в функции Ψ так, чтобы величина $|\Psi|^2 dV$ давала *абсолютную вероятность* обнаружения частицы в элементе объема пространства dV. Тогда $|\Psi|^2 = \Psi^*\Psi$ будет иметь смысл *плотности вероятности*, которую следует ожидать при попытке обнаружения частицы в пространстве. При этом Ψ будет определена все еще с точностью *до произвольного постоянного комплексного множителя, модуль которого, однако, равен единице*. При таком определении должно быть выполнено *условие нормировки*

$$\int |\Psi|^2 \, dV = 1, \tag{19.3}$$

где интеграл берется по всему бесконечному пространству. Оно означает, что во всем пространстве частица будет обнаружена с достоверностью.

Нормировка (19.3) может оказаться невозможной, если интеграл (19.3) *расходится*. Так будет, например, в случае плоской волны де Бройля, когда вероятность обнаружения частицы одинакова во всех точках пространства. Но такие случаи следует рассматривать как идеализации реальной ситуации, в которой частица не уходит на бесконечность, а вынуждена находиться в ограниченной области пространства. Тогда нормировка (19.3) не вызывает затруднений. Но предельным переходом от нее можно получить рациональную нормировку и в случае частицы, не локализованной в конечной области пространства (см. § 30).

8. Итак, непосредственный физический смысл связывается не с самой функцией Ψ, а с ее *модулем* $\Psi^*\Psi$ (или вообще с какими-то выражениями, билинейными по Ψ^* и Ψ). Почему же в квантовой теории оперируют с волновыми функциями Ψ, а не непосредственно с экспериментально наблюдаемыми величинами $\Psi^*\Psi$? Это необходимо для истолкования волновых свойств вещества — интерференции и дифракции. Здесь дело обстоит совершенно так же, как во всякой волновой теории. Она (во всяком случае в линейном приближении) принимает справедливость *принципа суперпозиции самих волновых полей*, а не их интенсивностей и, таким образом, достигает включения в теорию явлений интерференции и дифракции волн. Так и в квантовой механике принимается в качестве одного из основных постулатов *принцип суперпозиции волновых функций*, заключающийся в следующем. *Если* $\Psi_1(\mathbf{r},t)$ *и* $\Psi_2(\mathbf{r},t)$ *— волновые функции, описывающие какие-то два состояния частицы, то всякая их линейная комбинация с постоянными коэффициентами* $c_1\Psi_1 + c_2\Psi_2$ *представляет также волновую функцию той же частицы, описывающую какое-то состояние ее.* Найдя Ψ указанным путем, можно в дальнейшем определить и плотность вероятности $\Psi^*\Psi$ в состоянии Ψ.

Оправданием такого принципа суперпозиции является согласие с опытом вытекающих из него следствий. Является ли принцип суперпозиции точным законом природы, или он верен только в линейном приближении, этот вопрос не может считаться выясненным. Мы будем строить дальнейшую теорию в предположении, что принцип суперпозиции выполняется точно.

Подчеркнем особо, хотя это и содержится в изложенном выше, что физический смысл волновой функции Ψ связан не только с ее модулем, но и с ее *фазой*, определяемой мнимой частью этой функции. Если бы речь шла о волновой функции только одного состояния, то можно было бы ограничиться одним только модулем. Но если речь идет о наложении состояний, то происходит их интерференция, а она определяется относительной разностью фаз волновых функций, описывающих эти состояния.

9. В связи с изложенным становится понятным, почему при рассмотрении обычных волн, например электромагнитных, инвариантность

фазы волны $\omega t - \mathbf{kr}$ нами *доказывалась* (см. т. IV, § 107), тогда как в случае волн де Бройля *это не делалось*, а частота ω специально подбиралась так, чтобы указанная инвариантность была обеспечена (см. § 17, п. 2). Дело в том, что доказательство, приведенное в т. IV (§ 107), предполагало *возможность счета волн*, проходивших мимо наблюдателя. *Для волн де Бройля такая возможность исключена*, ибо наблюдатель может следить не за самой волной Ψ, а только за плотностью вероятности $\Psi^*\Psi$. Поэтому указанное доказательство к волнам де Бройля неприменимо.

10. Частота волны де Бройля ω и вообще частота волновой функции относятся к *принципиально ненаблюдаемым величинам*. Этим можно воспользоваться, чтобы перейти к квантовой механике *в нерелятивистской форме*. И в классической механике обширная область явлений охватывается в нерелятивистском приближении. То же может быть сделано и в квантовой механике. К тому же здесь переход к релятивистскому рассмотрению осложняется следующим обстоятельством. В сильных полях, когда энергия поля (например, γ-кванта) превосходит $2m_e c^2$, начинается *рождение пар электрон–позитрон*. То же наблюдается в аналогичных случаях и для других частиц. По этой причине *последовательная релятивистская квантовая механика не может быть теорией одного тела* (одной частицы). *Теория одного тела возможна только в нерелятивистском приближении*. Поэтому в дальнейшем мы ограничимся только нерелятивистской квантовой механикой.

В нерелятивистской квантовой механике мы будем по-прежнему пользоваться соотношениями

$$\mathscr{E} = \hbar\omega, \tag{19.4}$$

$$\mathbf{p} = \hbar\mathbf{k}. \tag{19.5}$$

Однако собственную энергию частицы $m_0 c^2$ учитывать не будем. Это значит, что, начиная с этого места, мы вводим новую частоту, отличающуюся от прежней частоты на постоянную. Для новой частоты сохраним прежнее обозначение ω. В частности, в случае свободного движения $\mathscr{E} = p^2/2m$ и закон дисперсии записывается в виде

$$\omega = \frac{\hbar}{2m} k^2 \tag{19.6}$$

вместо прежнего соотношения (19.2). Это приводит к новому выражению для фазовой скорости волн де Бройля:

$$v_\text{ф} = \frac{\omega}{k} = \frac{\hbar k}{2m} = \frac{p}{2m} = \frac{v}{2}. \tag{19.7}$$

Однако это не может отразиться на физических выводах теории, так как *фазовая скорость, как и сама частота ω волны де Бройля, относится к числу принципиально ненаблюдаемых величин*. Существенно, что физически наблюдаемые величины — плотность вероятности $\Psi^*\Psi$ и групповая скорость (17.8) — при новом выборе частоты остаются

неизменными (см. § 17, п. 3). Остаются неизменными, как это будет видно из дальнейшего, и все величины, доступные измерению на опыте.

11. Остановимся в заключение на общей характеристике всякой количественной физической теории. Такая характеристика особенно существенна при изучении квантовой механики. Как отмечал Л. И. Мандельштам (1879–1944), всякая теория состоит, вообще говоря, из двух дополняющих друг друга частей. Первая часть — это *математический аппарат теории*, т. е. уравнения между различными математическими символами, входящими в теорию. Вторая часть определяет *связь этих символов с природой*, с реальной действительностью. Без второй части теория иллюзорна — это чистая математика, а не естественная наука. Без первой части вообще нет количественной теории. *Только совокупность обеих частей составляет количественную физическую теорию.*

Классические теории всегда начинались со второй части. Смысл символов, которыми оперировала теория, считался известным заранее или устанавливался и уточнялся в процессе ее построения. Таковы были время, длина, масса, сила, электрический заряд, напряженности электрического и магнитного полей и т. д. Правда, точные научные определения соответствующих понятий иногда бессознательно подменялись наивными представлениями, заимствованными из повседневной жизни. Это приводило к трудностям. Так было, например, с понятиями пространства и времени, критический пересмотр которых привел к теории относительности. Но порядок построения теории всегда оставался прежним — сначала вырабатывались понятия и лишь потом устанавливались соответствующие уравнения. Квантовая механика пошла по другому пути. Сначала были установлены уравнения для каких-то символов (к числу их относится, например, волновая функция), физический смысл которых был совсем не ясен. Лишь потом занялись отысканием связи этих символов с реальной действительностью. Такой путь построения теории хотя и представляется противоестественным, но логически он допустим, если только связь с реальностью полностью установлена. По такому пути приходится идти и нам. Пока что мы связали волновую функцию Ψ с соответствующей плотностью вероятности $\Psi^*\Psi$. Но этим не исчерпывается физическая интерпретация всех понятий и величин, которые пришлось ввести в связи с функцией Ψ.

§ 20. Соотношение неопределенностей

1. В классической механике состояние материальной точки в каждый момент времени характеризуется ее *положением* и *импульсом*. Реальные микрочастицы — электроны, протоны, атомы, молекулы и пр. — более сложные объекты. *Нельзя характеризовать мгновенное состояние микрочастицы точными заданиями ее положения и импульса.* Причина этого в том, что всякая микрочастица проявляет и корпускулярные, и волновые свойства. Нельзя сказать, что в определенной точке пространства длина волны равна λ, если о волновом поле во всех остальных точках пространства ничего не известно. Длина волны есть *характеристика синусоиды*, а синусоида — *бесконечная периодическая*

кривая. Если из нее вырезать малый кусочек и удалить все остальные части, то оставшийся кусочек потеряет самое характерное свойство синусоиды — ее периодичность. Для кусочка, малого по сравнению с λ, понятие длины волны неприменимо. Ясно, что выражения «длина волны в данной точке пространства x равна λ» или «частота волнового процесса в данный момент времени t равна ω» не имеют никакого смысла — величина λ не является функцией x, а величина ω — функцией t.

С другой стороны, если какое-либо волновое образование занимает ограниченную область пространства, то его всегда можно представить синусоидами. Только одной синусоиды для этого недостаточно. Требуется волновой пакет — суперпозиция множества синусоид различных частот, которые усиливались бы в определенном интервале пространства и взаимно гасили бы друг друга вне этого интервала. Если длина волнового пакета равна Δx (ради простоты мы ограничиваемся одним измерением), то волновые числа k, необходимые для его образования, не могут занимать как угодно узкий интервал Δk. Минимальная ширина интервала Δk должна примерно удовлетворять соотношению

$$\Delta x \cdot \Delta k \gtrsim 2\pi \tag{20.1}$$

(см. т. IV, § 30). Это — *чисто волновое соотношение.* Из него следует, что в коротком радиосигнале (малое Δx) всегда представлены с заметной интенсивностью монохроматические волны с различными значениями λ. Такие сигналы будут приниматься приемниками, настроенными на различные волны. Если же требуется принимать сигналы, близкие к монохроматическим (малое $\Delta\lambda$), то они по необходимости должны быть длинными (большое Δx).

Напомним, как из элементарных соображений можно прийти к соотношению (20.1). Рассмотрим множество синусоид одинаковых амплитуд, волновые числа которых k последовательно возрастают от синусоиды к синусоиде на одну и ту же величину. В точке x фазы волн меняются от kx до $(k+\Delta k)x$, т. е. на величину $x\Delta k$. Если $x\Delta k = 2\pi$, то в этой точке все синусоиды взаимно гасят друг друга. Найдем ближайшую точку $x + \Delta x$, в которой будет происходить такое же гашение. В этой точке разность фаз между крайними синусоидами будет

$$(k+\Delta k)(x+\Delta x) - k(x+\Delta x) = x\Delta k + \Delta x \cdot \Delta k = 2\pi + \Delta x \cdot \Delta k.$$

Ближайшее гашение произойдет, когда $\Delta x \cdot \Delta k = 2\pi$. Таким образом, все волновое возмущение разобьется на отрезки одинаковой длины Δx, причем на концах каждого отрезка волновое поле обратится в нуль. Такой результат получился потому, что все синусоиды были взяты с одной и той же амплитудой. Но если воспользоваться синусоидами всевозможных амплитуд, то можно усилить возмущение в пределах только одного отрезка Δx, а вне этого отрезка всюду его погасить. Это следует из математической теоремы Фурье, причем необходимым условием является выполнение соотношения типа (20.1). Именно такой случай мы и будем иметь в виду.

2. Рассмотрим теперь волновой пакет из волн де Бройля, размеры которого и соответствующие пределы волновых чисел удовлетворяют условию (20.1). Согласно статистической интерпретации вероятность обнаружения частицы будет отлична от нуля только в пределах пакета. А чему будет равен импульс частицы? Каждой волне де Бройля с волновым вектором $\mathbf{k}$ соответствует значение импульса $\mathbf{p} = \hbar\mathbf{k}$. Определенного импульса для всего пакета не существует. Существует набор импульсов, заполняющих интервал от $\mathbf{p} = \hbar\mathbf{k}$ до $\mathbf{p} + \Delta\mathbf{p} = \hbar(\mathbf{k} + \Delta\mathbf{k})$. Неизвестно, какой импульс будет обнаружен в волновом пакете при измерении. В лучшем случае можно указать только его вероятность. При измерении импульс будет обнаружен с той или иной вероятностью между $\mathbf{p} = \hbar\mathbf{k}$ и $\mathbf{p} + \Delta\mathbf{p} = \hbar(\mathbf{k} + \Delta\mathbf{k})$. Поэтому, выражая $\mathbf{k}$ через $\mathbf{p}$, соотношение (20.1) можно переписать в виде

$$\Delta x \cdot \Delta p \gtrsim 2\pi\hbar = h. \tag{20.2}$$

Это соотношение называется *соотношением* или *принципом неопределенностей Гейзенберга для координаты и импульса частицы.* Оно определяет допустимый *принципиальный предел неточностей* δx и Δp, с которыми состояние частицы можно характеризовать классически, т. е. координатой x и импульсом p. Чем точнее x, тем с меньшей точностью возможно характеризовать p, и наоборот. Но соотношение Гейзенберга никоим образом нельзя толковать в том смысле, что частица в каждый момент времени имеет определенные значения x и p, но мы их принципиально не можем узнать с большей точностью, чем это позволяет соотношение неопределенностей (20.2). Такая агностическая точка зрения существовала, но она совсем не соответствует природе изучаемых микрообъектов. Истинный смысл соотношения (20.2) отражает тот факт, что в природе *объективно не существует* состояний частиц с точно определенными значениями обеих переменных x и p.

Принцип неопределенностей был сформулирован Гейзенбергом в 1927 г. и явился важным шагом в интерпретации закономерностей микромира и построении квантовой механики.

В частном случае неопределенности Δp может и не быть ($\Delta p = 0$). Так будет, например, в случае плоской монохроматической волны де Бройля. Но тогда, согласно соотношению неопределенностей, $\Delta x = \infty$, т. е. о месте, где будет локализована частица, ничего сказать нельзя. Она может быть с равной вероятностью обнаружена в любой точке пространства. Напротив, когда $\Delta x = 0$, то $\Delta p = \infty$. В этом случае волновая функция стягивается в точку. При локализации частица будет обнаружена в одной определенной точке (например, начале координат), но об импульсе, который будет найден при локализации, можно высказать только вероятностное утверждение. Можно показать, что в этом случае все значения импульса будут равновероятны.

3. В трехмерном случае классически частица характеризуется тремя прямоугольными координатами x, y, z и сопряженными им импульсами p_x, p_y, p_z. В этом случае соотношения неопределенностей

Гейзенберга выражаются тремя неравенствами

$$\Delta x \cdot \Delta p_x \geqslant h, \quad \Delta y \cdot \Delta p_y \geqslant h, \quad \Delta z \cdot \Delta p_z \geqslant h. \tag{20.3}$$

Никаких ограничений на произведения типа $\Delta x \cdot \Delta p_y$, $\Delta y \cdot \Delta p_z$, которые получаются от умножения неопределенностей координат на неопределенности импульсов, сопряженных с другими координатами, соотношения неопределенностей не накладывают. Величины x и p_y, x и p_z одновременно могут иметь и совершенно точные значения.

4. Соотношению (20.2) можно придать точную количественную форму, которую мы приведем без доказательства. В соотношении (20.2) величины Δx и Δp точно не определены. Но любая волновая функция Ψ позволяет определить *среднее значение координаты* $\overline{x}$, а ее спектральный состав — *среднее значение* $\overline{p}$. Из этих величин находятся отклонения от среднего $\Delta x = x - \overline{x}$, $\Delta p = p - \overline{p}$ и средние квадраты этих отклонений $\overline{\Delta x^2}$ и $\overline{\Delta p^2}$. Точное соотношение неопределенностей гласит:

$$\overline{\Delta x^2} \cdot \overline{\Delta p^2} \geqslant \hbar^2/4. \tag{20.4}$$

Как правило, мы не будем пользоваться этим соотношением, так как во всех принципиальных вопросах существенно знать лишь *порядок величины* $\Delta x \cdot \Delta p$, а не ее точное значение.

5. Мы распространим соотношения неопределенностей (20.3) и на случай *макроскопических тел*, хотя в этом случае и не существует никаких опытов, подтверждающих или опровергающих эти соотношения. Единственным оправданием соотношений неопределенностей в этих случаях служит наша уверенность в их универсальности и всеобщности. Поэтому мы поставим вопрос, как сказалось бы соотношение неопределенностей на движении макроскопического тела, если бы это соотношение в рассматриваемом случае было действительно верно? Возьмем маленький шарик с массой $m = 1$ г. Определим положение центра этого шарика с высокой точностью $\Delta x = 10^{-8}$ см, т. е. с точностью до размера атома. Тогда неопределенность импульса шарика будет $\Delta p \sim h/\Delta x \approx 6{,}63 \cdot 10^{-19}$ г $\cdot$ см/с, а неопределенность скорости $\Delta v = \Delta p/m \approx 6{,}63 \cdot 10^{-19}$ см/с. Такая точность недоступна никакому измерению, а потому и отступления от классического движения, вызываемые соотношением неопределенностей, далеко лежат за пределами возможностей эксперимента.

Совсем иначе обстоит дело с движением электрона в атоме. Имеет ли смысл говорить о классическом движении электрона по боровской орбите? Возьмем для определенности атом водорода и первую боровскую орбиту. Чтобы такое движение имело смысл, необходимо, чтобы неопределенность значения радиуса Δr была мала по сравнению с самим радиусом орбиты $r = \hbar^2/(me^2)$. Но в этом случае неопределенность значения радиального импульса будет

$$\Delta p_r \approx \frac{h}{\Delta r} \gg \frac{h}{r} = \frac{2\pi\hbar}{r} = 2\pi p,$$

что превосходит сам импульс электрона $p = \hbar/r$. Аналогичное справедливо и для других боровских орбит, если только квантовое число не очень велико. При таких условиях представление о движении по классическим орбитам теряет смысл. Поэтому квантовая механика при описании движений электронов в атомах отказалась от понятия траектории — этому понятию здесь реально ничего не соответствует. Так же обстоит дело и с другими элементарными частицами при движении в очень малых областях пространства.

6. Соотношение (20.2) проявляется при всякой попытке измерения точного положения или точного импульса частицы. Оказывается, что уточнение положения частицы сказывается на увеличении неточности в значении импульса, и наоборот. Иллюстрируем это на трех примерах.

Первый пример. Пусть движение электрона описывается плоской монохроматической волной де Бройля. Электрон в таком состоянии обладает вполне определенным импульсом, но его координата совершенно не определена. Для определения x-координаты электрона на пути волны перпендикулярно к ее распространению ставится непрозрачный экран со щелью ширины d (рис. 40). Пусть координатная плоскость XY расположена в плоскости экрана, причем ось X направлена перпендикулярно к щели. Если электрон прошел через щель, то в плоскости самой щели координата x будет зафиксирована с точностью $\Delta x \sim d$. Однако в результате дифракции на щели волновая функция электрона Ψ изменится. Она будет иметь максимумы и минимумы. Электрон же может быть обнаружен в любом месте, где $\Psi \neq 0$. Практически все волновое поле будет сосредоточено в пределах центрального максимума нулевого порядка. Его угловая ширина равна 2θ, причем $d\sin\theta = \lambda$. Этот центральный максимум мы приближенно можем принять за все поле, отбрасывая все его остальные части. В этом приближении после прохождения через щель неопределенность Δp_x импульса электрона получится порядка $\Delta p_x = p\sin\theta$. Умножая это выражение на $\Delta x = d$ и принимая во внимание, что $d\sin\theta = \lambda$ и $p = h/\lambda$, получим $\Delta x \cdot \Delta p_x \gtrsim h$, как и должно быть согласно соотношению Гейзенберга. Чем у́же щель, тем точнее будет измерена координата электрона, но зато потеряется точность в значении его импульса.

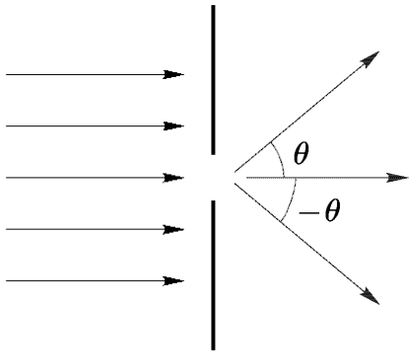


Рис. 40

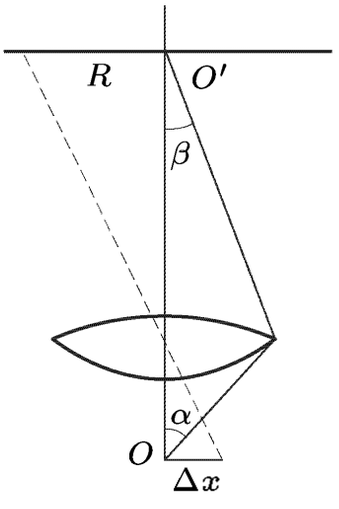


Рис. 41

Сужая щель до размеров $d < \lambda$, можно сделать неопределенность координаты сколь угодно малой. Однако при $d < \lambda$ волновое поле за щелью перестает быть плоской однородной волной де Бройля. Получается неоднородная волна, быстро затухающая на расстоянии порядка λ и меньше. В этом случае

наша оценка Δp неприменима. Однако соотношение неопределенностей, как показывает более точное исследование, остается в силе.

Второй пример (мысленный опыт Гейзенберга с микроскопом). Пусть частица находится под микроскопом (рис. 41). Для определения ее положения она освещается монохроматическим фотоном короткой длины волны λ. По месту попадания фотона на фотопластинку и судят о положении частицы. При освещении пучком света изображением частицы служит дифракционная картина со светлыми и темными кольцами и светлым кружком в центре. Но с одним фотоном дифракционную картину получить нельзя. При рассеянии на электроне фотон с той или иной вероятностью может попасть в любую точку потенциально возможной дифракционной картины, где интенсивность света отлична от нуля. Практически все фотоны могут попасть только в пределы центрального кружка. Интенсивностью остальных колец можно пренебречь. Радиус центрального кружка $R \approx \lambda/\beta$. В этом приближении положение точки попадания фотона в плоскости изображения может быть определено с точностью порядка R. Неточность положения Δx электрона в предметной плоскости найдется из условия синусов Аббе $R\beta = \Delta x \sin\alpha$, т. е. $\lambda = \Delta x \sin\alpha$. Чем меньше λ, тем точнее определяется положение частицы. Но при рассеянии фотона на электроне последний испытывает отдачу, в результате чего импульс электрона получает неконтролируемое приращение $\Delta p_x \sim (h/\lambda)\sin\alpha$. Чем меньше λ, тем больше это неконтролируемое приращение. Таким образом, при одновременном измерении x и p_x мы приходим к соотношению неопределенности

$$\Delta x \cdot \Delta p_x \sim h.$$

Третий пример. Рассмотрим измерение скорости движущегося макроскопического тела по доплеровскому смещению частоты при отражении монохроматического света от этого тела.

Пусть телом служит идеально отражающее плоское зеркало, движущееся в направлении нормали к своей поверхности, а свет (фотон) падает нормально на его поверхность. Допустим, что падающий фотон распространяется в направлении движения зеркала. Тогда отраженный фотон отразится в обратном направлении. На основании законов сохранения энергии и импульса

$$\hbar\omega_0 + \frac{1}{2}mv_0^2 = \hbar\omega + \frac{1}{2}mv^2, \tag{20.5}$$

$$\frac{\hbar\omega_0}{c} + mv_0 = -\frac{\hbar\omega}{c} + mv, \tag{20.6}$$

где m — масса тела, v_0 и v — его скорости до и после отражения фотона, ω_0 и ω — частоты падающего и отраженного фотонов. Переписав эти уравнения в виде

$$m(v^2 - v_0^2) = 2\hbar(\omega_0 - \omega), \tag{20.7}$$

$$m(v - v_0) = \frac{\hbar}{c}(\omega + \omega_0), \tag{20.8}$$

почленным делением находим

$$v + v_0 = 2c\,\frac{\omega_0 - \omega}{\omega_0 + \omega}.$$

Массу зеркала m можно считать бесконечно большой по сравнению с массой фотона. Тогда

$$v = v_0 = c\,\frac{\omega_0 - \omega}{\omega_0 + \omega}. \tag{20.9}$$

Измерив частоты ω_0 и ω, можно по этой формуле вычислить скорость зеркала v. Частоту ω_0 можно считать измеренной точно. Тогда ошибка Δv в определении скорости будет определяться неточностью измерения частоты ω. Чтобы измерить ω с точностью $\Delta\omega$, для измерения требуется минимальное время Δt, удовлетворяющее условию $\Delta\omega \cdot \Delta t \sim$ $\sim 2\pi$. На основании (20.9)

$$\Delta v = -2c\,\frac{\omega_0 \Delta\omega}{(\omega_0 + \omega)^2} \approx -\frac{c\Delta\omega}{2w_0}.$$

Так как моменты отражения фотона известны с ошибкой Δt, то неточность в скорости v поведет к ошибке в определении координаты зеркала

$$\Delta x \sim |\Delta v \cdot \Delta t| \sim \frac{c}{2\omega_0}\,|\Delta\omega \cdot \Delta t| \sim \frac{\pi c}{\omega_0}.$$

Но согласно (20.8) при взаимодействии с фотоном зеркало получает неконтролируемое изменение импульса $\Delta p \sim 2\omega_0\hbar/c$. Следовательно, опять получается соотношение неопределенностей $\Delta x \cdot \Delta p \sim 2\pi\hbar = h$.

7. Приведенные примеры показывают, что измерения в квантовой области принципиально отличаются от классических измерений. Конечно, и те и другие измерения сопровождаются ошибками. Однако классическая физика считала, что путем улучшения методики и техники измерений ошибки в принципе могут быть сделаны сколь угодно малыми. Напротив, *согласно квантовой физике существует принципиальный предел точности измерений. Он лежит в природе вещей и не может быть превзойден никаким совершенствованием приборов и методов измерений.* Соотношения неопределенностей Гейзенберга и устанавливают один из таких пределов. Взаимодействие между макроскопическим измерительным прибором и микрочастицей во время измерения принципиально нельзя сделать сколь угодно малым. Если измеряется, например, координата частицы, то измерение неизбежно приводит к принципиально неустранимому неконтролируемому искажению первоначального состояния частицы, а следовательно, и к неопределенности в значении импульса при последующем измерении. То же самое происходит, если порядок измерения координаты и импульса частицы поменять местами.

8. Отметим некоторые выводы, вытекающие из соотношения неопределенностей (20.2).

Прежде всего видно, что *состояние, в котором частица находится в полном покое, невозможно.* Далее, в макроскопической физике

импульс частицы определяется формулой $p = mv$. Для нахождения скорости v измеряют координаты частицы x_1 и x_2 в два близких момента времени t_1 и t_2. Затем находят частное $(x_2 - x_1)/(t_2 - t_1)$ и выполняют предельный переход $t_2 \to t_1$. Такой метод не годится для измерения мгновенной скорости микрочастицы. Действительно, предельный переход требует точного измерения x_1 и x_2. А точное измерение координаты существенно меняет импульс частицы. Ясно поэтому, что предельным переходом нельзя найти мгновенную скорость ни в одном положении частицы. Можно, правда, промежуток времени $t_2 - t_1$ взять длинным, а измерения x_1 и x_2 произвести с малой точностью. Тогда ошибки измерения мало скажутся на скоростях частицы и на значении дроби $(x_2 - x_1)/(t_2 - t_1)$. Но таким путем будет найдена *не истинная скорость частицы, а ее среднее значение на интервале времени* $t_2 - t_1$. Импульс микрочастицы можно определять по разности потенциалов, пройденной ею в ускоряющем электрическом поле, или по длине волны де Бройля λ, измеренной каким-либо дифракционным устройством.

В квантовой механике теряет смысл деление полной энергии $\mathscr{E}$ *на кинетическую и потенциальную.* Действительно, одна из этих величин зависит от импульсов, а другая от координат. Эти же переменные не могут иметь одновременно определенные значения. Энергия $\mathscr{E}$ должна определяться и измеряться лишь как *полная энергия без деления на кинетическую и потенциальную.* Об этом подробнее будет сказано ниже.

9. В классической теории не было параметра, определяющего размеры атома. Соотношение неопределенностей позволяет установить такой параметр. Рассмотрим для примера водородоподобный атом с зарядом ядра Ze. Рассуждая классически, напишем уравнение сохранения энергии

$$\frac{p^2}{2m} - \frac{Ze^2}{r} = \text{const}.$$

Если предположить, что электрон вначале находился на бесконечности практически в состоянии покоя, то следует положить const $= 0$. Этим определится p^2, а затем

$$p^2 r^2 = 2mZe^2 r.$$

Соотношение неопределенностей возьмем в точной форме (20.4). Так как во всяком случае $\Delta r < r$, $\Delta p < p$, то из него следует $p^2 r^2 > \hbar^2/4$, а потому $2mZe^2 r > \hbar^2/4$, откуда

$$r > \frac{1}{Z}\frac{\hbar^2}{8me^2} = \frac{1}{Z} \cdot 0{,}66 \cdot 10^{-9} \text{ см}. \qquad (20.10)$$

Порядок размера атома эта формула определяет правильно. Числовому коэффициенту не следует придавать особое значение, так как формула (20.10) носит только оценочный характер. В частности, оценка показывает, что в кулоновском поле ядра падение электрона на ядро

невозможно. Более того, нахождение электрона внутри атомного ядра несовместимо с соотношением неопределенностей Гейзенберга.

Если применить такую же оценку к определению размеров атомного ядра, то в формуле (20.10) вместо m надо было бы подставить массу протона. В результате для радиуса ядра получилась бы величина примерно в 2000 раз меньшая, чем (20.10). Но и это слишком много (размер ядра порядка 10^{-13} см). Это показывает, что *для образования ядра кулоновских сил недостаточно.* В ядре должны действовать более мощные — *ядерные* — силы, превосходящие кулоновские примерно на два порядка.

10. Наряду с соотношением (20.1) в волновой теории выводится также соотношение

$$\Delta t \cdot \Delta\omega \gtrsim 2\pi \tag{20.11}$$

(см. т. IV, § 29). Смысл этого соотношения состоит в том, что ограниченный во времени волновой процесс не может быть монохроматическим. Если процесс длится в течение времени Δt, то разброс частот $\Delta\omega$ входящих в него волн в лучшем случае удовлетворяет соотношению (20.11). Поэтому, если для наблюдения даже монохроматического процесса предоставлено малое время Δt, то частота процесса принципиально будет найдена в лучшем случае с ошибкой, подчиняющейся соотношению (20.11).

Если частоте ω сопоставить энергию по формуле $\mathscr{E} = \hbar\omega$, то формула (20.11) перейдет в

$$\Delta t \cdot \Delta\mathscr{E} \geqslant 2\pi\hbar = h. \tag{20.12}$$

Формула (20.12) называется *соотношением неопределенностей Гейзенберга для времени и энергии.*

Соотношение (20.12) означает, что чем короче время существования какого-то состояния или время, отведенное для его наблюдения, тем с меньшей определенностью можно говорить об энергии этого состояния. Наоборот, чем больше это время, тем с большей точностью определена энергия состояния. Если состояние стационарно, то оно может существовать бесконечно долго. Именно по этой причине энергия стационарного состояния имеет вполне определенное значение. Противоположным примером может служить нестабильная элементарная частица, распадающаяся за очень короткое время (скажем, порядка 10^{-20} с). Об определенной энергии такой частицы говорить не приходится. Поэтому при рассмотрении процесса ее распада не требуется налагать условие сохранения энергии.

ЗАДАЧИ

1. На какую кинетическую энергию $\mathscr{E}_{\text{кин}}$ должен быть рассчитан электронный (и протонный) ускоритель для исследования структур с линейными размерами $l \sim 1$ ферми (10^{-13} см)?

Ответ.

$$\mathscr{E}_{\text{кин}} > mc^2\sqrt{1+\left(\frac{\lambda}{l}\right)^2} - mc^2,$$

где $\lambda = h/(mc)$ — комптоновская длина волны электрона (протона). Для электрона

$$\mathscr{E}_{\text{кин}} > m_{\text{e}}c^2\frac{\lambda_{\text{e}}}{l} \approx 720 \text{ МэВ}.$$

Для протона

$$\mathscr{E}_{\text{кин}} > m_{\text{p}}c^2\sqrt{1+\left(\frac{\lambda_{\text{p}}}{l}\right)^2} - m_{\text{p}}c^2 \approx 600 \text{ МэВ}.$$

2. Среднее время жизни атома в возбужденном состоянии составляет около $\Delta t \sim 10^{-8}$ с. При переходе атома в нормальное состояние испускается фотон, средняя длина волны которого равна $\lambda = 500$ нм. Оценить ширину $\Delta\lambda$ и относительную ширину $\Delta\lambda/\lambda$ излучаемой спектральной линии, если не происходит ее уширения за счет других процессов. (Такая ширина называется естественной шириной спектральной линии.)

Ответ.

$$\Delta\lambda \sim \frac{\lambda^2}{c\Delta t} \sim 10^{-4} \text{ нм}, \quad \frac{\Delta\lambda}{\lambda} \sim \frac{\lambda}{c\Delta t} \sim 10^{-7}.$$

3. Принимая во внимание, что классическая (корпускулярная) механика является приближенным предельным случаем механики волновой, найти условие, при выполнении которого рассеяние α-частиц и электронов на атомных ядрах можно рассчитывать классически, как это делал Резерфорд при выводе своей формулы (9.3).

Решение. Обозначим через a радиус атома. Можно считать, что рассеивающаяся частица подвергается действию кулоновских сил ядра только в пределах сферы радиуса a с центром в центре атома. Такая идеализация неприменима только при очень далеких прицельных расстояниях, когда получаются слишком малые углы рассеяния, при которых экспериментальные исследования не производятся. Рассмотрим плоскопараллельный пучок частиц с поперечным сечением Δx, падающий на атом, и выясним, при каких условиях его распространение в атоме можно рассматривать классически, т. е. без учета дифракции частиц (как это делается в приближении геометрической оптики). Дифракционным расширением пучка, как известно из оптики (см. т. IV, § 6), можно пренебречь, если $l \ll \Delta x^2/\lambda$, где l — длина пучка. В рассматриваемой задаче под l следует понимать путь луча в кулоновском поле атома. По порядку величины $l \approx a$, $\Delta x \sim a$. Таким образом, условие применимости приближения геометрической оптики принимает вид $a \ll a^2/\lambda$, или $\lambda \ll a$. Подставив сюда выражение для длины волны де Бройля $\lambda = h/p = h/(2m\mathscr{E})^{1/02}$, находим искомое условие:

$$\mathscr{E} \gg \frac{h^2}{2ma^2}. \tag{20.13}$$

При нашем рассмотрении не учтено действие ядерных сил. Но радиус ядра $a_{\text{яд}}$, где и действуют ядерные силы, примерно в 10^5 раз меньше радиуса атома $a_{\text{ат}}$. Поэтому на ядро падает примерно в $(a_{\text{ат}}/a_{\text{яд}})^2 \approx 10^{10}$ меньше частиц, чем на атом. Точно так же не играют существенной роли и кулоновские силы в непосредственной близости от ядра. Этим наше приближение оправдано.

При численных расчетах формулу (20.13) удобно переписать в виде

$$\mathscr{E} \gg \frac{(hc)^2}{2mc^2a^2}, \tag{20.14}$$

пользуясь числовым значением $hc = 1{,}24 \cdot 10^{-4}$ эВ$\cdot$см. Энергию mc^2 следует брать в эВ, а длину a — в см. Для α-частицы $mc^2 = 3{,}75 \cdot 10^9$ эВ. Полагая $a = 10^{-8}$ см, получаем в этом случае

$$\mathscr{E} \gg 0{,}2 \text{ эВ}.$$

Энергия α-частицы порядка 1 МэВ, так что в случае α-частиц применение классической механики оправдано.

Для электрона $mc^2 = 0{,}511 \cdot 10^6$ эВ, а потому должно быть $\mathscr{E} \gg 120$ эВ. Таким образом, для электронов с кинетической энергией ~ 100 эВ классической механикой пользоваться уже нельзя.

Рассмотренную задачу можно было бы также решить на основе принципа неопределенностей Гейзенберга.

Глава IV

УРАВНЕНИЕ ШРЕДИНГЕРА. КВАНТОВАНИЕ

§ 21. Уравнение Шредингера

1. Плоская волна де Бройля

$$\Psi = Ce^{i(\mathbf{kr}-\omega t)} \tag{21.1}$$

является весьма специальным волновым образованием, соответствующим свободному равномерному движению частицы в определенном направлении и с определенным импульсом. Но частица, даже в свободном пространстве и в особенности в силовых полях, может совершать и другие движения, описываемые более сложными волновыми функциями. Основная задача волновой механики как раз и состоит в нахождении волновых функций и связанных с ними физических следствий в самых разнообразных условиях. Для ее решения служит *волновое уравнение*, найденное Шредингером в 1926 г. Это — *основное уравнение квантовой механики*, но оно справедливо только в *нерелятивистской квантовой механике*, т. е. в случае движений, медленных по сравнению со скоростью света в вакууме.

Уравнение Шредингера должно быть общим уравнением, т. е. должно быть пригодно для решения *всех*, а не только частных задач. Поэтому в него не должны входить значения параметров (например, начальные условия, конкретный вид силовых полей и пр.), выделяющие частные виды движения. В него могут входить мировые постоянные, например постоянная Планка. Могут входить массы и импульсы частиц, но их числовые значения не должны быть конкретизированы. Силовые поля, в которых движется частица, также должны быть представлены в общем виде. Здесь дело обстоит так же, как с уравнениями Ньютона или Максвелла, которые приспособлены для решения всех, а не только частных механических или электродинамических задач. Кроме того, надо потребовать, чтобы уравнение Шредингера было *линейно и однородно* по Ψ. Этим будет обеспечена справедливость принципа суперпозиции волновых функций, необходимость которого диктуется интерференцией и дифракцией волн вещества (см. § 19, п. 8).

2. При отыскании уравнения Шредингера заметим, что одним из решений его в свободном пространстве должна быть плоская волна де Бройля (21.1). Найдем дифференциальное уравнение, удовлетворяющее перечисленным выше условиям, решением которого является эта волна. Дифференцирование (21.1) по x дает

$$\frac{\partial \Psi}{\partial x} = ik_x \Psi, \quad \frac{\partial^2 \Psi}{\partial x^2} = -k_x^2 \Psi.$$

Такие же соотношения получим при дифференцировании по y и z. Сложением полученных вторых производных найдем

$$\nabla^2\Psi = -k^2\Psi = -\frac{p^2}{\hbar^2}\Psi. \tag{21.2}$$

Это — дифференциальное уравнение, но не то, которое мы ищем. Действительно, при выводе величина p предполагалась постоянной, а потому уравнение (21.2) описывает конкретное движение с заданным постоянным импульсом. Продифференцируем теперь (21.1) по времени при постоянной ω:

$$\frac{\partial\Psi}{\partial t} = -i\omega\Psi = -i\frac{\mathscr{E}}{\hbar}\Psi. \tag{21.3}$$

Это уравнение также не годится. Оно описывает движение частицы в свободном пространстве с постоянной кинетической энергией $\mathscr{E}$. Разделим, однако, почленно уравнение (21.2) на уравнение (21.3) и учтем, что в нерелятивистской механике $\mathscr{E} = p^2/2m$. Таким путем придем к однородному линейному уравнению

$$i\hbar\frac{\partial\Psi}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\Psi, \tag{21.4}$$

которое уже не содержит никаких индивидуальных параметров, выделяющих конкретное движение. Примем в качестве постулата, что уравнение (21.4) справедливо для любых движений частицы в свободном пространстве. Это уравнение и есть *уравнение Шредингера в отсутствие силовых полей.*

Обобщим теперь уравнение (21.4) на случай движений в силовых полях. Ограничимся случаем *потенциальных силовых полей*, которые, как и в классической механике, характеризуются *потенциальной функцией* или *потенциальной энергией* $U(\mathbf{r})$. Единственными непотенциальными силами, встречающимися в атомной механике, являются *силы магнитные*, но мы временно отвлечемся от их рассмотрения. Заметим теперь, что $\hbar\partial/\partial t$ имеет размерность энергии. Значит, одинаковую размерность имеют и величины $i\hbar\dfrac{\partial\Psi}{\partial t}$ и $U(\mathbf{r})\Psi$. Поэтому прибавление в правой части уравнения (21.4) слагаемого $U(\mathbf{r})\Psi$ не меняет размерности этого уравнения. Можно думать, что полученное таким путем уравнение

$$i\hbar\frac{\partial\Psi}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\Psi + U(\mathbf{r})\Psi \tag{21.5}$$

будет правильно учитывать влияние потенциального силового поля на движение частицы. Это и есть *уравнение Шредингера.*

Путь, которым мы пришли к уравнению Шредингера, конечно, не может служить доказательством этого уравнения. Но уравнение Шредингера — существенно новый принцип. Его нельзя логически вывести из старых принципов, в которых он не содержится. Единственным доказательством уравнения Шредингера является только опыт — опытная проверка всех выводимых из него следствий. Такую проверку уравнение Шредингера выдержало.

3. В уравнении (21.5) в неявной форме уже заложена двойственная — корпускулярно-волновая — природа вещества. Согласно интерпретации волновой функции Ψ частица не локализована. Она, как принято говорить, с определенной вероятностью «размазана» в пространстве. Казалось бы, что при написании уравнения (21.5) это обстоятельство с самого начала должно быть принято во внимание, т. е. под U следовало бы понимать потенциальную энергию частицы с учетом всех возможных положений ее и их вероятностей. На самом деле в уравнении (21.5) это не предполагается. Потенциальная функция $U(\mathbf{r})$ рассматривается в нем так же, как в классической физике, т. е. как *функция локализованной, в частности точечной, частицы в силовом поле.* Например, в атоме водорода для электрона в поле ядра полагают $U(r) = -e^2/r$, т. е. поступают так же, как если бы обе эти частицы были локализованы.

4. Уравнение Шредингера — уравнение *первого порядка* по времени. Отсюда следует, что заданием волновой функции Ψ во всем пространстве в какой-либо момент времени (например, принимаемый за начальный) однозначно определяется функция Ψ также во всем пространстве во все последующие моменты времени. Не следует смотреть на это утверждение как на выражение принципа причинности в квантовой механике. Ибо выражаемая им «причинность» относится к волновой функции Ψ. А волновая функция связана с реально наблюдаемыми объектами вероятностными соотношениями. Поэтому квантовая механика, по крайней мере в современной ее форме, является *принципиально статистической теорией.*

5. Уравнение Шредингера, как это требовалось с самого начала для выполнения принципа суперпозиции, *линейно и однородно относительно функции* Ψ. В точной математической форме принцип суперпозиции сводится к двум утверждениям. Во-первых, *если* Ψ_1 *и* Ψ_2 *— какие-либо два решения уравнения Шредингера, то и всякая линейная комбинация их* $\alpha_1\Psi_1 + \alpha_2\Psi_2$ *с постоянными* (*вообще говоря, комплексными*) *коэффициентами* α_1 *и* α_2 *есть также решение того же уравнения.* Во-вторых, *если волновые функции* Ψ_1 *и* Ψ_2 *описывают какие-либо два состояния системы, то и линейная комбинация* $\alpha_1\Psi_1 + \alpha_2\Psi_2$ *также описывает какое-то состояние той же системы.* Конечно, состояние частицы определяется не самими коэффициентами α_1 и α_2, а только их отношением α_1/α_2. Состояние не изменится, если оба коэффициента умножить на одну и ту же вещественную или комплексную постоянную. Это позволяет, например, функцию $\Psi = \alpha_1\Psi_1 + \alpha_2\Psi_2$ нормировать (если интеграл $\int \Psi^*\Psi dV$, взятый по всему пространству, сходится).

6. Особое значение в квантовой механике имеют *стационарные состояния.* Это — такие состояния, в которых все наблюдаемые физические параметры не меняются с течением времени. Сама волновая функция Ψ не относится к этим параметрам. Она *принципиально ненаблюдаема.* Не должны меняться во времени только *физически наблюдаемые величины,* которые могут быть образованы из Ψ по правилам

квантовой механики. Оказывается, что в стационарных состояниях

$$\Psi(\mathbf{r}, t) = \psi(\mathbf{r}) e^{-i\omega t}, \tag{21.6}$$

где частота ω постоянна, а функция $\psi(\mathbf{r})$ не зависит от времени. Не располагая сейчас правилами составления из Ψ принципиально наблюдаемых величин, проверим, что одна из таких величин, а именно плотность вероятности $\rho = \Psi^*\Psi$, в состоянии (21.6) во времени остается постоянной. Действительно,

$$\rho = \psi^*(\mathbf{r}) e^{i\omega t} \psi(\mathbf{r}) \cdot e^{-i\omega t} = \psi^*(\mathbf{r})\psi(\mathbf{r}),$$

а эта величина от времени действительно не зависит.

Для определения функций $\psi(\mathbf{r})$ в стационарных состояниях подставляем выражение (21.6) в уравнение (21.5) и находим

$$\hbar\omega\psi = \left[-\frac{\hbar^2}{2m}\nabla^2 + U(\mathbf{r})\right]\psi.$$

По аналогии со световыми квантами примем гипотезу, что величина $\hbar\omega$ представляет собой полную энергию частицы $\mathscr{E}$ в стационарном состоянии. Таким образом, для энергии в стационарном состоянии получается уравнение

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + U(\mathbf{r})\right]\psi = \mathscr{E}\psi. \tag{21.7}$$

Это уравнение не содержит времени и называется *уравнением Шредингера для стационарных состояний*. В отличие от него (21.5) называется *временным* или *общим уравнением Шредингера*. В отношении потенциальной функции $U(\mathbf{r})$, входящей в уравнение (21.7), полностью справедливы замечания, которые были сделаны в связи с уравнением (21.5). Функция $U(\mathbf{r})$ определяется классически, как если бы никакими волновыми свойствами частица не обладала.

Уравнение Шредингера для стационарных состояний, конечно, удовлетворяет принципу суперпозиции. Однако суперпозиция стационарных состояний с различными энергиями уже не будет стационарным состоянием.

Шредингер показал (см. § 22), что уравнение (21.7) полностью решает проблему квантования энергии системы. Для этого под $\mathscr{E}$ следует понимать энергию системы в стационарном состоянии, а относительно физического смысла самой волновой функции $\psi(\mathbf{r})$ никаких предположений вводить не требуется. Необходимо только наложить на решения $\psi(\mathbf{r})$ уравнения (21.7) некоторые *естественные условия*, которым они должны удовлетворять на бесконечности и в особых точках потенциальной функции $U(\mathbf{r})$. В следующем параграфе будет показано, что *такие решения существуют, вообще говоря, не при всяких значениях $\mathscr{E}$, а только при некоторых*. Это и есть избранные значения энергии в стационарных состояниях. В частности, для атома водорода получаются в точности те же значения $\mathscr{E}$, которые давала старая теория Бора. Это был первый крупный успех волновой механики, с которого началось ее дальнейшее бурное развитие.

7. Уравнение (21.7) в сочетании с принципом суперпозиции естественно приводит и *к правилу частот Бора.* С этой целью заметим, что всякий физический процесс характеризуется изменениями во времени каких-то реальных физических величин. Но в стационарных состояниях все реальные физические величины остаются постоянными. Поэтому волновая функция, описывающая состояние, в котором происходят реальные физические явления, должна быть обязательно *нестационарной.* Рассмотрим простейшее нестационарное состояние

$$\Psi = \Psi_1 + \Psi_2, \tag{21.8}$$

представляющее собой суперпозицию двух стационарных состояний

$$\Psi_1 = \psi_1(\mathbf{r})e^{-i\omega_1 t}, \quad \Psi_2 = \psi_2(\mathbf{r})e^{-i\omega_2 t}. \tag{21.9}$$

Вычислим в этом состоянии простейшую реально наблюдаемую величину — плотность вероятности ρ. Получим

$$\rho = \Psi^*\Psi = (\psi_1^*\psi_1 + \psi_2^*\psi_2) + \psi_2^*\psi_1 e^{i(\omega_2-\omega_1)t} + \psi_2\psi_1^* e^{-i(\omega_2-\omega_1)t}.$$

Учтем разность фаз, которая может существовать между ψ_1 и ψ_2. Для этого положим $\psi_1 = |\psi_1|e^{-i\delta_1}$, $\psi_2 = |\psi_2|e^{-i\delta_2}$, где δ_1 и δ_2 — величины вещественные. Тогда получим

$$\rho = (|\psi_1|^2 + |\psi_2|^2) + 2|\psi_1| \cdot |\psi_2| \cos[(\omega_2 - \omega_1)t + (\delta_2 - \delta_1)]. \tag{21.10}$$

Такова (ненормированная) плотность вероятности состояния $\Psi = \Psi_1 + \Psi_2$. Она содержит постоянный член $(|\psi_1|^2 + |\psi_2|^2)$ и интерференционный член $2|\psi_1| \cdot |\psi_2| \cos(\omega_{12}t + \delta_2 - \delta_1)$, гармонически колеблющийся с боровской частотой

$$\omega_{12} = \omega_2 - \omega_1 = \frac{\mathscr{E}_2 - \mathscr{E}_1}{\hbar}. \tag{21.11}$$

Полная плотность вероятности ρ может меняться от максимального значения $(|\psi_1| + |\psi_2|)^2$ до минимального $(|\psi_1| - |\psi_2|)^2$. Она содержит член, осциллирующий с боровской частотой ω_{12}. Поэтому приведенное рассуждение в сочетании с классической электродинамикой наводит на мысль (но отнюдь не доказывает), что с той же частотой должно происходить и излучение света. Действительно, если e — заряд частицы, то величина ρe имеет смысл плотности вероятности электрического заряда в пространстве. Если бы она была просто плотностью заряда, а не ее вероятностью, то получился бы классический случай, в котором заряд периодически колеблется во времени. По классическим представлениям такой заряд должен излучать. Правдоподобно ожидать излучения и в квантовом случае, где плотность заряда заменяется ее вероятностью.

Правда, интерференционный член в (21.10) имеет характер незатухающей стоячей волны. Для поддержания непрерывного излучения, если оно уходит от системы, требуется подводить энергию. Но и в классической физике, например при рассмотрении излучения при незатухающих колебаниях диполя Герца, положение такое же. Мы рассчитываем поле и интенсивность излучения, отвлекаясь от того, каким

механизмом поддерживается постоянство амплитуды и связанной с ней энергии колебаний.

Как и в первоначальной теории Бора, боровская частота ω_{12} появляется в результате квантовых переходов системы с одного энергетического уровня на другой.

§ 22. Уравнение Шредингера и квантование

1. Квантование энергии возникает потому, что на волновые функции, являющиеся решениями уравнения Шредингера (21.7), накладываются определенные естественные ограничения. При этих ограничениях уравнение (21.7) имеет решения, вообще говоря, не при всех, а только при избранных значениях параметра $\mathscr{E}$. Здесь дело обстоит аналогично тому, что имеет место в задаче о свободных колебаниях струны с закрепленными концами. Из-за закрепления концов эти колебания представляют собой стоячие волны с такими избранными частотами, что на длине струны укладывается целое число полуволн.

Естественные ограничения, накладываемые на решения уравнения Шредингера (21.7) (в несколько усиленной, но практически всегда выполняющейся форме), состоят в том, что *волновая функция* $\psi(\mathbf{r})$ *и ее первые пространственные производные должны быть конечны, однозначны и непрерывны даже в точках* (*а также линиях и поверхностях*) *разрыва потенциальной функции* $U(\mathbf{r})$. (Однозначность означает, что при обходе по любому замкнутому контуру функция $\psi(\mathbf{r})$ должна возвращаться к своему исходному значению.) Избранные значения параметра $\mathscr{E}$, для которых уравнение (21.7) имеет решения, удовлетворяющие перечисленным ограничениям, называются *собственными значениями величины* $\mathscr{E}$ для дифференциального уравнения (21.7), а соответствующие им решения — *собственными функциями* того же уравнения. *Собственные значения* $\mathscr{E}$ *и принимаются за возможные значения энергии в стационарных состояниях.* Собственные значения энергии $\mathscr{E}$ могут быть *дискретными*, а могут *непрерывно заполнять* конечный или бесконечный интервал. В первом случае говорят, что энергетический спектр *дискретный*, а во втором — *непрерывный.*

Поясним изложенное на частном случае одномерного движения частицы в потенциальном поле сил. Допустим, что частица движется вдоль оси X, а потенциальная функция $U(x)$ имеет вид симметричной «потенциальной ямы» (рис. 42 *а*) конечной глубины. Пусть $U(x)$ максимальна при $x = \pm\infty$ и принимает там одно и то же значение. Примем это значение за нуль отсчета энергии. Таким образом, предполагается, что $U(x)$ всюду отрицательна и при $x = \pm\infty$ обращается в нуль. В одномерном случае уравнение Шредингера (21.7) принимает вид

$$\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + (\mathscr{E} - U)\psi = 0. \tag{22.1}$$

Коэффициенты этого уравнения вещественны. Поэтому оно имеет вещественные решения, которыми и можно ограничиться. Правда, всякое решение уравнения (22.1) останется таковым, если его умножить

на постоянный множитель, который может быть и комплексным. Но комплексность не влияет на $\psi^*\psi$, а потому не сказывается ни на каких физических выводах теории.

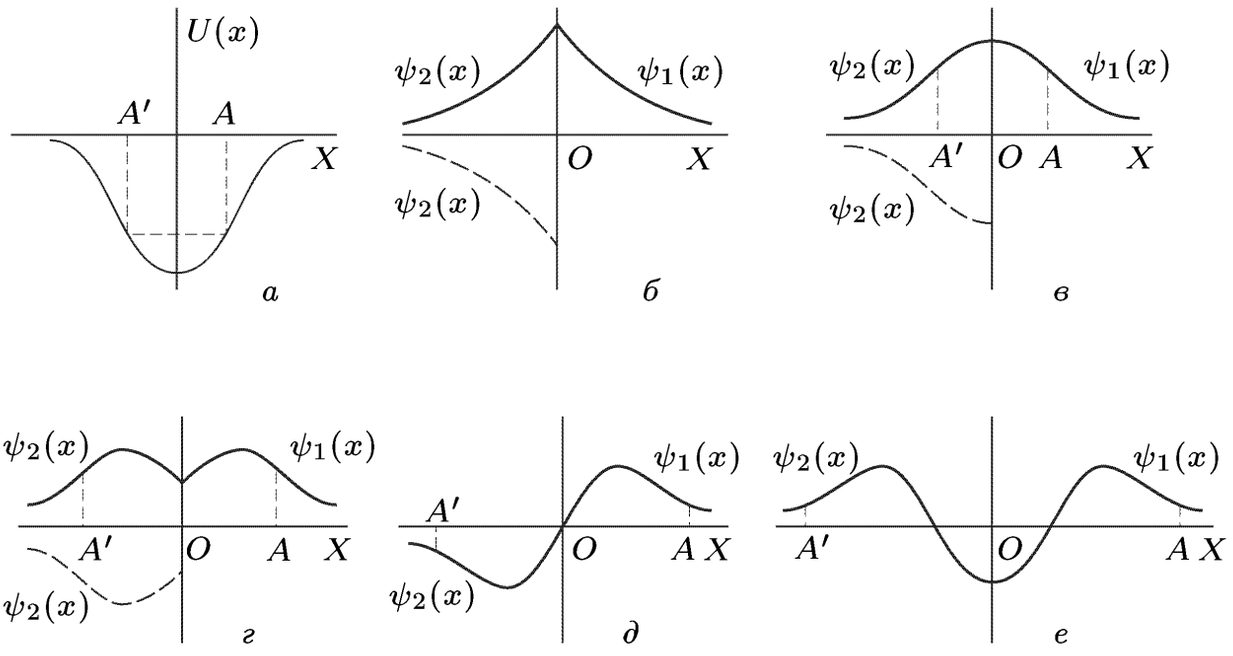


Рис. 42

2. Рассмотрим сначала случай, когда $\mathscr{E} < 0$. По классическим представлениям частица не может заходить в те области пространства, в которых $U > \mathscr{E}$, так как разность $\mathscr{E} - U$ равна $mv^2/2$, т. е. кинетической энергии, а она не может быть отрицательной. Частица может двигаться только между двумя точками, в которых $U = \mathscr{E}$. Достигнув одной из этих точек, частица должна повернуть обратно. Эти точки называются *точками поворота*. Согласно классической механике пространство вне точек поворота для частицы недостижимо. Но уравнение (22.1), если оно имеет решение, не равное тождественно нулю между точками поворота, должно и за точками поворота переходить в решение, не обращающееся тождественно в нуль. Действительно, решение уравнения (22.1), поскольку это уравнение второго порядка, однозначно определяется заданием ψ и $d\psi/dx$ в какой-либо одной точке пространства. Возьмем эту точку в области за точками поворота. Если допустить, что там $\psi = d\psi/dx \equiv 0$, то на основании сказанного функция ψ должна тождественно обращаться в нуль во всем пространстве, в частности и между точками поворота. Из изложенного следует, что не существует решения, которое было бы отлично от нуля между точками поворота и тождественно обращалось бы в нуль за этими точками.

Таким образом, плотность вероятности $\psi^*\psi$ отлична от нуля и за точками поворота. Значит, согласно квантовой механике, *существует конечная вероятность обнаружения частицы и в классически недостижимой области пространства, где* $U > \mathscr{E}$. Приведенное выше классическое рассуждение в квантовой механике неприменимо, поскольку в этом случае из-за принципа неопределенности теряет смысл разделение полной энергии на кинетическую и потенциальную. Состояние частицы с энергией $\mathscr{E}$ характеризуется единой волновой функцией

во всем бесконечном интервале $-\infty < x < +\infty$, а не только ее частью на интервале между классическими точками поворота.

3. При $x \to \pm\infty$ уравнение (22.1) асимптотически переходит в

$$\frac{d^2\psi}{dx^2} - \alpha^2\psi = 0, \tag{22.2}$$

где при $\mathscr{E} < 0$ α — положительная постоянная, равная $\alpha = \sqrt{-2m\mathscr{E}/\hbar^2}$. Последнему уравнению удовлетворяют функции $C_1 e^{-\alpha x}$ и $C_2 e^{+\alpha x}$ при произвольных значениях постоянных C_1 и C_2. Вторая функция обращается в бесконечность при $x = +\infty$, тогда как первая, $C_1 e^{-\alpha x}$, остается регулярной при всех положительных значениях x. Аналогично, при всех отрицательных значениях x остается регулярной только функция $C_2 e^{+\alpha x}$.

Рассмотрим теперь два решения уравнения (22.1): одно $\psi_1(x)$, которое при положительных x на бесконечности асимптотически переходит в $C_1 e^{-\alpha x}$, другое $\psi_2(x)$, которое при отрицательных x на бесконечности асимптотически переходит в $C_2 e^{+\alpha x}$. Искомая функция $\psi(x)$, представляющая решение уравнения (22.1) во всей бесконечной области $-\infty < x < +\infty$, должна получаться сшиванием обоих решений $\psi_1(x)$ и $\psi_2(x)$ при $x = 0$. При таком сшивании должны оставаться непрерывными сама функция $\psi(x)$ и ее первая производная $d\psi/dx$. Иными словами, должны выполняться два равенства:

$$\psi_1(x) = \psi_2(x), \quad \frac{d\psi_1(x)}{dx} = \frac{d\psi_2(x)}{dx} \quad \text{при } x = 0. \tag{22.3}$$

Если фиксировать постоянную C_1, то постоянной C_2 можно распорядиться так, чтобы соблюдалось первое равенство. Однако второе равенство, вообще говоря, соблюдаться не будет. Если же удовлетворить второму равенству, то, вообще говоря, не будет соблюдаться первое. Обоим равенствам можно одновременно удовлетворить лишь при определенных значениях параметра $\mathscr{E}$. Эти избранные значения и будут *собственными значениями энергии* частицы или уравнения (22.1).

Для исследования возможных собственных значений $\mathscr{E}$ заметим, что знаки функций $d^2\psi/dx^2$ и $(\mathscr{E} - U)\psi$ противоположны. Это непосредственно видно из уравнения (22.1). Поэтому при $(\mathscr{E} - U)\psi > 0$ кривая $\psi = \psi(x)$ обращена выпуклостью вверх, а при $(\mathscr{E} - U)\psi < 0$ — вниз. В точках поворота $\mathscr{E} - U = 0$, а также при $\psi = 0$ эта кривая испытывает перегиб.

4. Рассмотрим сначала случай, когда обе точки поворота совпадают между собой, т. е. когда $\mathscr{E} = U_{\text{мин}}$, где $U_{\text{мин}}$ — наименьшее значение потенциальной функции U, принимаемое ею при $x = 0$. Тогда $\mathscr{E} - U \equiv U_{\text{мин}} - U \leqslant 0$. Если функции $\psi_1(x)$ и $\psi_2(x)$ выбрать положительными соответственно при $x = +\infty$ и $x = -\infty$, то в этих точках $(\mathscr{E} - U)\psi_1 < 0$, $(\mathscr{E} - U)\psi_2 < 0$, т. е. обе вторые производные $d^2\psi_1/dx^2$ и $d^2\psi_2/dx^2$ на бесконечности будут положительны. Значит, при убывании $|x|$ обе кривые $\psi = \psi_1(x)$ и $\psi = \psi_2(x)$ начнут подниматься вверх. По мере убывания $|x|$ этот подъем только усилится, так как при этом величины $(\mathscr{E} - U)\psi_1$ и $(\mathscr{E} - U)\psi_2$, оставаясь отрицательными, будут непрерывно

возрастать по модулю. Значит, при убывании $|x|$ обе кривые $\psi = \psi_1(x)$ и $\psi = \psi_2(x)$ должны постоянно подниматься, будучи обращенными выпуклостями вниз (рис. 42 *б*). На кривой $\psi = \psi_1(x)$ первая производная $d\psi_1/dx$ отрицательна, а на кривой $\psi = \psi_2(x)$ — положительна. Поэтому невозможно одновременно удовлетворить обоим условиям (22.3). Действительно, если при $x = 0$ удовлетворяется первое условие, то второе условие удовлетворяться не может из-за различия знаков производных $d\psi_1/dx$ и $d\psi_2/dx$: при сшивании $\psi_1(x)$ и $\psi_2(x)$ на кривой $\psi = \psi_1 + \psi_2$ появится угловая точка (рис. 42 *б*). Можно устранить разрыв первой производной, выбрав функции $\psi_1(x)$ и $\psi_2(x)$ противоположных знаков (на рис. 42 *б* отрицательная функция ψ_2 изображена штриховой линией). Но тогда в точке $x = 0$ окажется разрывной сама функция $\psi = \psi_1 + \psi_2$. Из изложенного следует, что значение $\mathscr{E} = U_{\text{мин}}$ не может быть собственным значением энергии.

5. Будем теперь непрерывно увеличивать параметр $\mathscr{E}$, т. е. поднимать вверх горизонтальную прямую $U(x) = \mathscr{E}$ на рис. 42 *а*. Тогда точки поворота A и A' начнут непрерывно раздвигаться, находясь на равных расстояниях от начала координат, в силу симметрии потенциальной ямы $U = U(x)$. Поскольку A и A' являются точками перегиба, обе кривые $\psi = \psi_1(x)$ и $\psi = \psi_2(x)$ между ними начнут загибаться вниз. Функции $\psi_1(x)$ и $\psi_2(x)$ выберем так, чтобы соблюдалось первое условие (22.3). При небольших значениях $\mathscr{E}$ обе кривые $\psi = \psi_1(x)$ и $\psi = \psi_2(x)$ при $x = 0$ по-прежнему будут пересекаться под некоторым углом. Конечно, ввиду симметрии потенциальной ямы $U = U(x)$ эти кривые зеркально симметричны относительно вертикальной оси координат. По мере увеличения параметра $\mathscr{E}$, т. е. по мере раздвигания точек поворота A и A', угол между ними уменьшается. Всегда наступит момент, когда в точке $x = 0$ касательная к одной, а следовательно, и к другой кривой сделается горизонтальной. Тогда обе кривые $\psi = \psi_1(x)$ и $\psi = \psi_2(x)$ плавно перейдут одна в другую (рис. 42 *в*). Соответствующее значение параметра $\mathscr{E} = \mathscr{E}_1$ и будет *первым или наименьшим собственным значением энергии.* В случае симметричной потенциальной ямы наименьшее собственное значение $\mathscr{E}_1$ всегда существует.

6. При дальнейшем увеличении параметра $\mathscr{E}$, а следовательно, и расстояния между точками поворота кривые $\psi = \psi_1(x)$ и $\psi = \psi_2(x)$, по-прежнему загибаясь вниз, начнут опускаться. Опять на вертикальной оси появится угловая точка (рис. 42 *г*), т. е. второе условие (22.3) перестанет выполняться. Однако в качестве решения уравнения Шредингера при $x < 0$ теперь можно взять новую функцию $\psi_2(x)$, отличающуюся от прежней функции $\psi_2(x)$ знаком. Она изображена на рис. 42 *г* штриховой кривой. Ясно, что эта штриховая кривая зеркально симметрична с прежней кривой $\psi = \psi_2(x)$ относительно горизонтальной координатной оси, а следовательно, симметрична с кривой $\psi = \psi_1(x)$ относительно начала координат. Когда угловая точка, упомянутая выше, опустится в начало координат, верхняя кривая $\psi = \psi_1(x)$ непрерывно сомкнется с нижней (штриховой) кривой, т. е. первое условие (22.3) будет выполнено. Соответствующая результирующая кривая

изображена на рис. 42 *д* уже всюду сплошной линией. При этом будет выполнено и второе условие (22.3), так как в начале координат $\psi_1 = \psi_2$, а потому оно является точкой перегиба результирующей кривой. Рассмотренному случаю соответствует *второе собственное значение энергии* $\mathscr{E} = \mathscr{E}_2$.

Третьему собственному значению $\mathscr{E} = \mathscr{E}_3$ соответствует волновая функция, представленная на рис. 42 *е*.

Продолжая этот процесс дальше, убеждаемся, что *собственные функции стационарных состояний имеют узлы, в которых они обращаются в нуль. Число узлов на единицу меньше номера соответствующего собственного значения энергии.*

7. Итак, при $\mathscr{E} < 0$ собственные значения энергии образуют *дискретный спектр.* Поскольку в этом случае волновая функция стационарного состояния на бесконечности асимптотически экспоненциально убывает, можно сказать, что при $\mathscr{E} < 0$ частица находится в практически ограниченной области пространства, т. е. совершает *финитное движение.* Напомним, что в классической механике условие финитности движения также имеет вид $\mathscr{E} < 0$ (см. т. I, § 25). Различие состоит только в том, что согласно классической механике частица не может проникать в пространство за точками поворота, тогда как по квантовой механике ее можно обнаружить и в этом пространстве, хотя и с вероятностью, быстро убывающей при удалении от точек поворота.

Число возможных стационарных состояний или энергетических уровней зависит от вида потенциальной функции $U(x)$. Оно может быть *конечным* или *бесконечным.* В частности, когда глубина симметричной потенциальной ямы достаточно мала, возможно всего *одно* стационарное состояние. Если же число дискретных энергетических уровней бесконечно велико, то, очевидно, с возрастанием номера уровня его энергия должна асимптотически приближаться к $\mathscr{E} = 0$, а расстояние между соседними уровнями — стремиться к нулю.

8. Отметим еще раз, что *не существует стационарного состояния с энергией* $\mathscr{E} = U_{\text{мин}}$. В противном случае частица все время находилась бы в точке $x = 0$, т. е. покоилась бы на дне потенциальной ямы, что противоречит принципу неопределенностей Гейзенберга (см. § 20, п. 8). Наименьшая энергия, которую может иметь частица в потенциальной яме, равна $\mathscr{E}_1$. Эта энергия называется *нулевой энергией.* Нулевая энергия в заданной потенциальной яме $U(x)$ не может быть отнята от частицы, поскольку она является наименьшей допустимой энергией. Чтобы ее изменить, надо изменить саму потенциальную яму.

Нулевая энергия проявляется во многих явлениях. Примером может служить гелий. При абсолютном нуле температуры его атомы не находятся в покое, а благодаря наличию нулевой энергии совершают так называемые *нулевые колебания.* Вблизи абсолютного нуля они еще достаточно интенсивны, тогда как силы молекулярного притяжения слабы. Этих сил недостаточно, чтобы жидкий гелий перевести в твердое состояние, даже при абсолютном нуле температуры. Требуется повысить давление на гелий до 24 атм или выше, чтобы он (при $T = 0$) перешел в твердое состояние.

9. Осталось рассмотреть решение уравнения (22.1) при положительных значениях параметра $\mathscr{E}$. В этом случае на бесконечности уравнение (22.1) асимптотически переходит в

$$\frac{d^2\psi}{dx^2} + \beta^2\psi = 0, \tag{22.4}$$

где $\beta = \sqrt{2m\mathscr{E}/\hbar^2}$, т. е. β — положительная постоянная. Оно имеет два линейно независимых решения $\sin\beta x$ и $\cos\beta x$, остающиеся конечными при $x = \pm\infty$. Поэтому любое решение $\psi(x)$ уравнения (22.4) также конечно при $x \to \pm\infty$, хотя при этом оно и не стремится к определенному пределу, а осциллирует. Если $U(x)$ непрерывна, то будет непрерывно и всякое решение уравнения (22.4) вместе со своей производной в любой точке интервала $-\infty < x < +\infty$. Взяв любое решение, осциллирующее на отрицательной бесконечности, и продолжив его в сторону положительных x, мы получим всюду непрерывное решение с непрерывной производной, осциллирующее на положительной бесконечности. Условие сшивания (22.1) выполняется автоматически.

Таким образом, каким бы ни был положительный параметр $\mathscr{E}$, любое решение уравнения (22.1) может быть волновой функцией. Это значит, что при $\mathscr{E} > 0$ энергетический спектр частицы *непрерывный*. Поскольку при $x = \pm\infty$ функция ψ остается конечной, частица с отличной от нуля вероятностью может уходить в бесконечность. Иными словами, при $\mathscr{E} > 0$ движение частицы будет *инфинитно*. Таково же условие инфинитности движения и в классической механике (см. т. I, § 25).

10. В заключение заметим, что естественные требования, накладываемые на решения уравнения Шредингера, о которых говорилось в начале этого параграфа, могут быть ослаблены. Достаточно ограничиться трехмерным случаем. Во-первых, в этом случае потенциальная функция $U(\mathbf{r})$ может в некоторой точке обращаться в бесконечность. Пусть такой точкой является начало координат $r = 0$. Тогда, как доказывается в квантовой механике, допустимы решения, которые в окрестности начала координат ведут себя как $\psi \sim 1/r^\alpha$ с положительным значением постоянной α. Однако должно быть $\alpha < 1$. Во-вторых, из физических соображений ясно, что требование однозначности должно налагаться не на саму функцию ψ, а *на плотность вероятности* $\psi^*\psi$. Но во всех вопросах, рассматриваемых в квантовой механике, оно сводится к однозначности ψ. На этих вопросах мы не можем останавливаться.

§ 23. Гармонический осциллятор

1. *Гармоническим осциллятором* в классической физике называют частицу, на которую действует сила, пропорциональная отклонению частицы из положения равновесия и направленная к нему. Осциллятор называется *одномерным*, если частица может двигаться только вдоль одной прямой. Последнюю мы примем за ось X, а положение равновесия — за начало координат. Потенциальная функция частицы имеет

вид

$$U = \frac{1}{2} kx^2, \tag{23.1}$$

где k — постоянная (*коэффициент упругости*), а x — отклонение частицы от положения равновесия. Графиком функции $U(x)$ является парабола (рис. 43). Согласно классической механике осциллятор совершает гармонические колебания с циклической частотой $\omega = \sqrt{k/m}$, где m — масса частицы.

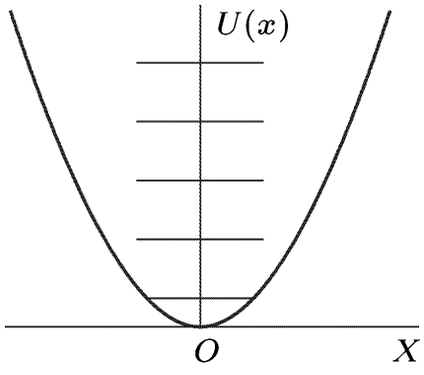


Рис. 43

В квантовой механике понятие силы не используется. Поэтому квантовый гармонический осциллятор следует определить как частицу с потенциальной функцией $U(x)$. Найдем энергии стационарных состояний осциллятора, следуя идеям предыдущего параграфа. Но здесь возникает следующая трудность. Функцию $U(x)$ нельзя нормировать так, чтобы она обращалась в нуль в бесконечности, так как при $x = \pm\infty$ она сама бесконечно велика. Но эта трудность искусственная. В реальных системах при возрастании $|x|$ начинают проявляться отступления от параболической формулы (23.1), так что $U(\pm\infty)$ становится конечной. Рассмотрим случай, когда $U(x)$ симметрична, так что $U(+\infty) = U(-\infty)$. Тогда методы предыдущего параграфа становятся применимыми. Но здесь удобнее за нуль $U(x)$ принять ее значение при $x = 0$. Мы проведем решение, предполагая, что формула (23.1) справедлива при любых x. Однако для реального осциллятора полученные результаты будут справедливы для не слишком больших значений $|x|$.

2. Уравнение Шредингера для одномерного гармонического осциллятора имеет вид

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + \frac{1}{2} kx^2\psi = \mathscr{E}\psi. \tag{23.2}$$

Если ввести безразмерные величины

$$\lambda = \frac{2\mathscr{E}}{\hbar\omega}, \quad \xi = x\sqrt{\frac{k}{\hbar\omega}}, \tag{23.3}$$

то оно преобразуется в

$$-\frac{d^2\psi}{d\xi^2} + \xi^2\psi = \lambda\psi. \tag{23.4}$$

При определенном значении параметра λ это уравнение имеет решение $\psi = e^{\alpha\xi^2}$, где α — постоянная, которая сейчас будет определена вместе с λ. Действительно,

$$\frac{d\psi}{d\xi} = 2\alpha\xi e^{\alpha\xi^2} = 2\alpha\xi\psi,$$

$$\frac{d^2\psi}{d\xi^2} = 2\alpha\psi + 2\alpha\xi\frac{d\psi}{d\xi} = (4\alpha^2\xi^2 + 2\alpha)\psi.$$

Подставляя эти значения в (23.4), получим

$$(1-4\alpha^2)\xi^2 - 2\alpha = \lambda,$$

причем это соотношение должно выполняться тождественно по ξ. Это будет тогда и только тогда, когда $1-4\alpha^2 = 0$, $\lambda = -2\alpha$, т.е. $\alpha = \pm 1/2$. Знак плюс следует отбросить, так как в этом случае функция $\psi = e^{\alpha\xi^2}$ обращалась бы в бесконечность при $\xi = \pm\infty$. Таким образом, получается решение

$$\psi = e^{-\xi^2/2}, \tag{23.5}$$

если $\lambda = 1$. Это решение не имеет узлов, а потому оно описывает *основное состояние гармонического осциллятора*. Ему соответствует нулевая энергия

$$\mathscr{E}_0 = \frac{\lambda}{2}\hbar\omega = \frac{\hbar\omega}{2}. \tag{23.6}$$

3. В стационарном состоянии с энергией $\mathscr{E}_n$ функция ψ должна иметь n узлов. Такое число узлов имеет функция

$$\psi = P_n(\xi)e^{-\xi^2/2}, \tag{23.7}$$

где $P_n(\xi)$ — полином n-й степени с некратными вещественными корнями. При избранных значениях параметра λ такая функция действительно является решением уравнения (23.4) и обращается в нуль на бесконечности. При таких значениях λ она и будет волновой функцией осциллятора. Дважды дифференцируя ее и подставляя $d^2\psi/dx^2$ в уравнение (23.4), получим

$$-P_n''(\xi) + \underline{2\xi P_n'(\xi)} + \underline{P_n(\xi)} = \underline{\lambda P_n(\xi)}. \tag{23.8}$$

Это соотношение должно выполняться тождественно по ξ. В нем все подчеркнутые члены являются полиномами степени n. Степень полинома $P_n''(\xi)$ на два меньше, т.е. равна $n-2$ ($n \geqslant 2$). Чтобы определить λ, достаточно сравнить коэффициенты при старших членах подчеркнутых полиномов. Если коэффициент при ξ^n в полиноме $P_n(\xi)$ равен a_n, то в полиноме $2\xi P_n'(\xi)$ соответствующий коэффициент равен $2na_n$. Поэтому необходимо, чтобы выполнялось соотношение $2n + 1 = \lambda$. Тогда

$$-P_n''(\xi) + 2\xi P_n'(\xi) = 2nP_n(\xi). \tag{23.9}$$

Полиномы, являющиеся решениями этого уравнения, называются *полиномами Чебышева–Эрмита*. Можно доказать (на чем мы не останавливаемся), что все корни полиномов Чебышева–Эрмита некратные и вещественные. Это легко доказать для небольших n, фактически находя сами полиномы и вычисляя их корни (см. задачу к этому параграфу).

Подставляя $\lambda = 2n+1$ в (23.3), находим энергетические уровни осциллятора:

$$\mathscr{E}_n = \hbar\omega\left(n + \frac{1}{2}\right), \quad n = 0, 1, 2, \ldots \tag{23.10}$$

Эти уровни *эквидистантны*, т. е. находятся на равных расстояниях друг от друга. На рис. 43 они изображены горизонтальными прямыми.

Классический осциллятор излучает свет только с *одной* частотой ω. Казалось бы, что в соответствии с правилом частот Бора в квантовом случае возможно излучение со всевозможными кратными частотами N_ω (N — целое число). На самом деле при излучении фотона этого не происходит. Из этого затруднения в старой квантовой теории Бор вышел, руководствуясь принципом соответствия. Чтобы исключить кратные частоты, на переходы между уровнями энергии осциллятора было наложено ограничение, называемое *правилом отбора.* Согласно этому правилу квантовое число n осциллятора при излучении и поглощении фотона может меняться только на ± 1, т. е.

$$\Delta n = \pm 1. \tag{23.11}$$

Это правило отбора выводится и в последовательной квантовой механике, не обращаясь ни к какому принципу соответствия. Квантовая механика позволяет вычислить вероятность перехода осциллятора с одного уровня на другой с излучением или поглощением фотона. Оказалось, что эта вероятность обращается в нуль, когда правило отбора (23.11) не соблюдается.

ЗАДАЧА

Найти полиномы Чебышева–Эрмита и волновые функции одномерного гармонического осциллятора для $n = 1, 2, 3, 4, 5$.

Р е ш е н и е. Ради примера рассмотрим случай $n = 4$. Задача сводится к решению уравнения (23.9) в виде полинома

$$P_4(\xi) = a_4\xi^4 + a_3\xi^3 + a_2\xi^2 + a_1\xi + a_0.$$

Подставляя это выражение в уравнение (23.9) и сравнивая коэффициенты, найдем, что оно удовлетворяется при любом значении a_4, как это и должно быть согласно общей теории. Далее, находим $a_2 = -3a_4$, $a_0 = -(1/4)a_2 = (3/4)a_4$, $a_3 = a_1 = 0$. Итак,

$$P_4(\xi) = a_4(\xi^4 - 3\xi^2 + 3/4).$$

Корни этого полинома

$$\xi = \pm\sqrt{\frac{3 \pm \sqrt{6}}{2}}$$

вещественны и некратны.

Аналогично,

$$P_1(\xi) = a_1\xi, \quad P_2(\xi) = a_2(\xi^2 - 1/2),$$
$$P_3(\xi) = a_3(\xi^3 - 3\xi/2), \quad P_5(\xi) = a_5(\xi^5 - 5\xi^3 + 15\xi/4).$$

Волновые функции получаются умножением этих полиномов на $e^{-\xi^2/2}$. Их обычно нормируют к единице, т. е. подчиняют условию

$$\int_{-\infty}^{+\infty} \left| P(\xi)e^{-\xi^2/2} \right|^2 d\xi = 1.$$

§ 24. Одномерные прямоугольные потенциальные ямы

1. Квантование на основе уравнения Шредингера (22.1) полезно уяснить на примере одномерной симметричной «потенциальной ямы» прямоугольной формы. Так называется потенциальная функция $U(x)$, принимающая на интервале $-a < x < +a$ постоянное значение $-U_0$ и обращающаяся в нуль вне этого интервала (рис. 44). Для этого случая легко получить точное решение уравнения Шредингера и на его основе рассмотреть задачу о квантовании энергии. Но этим значение прямоугольных потенциальных ям не исчерпывается. В ряде случаев (напри-

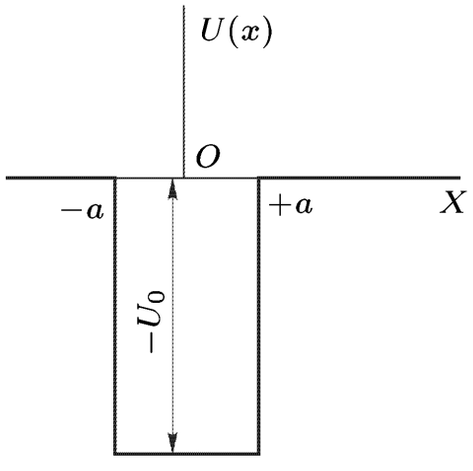


Рис. 44

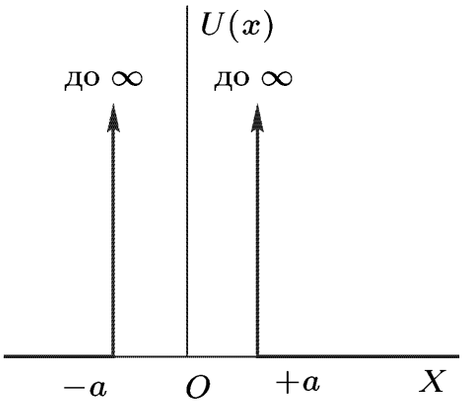


Рис. 45

мер, в ядерной физике) истинный ход потенциальной функции $U(x)$ неизвестен. Аппроксимируя $U(x)$ потенциальной ямой прямоугольной формы, получают в таких случаях не только качественные, но даже количественные результаты оценочного характера.

2. Наиболее простым в математическом отношении является случай *бесконечно глубокой* потенциальной ямы, когда величина U_0 обращается в бесконечность. В этом случае целесообразно за нуль потенциальной функции принять ее значение на «дне» потенциальной ямы, т. е. на интервале $-a < x < +a$. Тогда на «стенках» ямы (т. е. при $x = \pm a$) функция $U(x)$ будет претерпевать разрыв от 0 до $+\infty$. Такая потенциальная яма изображена на рис. 45.

Математическое упрощение задачи при переходе от ямы конечной глубины к бесконечно глубокой яме связано с тем, что в последнем случае вне интервала $-a < x < +a$, где U всюду бесконечно велика, функция ψ должна обращаться в нуль. Действительно, согласно классической физике, частица с конечной энергией $\mathscr{E}$ не может попасть в область, где $U(x) = +\infty$. В квантовой механике это утверждение заменяется требованием обращения в нуль плотности вероятности $\psi^*\psi$, а следовательно, и самой функции ψ. Таким образом, достаточно рассмотреть решение уравнения Шредингера только в интервале $-a < x < +a$, что и ведет к упрощению задачи.

Внутри интервала $-a < x < +a$ $U(x) = 0$ и уравнение (22.1) принимает вид

$$\frac{d^2\psi}{dx^2} + k^2\psi = 0, \tag{24.1}$$

где введено обозначение

$$k^2 = 2m\mathscr{E}/\hbar. \tag{24.2}$$

Не теряя общности, достаточно ограничиться положительными значениями k, что и предполагается ниже. Общее решение уравнения (24.2) имеет вид

$$\psi = A\cos kx + B\sin kx,$$

причем на стенках ямы $x = \pm a$ должно быть $\psi = 0$. Это дает

$$A\cos ka + B\sin ka = 0 \quad \text{при} \quad x = +a,$$

$$A\cos ka - B\sin ka = 0 \quad \text{при} \quad x = -a.$$

Если $A \neq 0$, то $A\cos ka = 0$ и, следовательно, $\cos ka = 0$, $\sin ka \neq 0$, $B = 0$. Наоборот, если $B \neq 0$, то $B\sin ka = 0$ и, следовательно, $\sin ka = 0$, $\cos ka \neq 0$, $A = 0$. Таким образом, все решения уравнения (24.2) распадаются на два класса:

1) с четными функциями

$$\psi = A\cos kx, \quad ka = \pi/2,\ 3\pi/2,\ 5\pi/2, \ldots;$$

2) с нечетными функциями

$$\psi = B\sin kx, \quad ka = 2(\pi/2),\ 4(\pi/2),\ 6(\pi/2), \ldots$$

Возможность $ka = 0$ во втором случае исключается, так как тогда было бы $\psi = 0$, что не имеет физического смысла. Постоянные A и B обычно определяются из условия нормировки $\int\limits_{-a}^{+a} |\psi|^2\, dx = 1$ (для разбираемого нами вопроса это не имеет значения). Тогда получается

$$\psi = \begin{cases} \dfrac{1}{\sqrt{a}}\cos\dfrac{n\pi x}{2} & \text{при нечетных } n, \\ \dfrac{1}{\sqrt{a}}\sin\dfrac{n\pi x}{2} & \text{при четных } n. \end{cases} \tag{24.3}$$

В обоих случаях $k = n\pi/2a$, так что при любом целом n

$$\mathscr{E} = \frac{\hbar}{2m}k^2 = \frac{\hbar\pi^2}{8ma^2}n^2. \tag{24.4}$$

Отсюда видно, что энергия квантуется. Энергетические уровни *дискретны*, при $U_0 = +\infty$ число их бесконечно велико. Так как значение $n = 0$ исключается, то энергия наинизшего уровня равна $\hbar\pi^2/(8ma^2)$. Это — нулевая энергия, необходимость которой следует из общих положений.

Против приведенного решения можно выдвинуть следующее возражение. На всякой поверхности разрыва потенциальной функции $U(x)$

должны выполняться граничные условия

$$\psi_1(x-0)=\psi_2(x+0), \quad \frac{d\psi_1(x-0)}{dx}=\frac{d\psi_2(x+0)}{dx}, \tag{24.5}$$

где $\psi_1(x)$ — функция $\psi(x)$ по одну сторону поверхности разрыва, а $\psi_2(x)$ — по другую (см. § 22, п. 1). В нашем случае внутри интервала $-a<x<+a$ $\psi=\psi_1$ дается выражениями (24.3), а вне этого интервала $\psi\equiv\psi_2=0$. Первое условие (24.5) выполняется, тогда как второе не выполняется. Таким образом, на стенках потенциальной ямы первая производная найденной нами функции $\psi(x)$ претерпевает разрыв непрерывности. Однако это противоречие с общими требованиями, которым должна удовлетворять функция $\psi(x)$, является только кажущимся и возникает в результате *математического перехода к пределу.* Во всяком реальном случае глубина ямы U_0 конечна, хотя и может быть очень большой. В этом случае вблизи стенки по обе стороны от нее $\psi(x)$ и $d\psi/dx$, вообще говоря, отличны от нуля, и условия (24.5) строго выполняются. Но при переходе к пределу бесконечно глубокой ямы они могут и не выполняться для предельных значений этих величин. Действительно, из соотношений (24.5) не следует, что должны выполняться и предельные соотношения

$$\lim\psi_1(x-0)=\lim\psi_2(x+0),$$

$$\frac{d}{dx}\lim\psi_1(x-0)=\frac{d}{dx}\lim\psi_2(x+0).$$

Это на самом деле и происходит с производными функции $\psi(x)$. Найденное нами решение относится не к реальной функции $\psi(x)$ при очень большом значении U_0, а к ее *предельному значению при* $U_0\to\infty$.

На этом примере с особой отчетливостью проявляется отмеченная выше аналогия между задачей о квантовании энергии и задачей о колебании струны с закрепленными концами. Действительно, в случае прямоугольной бесконечно глубокой потенциальной ямы обе задачи математически тождественны.

3. Рассмотрим теперь случай симметричной прямоугольной ямы *конечной* глубины (см. рис. 44). Потенциальную функцию $U(x)$ вне ямы примем равной нулю. Внутри ямы $U(x)=U_0<0$. За начало координат возьмем центр дна ямы O. Исследуем сначала случай, когда полная энергия $\mathscr{E}$ отрицательна, причем $U_0<\mathscr{E}<0$. Введем обозначения

$$k=+\sqrt{2m(\mathscr{E}-U_0)/\hbar^2}, \quad \alpha=+\sqrt{-2m\mathscr{E}/\hbar^2}. \tag{24.6}$$

Тогда уравнение Шредингера внутри ямы будет

$$\frac{d^2\psi}{dx^2}+k^2\psi=0, \tag{24.7}$$

а вне ямы

$$\frac{d^2\psi}{dx^2}-\alpha^2\psi=0. \tag{24.8}$$

Общее решение уравнения (24.7) имеет вид

$$\psi = A\cos kx + B\sin kx. \tag{24.9}$$

Решением уравнения (24.8) является $e^{\pm\alpha x}$. Здесь надо выбрать такой знак, чтобы решение обращалось в нуль при $x = \pm\infty$. Таким образом, вне ямы должно быть

$$\psi = Ce^{-\alpha x} \quad \text{при} \quad x > a,$$
$$\psi = De^{\alpha x} \quad \text{при} \quad x < -a.$$

Из соображений симметрии следует, что плотность вероятности $|\psi|^2$ должна быть симметричной функцией x относительно начала координат. Следовательно, должно быть $C^2 = D^2$, т. е. возможны два случая: $C = D$ и $C = -D$. Постоянные A, B, C, D надо выбрать так, чтобы на краях ямы функция ψ и ее производная $d\psi/dx$ были непрерывны. На границе $x = +a$ это дает

$$A\cos ka + B\sin ka = Ce^{-\alpha a},$$
$$-kA\sin ka + kB\cos ka = -\alpha Ce^{-\alpha a},$$

а на границе $x = -a$

$$A\cos ka - B\sin ka = De^{-\alpha a},$$
$$kA\sin ka + kB\cos ka = \alpha De^{-\alpha a}.$$

Отсюда

$$2A\cos ka = (C + D)e^{-\alpha a}, \quad 2B\sin ka = (C - D)e^{-\alpha a},$$
$$2kA\sin ka = \alpha(C + D)e^{-\alpha a}, \quad 2kB\cos ka = -\alpha(C - D)e^{-\alpha a}.$$

Если $A \neq 0$ и $C = D$, то

$$k\operatorname{tg} ka = \alpha. \tag{24.10}$$

Если же $B \neq 0$ и $C = -D$, то

$$k\operatorname{ctg} ka = -\alpha. \tag{24.11}$$

Эти условия не могут быть удовлетворены одновременно, так как в противном случае получилось бы $k^2 = -\alpha^2$, а это невозможно ввиду вещественности k и α. Решение, когда все коэффициенты A, B, C, D равны нулю, физического смысла не имеет. Таким образом, все возможные решения разделяются на два класса: решения с четной волновой функцией, когда $A \neq 0$, $B = 0$, $C = D$, и решения с нечетной волновой функцией, когда $A = 0$, $B \neq 0$, $C = -D$.

Уровни энергии найдутся путем графического или численного решения уравнения (24.10) или уравнения (24.11), в которых положительные величины k и α определяются выражениями (24.6). Для графического решения введем безразмерные величины

$$\xi = ak, \quad \eta = a\alpha. \tag{24.12}$$

Тогда

$$\xi^2 + \eta^2 = -\frac{2mU_0a^2}{\hbar^2}, \tag{24.13}$$

причем для решений с четной волновой функцией из (24.10) следует

$$\eta = \xi \operatorname{tg} \xi, \tag{24.10а}$$

а для решений с нечетной волновой функцией из (24.11) получаем

$$\eta = -\xi \operatorname{ctg} \xi. \tag{24.11а}$$

На рис. 46 *а* построены кривые $\eta = \xi \operatorname{tg} \xi$, на рис. 46 *б* — кривые $\eta = -\xi \operatorname{ctg} \xi$. Вертикальными штриховыми линиями изображены асимптоты этих кривых. Ввиду положительности ξ и η нужны

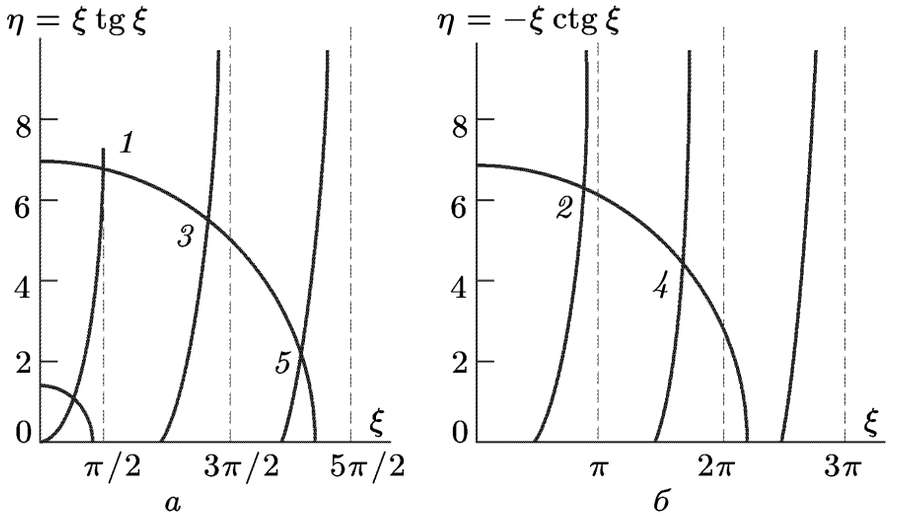


Рис. 46

только участки кривых, расположенные в положительном квадранте ($\xi > 0$, $\eta > 0$). Пересечем эти кривые окружностью (24.13), радиус которой $\sqrt{-2mU_0}\, a/\hbar$ должен считаться известным, поскольку известны величины U_0 и a. Координаты точек пересечения этой окружности с кривыми (24.10а) и (24.11а) дадут возможные значения ξ и η, а следовательно, k и α. После этого по формулам (24.6) легко найти значения $\mathscr{E}$. Число уровней всегда конечно и определяется глубиной $-U_0$ и шириной $2a$ потенциальной ямы. Например, если радиус окружности равен 7, то получается пять уровней. Точкам пересечения *1*, *3*, *5* соответствуют четные, а точкам *2*, *4* — нечетные волновые функции. Если $0 \leqslant -U_0a^2 \leqslant \hbar^2\pi^2/(8m)$, то имеется только одна точка пересечения, которой соответствует четная волновая функция. Так как величина k существенно отлична от нуля, то из (24.6) следует, что $\mathscr{E} > U_0$. Все уровни энергии, в том числе и самый низший, лежат выше дна потенциальной ямы. Опять наше решение приводит к необходимости существования нулевой энергии.

4. Остается рассмотреть случай, когда $\mathscr{E} > 0$. В этом случае величина α чисто мнимая: $\alpha = i\beta$. Вместо (24.8) получается уравнение

$$\frac{d^2\psi}{dx^2} + \beta^2\psi = 0. \tag{24.8а}$$

Его решения:

$$\psi = A' \cos \beta x + B' \sin \beta x \quad \text{при} \quad x > +a,$$
$$\psi = A'' \cos \beta x + B'' \sin \beta x \quad \text{при} \quad x < -a.$$

Оба решения остаются конечными при любых значениях x, в частности сколь угодно больших по абсолютной величине. Они содержат четыре произвольных постоянных A', B', A'', B''. Эти решения надо сшить с решением внутри интервала $-a < x < +a$, которое представляется формулой (24.9), чтобы при этом оставались непрерывными ψ и $d\psi/dx$ на обеих стенках потенциальной ямы. Таким путем получаются четыре линейных уравнения относительно коэффициентов A', B', A'', B'', содержащие A и B в качестве параметров. Этого как раз достаточно, чтобы выразить эти неизвестные коэффициенты через A и B. При этом A и B могут принимать любые значения. Отсюда следует, что при $\mathscr{E} > 0$ энергия не квантуется — энергетический спектр *непрерывен*. Волновая функция не стремится к нулю при $x \to \pm\infty$, т. е. движение частицы *инфинитно*, как того и требует общая теория.

§ 25. Квантование в случае сферически симметричного силового поля

1. В атомной физике более важен случай, когда потенциальная функция U не одномерна, а *сферически симметрична* относительно некоторого силового центра. Примером может служить положительно заряженное атомное ядро, в электрическом поле которого движется электрон. Силовой центр (например, атомное ядро) мы сначала будем считать *бесконечно тяжелым* и *неподвижным*. Начало координат поместим в силовом центре. Обозначим через $\mathbf{r}$ радиус-вектор, проведенный из начала координат к рассматриваемой частице. Тогда в случае сферической симметрии $U = U(r)$, где $r \equiv |\mathbf{r}|$. В этом случае волновая функция ψ, т. е. решение уравнения Шредингера (21.7), может зависеть не только от r, но и от угловых переменных, определяющих направление радиус-вектора $\mathbf{r}$. Однако мы ограничимся здесь только *сферически симметричными решениями*, т. е. решениями, зависящими только от $|\mathbf{r}|$; $\psi = \psi(r)$. Тогда уравнение Шредингера для стационарных состояний (21.7) запишется в виде

$$\frac{\hbar^2}{2m}\left(\frac{d^2\psi}{dr^2} + \frac{2}{r}\frac{d\psi}{dr}\right) + [\mathscr{E} - U(r)]\psi = 0. \tag{25.1}$$

Это уравнение отличается от исследованного выше одномерного уравнения (22.1) наличием дополнительного члена $\frac{2}{r}\frac{d\psi}{dr}$. Однако подстановкой

$$\psi = \chi/r \tag{25.2}$$

оно приводится к виду

$$\frac{\hbar^2}{2m}\frac{d^2\chi}{dr^2}+[\mathscr{E}-U(r)]\chi=0. \tag{25.3}$$

Это уравнение математически тождественно с уравнением (22.1) для одномерного случая. Поэтому все результаты, полученные в § 22, сохраняют силу для вспомогательной функции $\chi(r)$. Единственное отличие состоит в том, что при $r=0$ функция χ должна быть не только конечной, но и обращаться в нуль, так как в противном случае функция $\psi=\chi/r$ обращалась бы в бесконечность при $r=0$. Поэтому половина решений, полученных в § 22, должна быть исключена. Надо оставить только решения, изображающиеся кривыми, проходящими через начало координат (рис. 47). При $\mathscr{E}<0$ эти решения для положительных r представлены сплошными кривыми, а их продолжения в область от-

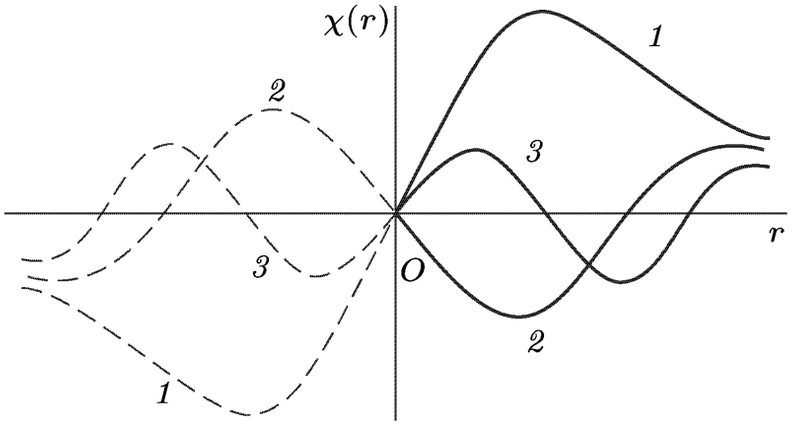


Рис. 47

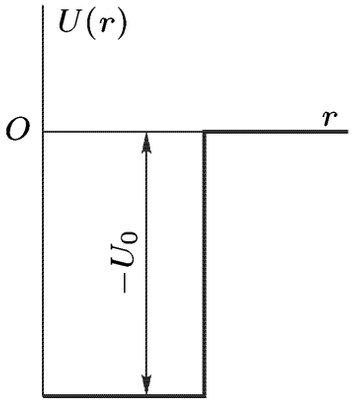


Рис. 48

рицательных r — штриховыми кривыми. Разумеется, в нашей задаче отрицательные r не имеют физического смысла и введены формально математически, чтобы получилась аналогия с одномерным случаем. Поэтому остается справедливой доказанная в § 22 теорема, что число узлов функции $\psi(r)$ (r существенно положительно) на единицу меньше номера соответствующего собственного значения. При этом точка $r=0$ за узел не считается. Рисунок 47 иллюстрирует это утверждение.

2. Частным случаем сферически симметричного силового поля является *трехмерная сферически симметричная потенциальная яма*, сечение которой плоскостью, проходящей через силовой центр, имеет прямоугольную форму (рис. 48). Так называется трехмерная потенциальная функция U, зависящая только от расстояния r до силового центра, которая определяется выражениями

$$U(r)=\begin{cases} -U_0 & \text{при } r<a,\\ 0 & \text{при } r>a. \end{cases} \tag{25.4}$$

Из изложенного выше следует, что в этом случае сферически симметричные волновые функции $\psi(r)$ и значения $\mathscr{E}$ в стационарных состояниях находятся так же, как и в случае одномерной прямоугольной потенциальной ямы (см. § 24). Различие состоит только в том, что

теперь четные решения уравнения (25.3) (если решения формально продолжить в сторону отрицательных r) должны быть исключены. Иными словами, при $\mathscr{E} < 0$ остаются только решения

$$\begin{aligned} \psi &= B \sin kr \quad \text{при } 0 < r < +a, \\ \psi &= Ce^{-\alpha r} \quad \text{при } r > +a, \end{aligned} \tag{25.5}$$

где k и α определяются прежними формулами (24.6). В соответствии с этим из двух формул (24.10а) и (24.11а) надо сохранить только вторую, т. е.

$$\eta = -\xi \operatorname{ctg} \xi, \tag{25.6}$$

причем ξ и η определяются прежними выражениями (24.12).

В случае одномерной симметричной ямы *всегда существует по крайней мере одно собственное значение* дискретного спектра энергии с четной волновой функцией. В случае сферически симметричной прямоугольной ямы этого может и не быть. Действительно, кривая $\eta = -\xi \operatorname{ctg} \xi$ первый раз пересекает горизонтальную ось координат в точке $\xi = \pi/2$, $\eta = 0$ (рис. 46 *б*). Из формулы (24.13) видно, что если

$$\left(\frac{\pi}{2}\right)^2 > -\frac{2mU_0a^2}{\hbar^2},$$

т. е.

$$-U_0 < \frac{\pi^2\hbar^2}{8ma^2}, \tag{25.7}$$

то эта кривая нигде не пересечется с окружностью (24.13). Это значит, что при условии (25.7) в потенциальной яме не появится *ни одного уровня дискретного спектра энергии.* Первый уровень появляется, когда крайняя левая кривая $\eta = -\xi \operatorname{ctg} \xi$ на рис. 46 *б* начинает пересекаться с соответствующей окружностью. Это происходит в точке $\xi = \pi/2$, $\eta = 0$. Из второго уравнения (24.6) следует, что в этом случае $\mathscr{E} = 0$, т. е. при возрастании глубины ямы первый уровень появляется на границе дискретного и непрерывного спектров. Расстояние этого уровня от дна потенциальной ямы равно $-U_0 = \hbar^2\pi^2/(8ma^2)$, как это легко получить из первой формулы (24.6). В рассматриваемом случае весь дискретный спектр состоит из одного только уровня нулевой энергии.

§ 26. Система двух взаимодействующих частиц

1. До сих пор мы рассматривали движение *одной* частицы в заданном силовом поле. Более реальным является случай *двух* частиц, взаимодействующих между собой. В классической механике в этом случае движение распадается на движение системы как целого в отсутствие внешних сил (*движение центра масс*) и на движение *одной частицы относительно другой* под действием сил взаимодействия между ними

(*относительное движение*). Последнее формально сводится к движению одной частицы с заменой ее истинной массы на *приведенную массу*

$$m = \frac{m_1 m_2}{m_1 + m_2}, \tag{26.1}$$

где m_1 и m_2 — массы первой и второй частиц. Совершенно так же обстоит дело и в квантовой механике при рассмотрении стационарных состояний.

Исходным служит уравнение Шредингера для стационарных состояний двух частиц. Оно является естественным обобщением уравнения (21.7) и имеет вид

$$\left[-\frac{\hbar^2}{2m_1}\nabla_1^2 - \frac{\hbar^2}{2m_2}\nabla_2^2 + U(\mathbf{r}_1 - \mathbf{r}_2)\right]\psi = \mathscr{E}\psi.$$

Здесь

$$\nabla_1^2 = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial y_1^2} + \frac{\partial^2}{\partial z_1^2}, \quad \nabla_2^2 = \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial y_2^2} + \frac{\partial^2}{\partial z_2^2}, \tag{26.2}$$

а через $\mathbf{r}_1(x_1, y_1, z_1)$ и $\mathbf{r}_2(x_2, y_2, z_2)$ обозначены координаты рассматриваемых частиц. Потенциальная функция взаимодействия $U(\mathbf{r}_1 - \mathbf{r}_2)$ зависит только от разностей координат первой и второй частиц. Обычно (в случае центральных сил) она является функцией только расстояния между ними $r \equiv |\mathbf{r}_1 - \mathbf{r}_2|$.

2. Преобразуем уравнение (26.2) к новым независимым переменным: координатам центра масс

$$\mathbf{R}(X, Y, Z) = \frac{m_1\mathbf{r}_1 + m_2\mathbf{r}_2}{m_1 + m_2} \tag{26.3}$$

и координатам первой частицы относительно второй

$$\mathbf{r}(x, y, z) = \mathbf{r}_1 - \mathbf{r}_2. \tag{26.4}$$

Величина ψ теперь является функцией шести переменных: $\psi = \psi(\mathbf{r}_1, \mathbf{r}_2)$. Но мы для сокращения выкладок (без нарушения общности) произведем их только для функции $\psi(x_1, x_2)$ двух переменных x_1 и x_2 и соответственно для $\psi(X, x)$, где

$$X = \frac{m_1 x_1 + m_2 x_2}{m_1 + m_2}, \quad x = x_1 - x_2.$$

На основании инвариантности полного дифференциала

$$\frac{\partial\psi}{\partial x_1}\,dx_1 + \frac{\partial\psi}{\partial x_2}\,dx_2 = \frac{\partial\psi}{\partial X}\,dX + \frac{\partial\psi}{\partial x}\,dx.$$

Подставим в правую часть

$$dX = \frac{m_1 dx_1 + m_2 dx_2}{m_1 + m_2}, \quad dx = dx_1 - dx_2.$$

Сравнением коэффициентов получаем

$$\frac{\partial\psi}{\partial x_1}=\left(\frac{m_1}{m_1+m_2}\,\frac{\partial}{\partial X}+\frac{\partial}{\partial x}\right)\psi,\quad \frac{\partial\psi}{\partial x_2}=\left(\frac{m_2}{m_1+m_2}\,\frac{\partial}{\partial X}+\frac{\partial}{\partial x}\right)\psi.$$

Мы представили результат дифференцирования в операторной форме. Это позволяет сразу написать выражения для вторых производных:

$$\frac{\partial^2\psi}{\partial x_1^2}=\left(\frac{m_1}{m_1+m_2}\,\frac{\partial}{\partial X}+\frac{\partial}{\partial x}\right)^2\psi=$$
$$=\frac{m_1^2}{(m_1+m_2)^2}\,\frac{\partial^2\psi}{\partial X^2}+2\frac{m_1}{m_1+m_2}\,\frac{\partial\psi}{\partial x}+\frac{\partial^2\psi}{\partial x^2},$$

$$\frac{\partial^2\psi}{\partial x_2^2}=\left(\frac{m_1}{m_1+m_2}\,\frac{\partial}{\partial X}+\frac{\partial}{\partial x}\right)^2\psi=$$
$$=\frac{m_2^2}{(m_1+m_2)^2}\,\frac{\partial^2\psi}{\partial X^2}-2\frac{m_2}{m_1+m_2}\,\frac{\partial\psi}{\partial x}+\frac{\partial^2\psi}{\partial x^2}.$$

Поделив эти соотношения соответственно на m_1 и m_2 и сложив, получим

$$\frac{1}{m_1}\,\frac{\partial^2\psi}{\partial x_1^2}+\frac{1}{m_2}\,\frac{\partial^2\psi}{\partial x_2^2}=\frac{1}{m_1+m_2}\,\frac{\partial^2\psi}{\partial X^2}+\left(\frac{1}{m_1}+\frac{1}{m_2}\right)\frac{\partial^2\psi}{\partial x^2}.$$

Теперь уже легко перейти к трем переменным, а затем представить уравнение Шредингера в виде

$$\left[-\frac{\hbar^2}{2(m_1+m_2)}\nabla_{\mathbf{R}}^2-\frac{\hbar^2}{2m}\nabla^2+U(\mathbf{r}_1-\mathbf{r}_2)\right]\psi=\mathscr{E}\psi,\qquad(26.5)$$

где $\nabla_{\mathbf{R}}^2$ и ∇^2 — операторы Лапласа соответственно в переменных X, Y, Z и x, y, z.

3. Оператор, действующий на ψ в левой части (26.5), распадается на сумму двух независимых членов, один из которых зависит только от $\mathbf{R}$, а другой только от $\mathbf{r}=\mathbf{r}_1-\mathbf{r}_2$. В соответствии с этим решение уравнения (26.5) сводится к решению двух уравнений:

$$-\frac{\hbar^2}{2(m_1+m_2)}\nabla_{\mathbf{R}}^2\psi(\mathbf{R})=\mathscr{E}_{\mathbf{R}}\psi(\mathbf{R}),\qquad(26.6)$$

$$-\frac{\hbar^2}{2m}\nabla^2\psi(\mathbf{r})+U(\mathbf{r})\psi(\mathbf{r})=\mathscr{E}_{\mathbf{r}}\psi(\mathbf{r}),\qquad(26.7)$$

где $\mathscr{E}_{\mathbf{R}}$ и $\mathscr{E}_{\mathbf{r}}$ — постоянные, удовлетворяющие условию $\mathscr{E}_{\mathbf{R}}+\mathscr{E}_{\mathbf{r}}=\mathscr{E}$. Из них первое описывает *свободное движение центра масс* системы — воображаемой частицы с массой m_1+m_2. Второе же описывает *относительное движение первой частицы относительно второй*, в нем истинная масса частицы заменена приведенной массой $m=$ $=m_1m_2/(m_1+m_2)$. В самом деле, произведение $\psi(\mathbf{r})\psi(\mathbf{R})$, в котором переменным является $\mathbf{R}$, а $\mathbf{r}$ рассматривается как параметр, т. е.

в сущности как постоянная, описывает то же движение центра масс, что и функция $\psi(\mathbf{R})$. Аналогично, то же произведение $\psi(\mathbf{r})\psi(\mathbf{R})$ будет описывать относительное движение, если за переменную принять $\mathbf{r}$, а $\mathbf{R}$ рассматривать как параметр. Если теперь (26.6) умножить на $\psi(\mathbf{r})$, а (26.7) на $\psi(\mathbf{R})$ и оба уравнения сложить почленно, то мы получим, что функция $\psi = \psi(\mathbf{r})\psi(\mathbf{R})$ будет *общим решением* уравнения (26.5). Таким образом, *общая волновая функция, описывающая независимые движения центра масс и относительное движение частиц, распадается на произведение двух функций от различных переменных*: $\psi(\mathbf{r})$ и $\psi(\mathbf{R})$. При этом, как и в классической механике полная энергия распадается на сумму энергии, связанной с движением центра масс системы, и энергии относительного движения частиц.

Если отвлечься от движения центра масс, считая его как бы неподвижным, то уравнение (26.6) отпадает. Остается только уравнение (26.7) для относительного движения частиц. Поэтому это уравнение мы будем писать просто в виде

$$-\frac{\hbar^2}{2m}\nabla^2\psi + U(\mathbf{r})\psi = \mathscr{E}\psi. \tag{26.8}$$

ЗАДАЧА

Дейтрон состоит из связанных протона и нейтрона Они удерживают друг друга посредством короткодействующих *ядерных сил*. Потенциальную функцию взаимодействия можно аппроксимировать пространственной потенциальной ямой прямоугольной формы с глубиной $-U_0$ и радиусом a (расстояние между центрами протона и нейтрона). Дейтрон имеет *только одно* связанное состояние. Энергия связи дейтрона, измеренная экспериментально, составляет 2,225 МэВ. Этого недостаточно для определения двух неизвестных U_0 и a. Зададим $a = 2 \cdot 10^{-13}$ см (это недалеко от истины). Мы не настаиваем, что это есть точное значение a. Наша цель — привести только схему расчета. Из этих данных определить глубину потенциальной ямы $-U_0$.

Р е ш е н и е. Представляет интерес только относительное движение протона и нейтрона. Поэтому можно воспользоваться уравнением (26.8), понимая под m приведенную массу системы. Если пренебречь различием масс протона и нейтрона, то приведенная масса будет $m/2$, где m — масса одной из частиц (например, протона). В дальнейших вычислениях используются следующие постоянные:

$$mc^2 = 938{,}28 \text{ МэВ}, \quad \hbar c = 1{,}97329 \cdot 10^{-11} \text{ МэВ} \cdot \text{см}.$$

Так как у дейтрона только одно связанное состояние, то его энергия в этом состоянии $\mathscr{E} = -2{,}225$ МэВ. Это позволяет по формулам (24.6) и (24.12) найти η. Только в формуле (24.6) m следует заменить на $m/2$. Это дает

$$\eta^2 = -\frac{m\mathscr{E}a^2}{\hbar^2} = -\frac{mc^2\mathscr{E}a^2}{\hbar^2c^2} = 0{,}21437, \quad \eta = 0{,}463099.$$

Величину ξ находим из уравнения

$$\eta = -\xi \operatorname{ctg} \xi.$$

Сначала решаем это уравнение грубо графически, пользуясь крайней левой кривой на рис. 46 *б*. Затем уточняем решение аналитически с использованием

интерполирования. Таким путем без труда находим

$$\xi = 1{,}81993.$$

Искомая глубина потенциальной ямы

$$-U_0 = \frac{\hbar^2 c^2}{mc^2 a^2}(\xi^2 + \eta^2) = 36{,}61 \text{ МэВ}.$$

§ 27. Квантование водородоподобного атома в сферически симметричном случае

1. Приведем еще один пример на квантование энергии атомной системы. Речь идет о водородоподобном атоме. Рассмотрим частный случай, когда волновая функция ψ электрона в атоме *сферически симметрична*, т. е. зависит только от радиуса r — расстояния электрона от атомного ядра. Такой случай не предусматривался старой теорией Бора. В ней всякое движение электрона вокруг ядра происходило по *плоским орбитам* и, следовательно, не могло быть сферически симметричным. Но в квантовой механике, в которой нет представления о движении электронов по орбитам, нет никаких препятствий для реализации сферически симметричных состояний атома. Из сферической симметрии следует, что в таких состояниях должна обращаться в нуль величина, соответствующая тому, что в классической механике называется моментом количества движения. В теории Бора нулевым моментом количества движения обладал бы электрон, движущийся прямолинейно вдоль радиуса. При таких движениях он неизменно претерпевал бы столкновения с атомным ядром. Старая теория Бора не давала удовлетворительного решения возникавшей здесь трудности, — чтобы избежать столкновений с ядром, она просто исключала возможность радиальных движений электрона. Понятно, что в квантовой механике подобной трудности не возникает.

2. Пусть Ze — заряд ядра. Естественно записать уравнение Шредингера в полярных координатах. В рассматриваемом случае сферической симметрии оно будет

$$\frac{d^2\psi}{dr^2} + \frac{2}{r}\frac{d\psi}{dr} + \left(\frac{q}{r} - \beta^2\right)\psi = 0, \tag{27.1}$$

где введены обозначения

$$\beta^2 = -\frac{2m\mathscr{E}}{\hbar^2}, \quad q = \frac{2mZe^2}{\hbar^2}. \tag{27.2}$$

Введем новую функцию $u(r)$ по формуле

$$\psi = \frac{u(r)}{r}\, e^{-\beta r}.$$

Тогда

$$\frac{d^2u}{dr^2} - 2\beta\frac{du}{dr} + \frac{q}{r}\,u = 0. \tag{27.3}$$

Ищем решение этого уравнения в виде ряда

$$u = \sum_{k=\gamma}^{\infty} a_k r^k, \tag{27.4}$$

где γ — постоянное число, пока что не определенное. Подставляя (27.4) в (27.3) и приравнивая члены с одинаковыми степенями, придем к соотношениям

$$\gamma(\gamma - 1) = 0, \tag{27.5}$$

$$k(k+1)a_{k+1} - 2\beta k a_k + q a_k = 0 \quad \text{при} \quad k \neq \gamma. \tag{27.6}$$

Из (27.5) следует, что либо $\gamma = 0$, либо $\gamma = 1$. Первая возможность исключается, так как при $\gamma = 0$ нулевой член ряда (27.4), т. е. a_0, был бы отличен от нуля. А в таком случае функция ψ при $r = 0$ обращалась бы в бесконечность как a_0/r, что противоречит общим требованиям, накладываемым на ψ в особых точках. Таким образом, разложение (27.4) должно начинаться с $k = 1$, а это значит, что $\gamma = 1$.

Исследуем теперь поведение ряда (27.4) на бесконечности. Из (27.6) получаем

$$\frac{a_{k+1}}{a_k} = \frac{2\beta k - q}{k(k+1)}. \tag{27.7}$$

Отсюда следует, что при $k \to \infty$

$$\frac{a_{k+1}}{a_k} \to \frac{2\beta}{k+1}.$$

Сравним разложение (27.4) с разложением показательной функции:

$$e^{2\beta r} = \sum_{k=0}^{\infty} c_k r^k = \sum_{k=0}^{\infty} \frac{1}{k!} (2\beta r)^k.$$

Коэффициенты c_k последнего разложения асимптотически ведут себя на бесконечности так же, как и коэффициенты a_k, ибо

$$\frac{c_{k+1}}{c_k} = \frac{2\beta}{k+1}.$$

Значит, на бесконечности сумма ряда (27.4) асимптотически ведет себя как показательная функция $e^{+2\beta r}$, а волновая функция $\psi(r)$ — как $e^{\beta r}/r$, т. е. при произвольно выбранном значении $\mathscr{E}$ функция $\psi(r)$ при $r = \infty$ обращается в бесконечность. Этого не будет только для таких значений $\mathscr{E}$, при которых ряд (27.4) обрывается, т. е. переходит в сумму конечного числа членов. Пусть, например, при $k = n$ числитель формулы (27.7) $2\beta k - q = 0$. Тогда, как видно из (27.7), a_{n+1} и все последующие коэффициенты будут равны нулю, т. е. ряд (27.4) оборвется. Следовательно, n-й энергетический уровень определится условием $2\beta n - q = 0$. Используя его, из (27.2) находим

$$\mathscr{E} = -\frac{mZ^2e^4}{2\hbar^2 n^2}, \tag{27.8}$$

что совпадает с соответствующей формулой теории Бора.

Изложенным еще не решается задача о спектре водородного и водородоподобного атомов, даже в ее наиболее грубой постановке. Чтобы объяснить спектральные серии, необходимо схему энергетических уровней дополнить *правилами отбора* при излучении фотонов. Оказывается, что переходы между найденными нами уровнями энергии, соответствующими сферически симметричным состояниям водородоподобного атома, являются *запрещенными*, т. е. не сопровождаются (дипольным) излучением. Для объяснения спектральных серий необходимо рассмотреть сферически несимметричные состояния водородоподобного атома и установить *правила отбора*. Это будет сделано ниже (см. § 39).

3. Для сравнения с теорией Бора найдем еще волновую функцию $\psi_1(r)$ основного состояния в сферически симметричном случае. Так называется стационарное состояние наименьшей энергии. Посмотрим, при каких значениях параметров $\mathscr{E}$ и a_1 уравнению (27.1) удовлетворяет экспоненциальная функция

$$\psi_1(r) = e^{-r/a_1}, \tag{27.9}$$

где $a_1 > 0$. Эта функция не имеет узлов. Поэтому, если $\psi_1(r)$ удовлетворяет уравнению Шредингера (27.1), то она и будет волновой функцией основного состояния. Дифференцируя $\psi_1(r)$ дважды по r и подставляя результаты в (27.1), получим

$$\frac{1}{a_1^2} - \frac{2}{a_1 r} + \frac{q}{r} - \beta^2 = 0.$$

Это соотношение должно выполняться тождественно по r, а потому должно быть

$$\frac{1}{a_1^2} = \beta^2, \quad \frac{2}{a_1} = q,$$

или на основании (27.2)

$$a_1 = \frac{\hbar}{mZe^2}, \quad \mathscr{E} = \mathscr{E}_1 = -\frac{mZ^2e^4}{2\hbar^2}. \tag{27.10}$$

Последнее выражение является частным случаем (27.8) при $n = 1$, как и должно быть для основного состояния. Параметр a_1 имеет размерность длины, при $Z = 1$ он обращается в боровский радиус. О физическом смысле этого параметра в квантовой механике будет сказано несколько ниже.

Функция $\psi_1(r)$ при $r = 0$ обращается в единицу, т. е. остается конечной. Следовательно, при $r = 0$ $u(r) = r\psi_1(r) = 0$, как того требует и общая теория.

Если функция ψ_1 нормирована, то $|\psi_1|^2$ дает объемную плотность вероятности обнаружения электрона

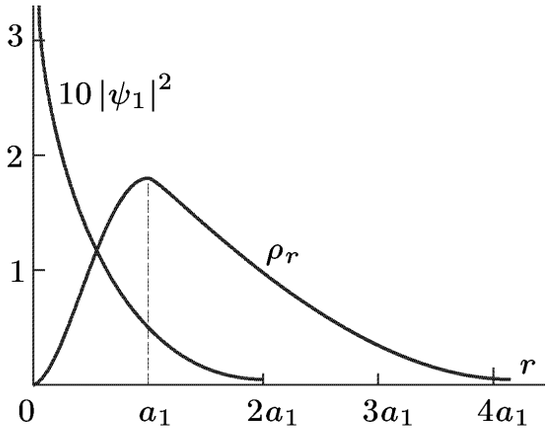


Рис. 49

в пространстве. Наряду с ней введем *радиальную плотность вероятности* ρ_r. Вероятность обнаружения электрона в сферическом слое между r и $r + dr$ равна объему этого слоя $4\pi r^2 dr$, умноженному на $|\psi_1|^2$, т. е. $4\pi r^2 |\psi_1|^2\, dr$. Эту вероятность можно представить в виде $\rho_r dr$. Величина ρ_r и есть радиальная плотность вероятности — произведение $\rho_r dr$ дает вероятность того, что электрон будет обнаружен на расстоянии от ядра между r и $r + dr$. Таким образом,

$$\psi_1 = Ce^{-r/a_1},$$

$$\rho_r = 4\pi C^2 r^2 e^{-2r/a_1}.$$

Интегрируя второе выражение по r в пределах от 0 до $+\infty$ и приравнивая результат единице, находим нормировочную постоянную C и таким путем получаем

$$\psi_1 = \sqrt{\frac{1}{\pi a_1^3}}\, e^{-r/a_1}, \quad \rho_r = \frac{4}{a_1^3}\, r^2 e^{-2r/a_1}. \tag{27.11}$$

На рис. 49 представлены графики кривых $|\psi_1|^2$ и ρ_r. Ординаты первой кривой увеличены в десять раз. Кривая ρ_r проходит через максимум при $r = a_1$. Следовательно, в квантовой механике радиус первой боровской орбиты надо истолковать как такое расстояние от ядра, на котором вероятность обнаружения электрона максимальна.

ЗАДАЧИ

1. Найти среднее расстояние $\overline{r}$, на каком будет обнаружен электрон от ядра атома, если последний находится в основном состоянии.

Ответ. $\overline{r} = (3/2)a_1$.

2. В той же задаче найти среднее значение $\overline{(1/r)}$ обратного расстояния электрона от ядра.

Ответ. $\overline{(1/r)} = 1/a_1$.

3. Найти средние значения потенциальной $\overline{U}$ и кинетической $\overline{\mathscr{E}}_{\text{кин}}$ энергий основного состояния водородоподобного атома.

Ответ. $\overline{U} = -Ze^2/a_1$; $\overline{\mathscr{E}}_{\text{кин}} = Ze^2/2a_1 = -\overline{U}/2$. Отметим, что такое же соотношение между U и $\mathscr{E}_{\text{кин}}$ получилось бы в классической механике для электрона, движущегося вокруг ядра по всякой круговой орбите.

4. Определить уровни энергии в сферически симметричном состоянии водородоподобного атома по числу узлов волновой функции, подобно тому как это было сделано в § 23 для гармонического осциллятора.

Решение. Волновые функции возбужденных состояний должны иметь узлы, число которых на единицу меньше номера соответствующего стационарного состояния. Этому условию для n-го стационарного состояния удовлетворяет выражение

$$\psi_n(r) = P_{n-1}(r)e^{-r/a_n},$$

где a_n — положительная постоянная, а $P_{n-1}(r)$ — полином степени $n - 1$, все корни которого вещественны и различны. Необходимо, чтобы функция

$\psi_n(r)$ удовлетворяла уравнению Шредингера (27.1). Простым дифференцированием находим

$$\frac{d\psi_n}{dr} = \left(-\frac{1}{a_n} P_{n-1} + P'_{n-1}\right) e^{-r/a_n},$$

$$\frac{d^2\psi_n}{dr^2} = \left(\frac{1}{a_n^2} P_{n-1} - \frac{2}{a_n} P'_{n-1} + P''_{n-1}\right) e^{-r/a_n}.$$

После подстановки в (27.1) получаем

$$\frac{1}{a_n^2} P_{n-1} - \underline{\frac{2}{a_n} P'_{n-1}} + P''_{n-1} - \underline{\frac{2}{a_n}\frac{P_{n-1}}{r}} + \frac{2P'_{n-1}}{r} + \underline{\frac{qP_{n-1}}{r}} - \beta^2 P_{n-1} = 0.$$

Это соотношение должно выполняться тождественно по r. Старшую степень r^{n-1} содержат только первое и последнее слагаемые. Поэтому должно быть

$$1/a_n^2 = \beta^2 \quad \text{или} \quad 1/a_n = \beta.$$

Степень r^{n-2} содержат только подчеркнутые члены. При этом при взятии производной P_{n-1} появляется коэффициент $(n-1)$. С учетом этого

$$-\frac{2(n-1)}{a_n} - \frac{2}{a_n} + q = 0, \quad \text{или} \quad \frac{2n}{a_n} = q.$$

Таким образом,

$$a_n = \frac{2n}{q} = n\frac{\hbar^2}{mZe^2} = na_1,$$

$$\mathscr{E} = -\frac{\hbar^2\beta}{2m} = -\frac{\hbar^2}{2ma^2} = \frac{mZ^2e^4}{2\hbar^2 n^2},$$

что совпадает с ранее полученными результатами.

Недостаток приведенного решения — в том, что мы не исследовали до конца, что наша функция $\psi_n(r)$ действительно является решением уравнения Шредингера. Для небольших n, подобно тому как это было сделано в § 23, нетрудно найти в явном виде полиномы $P_{n-1}(r)$ и соответствующие им постоянные a_n. Таким путем можно убедиться, что функции $\psi_n(r) = P_{n-1}(r)e^{-r/a_n}$ действительно удовлетворяют уравнению Шредингера. Можно проверить также, что все корни полинома $P_{n-1}(r)$ вещественные и некратные.

§ 28. Потенциальные барьеры

1. К задаче о квантовании энергии в потенциальных ямах примыкает задача о прохождении частицы через потенциальные барьеры. Ограничимся рассмотрением *одномерных потенциальных барьеров*, когда потенциальная функция U зависит только от одной координаты x. Потенциальным барьером такого типа называется ограниченная параллельными плоскостями область пространства, в которой потенциальная функция $U(x)$ больше, чем в примыкающих областях.

Начнем с простейшего идеализированного случая *прямоугольного потенциального барьера*, когда одна из его стенок удалена в бесконечность (рис. 50). Такой барьер может быть назван *ступенчатым*, так как потенциальная функция $U(x)$ в этом случае представляется

ступенчатой линией:

$$U(x) = \begin{cases} U_1 = \text{const} & \text{в области I,} \quad \text{где } x < 0, \\ U_2 = \text{const} & \text{в области II,} \quad \text{где } x > 0, \end{cases} \tag{28.1}$$

причем $U_2 > U_1$. На границу барьера слева с постоянной скоростью налетает частица или поток частиц. С классической точки зрения частица ведет себя по-разному в зависимости от того, будет ли ее полная энергия $\mathscr{E}$ больше или меньше U_2. В первом случае, когда $\mathscr{E} > U_2$, частица, достигнув границы барьера, будет продолжать движение в прежнем направлении, но с меньшей кинетической энергией. Во втором случае, когда $\mathscr{E} < U_2$, частица вообще не может проникнуть через границу барьера. Она отразится от него и начнет движение в обратном направлении с той же кинетической энергией.

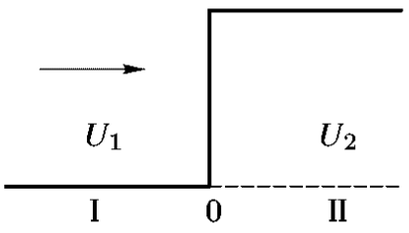


Рис. 50

2. Совсем иное решение задачи дает квантовая механика. Здесь движение частицы, хотя и символически, связано с распространением волны. Основное уравнение квантовой механики — уравнение Шредингера — описывает (и притом детерминистически) распространение именно *волн*, а не движение частиц. Переход же от поведения волн к движению частиц устанавливается *вероятностными законами.* Поэтому поставленная нами задача должна быть переформулирована, а затем решена для волн на основе уравнения Шредингера. Последнее мы будем записывать в виде

$$\frac{d^2\psi}{dx^2} + k^2\psi = 0, \tag{28.2}$$

где

$$k^2 = \frac{2m}{\hbar^2}(\mathscr{E} - U), \tag{28.3}$$

причем U имеет разные, но постоянные значения U_1 и U_2 по разные стороны границы барьера. Соответствующие им значения k обозначаются через k_1 и k_2.

Вместо потока частиц теперь надо предположить, что в области I к границе барьера распространяется плоская монохроматическая волна

$$\psi_1 = e^{i(k_1 x - \omega t)}.$$

Чтобы удовлетворялись граничные условия для ψ и $d\psi/dx$ на границе барьера, в области I должна существовать отраженная волна

$$\psi_1' = re^{-i(k_1 x + \omega t)},$$

а в области II — прошедшая волна

$$\psi_2 = de^{i(k_2 x - \omega t)}.$$

Амплитуда падающей волны принята равной единице, что, очевидно, не нарушает общности получаемых ниже результатов. Постоянные r

и d называются *амплитудными коэффициентами* отражения и пропускания волн. Для их определения заметим, что функция ψ и ее производная по x на границе барьера должны быть непрерывны. Это значит, что при $x = 0$ должны выполняться соотношения

$$\psi_1 + \psi_1' = \psi_2, \quad \frac{d}{dx}(\psi_1 + \psi_1') = \frac{d\psi_2}{dx},$$

или

$$1 + r = d, \quad k_1 - k_1 r = k_2 d.$$

Отсюда находим

$$r = \frac{k_1 - k_2}{k_1 + k_2}, \quad d = \frac{2k_1}{k_1 + k_2}. \tag{28.4}$$

Это такие же соотношения, но записанные в иной форме, которые были получены в оптике для *коэффициентов Френеля* (1788–1827) при нормальном падении света на границу раздела сред (см. т. IV, § 65, 69). Они справедливы не только при $U_2 > U_1$ (потенциальный барьер), но и при $U_2 < U_1$ (потенциальная яма).

3. Принципиальное отличие квантовомеханического решения от классического состоит в том, что в классической физике частица *локализована*, а в квантовой механике — нет. В классической физике говорят об энергии и состоянии частицы, когда она находится в определенном месте пространства, независимо от того, что происходит в остальных местах пространства. В квантовой механике это не так. Решение, даваемое квантовой механикой, — волна, есть понятие, относящееся *ко всему пространству*. Падающая волна органически связана с отраженной и прошедшей волнами. Нельзя выделить одну из этих волн, отвлекаясь от остальных. Полная энергия $\mathscr{E}$ относится не к какой-либо одной волне, а к *состоянию частицы в целом*, определяемому всеми тремя функциями ψ_1, ψ_1', ψ_2. Понимание этого обстоятельства позволяет избежать многих парадоксальных выводов, связанных с прохождением частиц через потенциальные барьеры.

Заметим еще, что задача об определении амплитудных коэффициентов отражения и пропускания волн есть чисто *детерминистическая задача*. Она формулируется и решается в стиле классической физики — на основе точно сформулированного уравнения Шредингера и соответствующих ему граничных условий. Но не эти коэффициенты определяют реальные величины, с которыми приходится иметь дело на опыте. На опыте измеряются *коэффициенты отражения и пропускания* не для волн, а *для частиц*. Они же связаны с амплитудными коэффициентами отражения и пропускания волн *вероятностными соотношениями*. Коэффициенты отражения и пропускания для частиц определяются ниже. Таким образом, отражение частиц от потенциального барьера и прохождение через него определяются вероятностными законами.

4. При сравнении квантовомеханического решения с классическим рассмотрим сначала случай $\mathscr{E} > U_2$. В этом случае все три волны — падающая, отраженная и прошедшая — *однородны*. Отличие квантового случая от классического состоит прежде всего в том, что в классическом случае нет отраженного потока частиц. В квантовом же

случае неизбежно появляется отраженная волна, а с ней и вероятность обнаружить частицу, движущуюся навстречу падающему потоку. Для однородной волны можно ввести понятие *плотности вероятности потока вещества*. В самом деле, *однородный поток не локализован*, он характеризуется определенной *плотностью импульса*, тогда как его координата совершенно не определена. Можно говорить и о *скорости распространения вероятности* такого потока. Она просто совпадает с классической скоростью и равна $v = p/m = \hbar k/m$. Наконец, плотность вероятности потока массы вещества равна $mv\psi^*\psi = \hbar k\psi^*\psi$. В падающей волне эта величина равна $\hbar k_1 \psi_1^* \psi_1 = \hbar k_1$. Аналогично, плотности вероятности потока вещества в отраженной и прошедшей волнах равны соответственно $|r|^2 \hbar k_1$ и $|d|^2 \hbar k_2$. Отношение плотности вероятности потока массы в отраженной волне к плотности вероятности потока массы в падающей волне называется *коэффициентом отражения частицы* R. Аналогично определяется *коэффициент пропускания частицы* D. Он называется также *пропускаемостью* или *прозрачностью барьера*. Для этих величин находим

$$R = |r|^2 = \left|\frac{k_1 - k_2}{k_1 + k_2}\right|^2, \quad D = \frac{k_2}{k_1}|d|^2 = \frac{4k_1 k_2}{(k_1 + k_1)^2}, \tag{28.5}$$

так что $R + D = 1$, в согласии с законом сохранения вещества.

5. Обратимся теперь к случаю, когда $\mathscr{E} < U_2$. В этом случае формулы (28.4), конечно, также остаются справедливыми. Остается справедливой и первая формула (28.5), поскольку отраженная волна по-прежнему однородна. Однако величина k_2 будет *чисто мнимой*, так что волна во второй области станет *неоднородной*. В первой же формуле (28.5) числитель и знаменатель будут величинами комплексно сопряженными. Значит, $R = 1$, т. е. отражение частиц становится *полным*, как и в аналогичном случае в оптике. Однако волна во второй области не исчезает. Действительно, полагая $k_2 = i\alpha$, для этой волны получаем

$$\psi_2 = \frac{2k_1}{k_1 + k_2} e^{-\alpha x} e^{-i\omega t}, \tag{28.6}$$

т. е. амплитуда волны в области II экспоненциально затухает при удалении от границы раздела областей. Глубина проникновения l определяется как расстояние, на котором плотность вероятности потока вещества убывает в e раз. Для нее получаем

$$l = \frac{1}{2\alpha} = \frac{\lambda_2}{4\pi}, \tag{28.7}$$

где

$$\lambda_2 = \frac{2\pi\hbar}{\sqrt{2m(U_2 - \mathscr{E})}} = \frac{h}{\sqrt{2m(U_2 - \mathscr{E})}}. \tag{28.8}$$

Таким образом, волна проникает в область II, несмотря на то, что она отражается полностью, а вероятность отражения частицы обращается в единицу. Разрешение возникающего здесь кажущегося парадокса в точности такое же, как и в случае полного отражения света

(см. т. IV, § 66). Наше решение относится к *стационарному состоянию*, поскольку оно основано на уравнении Шредингера именно для таких состояний. Проникновение же волны во вторую область происходит в переходный период, когда состояние во времени еще не установилось. В этот переходный период полного отражения волны еще не может быть. Исследование же переходного периода может быть осуществлено на основе уравнения Шредингера, но уже для нестационарных состояний (21.5).

6. Подчеркнем еще раз, что в найденном нами стационарном состоянии, описываемом тремя волновыми функциями ψ_1, ψ_1', ψ_2 частица не локализована. Она может с той или иной вероятностью находиться в любой точке пространства. Общим для всего этого состояния является параметр $\mathscr{E}$, названный нами *полной энергией* частицы. Следует с осторожностью отождествлять это понятие с полной энергией, как она понимается в классической механике. Так, мы уже указывали, что в квантовой механике не всегда имеет смысл деление полной энергии на кинетическую и потенциальную (см. § 20, п. 8). Чтобы определить параметр $\mathscr{E}$ надо произвести *измерение*, т. е. как-то воздействовать на частицу.

7. Рассмотрим сначала состояние частицы в части пространства I. Оно представляет собой суперпозицию двух плоских монохроматических волн ψ_1 и ψ_1', распространяющихся навстречу друг другу. Их волновые числа имеют определенные значения, одинаковые по величине. Поэтому одинаковы по величине и импульсы, соответствующие обеим волнам. Измеряя импульс, когда частица находится в части пространства I, мы найдем, что он с той или иной вероятностью равен либо $p_1 = \hbar k_1$, либо $p_1' = -\hbar k_1$. Согласно соотношению неопределенностей Гейзенберга определенность импульса в каждой волне связана с тем, что *частица не локализована.* Действительно, неопределенность координаты Δx бесконечно велика, и потому, согласно соотношению (20.2), неопределенность импульса Δp для каждой волны обращается в нуль. Учитывая соотношение $p = \hbar k$, формулу (28.3) можно переписать в виде

$$\mathscr{E} = \frac{p^2}{2m} + U, \tag{28.9}$$

т. е. $\mathscr{E}$, как и в классической механике, равна сумме кинетической и потенциальной энергии. Такое совпадение с классической механикой обусловлено тем, что потенциальная функция U_1 во всем пространстве I постоянна, т. е. это пространство свободно от сил. К тому же результату мы придем и в пространстве II, если только в этом пространстве $\mathscr{E} > U_2$, и, следовательно, волна однородна.

8. Рассмотрим теперь область II в случае, когда $\mathscr{E} < U$. Вероятность обнаружить частицу в области II в этом случае рассматривалась как парадокс. Основанием для этого является соотношение (28.9), из которого следует, что всегда $\mathscr{E} > U$, так как кинетическая энергия $p^2/2m$ существенно положительна. Однако, как уже неоднократно

подчеркивалось, формула (28.9) есть *соотношение классической механики* и неприменима при $\mathscr{E} < U$. В этом случае волна де Бройля неоднородна и обычные выражения импульса и кинетической энергии частицы теряют смысл. Однако обнаружить частицу в области II возможно, поскольку вероятность такого обнаружения не обращается в нуль, а лишь экспоненциально убывает по мере удаления от границы барьера в сторону положительных x. Обнаружить частицу — это значит указать границы, между которыми она окажется в результате обнаружения. Практически частицу можно обнаружить только в тонком поверхностном слое вблизи границы барьера, толщина которого порядка глубины проникновения l. Величина l и может быть принята за неопределенность координаты после обнаружения частицы. Неопределенность импульса обозначим через Δp. Тогда в силу соотношения неопределенностей (20.4)

$$\overline{\Delta p^2} \cdot l^2 \geqslant \frac{\hbar^2}{4}.$$

Подставляя сюда значение l из (28.7) и (28.8), получим

$$\overline{\frac{\Delta p^2}{2m}} \geqslant U_2 - \mathscr{E},$$

т. е. для локализации частицы в области II в слое толщины l ей необходимо сообщить кинетическую энергию, величина которой во всяком случае не меньше $U_2 - \mathscr{E}$, т. е. положительна. Такую энергию частица может, например, получить при освещении ее световым квантом достаточно короткой длины волны (эффект Комптона). Понятно, что такая локализация меняет квантовое состояние частицы. После взаимодействия со световым квантом волновая функция частицы будет отличаться от нуля только внутри выбранного нами слоя толщины l, обращаясь в нуль вне этого слоя.

Не обязательно, чтобы слой толщины l, в котором обнаруживается частица, располагался у края барьера. Он может быть расположен где угодно в области II. От его положения зависит лишь величина вероятности обнаружения частицы в слое. Но энергия, которую надо сообщить частице при ее локализации в слое толщины l, *зависит только от толщины слоя*, а не от его положения. Толщина же слоя определяется экспонентой $e^{-2\alpha l}$ и от положения слоя не зависит.

Иллюстрируем роль измерения еще на следующем примере. Частица должна быть локализована внутри слоя толщины l. С этой целью осветим ее пучком света, распространяющимся вдоль слоя перпендикулярно к оси X. Если произойдет рассеяние света, то это и будет означать, что частица в момент рассеяния была локализована внутри рассматриваемого слоя. Из оптики известно, что длина световой волны для локализации должна быть короче l, т. е. $\lambda < l$. Из формул (28.7) и (28.8) получаем

$$\lambda < \frac{h}{4\pi\sqrt{2m(U_2 - \mathscr{E})}},$$

или

$$(hc/\lambda)^2 \equiv (h\nu)^2 > 32\pi^2 mc^2(U_2 - \mathscr{E}).$$

Нерелятивистская механика применима к процессам, когда энергия светового кванта $h\nu$ мала по сравнению с собственной энергией частицы mc^2. Поэтому, разделив левую часть предыдущего неравенства на меньшую величину $h\nu$, а правую на бóльшую $32\pi^2 mc^2$, получим

$$h\nu > U_2 - \mathscr{E}.$$

Таким образом, для локализации должны применяться световые кванты, энергия которых во всяком случае не меньше разности между потенциальной и полной энергиями частицы. Это находится в согласии с тем, что было сказано выше.

9. Заканчивая рассмотрение ступенчатого барьера, выведем некоторые общие соотношения, связывающие амплитудные коэффициенты отражения и пропускания волн де Бройля на границе барьера. Если переменить на противоположные направления распространения всех волн де Бройля без изменения их амплитуд, то уравнение Шредингера и соответствующие ему граничные условия будут по-прежнему удовлетворены. Отсюда следует, что если возникло состояние, изображенное на рис. 51 *а*, то возможно также и состояние, изображенное на рис. 51 *б*. На этих рисунках каждая волна де Бройля представлена двумя символами. Первый из них представляет амплитуду, а второй — волновое число соответствующей волны, распространяющейся в положительном или отрицательном направлении оси X. Направления распространения волн обозначены стрелками. На рис. 51 *а* есть только одна, а на рис. 51 *б* — две падающие волны. Обозначим через r' и d' амплитудные

Рис. 51

коэффициенты отражения и пропускания, когда падающая волна распространяется справа налево из области II к области I. Падающая волна (r, k_1) дает отраженную волну $(r^2, -k_1)$. Падающая волна $(d, -k_2)$ возбуждает проходящую волну $(dd', -k_1)$. Обе возбужденные волны должны при наложении дать уходящую волну $(1, -k_1)$. Таким образом, должно быть

$$r^2 + dd' = 1. \tag{28.10а}$$

Аналогично, волна (r, k_1) возбуждает проходящую волну (rd, k_2), а волна $(d, -k_2)$ — отраженную волну (dr', k_2). Обе возбужденные волны должны гасить друг друга, т. е. $rd + dr' = 0$, откуда

$$r' = -r. \tag{28.10б}$$

Соотношения (28.10а) и (28.10б) справедливы как для однородных, так и для неоднородных волн. Они были уже получены в т. IV (§ 67)

для световых волн. Применим их к прямоугольному потенциальному барьеру (или яме) конечной ширины.

Барьер вместе с падающей волной изображен на рис. 52 *а*. Более детальная картина падающей, прошедших и отраженных волн представлена на рис. 52 *б*. Для волн и их направлений примем те же обозначения, что и на рис. 51. Амплитудные коэффициенты отражения и прохождения волн (слева направо) на первой границе обозначим через r_1 и d_1, на второй — через r_2 и d_2. Для обратного направления волн (справа налево) те же коэффициенты обозначим через r_1', d_1', r_2', d_2'. Результирующие амплитудные коэффициенты отражения и пропускания волн для всего барьера обозначим соответственно через r и d. Все волны, которые возникнут внутри и вне барьера, представлены на рис. 52 *б*. Внутри барьера в противоположных направлениях будут распространяться две волны: (a, k) и $(b, -k)$. На левой границе барьера, как видно из рисунка, сходятся четыре волны, а на правой — три волны. Написав граничные условия — непрерывность ψ и $d\psi/dx$ на каждой границе барьера, — мы получим четыре уравнения первой степени, из которых могут быть найдены все неизвестные амплитуды r, a, b, d. Однако выкладки упростятся, а результаты выразятся в более краткой и компактной форме, если поступить несколько иначе.

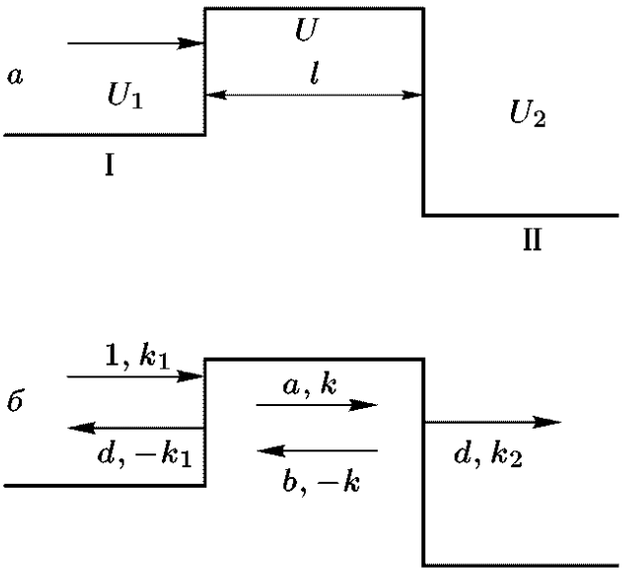


Рис. 52

10. Рассмотрим сначала условия на левой границе барьера. Примем эту границу за начало координат. К ней подходят две волны: $(1, k_1)$ слева и $(b, -k)$ справа. Обе волны отражаются от рассматриваемой границы барьера и частично проходят через нее. В результате наложения отраженной и прошедшей волн в первой области должна получиться результирующая отраженная волна $(r, -k_1)$, а внутри барьера — волна (a, k). Таким образом, должно быть

$$r = r_1 + d_1' b, \quad a = d_1 + r_1' b. \tag{28.11}$$

Аналогично поступаем на второй границе барьера. Только теперь начало координат надо перенести на вторую границу и соответственно этому преобразовать амплитуды сходящихся на ней волн. Амплитуды волн, изображенных на рис. 52 *б*, отнесены к началу координат, помещенному на левой границе. Соответствующие координаты обозначены через x, а координаты относительно начала, помещенного на второй границе, обозначим через x'. Эти координаты связаны соотношением $x = x' + l$. При прежнем выборе начала координат волны, сходящиеся

на правой границе барьера, представляются выражениями

$$ae^{ikx}, \quad be^{-ikx}, \quad de^{ik_2x}$$

(временной множитель $e^{-i\omega t}$ мы опускаем). При замене x через x' те же выражения преобразуются в

$$(ae^{ikl})e^{ikx'}, \quad (be^{-ikl})e^{-ikx'}, \quad (de^{ik_2l})e^{ik_2x'}.$$

Теперь роль амплитуд волн играют выражения, заключенные в круглых скобках. В результате условия на правой границе барьера принимают вид

$$de^{ik_2l} = d_2ae^{ikl}, \quad be^{-ikl} = r_2ae^{ikl}. \tag{28.12}$$

Из уравнений (28.11) и (28.12) можно найти все неизвестные r, d, a, b. Из них представляют интерес прежде всего r и d. С учетом соотношений $r_1' = -r_1$ и $r_1^2 + d_1d_1' = 1$ для них находим

$$r = \frac{r_1 + r_2e^{2ikl}}{1 + r_1r_2e^{2ikl}}, \quad d = \frac{d_1d_2e^{-i(k_2-k)l}}{1 + r_1r_2e^{2ikl}}. \tag{28.13}$$

11. Пользуясь этими формулами, можно рассчитать коэффициенты отражения и пропускания для частиц. Рассчитаем коэффициент пропускания D. Предположим, что вне барьера $U_1 < \mathscr{E}$ и $U_2 < \mathscr{E}$, тогда как внутри барьера $U > \mathscr{E}$. Тогда обе величины

$$k_1 = \frac{1}{\hbar}\sqrt{2m(\mathscr{E} - U_1)} \quad \text{и} \quad k_2 = \frac{1}{\hbar}\sqrt{2m(\mathscr{E} - U_2)}$$

вещественны и положительны, внутри же барьера $k = i\alpha$, где

$$\alpha = \frac{1}{\hbar}\sqrt{2m(U - \mathscr{E})}.$$

Далее,

$$r_1 = \frac{k_1 - k}{k_1 + k} = \frac{k_1 - i\alpha}{k_1 + i\alpha}, \quad r_2 = \frac{k - k_2}{k + k_2} = \frac{i\alpha - k_2}{i\alpha + k_2},$$

$$d_1 = \frac{2k_1}{k_1 + k} = \frac{2k_1}{k_1 + i\alpha}, \quad d_2 = \frac{2k}{k + k_2} = \frac{2i\alpha}{k_2 + i\alpha}.$$

Коэффициент пропускания частицы, согласно (28.5), вычисляется по формуле

$$D = \frac{k_2}{k_1}|d|^2 = \frac{k_2}{k_1}dd^*. \tag{28.14}$$

Простые, но несколько длинные вычисления приводят к результату

$$D = \frac{16k_1k_2\alpha^2}{(k_1^2 + \alpha^2)(k_2^2 + \alpha^2)(e^{2\alpha l} + e^{-2\alpha l}) + 2(\alpha^2 - k_1k_2)}. \tag{28.15}$$

В большинстве интересующих нас случаев экспонентой $e^{-2\alpha l}$ в знаменателе можно пренебречь. Допустим, например, что $U - \mathscr{E} = 50$ эВ $= 0{,}8\cdot10^{-10}$ эрг. Тогда для электрона $\alpha = 3{,}64\cdot10^8$ см$^{-1}$, $e^{2\alpha l} = 1{,}45\cdot10^3$, $e^{-2\alpha l} = 0{,}69\cdot10^{-3}$. Можно также пренебречь слагаемым $2(\alpha^2 - k_1k_2)$, так как оно того же порядка, что и коэффициент $(k_1^2 + \alpha^2)(k_2^2 + \alpha^2)$.

В результате получается простая формула

$$D = D_0 e^{-2\alpha l} = D_0 \exp\left[-\frac{2l}{\hbar}\sqrt{2m(U-\mathscr{E})}\right], \qquad (28.16)$$

где коэффициент D_0 слабо меняется с изменением l, $\mathscr{E}$, U_1, U_2. Его можно принять постоянным и в большинстве интересных случаев считать порядка единицы.

Если частица падает на барьер с одной стороны, то по классическим представлениям при $U > \mathscr{E}$ она не может появиться с другой стороны. Напротив, согласно квантовой механике это возможно. Частица как бы проходит по туннелю через классически запрещенную область $U > \mathscr{E}$. Это явление получило название *туннельного эффекта*. Если при этом $U_1 = U_2$ (симметричный прямоугольный потенциальный барьер), то кинетическая энергия, с которой частица появляется за барьером, равна начальной кинетической энергии, с которой она падала на барьер. Начало представления о туннельных переходах было заложено в 1927 г. Л. И. Мандельштамом и М. А. Леонтовичем (1903–1981). Они на основе уравнения Шредингера рассмотрели проблему квантования для ангармонического осциллятора, у которого потенциальная функция $U = (1/2)kx^2$ при $|x| < a$ и $U = \text{const}$ при $|x| > a$.

12. Мы рассмотрели потенциальный барьер прямоугольной формы (см. рис. 52). Это было сделано только с целью математического упрощения задачи. Ничего принципиального при этом потеряно не было. Для полноты рассмотрим теперь потенциальный барьер, в котором U является произвольной функцией x. Пример такого барьера приведен на рис. 53. Горизонтальная прямая $U(x) = \mathscr{E}$ пересекает кривую барьера в двух точках с абсциссами x_1 и x_2. Аппроксимируем кривую барьера над этой прямой ступенчатой линией. Тогда вся площадь, где $\mathscr{E} < U$, разобьется на заштрихованные прямоугольники, каждый из которых можно рассматривать как прямоугольный потенциальный барьер. Пусть ширина одного из таких прямоугольников равна dx, а высота $U(x)$. Если dx взять достаточно большим, то коэффициент пропускания такого прямоугольного барьера представится выражением

$$D_0' \exp\left[-\frac{2dx}{\hbar}\sqrt{2m(U-\mathscr{E})}\right].$$

Коэффициент пропускания всего барьера получится путем перемножения выражений такого вида. При этом показатели степеней сложатся, и мы придем к выражению вида

$$D = D_0 \exp\left[-\int_{x_1}^{x_2} \frac{2}{\hbar}\sqrt{2m(U-\mathscr{E})}\,dx\right]. \qquad (28.17)$$

Эту приближенную формулу можно обосновать и математически более строго.

Туннельным прохождением через потенциальный барьер объясняются многие физические явления. Сюда относятся разбираемые в следующем параграфе контактная разность потенциалов и холодная эмис-

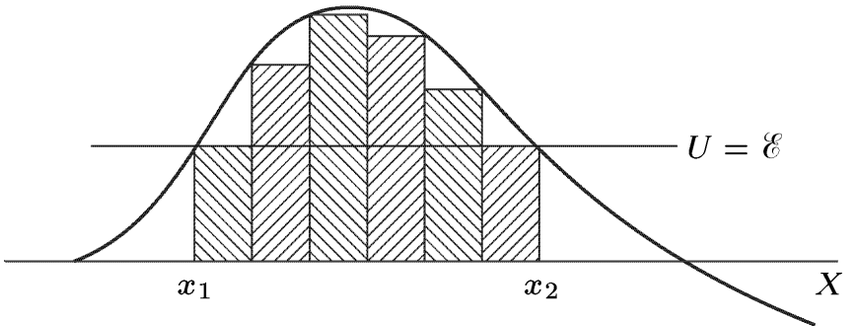


Рис. 53

сия электронов из металлов. Сюда же относятся такие явления, как α-распад, спонтанное деление атомных ядер, ядерные реакции, когда по классическим представлениям кинетической энергии сталкивающихся ядер недостаточно для преодоления кулоновского потенциального барьера между ними, и т. д. Эти явления будут разобраны в ядерной физике.

ЗАДАЧА

В прямоугольном барьере (или яме) $U_1 = U_2$. При каком условии частица не будет отражаться от потенциального барьера (ямы)?

Ответ. Полная энергия $\mathscr{E}$ должна быть больше потенциальной энергии U частицы внутри барьера (ямы). Толщина барьера (ямы) должна быть $l = (1/4)\lambda$, $(3/4)\lambda$, $(5/4)\lambda, \ldots$, где $\lambda = h/\sqrt{2m(\mathscr{E} - U)}$ — длина волны де Бройля внутри барьера (ямы).

§ 29. К объяснению контактной разности потенциалов. Холодная эмиссия электронов из металлов

1. Будем исходить из простейшей модели металла — модели *свободных электронов*. Согласно этой модели электроны в металле ведут себя как газ невзаимодействующих частиц, движущихся в свободном от поля пространстве. Электроны удерживаются в металле силами отталкивания, возникающими при приближении электронов к стенке металла. В такой модели металл можно рассматривать как потенциальную яму, в которой заперты электроны. Для простоты будем считать яму прямоугольной определенной глубины. Мы не можем дать удовлетворительное обоснование и указать границы применимости такой модели. Особое удивление вызывает то, что электрические силы взаимодействия между электронами не учитываются, хотя они отнюдь не малы. Возможность отвлекаться от таких сил, по-видимому, связана с тем, что взаимодействие между электронами не меняет числа энергетических уровней системы. Последнее определяется только *общим числом электронов*, а не силами взаимодействия между ними. Для

ряда явлений, по крайней мере при их качественном рассмотрении, существенно именно общее число энергетических уровней, а не их точное расположение. Разумеется, как и всякая модель, модель свободных электронов объясняет отнюдь не все свойства металлов. Однако ряд явлений объясняется этой моделью правильно, по крайней мере качественно.

На рис. 54 представлена модель металла в виде прямоугольной потенциальной ямы. Внутри металла (т. е. на дне потенциальной ямы) потенциальная функция принята равной нулю, на стенках ямы она скачкообразно меняется до постоянного значения $U_0 > 0$. Конечно, энергетические уровни электрона внутри ямы дискретны, хотя в макроскопических кусках металла и расположены очень густо. Собственно говоря, нельзя сказать, что в модели свободных электронов между электронами нет никакого взаимодействия. Оно имеется. Но это не есть силовое взаимодействие, а взаимодействие особого рода, которое не может быть понято в рамках классической механики. О нем подробно говорится в гл. VI. Такое взаимодействие проявляется в том, что в каждом квантовом состоянии системы может находиться не более одного электрона. Это положение называется *принципом Паули* (1900–1958). Под квантовым состоянием в рассматриваемом нами вопросе следует понимать энергетический уровень электрона (с одним уточнением: допустимое число электронов на энергетическом уровне должно быть удвоено из-за наличия у них спина; но это обстоятельство в разбираемых сейчас вопросах не играет роли, и мы его учитывать не будем).

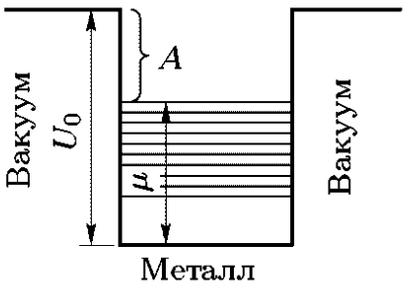


Рис. 54

Будем теперь последовательно заполнять потенциальную яму электронами в предположении, что температура системы равна абсолютному нулю. Первый электрон займет самый нижний (нулевой) уровень энергии. Второй электрон расположится на втором энергетическом уровне и т. д. Последний электрон соответствует такому состоянию металла, когда он сделается электрически нейтральным. Этот электрон займет наивысший уровень энергии μ, называемый *уровнем* или *энергией Ферми* (1901–1954). Таким образом, ниже уровня Ферми все энергетические уровни потенциальной ямы заполнены, а выше — свободны. Напомним, что при этом температура металла предполагается равной абсолютному нулю. Чтобы удалить электрон из металла с уровня Ферми, необходимо затратить работу, не меньшую

$$A = U_0 - \mu. \tag{29.1}$$

Это и есть *работа выхода* электрона из металла. Разумеется формула (29.1) остается справедливой и тогда, когда величины U_0 и μ отсчитываются не от дна ямы, а от произвольно выбранного уровня.

2. Обратимся теперь к объяснению *контактной разности потенциалов*, открытой еще Вольтой (1745–1827). Рассмотрим два разных металла I и II (рис. 55 *а*). Дно обеих потенциальных ям и все уровни

энергии условимся отсчитывать от одного и того же общего уровня. Дно потенциальной ямы первого металла, вообще говоря, не будет совпадать с дном потенциальной ямы второго металла. То же самое

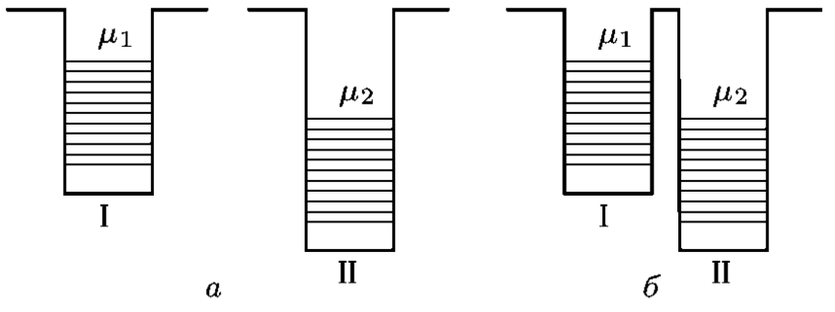


Рис. 55

относится к соответствующим уровням Ферми. Пусть, например, уровень Ферми первого металла расположен выше, чем у второго металла. Сблизим оба металла друг с другом, чтобы зазор между ними стал порядка атомных расстояний, т. е. 10^{-8} см (рис. 55 *б*). Тогда в зазоре между металлами образуется узкий потенциальный барьер, через который электроны с заметной вероятностью могут переходить из одного металла в другой. Переход электронов из металла I в металл II действительно будет осуществляться. Однако обратный переход из металла II в металл I невозможен, так как все уровни энергии, на которые могли бы переходить электроны из металла II, в металле I уже заполнены. В результате металл I будет терять электроны и заряжаться положительно, его потенциал начнет повышаться, а уровень Ферми понижаться. Наоборот, металл II, приобретая электроны, начнет заряжаться отрицательно, его потенциал будет уменьшаться, а уровень Ферми подниматься. Статистическое равновесие установится, когда уровни Ферми обоих металлов сравняются. Но это есть как раз то условие, на основе которого в томе III (§ 104) было подробно рассмотрено возникновение контактной разности потенциалов, как внутренней, так и внешней. Поэтому нет надобности продолжать дальнейшее изложение, а достаточно ограничиться ссылкой на указанный параграф тома III. Здесь же важно было подчеркнуть только то, что процесс установления равновесного состояния осуществляется путем *туннельных переходов* электронов через потенциальный барьер.

3. Перейдем теперь к рассмотрению эмиссии электронов из металлов. Когда температура металла делается достаточно высокой (выше ~ 1000 °C), появляются быстрые электроны, способные преодолевать задерживающий потенциал и выходить из металла. Это — *термоэлектронная эмиссия* (см. т. III, § 101). Однако эмиссия электронов может происходить и из *холодного* металла. Для этого нормально к поверхности металла надо приложить сильное электрическое поле (порядка 10^6 В/см), направленное к металлу. Такая эмиссия называется *холодной*. Объяснение этого явления, в общих чертах согласующееся с опытом, основано на теории прохождения электронов через потенциальный барьер.

В отсутствие внешнего электрического поля потенциальная энергия электрона представляется на рис. 56 ступенчатой линией $AOBC$, причем начало координат O помещено на стенке металла. Внутри металла потенциальная энергия принята равной нулю, вне металла она постоянна и равна C. Если наложить внешнее электрическое поле E, направленное к металлу, то в металл оно не проникнет, и потенциальная энергия электрона в металле по-прежнему будет равна нулю. Снаружи же металла к потенциальной энергии C добавится потенциальная энергия электрона во внешнем электрическом поле, равная $-eEx$ (заряд электрона обозначен через $-e$). Она изображена наклонной прямой BM. В результате полная потенциальная функция электрона во внешнем поле представляется выражениями

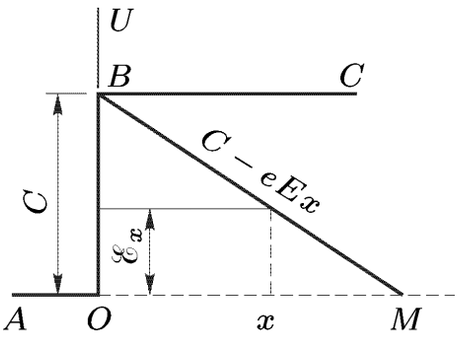


Рис. 56

$$U(x) = \begin{cases} 0 & \text{при } x < 0, \\ C - eEx & \text{при } x > 0. \end{cases}$$

Между металлом и вакуумом возникает потенциальный барьер OBM. Выделим в металле группу электронов с энергией, близкой к $\mathscr{E}_x$. Проницаемость барьера для электронов с такой энергией найдется по формуле (28.17), в которой следует положить $x_1 = 0$. Здесь x_2 найдется из уравнения $C - eEx_2 = \mathscr{E}_x$, которое дает $x_2 = (C - \mathscr{E}_x)/eE$. Задача сводится к вычислению интеграла

$$S = \int_0^{x_2} \sqrt{2m[U(x) - \mathscr{E}_x]}\, dx = \int_0^{x_2} \sqrt{2m(C - eEx - \mathscr{E}_x)}\, dx = \\ = \frac{2}{3}\sqrt{2m}\, \frac{(C - \mathscr{E}_x)^{3/2}}{eE}.$$

Таким образом, коэффициент прозрачности барьера для электронов с энергией $\mathscr{E}_x$ выражается формулой

$$D(\mathscr{E}_x) = D_0 \exp\left\{-\frac{4}{3}\,\frac{\sqrt{2m}}{\hbar}\,\frac{(C - \mathscr{E}_x)^{3/2}}{eE}\right\}. \tag{29.2}$$

Коэффициент этот имеет несколько разные значения для различных $\mathscr{E}_x$. Можно ввести средний или *эффективный коэффициент прозрачности барьера* путем соответствующего усреднения по $\mathscr{E}_x$ (чтобы получился тот же ток эмиссии). Всякое усреднение сводится к усреднению выражения вида $D_0 \exp[-f(\mathscr{E}_x)/E]$, где смысл функции $f(\mathscr{E}_x)$ легко устанавливается сравнением с формулой (29.2). Поскольку усреднение производится по $\mathscr{E}_x$ при фиксированном E, усреднению фактически подлежит функция $f(\mathscr{E}_x)$. Эта функция положительна, так как $C > \mathscr{E}_x$, а потому после усреднения ее можно представить в виде

экспоненциального выражения. В результате для усредненного коэффициента прозрачности барьера получаем

$$\overline{D} = \overline{D}_0 e^{-E_0/E}, \tag{29.3}$$

где $\overline{D}_0$ и E_0 — постоянные, зависящие от рода металла. Ток холодной эмиссии выражается формулой

$$I(E) = I_0 \overline{D} = A e^{-E_0/E}. \tag{29.4}$$

Именно такая зависимость тока холодной эмиссии от напряженности внешнего приложенного электрического поля была экспериментально подтверждена П. И. Лукирским.

Классическую теорию холодной эмиссии и ее сравнение с опытом см. в задаче к этому параграфу.

ЗАДАЧА

Холодную эмиссию электронов из металла пытались объяснить классически влиянием силы электрического изображения $e^2/4x$, с которой электрон притягивается к поверхности металла (см. т. III, § 23, п. 2). С учетом этой силы потенциальная энергия электрона вблизи поверхности металла на расстоянии x от нее представляется выражением $U = C - eEx - e^2/4x$. Учет этой силы понижает потенциальный барьер, который должен преодолеть электрон, чтобы выйти из металла, т. е. уменьшает работу выхода. Рассчитать это уменьшение и оценить напряженность внешнего электрического поля, начиная с которого должна была бы происходить холодная эмиссия. Провести численный расчет для вольфрама. Работа выхода для вольфрама $A = 4{,}5$ эВ.

Р е ш е н и е. Функция U достигает максимума при $x = (1/2)\sqrt{e/E}$, который равен

$$U_{\text{макс}} = C - \sqrt{Ee^3}.$$

Отсюда видно, что сила электрического изображения уменьшает высоту барьера, а с ней и величину работы выхода на величину $\sqrt{Ee^3}$. Новая работа выхода $A' = A - \sqrt{Ee^3}$. Холодная эмиссия начинается, когда $A' = 0$. Это дает

$$E = \frac{A^2}{e^3} = \frac{V^2}{e}. \tag{29.5}$$

Последнее выражение получается из предыдущего, если работу выхода выразить через соответствующее напряжение V по формуле $A = eV$. Для вольфрама $V = 4{,}5$ В $= 1{,}5 \cdot 10^{-2}$ СГСЭ,

$$E = \frac{2{,}25 \cdot 10^{-4}}{4{,}8 \cdot 10^{-10}} = 0{,}469 \cdot 10^6 \text{ СГСЭ } = 1{,}41 \cdot 10^8 \text{ В/см}.$$

Между тем Милликен получал сильные токи холодной эмиссии уже при $E \approx 4 \cdot 10^6$ В/см.

Глава V

ДАЛЬНЕЙШЕЕ ПОСТРОЕНИЕ КВАНТОВОЙ МЕХАНИКИ И СПЕКТРЫ

§ 30. Операторный метод

1. *Операторный метод* широко распространен в большинстве исследований квантовой механики, а потому необходимо сообщить краткие сведения о нем. Это тем более необходимо для того, чтобы дать законченную форму связи символов квантовой механики с реально наблюдаемыми величинами.

Под *оператором* понимают символ, который при действии на функцию некоторых переменных дает новую функцию тех же переменных. Примерами такого действия может служить умножение на x или на какую-либо функцию $f(x)$. Рассматриваемые с этой точки зрения символы x и $f(x)$ являются операторами. Для отличия от чисел их обозначают через $\widehat{x}$ и $\widehat{f}(x)$, т. е. ставят шляпку над x и $f(x)$. Другим примером оператора может служить дифференцирование по x, т. е. $\dfrac{\partial}{\partial x}, \dfrac{\partial^2}{\partial x^2}, \ldots$

Операторы можно складывать. Под *суммой операторов* $\widehat{A}+\widehat{B}$ понимают такой оператор, действие которого на любую функцию $f(x)$ дает результат $\widehat{A}f(x) + \widehat{B}f(x)$. Под *произведением операторов* $\widehat{A}\widehat{B}$ понимают оператор, результат действия которого на любую функцию $f(x)$ равен $\widehat{A}[\widehat{B}f(x)]$. Здесь функция $f(x)$ сначала подвергается действию оператора $\widehat{B}$, а затем на полученный результат действует оператор $\widehat{A}$. Частным случаем произведения операторов является произведение оператора $\widehat{A}$ на число λ, т. е. либо $\lambda\widehat{A}$, либо $\widehat{A}\lambda$, ибо всякое число можно рассматривать как частный случай оператора. В алгебре операторов не всегда соблюдается коммутативный закон относительно умножения. Это значит, что не всегда $\widehat{A}\widehat{B} = \widehat{B}\widehat{A}$. Если такое равенство соблюдается, то говорят, что операторы $\widehat{A}$ и $\widehat{B}$ *коммутируют* друг с другом. Иначе их называют *коммутирующими операторами.* В противном случае операторы A и B *не коммутируют* и называются *некоммутирующими* или антикоммутирующими. Примером некоммутирующих операторов могут служить умножение на x и дифференцирование по x. Действительно,

$$\left(x\frac{\partial}{\partial x}\right)f = x\frac{\partial f}{\partial x}, \quad \left(\frac{\partial}{\partial x}\,x\right)f = \frac{\partial}{\partial x}\,(xf) = f + x\frac{\partial f}{\partial x},$$

так что

$$\frac{\partial}{\partial x}\,x - x\frac{\partial}{\partial x} = 1. \tag{30.1}$$

Эти определения позволяют по заданным операторам $\widehat{A}$ и $\widehat{B}$ строить другие операторы $\widehat{L}(\widehat{A}, \widehat{B})$, являющиеся их функциями. Определение это имеет смысл только для *целых рациональных функций* операторов $\widehat{A}$ и $\widehat{B}$. Достаточность такого ограничения в этом построении связана с тем, что именно при таком ограничении в классической физике определяют новые физические величины через другие, ранее введенные физические величины.

Сложение и умножение операторов производится по обычным алгебраическим правилам сложения и умножения чисел. Единственное отличие состоит в том, что при умножении операторов не всегда можно переставлять порядок сомножителей. Например, всегда

$$(\widehat{A} + \widehat{B})^2 = (\widehat{A} + \widehat{B})(\widehat{A} + \widehat{B}) = \widehat{A}^2 + \widehat{B}\widehat{A} + \widehat{A}\widehat{B} + \widehat{B}^2.$$

В общем виде было бы неправильно писать

$$(\widehat{A} + \widehat{B})^2 = \widehat{A}^2 + 2\widehat{A}\widehat{B} + \widehat{B}^2.$$

Такая формула верна только тогда, когда операторы $\widehat{A}$ и $\widehat{B}$ коммутируют между собой, ибо при $\widehat{B}\widehat{A} = \widehat{A}\widehat{B}$ она получается из предыдущей. Но в случае некоммутирующих операторов эта формула неверна, ибо в этом случае $\widehat{B}\widehat{A} \neq \widehat{A}\widehat{B}$.

Оператор $\widehat{A}$ называется *линейным*, если для любых двух функций f и φ и любых постоянных λ и μ соблюдается соотношение

$$\widehat{A}(\lambda f + \mu\varphi) = \lambda\widehat{A}f + \mu\widehat{A}\varphi.$$

В квантовой механике применяются только линейные операторы. В противном случае нарушался бы *принцип суперпозиции состояний.*

2. Предположим теперь, что многократно производится измерение координаты x частицы, причем частица, поскольку это позволяет опыт, всякий раз приводится в *одинаковые макроскопические условия.* Тогда состояние частицы в этих опытах можно характеризовать волновой функцией $\Psi(x)$, которую мы ради простоты будем считать функцией только одной пространственной координаты x. Среднее значение координаты, которое будет найдено в результате измерений, можно записать в виде

$$\langle x \rangle = \int x\Psi^*\Psi\,dx,$$

ибо $\Psi^*\Psi\,dx$ есть вероятность того, что частица будет обнаружена в интервале $x, x + dx$. При этом необходимо оговорить, что функция $\Psi(x)$ всюду конечна, отлична от нуля в ограниченной области пространства и нормирована к единице, т. е.

$$\int \Psi^*\Psi\,dx = 1, \tag{30.2}$$

где интегрирование производится по всему пространству, в котором Ψ отлична от нуля. Выражение для среднего значения $\langle x\rangle$ мы предпочитаем записать в виде

$$\langle x\rangle = \int \Psi^* \widehat{x} \Psi \, dx. \tag{30.3}$$

Совершенно так же вычисляется среднее значение функции $f(x)$, т. е. по формуле

$$\langle f(x)\rangle = \int \Psi^*(x) \widehat{f}(x) \Psi(x) \, dx, \tag{30.4}$$

в которой $f(x)$ рассматривается как оператор.

Если состояние Ψ меняется во времени, то формула (30.4) дает среднее значение для определенного момента времени. В этом случае в функции $\Psi(x, t)$ следует время t рассматривать как *параметр*, т. е. при взятии интеграла его следует считать постоянным.

3. Как же вычислять по волновой функции $\Psi(x)$ средние значения импульса частицы или средние значения целых рациональных функций от импульса? Будем предполагать, что в каждый момент времени функция $\Psi(x)$ всюду конечна и отлична от нуля в ограниченной области пространства, а потому может быть нормирована согласно формуле (30.2). В целях математического упрощения применим искусственный прием. Заменим истинную волновую функцию $\Psi(x)$ другой периодической функцией $\Phi(x)$ с периодом l, так что при любом x $\Phi(x + l) = \Phi(x)$. Функция $\Psi(x)$ отлична от нуля только на небольшом участке где-то в середине интервала $0 < x < l$ (основного периода). В этом интервале обе функции $\Psi(x)$ и $\Phi(x)$ совпадают. Вне интервала $0 < x < l$ функция $\Psi(x)$ обращается в нуль. Поэтому из нормировки (30.2) следует нормировка для функции $\Phi(x)$:

$$\int_0^l \Phi^*(x) \Phi(x) \, dx = 1. \tag{30.5}$$

Указанная замена не может существенно отразиться на явлениях, протекающих в рассматриваемое время в рассматриваемой нами области пространства. Действительно, различие между функциями $\Phi(x)$ и $\Psi(x)$ относится только к достаточно удаленным областям пространства, которые не могут оказать существенное влияние на протекание изучаемого нами явления. В пределе, когда $l \to \infty$, небольшие искажения явлений, вызванные заменой Ψ на Φ, совсем исчезают.

Периодическую функцию $\Phi(x)$ разложим в ряд Фурье:

$$\Phi(x) = \sum_{n=-\infty}^{n=+\infty} c_n e^{ik_n x}, \tag{30.6}$$

где

$$k_n = \frac{2\pi}{l} n. \tag{30.7}$$

Чтобы определить коэффициент c_m этого ряда, надо обе части (30.6) умножить на $e^{-ik_m x}$ и проинтегрировать по x от 0 до l. При этом

$$\int_0^l e^{i(k_n-k_m)x}\,dx = \begin{cases} l, & \text{при } n=m, \\ \dfrac{1}{i(k_n-k_m)}\,e^{i(k_n-k_m)x}\Big|_0^l = 0, & \text{при } n\neq m, \end{cases}$$

так как

$$e^{i(k_n-k_m)l} = e^{i2\pi(n-m)} = 1 = e^0.$$

С учетом этого получаем

$$c_m = \frac{1}{l}\int_0^l \Phi(x)e^{-ik_m x}\,dx = \frac{1}{l}\int_0^l \Psi(x)e^{-ik_m x}\,dx. \tag{30.8}$$

С учетом того же условия нормировка приводится к виду

$$\int_0^l \Phi^*\Phi\,dx = \int_0^l \Psi^*\Psi\,dx = l\sum_{n=-\infty}^{n=+\infty}|c_n|^2 = 1. \tag{30.9}$$

Выведем еще одну вспомогательную формулу. Имеем

$$\int_0^l \Phi^*\frac{\partial}{\partial x}\Phi\,dx = \int_0^l \sum_m\sum_n c_m^* c_n e^{-ik_m x}\frac{\partial}{\partial x}e^{ik_n x}\,dx =$$

$$= i\int_0^l \sum_m\sum_n c_m^* c_n k_n e^{i(k_n-k_m)x}\,dx,$$

или после перестановки порядка суммирования и интегрирования

$$\int_0^l \Phi^*\frac{\partial}{\partial x}\Phi\,dx = i\sum_m\sum_n c_m^* c_n k_n\int_0^l e^{i(k_n-k_m)x}\,dx.$$

Входящий сюда интеграл уже был вычислен выше. Воспользовавшись вычисленным значением, получим

$$-i\int_0^l \Phi^*\frac{\partial}{\partial x}\Phi\,dx = l\sum|c_n|^2 k_n. \tag{30.10}$$

4. До сих пор мы не обращали внимания на зависимость функции Ψ, а с ней и функции Φ, от времени t. Наши вычисления в сущности, относились к функциям $\Psi(x,t)$ и $\Phi(x,t)$ при фиксированном значении t, — время t рассматривалось как параметр. Временная зависимость определится из требования, чтобы функции Ψ и Φ удовлетворяли временному уравнению Шредингера (21.5). Такому условию удовлетворяет

функция

$$\Phi(x,t)=\sum_n c_n e^{i(k_n x-\omega_n t)}, \tag{30.11}$$

где частоты ω_n определяются законом дисперсии $\omega_n=\omega_n(k_n)$ по формуле (19.6). Этот ряд представляет собой разложение функции $\Phi(x,t)$ по плоским волнам де Бройля [1]).

Волне де Бройля $e^{i(k_n x-\omega_n t)}$ соответствует импульс $p_n=\hbar k_n$. Значения импульса дискретны. Но эта дискретность искусственная и получилась в результате замены истинной волновой функции $\Psi(x,t)$ на вспомогательную периодическую функцию $\Phi(x,t)$. *Истинные значения импульса непрерывны.* И действительно, чем длиннее взять период l, тем меньше расстояние между соседними значениями дискретного спектра импульса. В пределе, когда $l\to\infty$, это расстояние стремится к нулю, так что фактически импульс становится величиной, меняющейся непрерывно. Измерение импульса в состоянии $\Phi(x,t)$ дает одно из значений p_n. Вероятность этого значения, в силу условия нормировки (30.9), равна $l|c_n|^2$. Поэтому среднее значение импульса, которое получится в результате измерения, будет равно

$$\langle p\rangle=\sum l|c_n|^2 p_n, \tag{30.12}$$

или в силу соотношения (30.10)

$$\langle p\rangle=\int_0^l \Phi^*\left(-i\hbar\frac{\partial}{\partial x}\right)\Phi\,dx.$$

Теперь можно выполнить предельный переход к $l\to\infty$ и получить формулу

$$\langle p\rangle=\int \Psi^*\left(-i\hbar\frac{\partial}{\partial x}\right)\Psi\,dx, \tag{30.13}$$

в которой всякая неопределенность, связанная с введением вспомогательной функции Φ, исчезла. При этом интегрирование распространяется уже по всему бесконечному пространству, так как большие значения x не вносят никакого вклада в интеграл.

Рассуждая аналогичным образом, можем без труда получить для произвольного целого положительного n

$$\langle p^n\rangle=\int \Psi^*\left(-i\hbar\frac{\partial}{\partial x}\right)^n\Psi\,dx, \tag{30.14}$$

[1]) Произвольную функцию $\Phi(x,t)$, периодичную по x, можно разложить в ряд Фурье по x, коэффициенты которого будут функциями t. Распространено ошибочное мнение, что это и есть разложение по плоским волнам де Бройля. Ошибка состоит в том, что периодическая функция $\Phi(x,t)$ не может быть произвольной, а должна удовлетворять уравнению Шредингера (21.5). Это накладывает ограничения на коэффициенты разложения как функции времени t.

а для целой рациональной функции импульса

$$\langle F(p)\rangle = \int \Psi^* F(\widehat{p})\Psi\, dx, \tag{30.15}$$

где через $\widehat{p}$ обозначен оператор

$$\widehat{p} \equiv \widehat{p}_x = -i\hbar\frac{\partial}{\partial x}, \tag{30.16}$$

называемый *оператором импульса*, точнее — оператором проекции импульса $\widehat{p}_x$.

5. Квантовая механика постулативно обобщает полученные результаты на любые физические величины, являющиеся функциями координат и импульсов. Иначе говоря, она полагает

$$\langle F(x,p)\rangle = \int \Psi^*(x) F(\widehat{x},\widehat{p})\Psi(x)\, dx. \tag{30.17}$$

Здесь $F(x,p)$ — целая рациональная функция координат и импульсов, как она определяется классически, а $F(\widehat{x},\widehat{p})$ — соответствующий ей оператор. Формула (30.17) и может быть положена в основу введения операторов в квантовую механику. Заметим, что поскольку мы располагаем операторами $\widehat{x}$ и $\widehat{p}$ в прямоугольных координатах, при нахождении оператора $F(\widehat{x},\widehat{p})$ надо исходить из соответствующей классической формулы также в прямоугольных координатах или в векторной форме.

Нелишне подчеркнуть, что под x и p в формуле (30.17) нельзя понимать значения координаты x и импульса p, полученные в результате одного и того же измерения. Такое понимание противоречит принципу неопределенностей Гейзенберга. Под x и p следует понимать координату и импульс в классическом смысле. Квантовая механика заменяет эти величины операторами $\widehat{x}$ и $\widehat{p}$ и вводит новые операторы $F(\widehat{x},\widehat{p})$. Оператор $F(\widehat{x},\widehat{p})$ получается в результате применения к $\widehat{x}$ и $\widehat{p}$ тех же операций сложения и умножения, с помощью которых в классической физике по значениям x и p находится значение функции $F(x,p)$. Здесь нет еще никакой статистичности, свойственной квантовой механике. Статистичность появляется при переходе к формуле (30.17), ибо она дает только среднее значение функции $F(x,p)$, а не ее истинное значение (которое в квантовой механике, вообще говоря, может не иметь никакого смысла из-за невозможности характеризовать состояние частицы одновременным заданием x и p). Получение оператора $F(\widehat{x},\widehat{p})$ из классической функции $F(x,p)$ обусловливает, наряду с другими соображениями, тесною связь между классической и квантовой механиками. Получается парадоксальное утверждение, что обоснование квантовой механики *принципиально невозможно* без механики классической, хотя квантовая механика и является более общей теорией, в которой классическая механика содержится как предельный частный случай. Этот предельный случай получается из квантовой механики, когда постоянная Планка $\hbar$ пренебрежимо мала по сравнению со всеми величинами той же размерности, играющими роль в рассматриваемом явлении.

Все полученные результаты выведены для одномерного случая. Это сделано только в целях простоты и сокращения записи формул. Но эти результаты без труда обобщаются и на трехмерный случай. Так, оператором трехмерного импульса частицы является символический вектор

$$\widehat{\mathbf{p}} = -i\hbar\left(\frac{\partial}{\partial x}\,\mathbf{i} + \frac{\partial}{\partial y}\,\mathbf{j} + \frac{\partial}{\partial z}\,\mathbf{k}\right) = -i\hbar\nabla, \tag{30.18}$$

а формула для среднего значения такого импульса принимает вид

$$\langle \mathbf{p} \rangle = \int \Psi^*(-i\hbar\nabla)\Psi\, dV, \tag{30.19}$$

где интегрирование производится по всему пространству V, а функция Ψ предполагается нормированной к единице:

$$\int \Psi^*\Psi\, dV = 1. \tag{30.20}$$

Таким образом, всякой классической величине $F(\mathbf{r}, \mathbf{p})$ квантовая механика сопоставляет оператор $F(\widehat{\mathbf{r}}, \widehat{\mathbf{p}})$, получающийся заменой классических величин $\mathbf{r}$ и $\mathbf{p}$ на соответствующие операторы $\widehat{\mathbf{r}}$ и $\widehat{\mathbf{p}}$. При этом связь с реально наблюдаемыми величинами устанавливается статистически с помощью формулы

$$\langle F(\mathbf{r}, \mathbf{p}) \rangle = \int \Psi^* F(\widehat{\mathbf{r}}, \widehat{\mathbf{p}})\Psi\, dV. \tag{30.21}$$

6. Поставим теперь вопрос, не существует ли таких состояний, что при измерении величины L, соответствующей оператору $\widehat{L}$, всегда получается определенное значение L. Легко видеть, что такому условию удовлетворяют волновые функции, являющиеся решениями уравнения

$$\widehat{L}\Psi = L\Psi. \tag{30.22}$$

Действительно, в этом случае

$$\langle L \rangle = \int \Psi^* \widehat{L}\Psi\, dx = \int \Psi^* L\Psi\, dx = L\int \Psi^*\Psi\, dx = L,$$

т. е. среднее значение $\langle L \rangle$ всегда равно L. А это возможно тогда и только тогда, когда результат каждого измерения равен L. Мы доказали достаточность условия (30.22). Немного сложнее доказывается и его необходимость, но на этом мы не будем останавливаться.

Функции Ψ, удовлетворяющие уравнению (30.22), называются *собственными функциями* оператора $\widehat{L}$, а числа L — его *собственными значениями*. В квантовой механике принимается, что *при измерении физической величины могут получиться (с той или иной вероятностью) только собственные значения соответствующего ей оператора*. Так как физические величины *существенно вещественны*, то операторы $\widehat{L}$ физических величин должны быть такими, чтобы их собственные значения были также *вещественными*. Но мы не будем

углубляться в обсуждение условий, при которых это требование выполняется.

Уравнение (30.22) является обобщением на случай любых физических величин правила квантования энергии, рассмотренного в предыдущей главе. Чтобы убедиться в этом, найдем оператор $\widehat{H}$, соответствующий полной энергии частицы. Согласно изложенному такой оператор представляется суммой операторов кинетической и потенциальной энергий, т. е.

$$\widehat{H} = \frac{1}{2m}\widehat{p}^2 + \widehat{U} = -\frac{\hbar^2}{2m}\nabla^2 + U. \tag{30.23}$$

Следовательно, (30.22) переходит в

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + U\right)\psi - \mathscr{E}\psi.$$

Это уравнение Шредингера (21.7) для стационарных состояний. Таким образом, сокращенно его можно записать в символической форме

$$\widehat{H}\psi = \mathscr{E}\psi, \tag{30.24}$$

отличающейся от (30.22) только обозначениями. Общее уравнение Шредингера (21.5) для нестационарных состояний также можно записать символически:

$$i\hbar\frac{\partial\Psi}{\partial t} = \widehat{H}\Psi. \tag{30.25}$$

Для полноты заметим, что *уравнение* (30.25) *справедливо не только в случае потенциальных сил, но и в случае, когда силы потенциалом не обладают* (например, магнитные силы). Требуется только, чтобы соответствующие классические уравнения могли быть записаны в форме *уравнений Гамильтона*. В этом случае $\widehat{H}$ называют *оператором Гамильтона* или *гамильтонианом*. Если силы потенциальны, то гамильтониан тождественно совпадает с оператором энергии.

Приведем второй пример на применение уравнения (30.22). Найдем собственные функции и собственные значения оператора импульса. Ограничиваясь одномерным случаем, положим $\widehat{L} = \widehat{p} = -i\hbar\ \partial/\partial x$ и получим

$$-i\hbar\frac{\partial\Psi}{\partial x} = p\Psi.$$

Отсюда

$$\Psi = C(t)e^{ipx/\hbar} = C(t)e^{ikx}.$$

Чтобы удовлетворить общему уравнению Шредингера (30.25), следует положить $C(t) = Ce^{-i\omega t}$, т. е.

$$\Psi = Ce^{i(kx-\omega t)}. \tag{30.26}$$

Таким образом, собственными функциями оператора импульса являются плоские волны де Бройля. Параметр p может принимать любые значения, т. е. спектр собственных значений оператора $\widehat{p}$ непрерывный.

В связи с этим заметим, что к уравнению (30.22) мы пришли на основе уравнения (30.21). Наш вывод последнего предполагал выполнение условия нормировки (30.20), т. е. обращения функции Ψ в нуль на бесконечности. Собственные функции (30.26) этому условию не удовлетворяют. Да и в случае оператора энергии при $\mathscr{E} > 0$ собственные значения образуют непрерывный спектр, и нормировка (30.20) не может быть выполнена. В этих случаях наше доказательство формулы (30.22) не проходит. Однако сама формула (30.22) остается верной. Можно так обобщить нормировку (30.20), чтобы распространить доказательство и на такие случаи. Но для этого надо пользоваться *обобщенными функциями*. На этом формальном вопросе мы останавливаться не будем. В физике в принципе достаточно ограничиться волновыми функциями, обращающимися в нуль на бесконечности, для которых нормировка (30.20) всегда выполняется.

7. Остановимся в заключение еще на одном вопросе, специфическом только для квантовой, но не классической механики. Пусть $\widehat{A}$ и $\widehat{B}$ — два квантовомеханических оператора, каждому из которых соответствует свой спектр собственных значений. Всегда ли существует состояние Ψ, в котором оба оператора имеют определенные собственные значения A и B? Иными словами, существует ли состояние Ψ, в котором обе величины A и B измеримы одновременно? Для ответа на этот вопрос допустим, что Ψ_n является собственной функцией как оператора $\widehat{A}$, так и оператора $\widehat{B}$, т. е.

$$\widehat{A}\Psi_n = A_n\Psi_n, \quad \widehat{B}\Psi_n = B_n\Psi_n,$$

где A_n и B_n — числа, представляющие собой собственные значения операторов $\widehat{A}$ и $\widehat{B}$ в одном и том же состоянии Ψ_n. Умножим первое равенство слева на оператор $\widehat{B}$. Получим

$$\widehat{B}\widehat{A}\Psi_n = \widehat{B}A_n\Psi_n = A_n\widehat{B}\Psi_n = A_nB_n\Psi_n.$$

Аналогично

$$\widehat{A}\widehat{B}\Psi_n = B_nA_n\Psi_n.$$

Отсюда $(\widehat{A}\widehat{B} - \widehat{B}\widehat{A})\Psi_n = 0$. На этом основании нельзя еще заключить, что $\widehat{A}\widehat{B} - \widehat{B}\widehat{A} = 0$, так как Ψ_n — не произвольная функция, а лишь одна из общих собственных функций операторов $\widehat{A}$ и $\widehat{B}$ [1]).

Допустим, однако, что каждая собственная функция оператора $\widehat{A}$ является также собственной функцией оператора $\widehat{B}$ и наоборот. Суще-

---

[1]) Например, из равенства $\left(\dfrac{\partial^2}{\partial x^2} - \dfrac{\partial^3}{\partial x^3}\right)x = 0$ не следует, что $\dfrac{\partial^2}{\partial x^2} - \dfrac{\partial^3}{\partial x^3} = 0$.

ствует математическая теорема, которую мы доказывать не будем, что произвольная волновая функция Ψ может быть разложена по собственным функциям оператора $\widehat{A}$ (или, что то же самое, оператора $\widehat{B}$), т. е.

$$\Psi = \sum c_n \Psi_n$$

(предполагается, что спектр дискретный, что несущественно). Из этой формулы и из соотношения $(\widehat{A}\widehat{B} - \widehat{B}\widehat{A})\Psi_n = 0$ следует

$$(\widehat{A}\widehat{B} - \widehat{B}\widehat{A})\Psi = 0.$$

Теперь уже ввиду произвольности Ψ можно заключить, что

$$\widehat{A}\widehat{B} = \widehat{B}\widehat{A}, \tag{30.27}$$

т. е. операторы $\widehat{A}$ и $\widehat{B}$ коммутативны. Действительно, единственный оператор, обращающий в нуль произвольную функцию, есть оператор умножения на нуль.

Итак, *если все собственные функции операторов $\widehat{A}$ и $\widehat{B}$ совпадают, то эти операторы коммутируют.* Справедлива и обратная теорема: *если операторы $\widehat{A}$ и $\widehat{B}$ коммутируют, то совпадают и их собственные функции.* Эту теорему мы также примем без доказательства.

Приведенной теореме можно придать и другую формулировку. *Две величины A и B измеримы одновременно, вообще говоря, тогда и только тогда, когда соответствующие им операторы $\widehat{A}$ и $\widehat{B}$ коммутируют.* Это правило может нарушаться только в отдельных исключительных случаях (см. § 31, пп. 1 и 4).

Например, координаты x и y можно измерить одновременно, так как операторы $\widehat{x}$ и $\widehat{y}$ коммутируют. Напротив, координата x и соответствующий ей импульс p_x одновременно измерены быть не могут, поскольку операторы $\widehat{x}$ и $\widehat{p}_x$ не коммутируют, как это видно из формулы (30.1). Именно этого требует принцип неопределенностей Гейзенберга. Координата x и импульс p_y, соответствующий другой координате y, измеримы одновременно, ибо операторы $\widehat{x}$ и $\widehat{p}_y = -i\hbar\partial/\partial y$ коммутируют, поскольку при дифференцировании по y координата x ведет себя как постоянная.

§ 31. Момент импульса частицы

1. Момент импульса частицы $\mathbf{l}$ относительно начала координат O в классической механике определяется векторным произведением $[\mathbf{rp}]$. Такое определение в квантовой механике не имеет смысла, поскольку не существует состояния, в котором бы оба вектора $\mathbf{r}$ и $\mathbf{p}$ имели определенные значения. В квантовой механике векторному произведению $[\mathbf{rp}]$ соответствует оператор

$$\mathbf{l} = [\widehat{\mathbf{r}}\,\widehat{\mathbf{p}}]. \tag{31.1}$$

Раскрывая это векторное произведение и соблюдая при этом порядок расположения операторов координат и проекций вектора импульса, найдем операторы проекций момента импульса на координатные оси X, Y, Z:

$$\begin{aligned}
\widehat{l}_x &= (\widehat{y}\widehat{p}_z - \widehat{z}\widehat{p}_y) = i\hbar\left(z\frac{\partial}{\partial y} - y\frac{\partial}{\partial z}\right),\\
\widehat{l}_y &= (\widehat{z}\widehat{p}_x - \widehat{x}\widehat{p}_z) = i\hbar\left(x\frac{\partial}{\partial z} - z\frac{\partial}{\partial x}\right),\\
\widehat{l}_z &= (\widehat{x}\widehat{p}_y - \widehat{y}\widehat{p}_x) = i\hbar\left(y\frac{\partial}{\partial x} - x\frac{\partial}{\partial y}\right).
\end{aligned} \tag{31.2}$$

Через эти проекции сам оператор вектора момента импульса выражается формулой

$$\widehat{\mathbf{l}} = \mathbf{i}\widehat{l}_x + \mathbf{j}\widehat{l}_y + \mathbf{k}\widehat{l}_z. \tag{31.3}$$

Смысл этого векторного оператора выяснится, если написать результат действия его на произвольную функцию ψ. Это есть

$$\widehat{\mathbf{l}}\psi = \mathbf{i}(\widehat{l}_x\psi) + \mathbf{j}(\widehat{l}_y\psi) + \mathbf{k}(\widehat{l}_z\psi), \tag{31.4}$$

т. е. вектор с составляющими $\widehat{l}_x\psi$, $\widehat{l}_y\psi$, $\widehat{l}_z\psi$. Таким образом, произвольной волновой функции ψ соответствует вектор, определяемый выражением (31.4). Возникает, однако, вопрос, существует ли такая функция ψ, для которой все три проекции вектора (31.4) имеют определенные значения, т. е. одновременно выполняются все три равенства:

$$\widehat{l}_x\psi = l_x\psi, \quad \widehat{l}_y\psi = l_y\psi, \quad \widehat{l}_z\psi = l_z\psi. \tag{31.5}$$

Для ответа на этот вопрос надо найти правила коммутации операторов $\widehat{l}_x$, $\widehat{l}_y$, $\widehat{l}_z$. Перемножая операторы $\widehat{l}_x$ и $\widehat{l}_y$ и сохраняя порядок их расположения, получим

$$\begin{aligned}
\widehat{l}_x\widehat{l}_y &= -\hbar^2\left(z\frac{\partial}{\partial y} - y\frac{\partial}{\partial z}\right)\left(x\frac{\partial}{\partial z} - z\frac{\partial}{\partial x}\right) =\\
&= -\hbar^2\left(z\frac{\partial}{\partial y}\,x\frac{\partial}{\partial z} - y\frac{\partial}{\partial z}\,x\frac{\partial}{\partial z} - z\frac{\partial}{\partial y}\,z\frac{\partial}{\partial x} + y\frac{\partial}{\partial z}\,z\frac{\partial}{\partial x}\right) =\\
&= -\hbar^2\left(zx\frac{\partial^2}{\partial y\,\partial z} - yx\frac{\partial^2}{\partial z^2} - z^2\frac{\partial^2}{\partial y\,\partial x} + yz\frac{\partial^2}{\partial z\,\partial x} + y\frac{\partial}{\partial x}\right).
\end{aligned}$$

Аналогично

$$\widehat{l}_y\widehat{l}_x = -\hbar^2\left(zy\frac{\partial^2}{\partial x\,\partial z} - xy\frac{\partial^2}{\partial z^2} - z^2\frac{\partial^2}{\partial x\,\partial y} + xz\frac{\partial^2}{\partial z\,\partial y} + x\frac{\partial}{\partial y}\right).$$

Операции дифференцирования по двум независимым переменным перестановочны, т. е., например, $\partial^2/\partial x\,\partial y = \partial^2/\partial y\,\partial x$. Поэтому

почленным вычитанием предыдущих равенств получим

$$\hat{l}_x\hat{l}_y - \hat{l}_y\hat{l}_x = -\hbar^2\left(y\frac{\partial}{\partial x} - x\frac{\partial}{\partial y}\right) = i\hbar\hat{l}_z.$$

Аналогично получаются и два остальных правила коммутации. Итак,

$$\begin{aligned}
\hat{l}_y\hat{l}_z - \hat{l}_z\hat{l}_y &= i\hbar\hat{l}_x,\\
\hat{l}_z\hat{l}_x - \hat{l}_x\hat{l}_z &= i\hbar\hat{l}_y,\\
\hat{l}_x\hat{l}_y - \hat{l}_y\hat{l}_x &= i\hbar\hat{l}_z.
\end{aligned} \tag{31.6}$$

Таким образом, *любые две проекции оператора момента не коммутируют между собой. Поэтому не существует состояния, в котором бы все три и даже какие-либо две из трех проекций* $\hat{l}_x$, $\hat{l}_y$, $\hat{l}_z$ *имели определенные значения* (см. § 30, п. 7). Исключением является только случай, когда все три проекции одновременно равны нулю. Значит, *не существует состояния, в котором бы и сам вектор момента импульса имел определенное значение, т. е. был бы полностью определен как по величине, так и по направлению.* Иными словами, *оператор момента* $\hat{\mathbf{l}}$ *не имеет собственных функций и соответствующих им векторных собственных значений.* Связь векторного оператора момента с реальной действительностью в общем виде устанавливается *статистически* — с помощью формулы (30.21), которая позволяет находить средние значения, получаемые при измерении самого момента импульса, а не соответствующего ему оператора.

2. Какими же физическими величинами (а не их операторами) характеризуется в квантовой механике момент импульса частицы? Для решения этого вопроса выразим прежде всего операторы проекций момента импульса через сферические (полярные) координаты. Вопрос этот может быть просто решен аналитически. Но аналитическое решение довольно громоздко. Мы предпочитаем привести геометрическое решение. Начало координат прямоугольной XYZ и сферической систем координат поместим в общей точке O (рис. 57). В сферических координатах положение точки A характеризуется расстоянием ее r от точки O и двумя углами θ (полярный угол) и φ (азимут). В точке A проведем касательную AB к меридиану ZAN и касательную AC к соответствующей параллели. Введем вспомогательную прямоугольную систему координат $\xi\eta\zeta$ с осями $O\xi$ и $O\eta$,

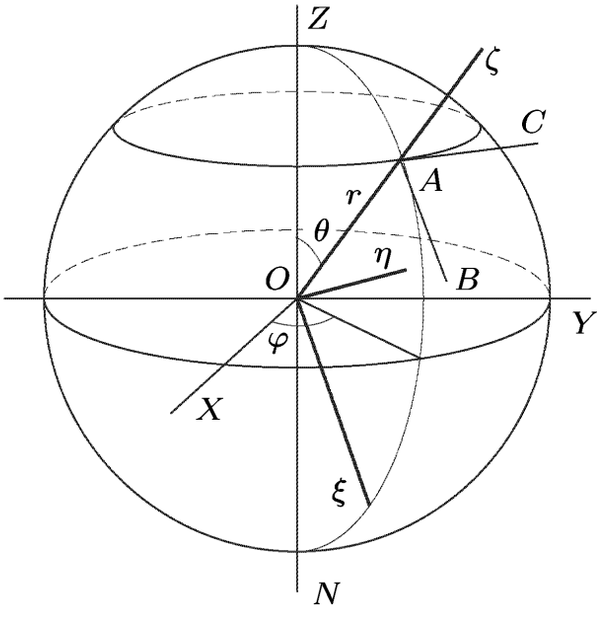


Рис. 57

параллельными соответственно AB и AC, и осью $O\xi$, направленной вдоль радиуса OA. Предполагая, что (классическая) частица находится в точке A, получаем для нее $\xi = \eta = 0$, $\zeta = r$. Поэтому, заменяя в формулах (31.2) x, y, z на ξ, η, ζ, можем написать

$$\widehat{l}_\xi = i\hbar r\frac{\partial}{\partial \eta}, \quad \widehat{l}_\eta = -i\hbar r\frac{\partial}{\partial \xi}, \quad \widehat{l}_\zeta = 0.$$

Но, как видно из рис. 57, $d\xi = r\,d\theta$ (при постоянных r и η), $d\eta = r\sin\theta\,d\varphi$ (при постоянных r и ξ). Поэтому

$$\widehat{l}_\xi = i\frac{\hbar}{\sin\theta}\frac{\partial}{\partial \varphi}, \quad \widehat{l}_\eta = -i\hbar\frac{\partial}{\partial \theta}, \quad \widehat{l}_\zeta = 0.$$

Теперь легко перейти к исходной прямоугольной системе координат X, Y, Z, пользуясь обычным правилом преобразования проекций вектора. Применимость такого правила к операторам обусловлена тем, что направляющие косинусы, определяющие расположение координатных систем XYZ и $\xi\eta\zeta$ относительно друг друга, следует рассматривать как величины постоянные, на которые операторы $\partial/\partial\theta$ и $\partial/\partial\varphi$ не действуют. Запишем нужные направляющие косинусы в виде таблицы:

	ξ	η	ζ
X	$\cos\theta\cos\varphi$	$-\sin\varphi$	
Y	$\cos\theta\sin\varphi$	$\cos\varphi$	
Z	$-\sin\theta$	0	$\cos\theta$

Тогда

$$\widehat{l}_x = \widehat{l}_\xi\cos\theta\cos\varphi - \widehat{l}_\eta\sin\varphi,$$
$$\widehat{l}_y = \widehat{l}_\xi\cos\theta\sin\varphi + \widehat{l}_\eta\cos\varphi,$$
$$\widehat{l}_z = -\widehat{l}_\xi\sin\theta,$$

или

$$\widehat{l}_x = i\hbar\left(\sin\varphi\frac{\partial}{\partial\theta} + \operatorname{ctg}\theta\cos\varphi\frac{\partial}{\partial\varphi}\right),$$
$$\widehat{l}_y = i\hbar\left(-\cos\varphi\frac{\partial}{\partial\theta} + \operatorname{ctg}\theta\sin\varphi\frac{\partial}{\partial\varphi}\right), \tag{31.7}$$
$$\widehat{l}_z = -i\hbar\frac{\partial}{\partial\varphi}.$$

Отметим существенную разницу между классическим моментом импульса частицы $[\mathbf{rp}]$ и соответствующим ему квантовым оператором момента (31.3). Классический момент зависит от радиуса-вектора $\mathbf{r}$, т. е. от выбора начала координат O, относительно которого берется момент импульса. Оператор момента, как видно из формул (31.7), не содержит $\mathbf{r}$, а зависит только от углов θ и φ. Это значит, что оператор момента (31.3) не зависит от выбора начала координат, а зависит

только от направления координатных осей. Поэтому его лучше называть *оператором углового* или *вращательного момента* частицы. При рассмотрении углового момента, в отличие от классического момента импульса, не надо указывать, относительно какого начала он берется. Разумеется, не зависят от выбора начала координат и собственные значения операторов проекций и квадрата углового момента, о которых говорится ниже.

3. Теперь поставим вопрос, может ли одна из проекций углового момента иметь определенное значение? Ясно, что если вопрос решается в положительном (или отрицательном) смысле для какой-либо одной из проекций, то он должен решаться в том же смысле и для каждой из остальных двух проекций, а также для проекции углового момента вдоль любого произвольно избранного направления. Это непосредственно следует из изотропии пространства, т. е. из эквивалентности всех направлений в пространстве. Не могут иметь в одном и том же состоянии определенные значения проекции углового момента вдоль двух различных направлений. Избранное направление можно поэтому взять произвольно. Такое направление обычно принимают за ось Z, так как в этом случае оператор $\widehat{l}_z$ выражается наиболее простой формулой. Для решения поставленного вопроса служит уравнение

$$\widehat{l}_z\psi = l_z\psi,$$

или

$$-i\hbar\frac{\partial\psi}{\partial\varphi} = l_z\psi, \tag{31.8}$$

где l_z — постоянная. Его решением является

$$\psi = C(r,\theta)\exp\left(i\frac{l_z}{\hbar}\varphi\right). \tag{31.9}$$

В силу требуемой однозначности ψ эта функция должна возвращаться к своему исходному значению, когда аргумент φ получает приращение 2π, ибо такое приращение возвращает исходную точку в начальное положение. Таким образом, должно быть

$$\exp\left(i\frac{l_z}{\hbar}\varphi\right) = \exp\left\{i\frac{l_z}{\hbar}(\varphi+2\pi)\right\}.$$

Так как показательная функция периодична с периодом $2\pi i$, то это равенство может выполняться только при условии

$$i\frac{l_z}{\hbar}2\pi = m\cdot 2\pi i,$$

или

$$l_z = m\hbar, \tag{31.10}$$

где m — целое положительное или отрицательное число, включая и нуль ($m = 0, \pm1, \pm2, \ldots$). Равенство (31.10) означает, что проекция углового момента на любое направление квантуется. Для сокращения за единицу углового момента обычно принимают постоянную $\hbar$. При

таком соглашении говорят, что проекция углового момента на избранное направление может принимать только целочисленные значения $m = 0, \pm 1, \pm 2, \ldots$ Проекцию углового момента на ось Z принято обозначать через m_z. Таким образом, $m_z = m$.

Правило (31.10) по своей форме аналогично соответствующему правилу квантования момента импульса (13.6) в теории Бора. Однако между этими двумя правилами есть глубокое различие. В формуле (13.6) под L следует понимать полный момент импульса частицы (электрона), тогда как в (31.10) речь идет только об одной проекции момента импульса на какое-либо направление, а самого вектора момента импульса, как точно определенной величины, вообще не существует.

4. Нетрудно непосредственно проверить, что собственная функция оператора $\widehat{l}_z$, т. е. функция (31.9), не может быть одновременно собственной функцией ни оператора $\widehat{l}_x$, ни оператора $\widehat{l}_y$. Допустим противоположное, что (31.9) является собственной функцией, например, оператора $\widehat{l}_x$. Для этой функции имеем

$$\widehat{l}_x\psi = i\hbar\left(\sin\varphi\frac{\partial C}{\partial\theta} + im\operatorname{ctg}\theta\cos\varphi C\right)e^{im\varphi}.$$

Этот результат не может быть представлен в виде $l_x\psi$, как мы предположили (l_x — постоянное число). Чтобы убедиться в этом, достаточно рассмотреть частный случай, когда $\varphi = 0$. Тогда

$$\widehat{l}_x\psi = -\hbar m\operatorname{ctg}\theta\psi.$$

С другой стороны, по нашему предположению должно быть

$$\widehat{l}_x\psi = l_x\psi.$$

И оба эти соотношения должны соблюдаться при любых значениях угла θ, что, очевидно, невозможно. Такое же рассуждение применимо для оператора $\widehat{l}_y$. Это подтверждает прежний результат, что не существует волновой функции Ψ, которая была бы одновременно собственной функцией операторов $\widehat{l}_x$, $\widehat{l}_y$, $\widehat{l}_z$. Исключением является только случай, когда функция Ψ сферически симметрична, т. е. зависит только от r. В этом случае Ψ будет собственной функцией всех трех операторов $\widehat{l}_x$, $\widehat{l}_y$, $\widehat{l}_z$, а все три проекции углового момента m_x, m_y, m_z обратятся в нуль.

5. Второй величиной наряду с проекцией m_z, характеризующей величину углового момента, является *квадрат полного углового момента*. Его принято обозначать через $\mathbf{l}^2$. Но это не есть квадрат вектора $\mathbf{l}$ (которого не существует), а *собственное значение квадрата оператора углового момента*, т. е.

$$\widehat{\mathbf{l}}^2 = (\widehat{l}_x\mathbf{i} + \widehat{l}_y\mathbf{j} + \widehat{l}_z\mathbf{k})^2 = \widehat{l}_x^2 + \widehat{l}_y^2 + \widehat{l}_z^2. \tag{31.11}$$

Поэтому общепринятое обозначение $\mathbf{l}^2$ неудачно, но мы не будем его менять. Однако истинный смысл величины $\mathbf{l}^2$, как собственного значения квадрата оператора углового момента $\widehat{\mathbf{l}}^2$ надо постоянно иметь в виду.

Чтобы убедиться, что величины $\mathbf{l}^2$ и m_z могут быть измерены в одном и том же состоянии, надо показать, что операторы $\widehat{\mathbf{l}}^2$ и $\widehat{l}_z$ коммутируют друг с другом. Для этого пишем

$$\widehat{\mathbf{l}}^2\widehat{l}_z = (\widehat{l}_x^2 + \widehat{l}_y^2 + \widehat{l}_z^2)\widehat{l}_z = \widehat{l}_x(\widehat{l}_x\widehat{l}_z) + \widehat{l}_y(\widehat{l}_y\widehat{l}_z) + \widehat{l}_z^3,$$

или в силу соотношений коммутации (31.6)

$$\widehat{\mathbf{l}}^2\widehat{l}_z = \widehat{l}_x(\widehat{l}_z\widehat{l}_x - i\hbar\widehat{l}_y) + \widehat{l}_y(\widehat{l}_z\widehat{l}_y + i\hbar\widehat{l}_x) + \widehat{l}_z^3.$$

Аналогично

$$\widehat{l}_z\widehat{\mathbf{l}}^2 = (\widehat{l}_x\widehat{l}_z + i\hbar\widehat{l}_y)\widehat{l}_x + (\widehat{l}_y\widehat{l}_z - i\hbar\widehat{l}_x)\widehat{l}_y + \widehat{l}_z^3.$$

Почленным вычитанием находим

$$\widehat{\mathbf{l}}^2\widehat{l}_z - \widehat{l}_z\widehat{\mathbf{l}}^2 = 0, \tag{31.12}$$

что и требовалось доказать. Разумеется, такое же соотношение коммутации справедливо и для операторов $\widehat{l}_x$ и $\widehat{l}_y$.

Итак, *существует состояние, в котором одновременно имеют определенные значения квадрат углового момента* $\mathbf{l}^2$ *и одна из его проекций на избранное направление.* Обычно это направление принимают за ось Z.

Рассмотрим какое-либо состояние Ψ, в котором величины $\mathbf{l}^2$ и m_z (а следовательно, и m_z^2) одновременно имеют определенные значения. Докажем, что в этом состоянии всегда (за исключением случая $l^2 = 0$) $\widehat{\mathbf{l}}^2 > m_z^2$. С этой целью рассмотрим операторное равенство

$$\widehat{\mathbf{l}}^2 - \widehat{l}_z^2 = \widehat{l}_x^2 + \widehat{l}_y^2.$$

Для оператора $\widehat{\mathbf{l}}^2 - \widehat{l}_z^2$ функция Ψ является собственной с собственным значением $\mathbf{l}^2 - m_z^2$. То же значение имеет и соответствующая средняя величина. Поэтому, производя усреднение по формуле (30.21), получим

$$\mathbf{l}^2 - m_z^2 = \int \Psi^*(\widehat{l}_x^2 + \widehat{l}_y^2)\Psi\, dV.$$

Но интеграл справа существенно положителен, так как он представляет среднее значение существенно положительной величины $l_x^2 + l_y^2$ в состоянии Ψ. Итак,

$$\mathbf{l}^2 > m_z^2. \tag{31.13}$$

Поэтому угловой момент не может ориентироваться точно вдоль оси Z. В любом состоянии он всегда содержит проекции l_x и l_y, которые, однако, в рассматриваемом состоянии остаются неопределенными. Это — уже известный нам факт, согласно которому не существует состояния, в котором все три проекции l_x, l_y, l_z имеют определенные значения.

Про этот факт иногда говорят, что в собственном состоянии, где $\mathbf{l}^2$ имеет определенное значение, угловой момент сохраняет свою длину, равную $\sqrt{\mathbf{l}^2}$, но его направление испытывает флуктуации. Вряд ли этот способ выражения можно признать удачным, ибо он предполагает, что существует какой-то вектор $\mathbf{l}$, имеющий в каждый момент времени определенную длину и направление, но это направление беспорядочно и непрерывно меняется во времени. На самом деле, как мы видели, такого вектора не существует, а потому картина его флуктуации не соответствует действительности.

6. Остается определить собственные значения $\mathbf{l}^2$ оператора квадрата углового момента $\widehat{\mathbf{l}}^2$. Это можно сделать с помощью одних только правил коммутации (31.6). Приведем сначала эти правила к другому виду, более удобному для нашей цели. Введем два оператора:

$$\widehat{l}_+ = \widehat{l}_x + i\widehat{l}_y, \quad \widehat{l}_- = \widehat{l}_x - i\widehat{l}_y. \tag{31.14}$$

Тогда из (31.6) нетрудно получить

$$\widehat{l}_+\widehat{l}_- - \widehat{l}_-\widehat{l}_+ = 2\hbar\widehat{l}_z, \quad \widehat{l}_z\widehat{l}_+ - \widehat{l}_+\widehat{l}_z = \hbar\widehat{l}_+, \quad \widehat{l}_z\widehat{l}_- - \widehat{l}_-\widehat{l}_z = -\hbar\widehat{l}_-. \tag{31.15}$$

Далее

$$\widehat{\mathbf{l}}^2 = \left(\frac{\widehat{l}_+ + \widehat{l}_-}{2}\right)^2 + \left(\frac{\widehat{l}_+ - \widehat{l}_-}{2i}\right)^2 + \widehat{l}_z^2$$

или

$$\widehat{\mathbf{l}}^2 = \widehat{l}_+\widehat{l}_- + \widehat{l}_z^2 - \hbar\widehat{l}_z = \widehat{l}_-\widehat{l}_+ + \widehat{l}_z^2 + \hbar\widehat{l}_z. \tag{31.16}$$

Так как $\mathbf{l}^2$ есть величина ограниченная, то из (31.13) следует, что m_z^2, а потому и m_z суть также величины ограниченные. Обозначим через l наибольшее положительное значение проекции m_z при заданном значении квадрата момента $\mathbf{l}^2$. Пусть Ψ — общая волновая функция операторов $\widehat{\mathbf{l}}^2$ и $\widehat{l}_z$, причем собственное значение оператора $\widehat{l}_z$ равно m. В этом предположении

$$\widehat{\mathbf{l}}^2\Psi = \mathbf{l}^2\Psi, \quad \widehat{l}_z\Psi = \hbar l\Psi. \tag{31.17}$$

Из соотношений коммутации (31.15) для такой функции получаем

$$\widehat{l}_z\widehat{l}_\pm\Psi = (\widehat{l}_\pm\widehat{l}_z \pm \hbar\widehat{l}_\pm)\Psi = \hbar(l \pm 1)l_\pm\Psi.$$

Отсюда видно, что функции $\widehat{l}_+\Psi$ и $\widehat{l}_-\Psi$ являются собственными функциями оператора $\widehat{l}_z$, имеющими собственные значения $\hbar(l+1)$ и $\hbar(l-1)$ соответственно. Но величина $\hbar(l+1)$ не может быть собственным значением оператора $\widehat{l}_z$, так как по предположению наибольшее собственное значение этого оператора равно $\hbar l$. Таким образом, равенство

$$\widehat{l}_z\widehat{l}_+\Psi = \hbar(l+1)\widehat{l}_+\Psi$$

невозможно. Но это равенство логически следует из соотношений коммутации (31.15) и уравнения $\widehat{l}_z\Psi = l\Psi$. Избежать противоречия можно только тогда, когда $\widehat{l}_+\Psi = 0$, ибо в этом случае указанное равенство,

конечно, будет удовлетворено. Но из этого следует, что $\widehat{l}_-\widehat{l}_+\Psi = 0$, или в силу соотношения (31.16)

$$(\widehat{\mathbf{l}}^2 - \widehat{l}_z^2 - \hbar\widehat{l}_z)\Psi = 0.$$

Но в силу (31.17) $\widehat{l}_z^2\Psi = \hbar^2 l^2\Psi$, $\hbar\widehat{l}_z\Psi = \hbar^2 l\Psi$, так что

$$(\widehat{\mathbf{l}}^2 - \hbar^2 l^2 - \hbar^2 l)\Psi = 0,$$

или

$$\widehat{\mathbf{l}}^2\Psi = \hbar^2 l(l+1)\Psi.$$

Значит Ψ есть собственная функция оператора квадрата углового момента с собственным значением

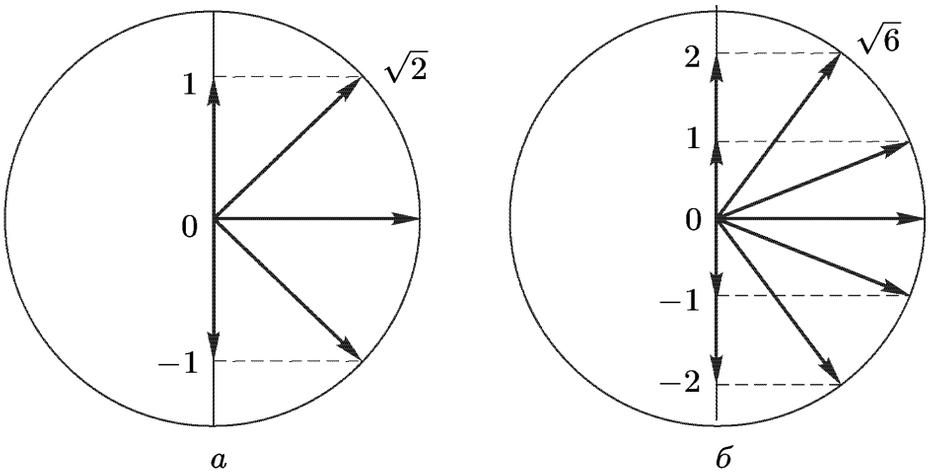
$$\mathbf{l}^2 = \hbar^2 l(l+1). \tag{31.18}$$

7. Пусть квадрат углового момента $\mathbf{l}^2$ имеет определенное значение $l(l+1)$. Во скольких состояниях может реализоваться такая ситуация, если в них проекция m также имеет определенное значение? Очевидно, в таких состояниях m может принимать следующие значения:

$$m = -l, -(l-1), \ldots, -1, 0, +1, \ldots, +(l-1), +l,$$

т. е. всего таких состояний будет $2l+1$.

Полученные результаты, определяющие возможные значения l_z и $\mathbf{l}^2$, называют *пространственным квантованием.* Этот термин заимствован из старой теории Бора, в которой пространственное квантование определяло возможные направления вектора углового момента $\mathbf{l}$ в пространстве. С точки зрения квантовой механики термин «пространственное квантование» вряд ли удачен, так как «вектор» $\mathbf{l}$ принципиально не имеет определенных направлений в пространстве. Для наглядности пространственное квантование обычно представляют графически на *векторных диаграммах.* По оси Z откладывают возможные значения m, рассматривая их как проекции вектора $\mathbf{l}$ длины $\sqrt{l(l+1)}$, имеющего дискретные направления в пространстве. В качестве примера на рис. 58 приведены векторные диаграммы для случаев $l = 1$



а *б*

Рис. 58

и $l = 2$ (за единицу углового момента принята постоянная Планка $\hbar$). Эти диаграммы нельзя понимать буквально. Они правильно передают только два факта: возможные значения проекции m и возможные значения квадрата углового момента $\mathbf{l}^2$.

Квантовое число l по причинам, которые выяснятся ниже, называют *орбитальным квантовым числом*, а число m — магнитным квантовым числом.

8. В классической механике кинетическая энергия вращающегося твердого тела определяется формулой

$$\mathscr{E} = \frac{\mathbf{l}^2}{2I},$$

где I — момент инерции тела относительно соответствующей оси вращения. Такая же формула справедлива и в квантовой механике. Различие состоит в том, что здесь величина $\mathbf{l}^2$ квантуется, принимая дискретные значения $\mathbf{l}^2 = \hbar^2 l(l+1)$. Неизменяемая вращающаяся система в квантовой механике называется *ротатором*. Таким образом, энергетические уровни ротатора дискретны и определяются формулой

$$\mathscr{E}_r = \frac{\hbar^2}{2I}\, l(l+1). \tag{31.19}$$

Такой формулой мы уже пользовались при качественном рассмотрении вращательной теплоемкости молекул (см. т. II, § 69).

Впрочем, необходимо заметить, что идеализированное представление о ротаторе, как и представление об идеально твердом теле, несовместимо с принципом неопределенностей Гейзенберга. В самом деле, в модели идеально твердого тела его размеры в любом направлении (например, в направлении оси X) строго определены и не могут меняться во времени, т. е. неопределенность координаты $\Delta x = 0$. Но тогда для импульса в том же направлении из соотношения неопределенностей следует, что $\Delta p_x = \infty$. В теле неизбежно возникнут колебания, меняющие величину момента инерции I. Например, вращающуюся молекулу можно для некоторых целей рассматривать как жесткий ротатор и пользоваться формулой (31.19), если изменения I, связанные с вращением, невелики.

§ 32. Сложение угловых моментов

1. Понятие углового момента можно распространить и на системы частиц. Ради простоты будем предполагать, что система состоит только из двух частиц: 1 и 2. Координаты (радиус-вектор) первой частицы обозначим через $\mathbf{r}_1$, второй — через $\mathbf{r}_2$. Оператором углового момента $\widehat{\mathbf{l}}$ системы называют сумму операторов угловых моментов ее частей $\widehat{\mathbf{l}}_1$ и $\widehat{\mathbf{l}}_2$:

$$\widehat{\mathbf{l}} = \widehat{\mathbf{l}}_1 + \widehat{\mathbf{l}}_2. \tag{32.1}$$

Так же определяется и оператор проекции углового момента системы на избранное направление. Например,

$$\hat{l}_z = \hat{l}_{1z} + \hat{l}_{2z}. \tag{32.2}$$

Будем предполагать, что частицы не взаимодействуют между собой. Тогда волновая функция первой частицы $\Psi_1(\mathbf{r}_1)$ будет одна и та же независимо от того, есть вторая частица или нет. Эту функцию можно умножить на произвольную постоянную, которая может зависеть от $\mathbf{r}_2$ как от параметра. В частности, за нее можно принять волновую функцию $\Psi_2(\mathbf{r}_2)$ второй частицы. Тогда волновая функция первой частицы представится произведением $\Psi = \Psi_1(\mathbf{r}_1)\Psi_2(\mathbf{r}_2)$ Но очевидно из тех же соображений, что в таком же виде представится и волновая функция второй частицы. Поэтому $\Psi = \Psi_1\Psi_2$ можно рассматривать как волновую функцию системы невзаимодействующих частиц 1 и 2. При этом для нее будет сохранена нормировка

$$\int |\Psi|^2\, dV_1\, dV_2 = \int |\Psi_1|^2\, dV_1 \int |\Psi_2|^2\, dV_2 = 1. \tag{32.3}$$

2. Если ограничиться действием операторов только на функцию $\Psi = \Psi_1\Psi_2$, то из доказанного следует, что операторы $\hat{\mathbf{l}}_1$ и $\hat{\mathbf{l}}_2$ перестановочны. В самом деле, поскольку оператор $\hat{\mathbf{l}}_1$ действует только на функцию Ψ_1 и не действует на функцию Ψ_2, а оператор $\hat{\mathbf{l}}_2$, наоборот, действует только на Ψ_2 и не действует на Ψ_1, можно написать

$$\hat{\mathbf{l}}_1\hat{\mathbf{l}}_2\Psi = \hat{\mathbf{l}}_1\hat{\mathbf{l}}_2(\Psi_1\Psi_2) = \hat{\mathbf{l}}_1(\Psi_1\hat{\mathbf{l}}_2\Psi_2) = (\hat{\mathbf{l}}_2\Psi_2)(\hat{\mathbf{l}}_1\Psi_1) = (\hat{\mathbf{l}}_1\Psi_1)(\hat{\mathbf{l}}_2\Psi_2).$$

К тому же результату приводит и действие оператора $\hat{\mathbf{l}}_2\hat{\mathbf{l}}_1$. Значит, для функций вида $\Psi = \Psi_1\Psi_2$ справедливо операторное равенство $\hat{\mathbf{l}}_1\hat{\mathbf{l}}_2 = \hat{\mathbf{l}}_2\hat{\mathbf{l}}_1$, что и доказывает наше утверждение. Таким же путем докажем для функций того же вида и перестановочность операторов проекций углового момента в случае системы независимых частиц, например $\hat{l}_{1x}$ и $\hat{l}_{2x}$. Из доказанного следует, что правила коммутации (31.6) и все следствия из них для отдельной частицы распространяются без всяких изменений и на системы независимых частиц.

Ввиду коммутации операторов $\hat{\mathbf{l}}_1$ и $\hat{\mathbf{l}}_2$ оператор квадрата углового момента $\hat{\mathbf{l}}^2$ будет равен

$$\hat{\mathbf{l}}^2 = (\hat{\mathbf{l}}_1 + \hat{\mathbf{l}}_2)^2 = \hat{\mathbf{l}}_1^2 + 2(\hat{\mathbf{l}}_1\hat{\mathbf{l}}_2) + \hat{\mathbf{l}}_2^2. \tag{32.4}$$

Теперь выясним, как коммутирует оператор квадрата углового момента $\hat{\mathbf{l}}^2 = \hat{l}_x^2 + \hat{l}_y^2 + \hat{l}_z^2$ с оператором одной из проекций его, например $\hat{l}_x = \hat{l}_{1x} + \hat{l}_{2x}$. Очевидно, операторы $\hat{l}_{1x}$ и $\hat{l}_{2x}$ коммутируют с $\hat{\mathbf{l}}_1^2$ и с $\hat{\mathbf{l}}_2^2$. Остается только проверить коммутативность операторов $\hat{\mathbf{l}}_1\hat{\mathbf{l}}_2$

и $\widehat{l}_x$. Имеем

$$(\widehat{\mathbf{l}}_1\widehat{\mathbf{l}}_2)\widehat{l}_x = (\widehat{l}_{1x}\widehat{l}_{2x} + \widehat{l}_{1y}\widehat{l}_{2y} + \widehat{l}_{1z}\widehat{l}_{2z})(\widehat{l}_{1x} + \widehat{l}_{2x}) =$$
$$= (\widehat{l}_{1x}\widehat{l}_{2x} + \widehat{l}_{1y}\widehat{l}_{2y} + \widehat{l}_{1z}\widehat{l}_{2z})\widehat{l}_{1x} + (\widehat{l}_{1x}\widehat{l}_{2x} + \widehat{l}_{1y}\widehat{l}_{2y} + \widehat{l}_{1z}\widehat{l}_{2z})\widehat{l}_{2x}.$$

Аналогично

$$\widehat{l}_x(\widehat{\mathbf{l}}_1\widehat{\mathbf{l}}_2) = \widehat{l}_{1x}(\widehat{l}_{1x}\widehat{l}_{2x} + \widehat{l}_{1y}\widehat{l}_{2y} + \widehat{l}_{1z}\widehat{l}_{2z}) + \widehat{l}_{2x}(\widehat{l}_{1x}\widehat{l}_{2x} + \widehat{l}_{1y}\widehat{l}_{2y} + \widehat{l}_{1z}\widehat{l}_{2z}).$$

Если воспользоваться правилами коммутации (31.6) и тем обстоятельством, что операторы $\widehat{\mathbf{l}}_1$ и $\widehat{\mathbf{l}}_2$ действуют на волновые функции разных частиц, то последнее выражение легко привести к предыдущему. Итак, оператор $\widehat{\mathbf{l}}^2 = (\widehat{\mathbf{l}}_1 + \widehat{\mathbf{l}}_2)^2$ коммутирует с операторами проекций $\mathbf{l}$ на любое направление. Поэтому существует состояние системы, в котором $\mathbf{l}^2$ и одна из проекций на любое направление имеют определенные значения.

3. Состояние первой частицы можно характеризовать значениями l_1 и m_1, второй — значениями l_2 и m_2. Число l_1 определяет квадрат углового момента первой частицы $l_1(l_1+1)$, число l_2 — квадрат такого же момента второй частицы $l_2(l_2+1)$. Числа m_1 и m_2 определяют проекции на ось Z угловых моментов $\mathbf{l}_1$ и $\mathbf{l}_2$ (в единицах $\hbar$). Очевидно, совокупность чисел l_1, l_2, m_1, m_2 характеризует некоторое состояние системы обеих независимых частиц. Волновую функцию такого состояния будем обозначать через $\varphi_{l_1 l_2 m_1 m_2}$. Определим число состояний такого типа, т. е. число линейно независимых функций φ при заданных значениях l_1 и l_2. При заданном l_1 число m_1 может принимать $(2l_1+1)$ значений (см. § 31, п. 7), при заданном l_2 число значений m_2 равно $(2l_2+1)$. Таким образом, при заданных l_1 и l_2 искомое число состояний с волновыми функциями типа φ равно $(2l_1+1)(2l_2+1)$. Из таких $(2l_1+1)(2l_2+1)$ состояний путем их линейных комбинаций можно составить любое состояние системы с заданными l_1 и l_2.

Но линейно независимые состояния, из которых может быть составлено любое состояние системы с заданными l_1 и l_2, можно выбрать и иначе. Должно оставаться постоянным лишь общее число таких линейно независимых состояний, т. е. это число по-прежнему должно быть равно $(2l_1+1)(2l_2+1)$. Ввиду справедливости для всей системы правил коммутации (31.6) существуют при заданных l_1 и l_2 состояния всей системы с определенными значениями квадрата $l(l+1)$ полного углового момента $\mathbf{l}$ и его проекции m на ось Z. Волновые функции таких состояний будем обозначать через $\psi_{l_1 l_2 l m}$. Из них путем линейных комбинаций можно составить волновую функцию любого состояния с заданными l_1 и l_2. Поэтому число линейно независимых функций типа $\psi_{l_1 l_2 l m}$ с заданными l_1 и l_2 должно быть равно $(2l_1+1)(2l_2+1)$.

Убедимся в этом путем прямого вычисления. Из формулы (32.2) непосредственно следует, что если проекции m_1 и m_2 имеют определенные значения, то и проекция m также имеет определенное значение m, причем $m = m_1 + m_2$. Ради определенности будем предполагать, что

$l_1 > l_2$. Тогда при заданных l_1 и l_2 возможные положительные значения m, получаемые таким путем, представятся следующей схемой:

m_1	m_2	m
l_1	l_2	$l_1 + l_2$
l_1 $l_1 - 1$	$l_2 - 1$ l_2	$l_1 + l_1 - 1$
l_1 $l_1 - 1$ $l_1 - 2$	$l_2 - 2$ $l_2 - 1$ l_3	$l_1 + l_2 - 2$
...	...	...
l_1	$-l_2$	$l_1 - l_2$

Отберем теперь всевозможные состояния (при заданных l_1 и l_2, для которых максимальные значения проекции m соответственно равны $(l_1 + l_2)$, $(l_1 + l_2 - 1)$, ... Это будут состояния с определенным значением l, равным

$$l = (l_1 + l_2),\ l_1 + (l_2 - 1),\ \ldots,\ l_1 - l_2.$$

Число таких состояний равно $2l_2+1$. В каждом из этих состояний m может принимать $(2l+1)$ значений. Таким образом, число всевозможных состояний (при заданных l_1 и l_2 с линейно независимыми функциями типа $\psi_{l_1 l_2 l m}$ будет

$$2(l_1 + l_2) + 1 + 2(l_1 + l_2 - 1) + 1 + \ldots + 2(l_1 - l_2) + 1.$$

Это — арифметическая прогрессия с разностью -2 и общим числом членов $2l_2 + 1$. Ее сумма равна

$$\frac{2(l_1 + l_2) + 1 + 2(l_1 - l_2) + 1}{2}(2l_2 + 1) = (2l_1 + 1)(2l_2 + 1),$$

что и требовалось доказать.

4. Полученные результаты относятся не только к сложению угловых моментов двух невзаимодействующих частиц. Они распространяются без всяких изменений и на произвольные сложные системы, состоящие из двух невзаимодействующих частей 1 и 2. Квадраты их угловых моментов (если они имеют определенные значения) определяются выражениями $L_1(L_1+1)$ и $L_2(L_2+1)$, где L_1 и L_2 — целые положительные числа. Соответствующие проекции на ось Z (если таковые также имеют определенные значения) могут принимать значения:

$$M_1 = -L_1,\ -(L_1 - 1),\ \ldots,\ +(L_1 - 1),\ +L_1,$$

$$M_2 = -L_2,\ -(L_2 - 1),\ \ldots,\ +(L_2 - 1),\ +L_2.$$

Тогда квадрат результирующего углового момента всей системы может принимать значения $L(L + 1)$, где L меняется в пределах

$$L = L_1 + L_2,\ L_1 + L_2 - 1,\ \ldots,\ L_1 - L_2, \tag{32.5}$$

причем предполагается, что $L_1 > L_2$. Соответствующие проекции на ось Z могут принимать целочисленные значения от $M = -L$ до $M = +L$. Полученный результат называется *правилом сложения угловых моментов.*

В полученном состоянии сложной системы имеют определенные значения также скалярные произведения $\mathbf{L}_1\mathbf{L}_2$, $\mathbf{L}\mathbf{L}_1$ и $\mathbf{L}\mathbf{L}_2$, т. е. собственные значения соответствующих операторов $\widehat{\mathbf{L}}_1\widehat{\mathbf{L}}_2$, $\widehat{\mathbf{L}}\widehat{\mathbf{L}}_1$ и $\widehat{\mathbf{L}}\widehat{\mathbf{L}}_2$. Это следует из формулы (32.4), если написать ее для операторов $\widehat{\mathbf{L}}$, $\widehat{\mathbf{L}}_1$, $\widehat{\mathbf{L}}_2$. Например,

$$\widehat{\mathbf{L}}_1\widehat{\mathbf{L}}_2 = \frac{1}{2}[\widehat{\mathbf{L}}^2 - \widehat{\mathbf{L}}_1^2 - \widehat{\mathbf{L}}_2^2], \tag{32.6}$$

или, переходя к собственным значениям,

$$\mathbf{L}_1\mathbf{L}_2 = \frac{1}{2}[L(L+1) - L_1(L_1+1) - L_2(L_2+1)]. \tag{32.7}$$

Аналогично

$$\mathbf{L}\mathbf{L}_1 = \frac{1}{2}[L(L+1) + L_1(L_1+1) - L_2(L_2+1)], \tag{32.8}$$

$$\mathbf{L}\mathbf{L}_2 = \frac{1}{2}[L(L+1) + L_2(L_2+1) - L_1(L_1+1)]. \tag{32.9}$$

5. Изложенные результаты принято представлять на векторных диаграммах. Складываемые векторы $\mathbf{L}_1$ и $\mathbf{L}_2$ изображаются стрелками с длинами $\sqrt{L_1(L_1+1)}$ и $\sqrt{L_2(L_2+1)}$, а результирующий вектор $\mathbf{L}$ — стрелкой с длиной $\sqrt{L(L+1)}$. В качестве примера на рис. 59 приведена векторная диаграмма для $L_1 = 2$ и $L_2 = 1$ при различных углах

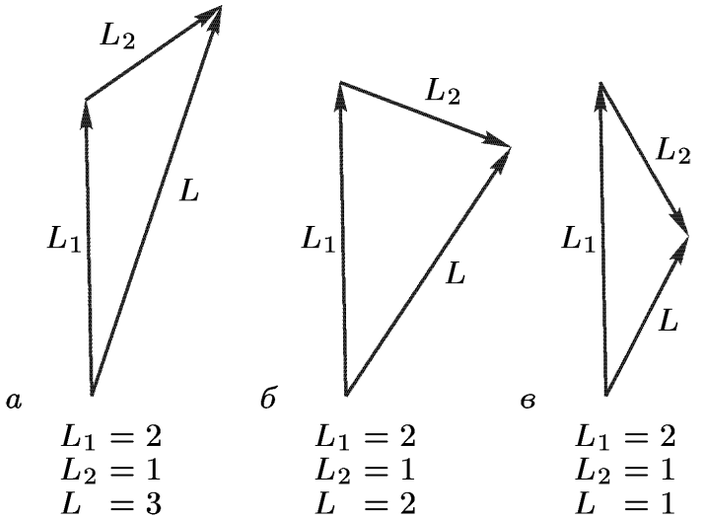


Рис. 59

между векторами $\mathbf{L}_1$ и $\mathbf{L}_2$. Получается всего три возможных случая, в соответствии с тем, что L может принимать значения $L_1 + L_2 = 3$, $L_1 + L_2 - 1 = 2$, $L_1 + L_2 - 2 = 1$. Такая диаграмма правильно передает длины всех векторов и их скалярные произведения. Но она не отражает

истинную квантовую природу угловых моментов, поскольку последние, имея определенные длины, не имеют определенных направлений в пространстве.

Если системы 1 и 2 не взаимодействуют и нет внешних сил, то сохраняется не только результирующий угловой момент системы $\mathbf{L}$, но и оба момента $\mathbf{L}_1$ и $\mathbf{L}_2$. При наличии взаимодействия между системами 1 и 2 моменты $\mathbf{L}_1$ и $\mathbf{L}_2$ в отдельности не сохраняются, а сохраняется только общий момент $\mathbf{L}$ (при отсутствии внешних сил). Если взаимодействие слабое, то по аналогии с классической механикой следует ожидать, что длины векторов $\mathbf{L}_1$ и $\mathbf{L}_2$ при этом практически не будут изменяться. На векторной диаграмме векторы $\mathbf{L}_1$ и $\mathbf{L}_2$ будут совершать прецессию, т. е. вращаться вокруг вектора $\mathbf{L}$ с одной и той же угловой скоростью. К такому же выводу приводит и последовательное квантовое рассмотрение.

§ 33. Квантование водородного атома в общем случае

1. В § 27 была рассмотрена задача о квантовании водородного (или водородоподобного) атома в предположении, что волновая функция ψ радиально симметрична, т. е. зависит только от r. В таком случае угловой момент электрона в атоме равен нулю, так как оператор момента действует только на угловые переменные θ и φ, но не действует на r. Уравнение же Шредингера для стационарных состояний записывается в виде

$$\widehat{H}_r\psi = \mathscr{E}\psi, \tag{33.1}$$

где оператор $\widehat{H}_r$ определяется выражением

$$\widehat{H}_r = -\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r}\right) + U(r) \tag{33.2}$$

и описывает только радиальное квантовое движение электрона в атоме. Учтем теперь зависимость функции ψ также от угловых координат θ и φ. Для этого к оператору $\widehat{H}_r$ надо добавить оператор $\dfrac{1}{2mr^2}\widehat{\mathbf{l}}^2$, соответствующий кинетической энергии вращения электрона вокруг ядра:

$$\widehat{H} = \widehat{H}_r + \frac{1}{2mr^2}\widehat{\mathbf{l}}^2. \tag{33.3}$$

Ясно, что оператор $\widehat{H}_r$ коммутирует с операторами $\widehat{\mathbf{l}}^2$ и $\widehat{l}_z$, поскольку последние не действуют на r, а действуют только на угловые переменные θ и φ. То же относится к оператору $[1/(2mr^2)]\widehat{\mathbf{l}}^2$, так как наличие множителя $1/(2mr^2)$ не отражается на такой коммутации. Следовательно, и полный оператор $\widehat{H}$ коммутирует с $\widehat{\mathbf{l}}^2$ и $\widehat{l}_z$, и стационарное состояние электрона в водородном или водородоподобном атоме

можно характеризовать его энергией $\mathscr{E}$, квадратом углового момента $\mathbf{l}^2$ и его проекцией l_z на избранное направление — ось Z.

Уравнение Шредингера для стационарных состояний теперь запишется в виде

$$\frac{\partial^2\psi}{\partial r^2}+\frac{2}{r}\frac{\partial\psi}{\partial r}+\frac{2m}{\hbar^2}\left(\mathscr{E}-U-\frac{\widehat{\mathbf{L}}^2}{2mr^2}\right)\psi=0. \tag{33.4}$$

Здесь применяется частное дифференцирование по r, поскольку волновая функция ψ может зависеть не только от r, но и от угловых переменных θ и φ. Но какова бы ни была зависимость от θ и φ, для стационарных состояний с определенным значением квадрата углового момента $\widehat{\mathbf{L}}^2\psi=\mathbf{L}^2\psi=\hbar^2 l(l+1)\psi$. Поэтому в таких случаях

$$\frac{\partial^2\psi}{\partial r^2}+\frac{2}{r}\frac{\partial\psi}{\partial r}+\frac{2m}{\hbar^2}\left(\mathscr{E}-U-\frac{\hbar^2 l(l+1)}{2mr^2}\right)\psi=0. \tag{33.5}$$

Это уравнение отличается от уравнения (33.1) наличием в скобках дополнительного члена $-\hbar^2 l(l+1)/(2mr^2)$. Формально уравнение (33.5) имеет вид уравнения Шредингера в радиально-симметричном силовом поле с потенциальной силовой функцией

$$U(r)+\frac{\hbar^2 l(l+1)}{2mr^2}.$$

На второе слагаемое этого выражения можно смотреть как на потенциальную функцию электрона в поле центробежной силы, причем само уравнение (33.5) можно рассматривать как уравнение движения электрона во «вращающейся» системе отсчета.

Поскольку, однако, независимые переменные θ и φ в уравнение (33.5) не входят, волновая функция ψ должна иметь вид $f(\theta,\varphi)\psi(r)$. Во всех вычислениях функция $f(\theta,\varphi)$ будет всюду входить в виде множителя, который не зависит от r, т. е. ведет себя как постоянная. Поэтому ради краткости множитель $f(\theta,\varphi)$ мы будем всюду опускать, т. е. рассуждать так, как если бы функция ψ зависела только от r. Это, очевидно, не нарушает общности рассуждений и их результатов.

2. До сих пор явный вид потенциальной функции $U(r)$ не использовался. Имея теперь в виду водородоподобный атом, положим $U(r)=-Ze^2/r$ и введем такие же обозначения, как в § 27, т. е.

$$\beta^2=-\frac{2m\mathscr{E}}{\hbar^2},\quad q=\frac{2mZe^2}{\hbar^2}. \tag{33.6}$$

Тогда уравнение (33.5) запишется в виде

$$\frac{\partial^2\psi}{\partial r^2}+\frac{2}{r}\frac{\partial\psi}{\partial r}+\left(\frac{q}{r}-\beta^2-\frac{l(l+1)}{r^2}\right)\psi=0. \tag{33.7}$$

Для его исследования применяем такой же метод, что и для исследования уравнения (27.1), т. е. вводим новую функцию $u(r)$ с помощью соотношения

$$\psi = \frac{u(r)}{r}\, e^{-\beta r}. \tag{33.8}$$

Тогда

$$\frac{d^2u}{dr^2} - 2\beta\frac{du}{dr} + \left[\frac{q}{r} - \frac{l(l+1)}{r^2}\right]u = 0. \tag{33.9}$$

Ищем решение этого уравнения в виде ряда

$$u = \sum_{k=\gamma}^{\infty} a_k r^k \tag{33.10}$$

и путем сравнения коэффициентов находим

$$\begin{gathered}\gamma(\gamma-1) - l(l+1) = 0,\\ [k(k+1) - l(l+1)]\, a_{k+1} = (2\beta k - q)a_k \text{ при } k \neq \gamma.\end{gathered} \tag{33.11}$$

Из первого уравнения получается либо $\gamma = l+1$, либо $\gamma = -l$. Значение $\gamma = -l$ должно быть отброшено по тем же соображениям, которые применялись при решении аналогичного вопроса в § 27. Таким образом, следует воспользоваться значением $\gamma = l+1$.

Для исследования сходимости ряда (33.10) из формулы (33.11) находим

$$\frac{a_{k+1}}{a_k} = \frac{2\beta k - q}{k(k+1) - l(l+1)}.$$

Асимптотически

$$\frac{a_{k+1}}{a_k} = \frac{2\beta}{k+1}.$$

Это выражение в точности совпадает с соответствующим выражением из § 27. Поэтому, как и раньше, заключаем, что ряд (33.10) должен обрываться. Из условия обрыва получаем прежнюю формулу (27.8) для энергии атома:

$$\mathscr{E} = -\frac{mZ^2e^4}{2\hbar^2 n^2}. \tag{33.12}$$

3. Из изложенного следует, что значения энергии в стационарных состояниях водородоподобного атома зависят только от главного квантового числа n. Но состояния с заданным n (т. е. с заданной энергией $\mathscr{E}$) могут отличаться одно от другого различными значениями квантовых чисел l и m (не путать с массой). Таким образом, одному и тому же значению $\mathscr{E}$ соответствуют несколько различных квантовых состояний. В этом случае говорят, что состояние с энергией $\mathscr{E}$ *вырождено*. Энергетический уровень $\mathscr{E}$ называют также вырожденным.

Число независимых состояний, суперпозицией которых может быть получено заданное состояние с энергией $\mathscr{E}$, называется *степенью* или

кратностью вырождения. Найдем степень вырождения для водородоподобного атома в состоянии с заданным главным квантовым числом n.

Рассмотрим сначала состояния, в которых (наряду с n) имеет определенное значение и число l. Воспользуемся формулой $\gamma = l + 1$, которая до сих пор еще не принималась во внимание. Имея в виду, что ряд (33.10) должен обрываться на члене n-й степени, запишем его в виде конечной суммы:

$$u = \sum_{k=l+1}^{n} a_k r^k = r^{l+1} \sum_{\alpha=0}^{\alpha=n-l-1} a_{l+\alpha+1} r^\alpha. \tag{33.13}$$

Отсюда видно, что главное квантовое число n имеет смысл старшего показателя степени в полиноме (33.13). Число l называется *орбитальным квантовым числом.* Оно определяет квадрат углового момента,

$$\mathbf{l}^2 = l(l+1) \quad (\text{в единицах } \hbar^2).$$

Фиксируя n, подсчитаем число квантовых состояний, отличающихся одно от другого значениями l. Наименьшее значение l есть $l = 0$, наибольшее $l = n - 1$, так как в этом случае сумма (33.13) сводится к одному члену. Следовательно, при заданном n число l может принимать значения

$$l = 0,\ 1,\ 2,\ \ldots,\ (n-1), \tag{33.14}$$

т. е. всего n значений и соответствующих им квантовых состояний с определенными n и l. Функция u, как видно из (33.13), имеет $n_r = n - l - 1$ узлов, если исключить из рассмотрения узел $r = 0$. Но число узлов, как известно (см. § 25, п. 1), определяет номер волновой функции с заданным n. Поэтому главное квантовое число n можно также определить соотношением

$$n = n_r + l + 1. \tag{33.15}$$

Квантовое число n_r было введено еще Зоммерфельдом в старой квантовой теории и получило название *радиального квантового числа.*

Заметим теперь, что в состоянии с определенным l может иметь различные значения квантовое число m, определяющее проекцию углового момента на ось Z (оно называется *магнитным*). Именно:

$$m = -l,\ -(l-1),\ \ldots,\ -1,\ 0,\ +1,\ \ldots,\ +(l-1),\ +l, \tag{33.16}$$

т. е. всего $2l + 1$ значений. Поэтому полное число квантовых состояний, с помощью которых может реализоваться состояние с заданным n, равно

$$N = \sum_{l=0}^{l=n-1} (2l+1) = n^2. \tag{33.17}$$

В действительности, как будет показано в § 36, это число следует удвоить из-за наличия спина электрона. Таким образом, кратность вырождения энергетического уровня в водородоподобном атоме равна $2n^2$.

§ 34. Энергетические уровни и спектральные серии щелочных металлов

1. В атомах щелочных металлов (литий, натрий, калий, рубидий, цезий) электронная оболочка содержит один наружный (валентный) электрон, сравнительно слабо связанный с ядром атома. То же относится к ионизованным атомам, если только они содержат по одному наружному (валентному) электрону (однократно ионизованный атом гелия, двукратно ионизованный атом лития, трехкратно ионизованный атом бериллия и т. д.). Переходы между энергетическими уровнями валентного электрона сопровождаются излучением или поглощением квантов сравнительно низких частот — из оптической области спектра. Остальные $Z-1$ электронов (Z — заряд ядра, выраженный в элементарных зарядах) вместе с ядром образуют сравнительно прочный остов, в электрическом поле которого движется валентный электрон. Изменения энергии квантовых уровней остова сравнительно велики и порождают рентгеновские спектры. Этот вопрос здесь рассматриваться не будет. Мы сосредоточим внимание только на излучении и поглощении света, связанных с поведением валентного электрона.

При такой постановке вопроса атом щелочного металла может рассматриваться как *одноэлектронный атом*, в котором роль ядра играет указанный остов. Последний можно характеризовать каким-то эффективным зарядом $Z_a e$. Для нейтрального атома $Z_a = Z - 1$, для однократно ионизованного $Z_a = Z - 2$, для двукратно ионизованного $Z_a = Z - 3$ и т. д. Если удалить валентный электрон, то распределение электрических зарядов в остове и его электрическое поле сделаются сферически симметричными.

2. Стационарные состояния валентного электрона в таком поле определяются теми же тремя квантовыми числами, что и в атоме водорода, а именно *главным квантовым числом* n, *орбитальным квантовым числом* l и *магнитным квантовым числом* m. Квантовое число l определяет квадрат углового момента электрона

$$\mathbf{l}^2 = \hbar^2 l(l+1),$$

магнитное число m определяет проекцию углового момента на избранное направление, обычно принимаемое за ось Z:

$$m_z = l_z m\hbar.$$

При заданном l число m может принимать $2l+1$ значений, а именно

$$m = -l,\ -(l-1),\ \ldots,\ -1,\ 0,\ +1,\ \ldots,\ (l-1),\ l.$$

Главное квантовое число определяется формулой $n = n_r + l + 1$, где n_r — так называемое *радиальное квантовое число*, равное числу узлов волновой функции ψ вдоль радиуса (точка $r = 0$ за узел не считается). При заданном n число l может принимать следующие значения:

$$l = 0,\ 1,\ 2,\ \ldots,\ (n-1).$$

Таким образом, получается всего n^2 независимых квантовых состояний, с помощью которых может быть реализовано любое состояние с заданным значением главного квантового числа n. Мы увидим дальше, что три квантовых числа n, l, m должны быть дополнены четвертым — *спиновым* — квантовым числом m_s, которое может принимать два значения $m_s = \pm 1/2$. От этого общее число независимых квантовых состояний удваивается. Но от спина электрона мы в этом параграфе отвлечемся.

Различные состояния электрона в атоме принято обозначать малыми буквами латинского алфавита в зависимости от значения числа l и в соответствии со следующей схемой:

Квантовое число	0	1	2	3	4	5	6	7
Обозначение состояния	s	p	d	f	g	h	i	k

Говорят, например, об s-состояниях и s-электронах, p-состояниях и p-электронах и т. д. Такая терминология сложилась под влиянием ранних спектроскопических исследований, когда еще не существовало представления не только о квантовых состояниях, но и о строении самого атома.

3. Обращаемся к рассмотрению поведения валентного электрона в электрическом поле ядерно-электронного остова. Внешний электрон, действуя на остов, искажает распределение зарядов и электрическое поле остова. В первом приближении поле остова можно рассматривать как наложение поля точечного заряда $Z_a e$ и поля точечного диполя, расположенных в центре атома. При этом ось диполя направлена к внешнему электрону. Поэтому движение последнего происходит так, как если бы поле остова, хотя и искажалось электроном, но сохранялось сферически симметричным. Соответствующая потенциальная функция может быть представлена в виде

$$U = -\frac{Z_a e^2}{r} - C\frac{Z_a e^2}{r^2}, \tag{34.1}$$

где C — постоянная. На это выражение следует смотреть как на разложение функции U по степеням $1/r$, оборванное на втором члене. В соответствии с этим член $-CZ_a e^2/r^2$ надо рассматривать как поправку к основному члену $-Z_a e^2/r$. Таким образом, в принятом приближении все отличие от водородоподобного атома состоит в том, что к потенциальной функции добавляется член $-C(Z_a e^2/r^2)$. В уравнении (33.5) этот член можно объединить с центробежной энергией и полученную сумму представить в виде

$$\frac{\hbar^2 l(l+1)}{2mr^2} - C\frac{Z_a e^2}{r^2} = \frac{\hbar^2 l^*(l^*+1)}{2mr^2}, \tag{34.2}$$

где постоянное число l^* определяется квадратным уравнением

$$l^*(l^*+1) = l(l+1) - \frac{2m}{\hbar^2} C Z_a e^2. \tag{34.3}$$

В результате мы снова придем к уравнению (33.7), в котором число l надо заменить числом l^*, т. е. к уравнению

$$\frac{\partial^2\psi}{\partial r^2}+\frac{2}{r}\frac{\partial\psi}{\partial r}+\left(\frac{q^*}{r}-\beta^2-\frac{l^*(l^*+1)}{r^2}\right)\psi=0, \tag{34.4}$$

причем теперь

$$q^*=2mZ_ae^2/\hbar^2. \tag{34.5}$$

В отличие от l, число l^*, вообще говоря, не целое. Но это не имеет никакого значения для применения метода, изложенного в предыдущем параграфе. По-прежнему функцию $u(r)$ следует искать в виде ряда (33.10). Для l^* из (34.3) получается выражение

$$l^*=-\frac{1}{2}\pm\sqrt{\left(l+\frac{1}{2}\right)^2-\frac{2m}{\hbar^2}CZ_ae^2}. \tag{34.6}$$

При этом положительную величину $(2m/\hbar^2)CZ_ae^2$ следует рассматривать как поправку к основному члену $(l+1/2)^2$. Когда эта поправка обращалась в нуль, мы видели, что перед квадратным корнем следует брать знак плюс. Следовательно, то же надо делать и тогда, когда поправка отлична от нуля. При отсутствии поправки $\gamma=l+1$, при наличии таковой $\gamma=l^*+1$. Таким образом, все отличие атома щелочного металла от водородоподобного сводится к замене числа l на число l^*. Поэтому энергетические уровни атомов щелочных металлов и сходных с ними ионов должны определяться формулой

$$\mathscr{E}=-\frac{mZ_a^2e^4}{2\hbar^2(n_r+l^*+1)^2}.$$

Вводя главное квантовое число (33.15), этой формуле можно придать вид

$$\mathscr{E}=-\frac{mZ_a^2e^4}{2\hbar^2(n+\Delta)^2}, \tag{34.7}$$

где $\Delta=l^*-l$.

4. Величина поправки Δ, как видно из формулы (34.6), зависит от орбитального квантового числа l. Постоянная C также зависит от l. Качественно происхождение последней зависимости особенно ясно с точки зрения модельной теории Бора. Максимальному значению l в этой теории соответствует круговая орбита, минимальному — наиболее вытянутая эллиптическая орбита. В последнем случае орбита валентного электрона глубоко проникает внутрь атомного остова и поэтому сильно возмущает его поле. В случае круговых орбит такого проникновения нет, и возмущение получается меньше. Квантовая механика, отвергая в принципе наглядные классические траектории, по существу дает такое же объяснение. С ее точки зрения дело сводится к тому, что вероятность обнаружения электрона глубоко в атомном остове больше при малых l, чем при больших. В самом деле, при больших l плотность

вероятности обнаружения электрона в атоме ближе к сферически симметричной, чем при малых. При максимальных значениях $l = n - 1$ она становится просто сферически симметричной. При малых l дело обстоит наоборот.

В предыдущем параграфе было показано, что в водородном или водородоподобном атоме при одном и том же n различным значениям l, m соответствует один и тот же уровень энергии. Таким образом, энергия зависит только от n, т. е. имеет место *вырождение по обоим квантовым числам* l и m. Такое вырождение — случайное и связано с тем, что электрическое поле ядра атома водорода, т. е. протона, — кулоновское, поскольку протон можно рассматривать как точечный заряд. В атоме щелочного металла валентный электрон находится в электрическом поле атомного остова. Заряд последнего — не точечный, хотя и распределен в нем сферически симметрично. Электрическое поле остова уже не обратно пропорционально квадрату расстояния до центра остова. Благодаря этому и получается зависимость энергии электрона не только от главного квантового числа n, но и от орбитального числа l. Иными словами, в некулоновском центрально-симметричном поле вырождение по l снимается. Вырождение по m остается, так как энергия не может зависеть от m ввиду изотропии пространства. С этим и связано отличие спектральных термов щелочных металлов от термов атома водорода.

5. Уровням энергии (34.7), согласно формуле (12.3), в атомах щелочных металлов соответствуют спектральные термы

$$T = -\frac{\mathscr{E}}{ch} = -\frac{\mathscr{E}}{2\pi c\hbar} = \frac{Z_a^2 R}{(n+\Delta)^2}, \tag{34.8}$$

где

$$R = \frac{2m\pi^2 e^4}{ch^3} = \frac{me^4}{4c\pi\hbar^3}. \tag{34.9}$$

Такой вид термов для щелочных металлов (т. е. при $Z_a = Z - 1$) был эмпирически установлен Ридбергом в конце XIX века. Эти термы имеют водородоподобный вид. Они отличаются от водородных термов поправкой Δ, которая для водорода равна нулю. Замечательно (и это поразило первых исследователей), что R есть та же постоянная Ридберга, которая входит в выражения для водородных термов.

Спектральные термы щелочных атомов характеризуются двумя квантовыми числами: главным квантовым числом n и орбитальным квантовым числом l. Главное квантовое число ставится впереди и обозначается цифрой, за ним следует значение числа l, обозначаемое буквой, в соответствии со схемой п. 2. Например, $3s$ обозначает терм с $n = 3$, $l = 0$; символ $5d$ обозначает терм с $n = 5$, $l = 2$ и т. д. Таким образом, для щелочных металлов получаются следующие обозначения термов:

$$ns = \frac{R}{(n+s)^2}, \quad np = \frac{R}{(n+p)^2}, \quad nd = \frac{R}{(n+d)^2}, \tag{34.10}$$

и т. д. Числовые значения поправки к числу n в правой части обозначены здесь через s, p, d, ... Их не надо смешивать с теми же символами, стоящими слева (здесь они обозначают соответственно $l = 0$, $l = 1$, $l = 2$, ...). Такая терминология исторически сложилась еще при чисто эмпирических исследованиях спектральных закономерностей. Число Z_a здесь положено равным единице, так как имеются в виду не ионы, а нейтральные атомы.

6. Самый низкий уровень энергии водородоподобного атома занимает электрон с главным квантовым числом $n = 1$. Это же главное квантовое число может иметь еще один электрон (подробнее см. § 38). С присоединением его в случае $Z = 2$ получается нейтральный атом гелия. Больше двух электронов с главным квантовым числом $n = 1$ быть не может. Такие два электрона образуют замкнутую оболочку гелия — то, что выше было названо атомным остовом. Если присоединить третий электрон, то он начинает занимать следующую оболочку с главным квантовым числом $n = 2$. Тогда при $Z = 3$ получится щелочной элемент литий. Построение этой замкнутой оболочки заканчивается на элементе неоне. Далее идут щелочные элементы, с которых начинается построение следующих замкнутых оболочек: натрий ($n_{\text{мин}} = 3$), калий ($n_{\text{мин}} = 4$), рубидий ($n_{\text{мин}} = 5$), цезий ($n_{\text{мин}} = 6$).

7. Путем комбинаций различных термов возникают спектральные линии в соответствии с комбинационным принципом Ритца. Только не все комбинации, сопровождающиеся излучением (или поглощением) спектральных линий, разрешены. Уже эмпирически давно было замечено, что в щелочных металлах при излучении (и поглощении) света термы s комбинируют только с термами p; термы p комбинируют только с термами s и d, термы d комбинируют только с термами p и f, и т. д. На атомном языке это означает, что при соответствующих квантовых переходах, сопровождающихся излучением света, число l может меняться только на единицу:

$$\Delta l = \pm 1.$$

Этот и аналогичные результаты получили название *правил отбора.* Остальные комбинации *запрещены.* Это не значит, что соответствующие переходы вообще невозможны. Правила отбора относятся только к *дипольному излучению и поглощению света* и не относятся к другим процессам, которые могут происходить в атомах. Например, при ударе возможны переходы с какого-либо уровня s на уровни d, f, g и т. д. Однако при этом не происходит изменения дипольного момента атома, сопровождающееся излучением света. Более того, и «запрещенные переходы» могут сопровождаться испусканием спектральных линий. Но это не будет *дипольное излучение*, при котором меняется дипольный момент атома, а, например, *квадрупольное* или *октупольное* излучения, возникающие из-за изменения квадрупольного и октупольного моментов. К такого рода излучениям правило отбора $\Delta l = \pm 1$ не относится. Заметим еще, что на изменения главного квантового числа n никакие ограничения не накладываются.

Квантовая механика сняла покров таинственности, с которой до нее воспринимались правила отбора. Она поставила и разрешила более общую задачу: найти *вероятность перехода* атомной системы с излучением света из одного квантового состояния в другое. Оказалось, что при несоблюдении правил отбора соответствующая вероятность обращается в нуль. Это и приводит к правилам отбора или, лучше, *правилам запрета* (для дипольного излучения света). Эти правила можно получить, и не прибегая к конкретному вычислению соответствующих вероятностей, а из общих соображений. Надо воспользоваться тем, что законы природы *зеркально симметричны*, т. е. *инвариантны относительно операции инверсии* — замене всех трех направлений координатных осей на противоположные. Но на этом вопросе мы остановимся в главах, посвященных ядерной физике.

8. С учетом правила отбора $\Delta l = \pm 1$ легко понять, что в спектрах щелочных элементов должны получаться (и действительно получались на опыте еще до всякой теории) следующие *спектральные серии*:

главная серия

$$\overline{\nu} = ns - mp; \tag{34.11}$$

первая побочная, или *диффузная*, *серия*

$$\overline{\nu} = np - md; \tag{34.12}$$

вторая побочная, или *резкая*, *серия*

$$\overline{\nu} = np - ms; \tag{34.13}$$

серия Бергмана, или *фундаментальная*, *серия*

$$\overline{\nu} = nd - mf; \tag{34.14}$$

Здесь $\overline{\nu} = 1/\lambda$ — *спектроскопическое волновое число*. Число n в каждой серии сохраняется постоянным, число m в главной серии должно пробегать значения $m = n, n+1, n+2, \ldots$, а в остальных сериях — значения $m = n+1, n+2, \ldots$, чтобы получились все линии серий. Поправки Δ в переменных членах в пределах каждой серии остаются практически постоянными, но меняются от серии к серии.

Схема квантовых переходов и соответствующие им спектральные линии представлены на рис. 60 для лития и на рис. 61 для натрия. Длины волн спектральных линий даны в ангстремах (1 Å = 0,1 нм). В первом столбце (s) на рис. 60 изображены энергетические уровни лития в s-состоянии при различных значениях главного квантового числа, во втором — то же при различных n, но при одних и тех же значениях p, в третьем — при одних и тех же значениях d и т. д. В случае кулоновского поля получились бы энергетические уровни, зависящие только от n, но не зависящие от l. Для щелочных металлов такое вырождение энергетических уровней снимается из-за отличия электрического поля от кулоновского. Детали спектральной диаграммы натрия, а также вопросы, касающиеся спектральной символики, выяснятся в § 40.

Главная серия возникает в результате квантовых переходов валентного электрона из различных p-состояний на наиболее глубокий s-уровень. На этом уровне атом находится в невозбужденном состоянии

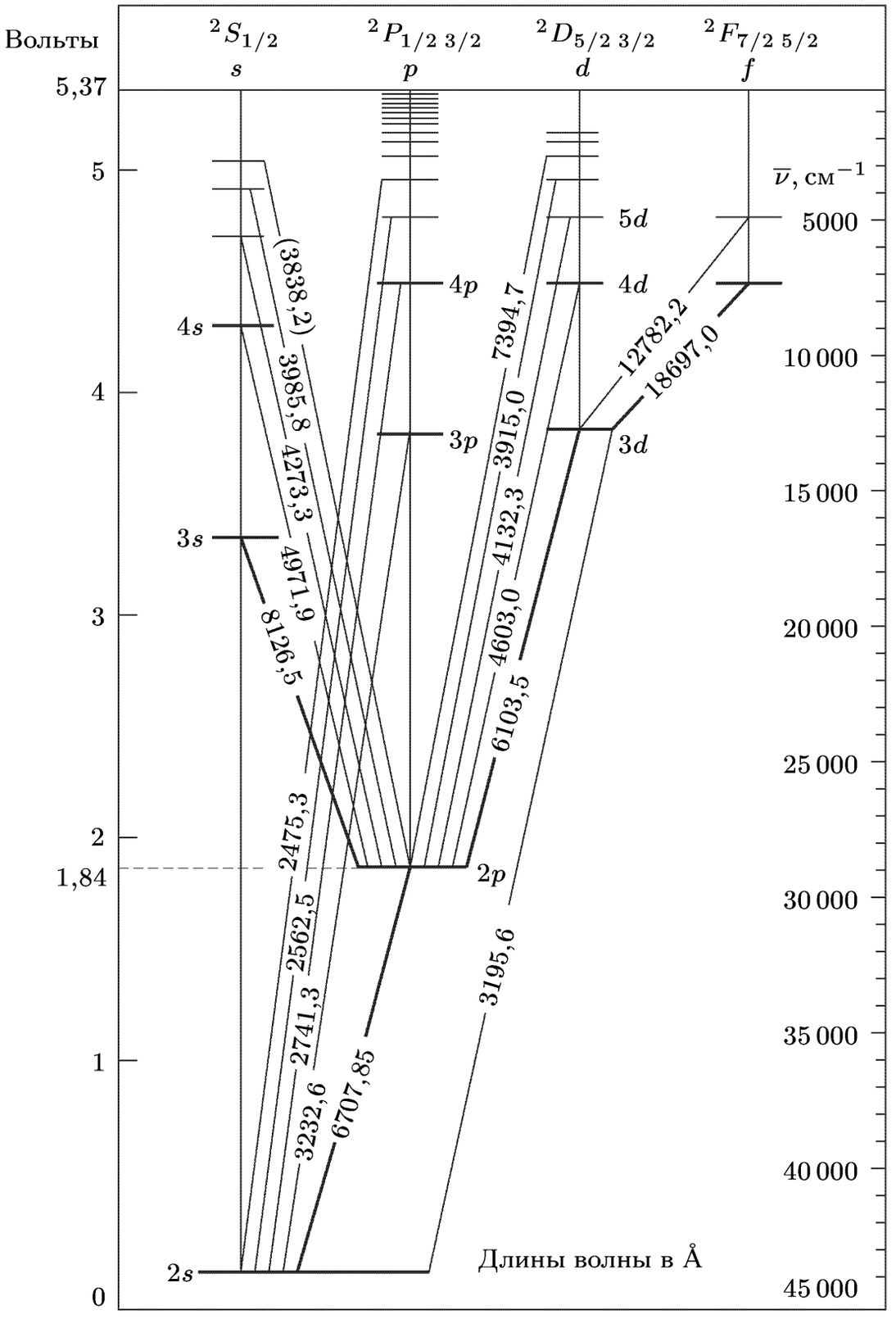


Рис. 60

и может переходить в энергетически более высокие состояния. Поэтому главная серия наблюдается не только в *испускании*, но и в *поглощении*. На рис. 62 показана в спектре поглощения главная серия паров натрия. Приведена только коротковолновая часть серии, начиная с пятого чле-

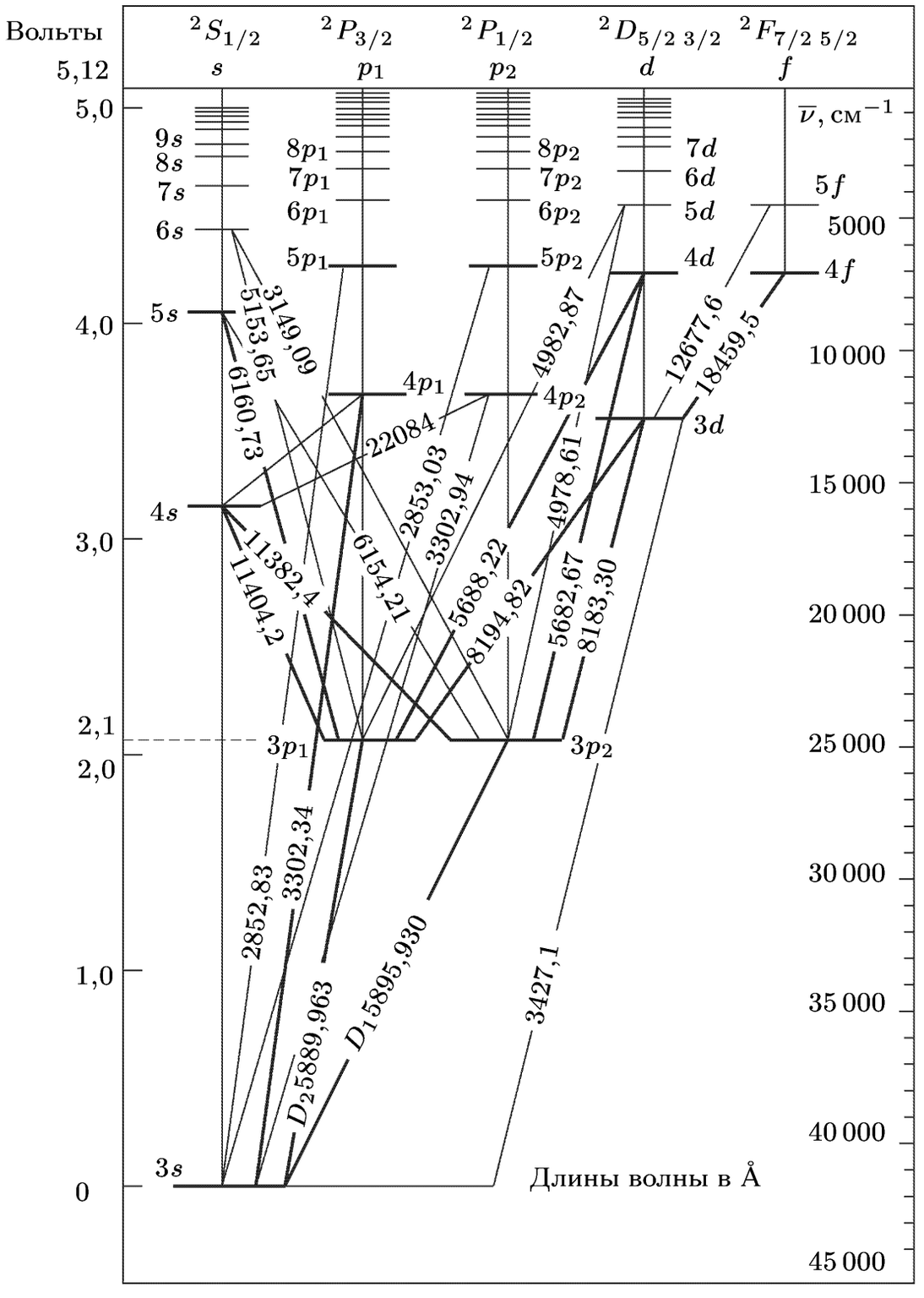

Вольты
5,12
5,0
4,0
3,0
2,1
2,0
1,0
0
$^2S_{1/2}$ s
$^2P_{3/2}$ p_1
$^2P_{1/2}$ p_2
$^2D_{5/2\,3/2}$ d
$^2F_{7/2\,5/2}$ f
$\overline{\nu}$, см$^{-1}$
5000
10 000
15 000
20 000
25 000
30 000
35 000
40 000
45 000
$9s$
$8s$
$7s$
$6s$
$5s$
$4s$
$3s$
$8p_1$
$7p_1$
$6p_1$
$5p_1$
$4p_1$
$3p_1$
$8p_2$
$7p_2$
$6p_2$
$5p_2$
$4p_2$
$3p_2$
$7d$
$6d$
$5d$
$4d$
$3d$
$5f$
$4f$
3149,09
5153,95
6160,73
22084
2853,03
3302,94
4982,87
4978,61
12677,6
18459,5
11382,4
11404,2
6154,21
5688,22
8194,82
5682,67
8183,30
2852,83
3302,34
$D_2$5889,963
$D_1$5895,930
3427,1
Длины волны в Å

Рис. 61

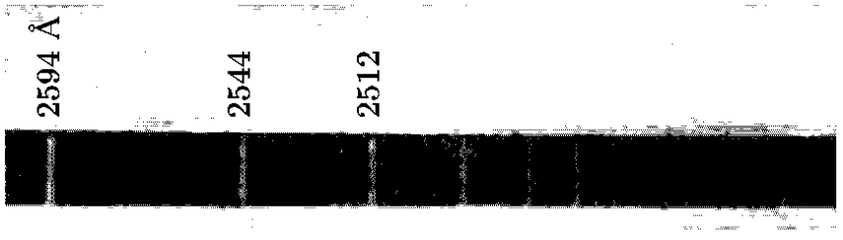

2594 Å
2544
2512

Рис. 62

на ($\lambda = 2594$ Å). Главная серия содержит наиболее характерную для рассматриваемого элемента резонансную линию

$$\overline{\nu}_{\text{рез}} = ns - np. \tag{34.15}$$

Таковы, например, красная линия лития и желтая линия натрия.

Первая побочная (диффузная) серия возникает в результате переходов валентного электрона из различных d-состояний на наиболее глубокий p-уровень, вторая побочная (резкая) — из различных s-состояний на тот же самый глубокий p-уровень. Оправдание названий «диффузная» и «резкая» будет дано в § 40.

ЗАДАЧИ

1. Будут ли в принципе одинаковы спектральные линии атомов, ядра которых имеют одинаковые заряды, но отличаются размерами и формой?

Р е ш е н и е. Спектры будут несколько (хотя и очень мало) отличаться друг от друга, так как по волновой механике поведение электрона определяется волновым уравнением во всем пространстве, где существует силовое поле.

2. Для лития значение терма $2p$, вычисленное из пределов побочных серий, равно $2p = 28\,581{,}4$ см$^{-1}$. Длины волн линий $2p - 3d$ и $3d - 4f$ равны соответственно $\lambda_1 = 6103{,}77$ Å и $\lambda_2 = 18\,697{,}0$ Å. Вычислить длину волны $2p - 4f$.

Р е ш е н и е. Спектроскопические волновые числа линий $2p - 3d$ и $3d - 4f$ равны соответственно

$$1/\lambda_1 = \overline{\nu}_1 = 16\,383{,}3 \text{ см}^{-1}, \quad 1/\lambda_2 = \overline{\nu}_2 = 5348{,}4 \text{ см}^{-1}.$$

Отсюда $3d = 2p - \overline{\nu}_1 = 12\,198{,}1$ см$^{-1}$, $4f = 3d - \overline{\nu}_2 = 6849{,}7$ см$^{-1}$; $2p - 4f = 21\,731{,}7$ см$^{-1}$, $\lambda = 4602{,}8$ Å.

§ 35. Магнетизм атомов

1. Со времени Ампера (1775–1836) магнетизм был сведен к *электрическим токам*, которые, по его представлениям, циркулируют внутри мельчайших частиц вещества (атомов и молекул). Природа этих токов была установлена с появлением электронных представлений о строении вещества и теории Бора. Считалось, что *амперовы молекулярные токи* создаются электронами, вращающимися вокруг ядра атома. Однако классическая физика до введения квантовых представлений была не в состоянии объяснить не только движение электронов вокруг ядра, но и сам факт существования атомов. Методами статистической физики было строго показано, что *с классической точки зрения в установившемся состоянии вещество не может быть намагничено, т. е. не может иметь отличный от нуля магнитный момент* (Бор, Лоренц, Ван-Лёвен; см. т. III, § 75). Это не значит, что его нельзя намагнитить вообще. Электрические заряды можно привести во вращение, т. е. возбудить в веществе круговые токи. А в таком случае появится магнитный момент, т. е. намагничивание вещества. Смысл приведенного утверждения состоит в том, что *если намагниченное вещество предоставить самому себе, поддерживая температуру его постоянной*,

то оно самопроизвольно придет в равновесное состояние, в котором всякая намагниченность исчезнет, даже если вещество помещено в магнитное поле. Это не согласуется с фактами.

Понимание природы магнетизма пришло только после создания квантовой механики. Магнетизм, как и существование атомов и молекул, оказался *квантовым эффектом.* Классические теории намагничивания (Ланжевен) имели известный успех, и притом немалый, только потому, что они молчаливо вводили допущения *существенно квантового характера*, а именно существование у атомов готовых магнитных моментов, или стационарных орбит, по которым вращаются электроны. А это, в сущности, и должна была бы объяснить теория.

2. Поскольку электроны, образующие оболочку атома, заряжены и обладают массами, с их движением в оболочке (оно называется орбитальным) связан не только момент количества движения, но и *магнитный момент атома.* Связь между этими двумя моментами уже рассматривалась в т. III (§ 75) — в той мере, как это можно было сделать до введения квантовых представлений. Та же связь сохраняется и в квантовой механике. Но ее смысл, а потому и обоснование — несколько иные, чем в классической механике, так как понятие момента количества движения (углового момента) не может быть перенесено автоматически из классической теории в квантовую. Это делается посредством введения соответствующего оператора. Так же надо поступить и с понятием магнитного момента. Отправным пунктом при этом должно служить классическое рассмотрение, с которого мы и начнем.

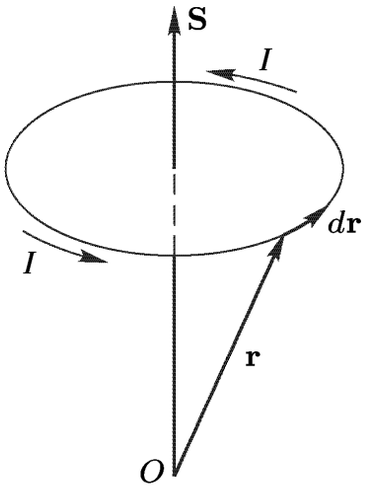


Рис. 63

Согласно электродинамике (см. т. III, § 75) замкнутый виток постоянного тока I (рис. 63) обладает магнитным моментом

$$\mathfrak{m} = \frac{I}{c}\mathbf{S}, \tag{35.1}$$

где $\mathbf{S}$ — вектор площади, натянутой на контур тока. Этот вектор выражается формулой

$$\mathbf{S} = \frac{1}{2}\oint[\mathbf{r}\,d\mathbf{r}]$$

и не зависит от выбора начала координат O, поскольку контур тока замкнут. Направление обхода контура предполагается совпадающим с направлением тока. Оно находится в правовинтовом соотношении с вектором $\mathbf{S}$. Таким образом, магнитный момент замкнутого постоянного тока можно представить в виде

$$\mathfrak{m} = \frac{1}{2c}\oint I[\mathbf{r}\,d\mathbf{r}]. \tag{35.2}$$

Но ток I образуется движущимися зарядами. Последние и являются непосредственными создателями магнитного момента $\mathfrak{m}$. Каждый заряд, если он движется, создает магнитный момент. Полный магнитный момент тела образуется векторной суперпозицией магнитных моментов отдельных зарядов, движущихся в нем. Преобразуем поэтому контурный интеграл (35.2) в интеграл по всем движущимся зарядам тела. Пусть dq — заряд, проходящий за время dt через поперечное сечение витка с током (в случае постоянного тока эта величина не зависит от того, в каком месте взято сечение витка). Тогда $I = dq/dt$,

$$\mathfrak{m} = \frac{1}{2c}\oint \frac{dq}{dt}\,[\mathbf{r}\,d\mathbf{r}] = \frac{1}{2c}\oint\left[\mathbf{r}\frac{d\mathbf{r}}{dt}\right] dq.$$

В этой формуле интегрирование производится еще по $d\mathbf{r}$, так что интеграл остается контурным. Выберем, однако, элемент контура $d\mathbf{r}$ так, чтобы за время dt заряд dq перемещался на $d\mathbf{r}$. Тогда $d\mathbf{r} = v\,dt$, и мы получим

$$\mathfrak{m} = \frac{1}{2c}\int[\mathbf{rv}]\,dq = \frac{1}{2\mu c}\int[\mathbf{rp}]\,dq, \tag{35.3}$$

где $\mathbf{v}$ — скорость, $\mathbf{p}$ — импульс, а μ — масса, связанная с движущимся зарядом dq. (Для массы используется обозначение μ, так как через m обозначается магнитное квантовое число.)

Но при сделанном выборе dq есть как раз заряд, содержащийся в рассматриваемый момент времени на элементе контура $d\mathbf{r}$. При таком истолковании заряда dq время dt выпало из формулы (35.3). Из нее выпало и всякое упоминание о витке с постоянным током (поэтому-то и опущен кружок у знака интеграла). Осталась только система зарядов, каждый из которых, помимо своей величины, характеризуется положением и скоростью движения. Только это и существенно для создания магнитного момента тела. Как создается система зарядов и ее состояние — это не имеет значения.

Формула (35.3) и представляет магнитный момент тела как суперпозицию магнитных моментов движущихся зарядов. Ее можно обобщить и записать в виде

$$\mathfrak{m} = \frac{1}{2c}\sum_i q_1[\mathbf{r}_i\mathbf{v}_i], \tag{35.4}$$

предполагая, что имеется в виду система точечных зарядов q_i, движущихся в рассматриваемый момент со скоростями v_i. Никаких предположений о характере движения при этом не вводится.

3. Классическое выражение (35.4) для магнитного момента системы движущихся зарядов зависит от выбора начала координат. Действительно, если $\mathbf{a}$ — радиус-вектор нового (штрихованного) начала относительно старого (нештрихованного), то для всех зарядов $\mathbf{r}_i = \mathbf{r}'_i + \mathbf{a}$, так что

$$\mathfrak{m} = \mathfrak{m}' + \frac{1}{2c}\sum_i q_i[\mathbf{a}\mathbf{v}_i].$$

Отсюда видно, что старый $\mathfrak{m}$ и новый $\mathfrak{m}'$ магнитные моменты только тогда будут всегда одинаковы, когда для любого вектора $\mathbf{a}$ векторное произведение $\left[\mathbf{a}\sum q_i\mathbf{v}_i\right]$ обращается в нуль. В частности, это имеет место для всякого замкнутого неподвижного витка постоянного тока, так как тогда $\sum q_i\mathbf{v}_i = 0$.

4. Для одиночного точечного заряда, движущегося со скоростью $\mathbf{v}$,

$$\mathfrak{m} = \frac{q}{2c}[\mathbf{rv}] = \frac{q}{2\mu c}[\mathbf{rp}], \tag{35.5}$$

где μ — масса, а $\mathbf{p}$ — импульс частицы, несущей этот заряд. Таким образом, классическая физика приводит к соотношению

$$\mathfrak{m} = \Gamma\mathbf{L}, \tag{35.6}$$

где

$$\Gamma = \frac{q}{2\mu c}. \tag{35.7}$$

Эти формулы более примитивным путем уже были получены в т. III (см. § 75). Для электрона $q = -e$,

$$\Gamma = -\frac{e}{2\mu_e c}. \tag{35.8}$$

В этом случае отношение Γ магнитного момента электрона к механическому называется *гиромагнитным отношением* для орбитального движения электрона.

Заметим еще, что при выводе всех полученных соотношений применялась *нерелятивистская механика* (зависимость массы от скорости не учитывалась), а частицы считались точечными. Впрочем, частицы могут быть и протяженными, так как их можно мысленно разбить на малые части и рассматривать последние как точки. Однако чтобы отношение $\mathfrak{m}/L$ не изменилось, необходимо предположить, что заряды и массы распределены в пространстве по одному и тому же закону. Для заряженного шарика, например, вращающегося вокруг диаметра с нерелятивистской скоростью, классическая физика приводит к формулам (35.7) и (35.8) независимо от того, как распределены в нем заряды и массы; важно только, чтобы обе величины были распределены *одинаково*. Но, конечно, результат получится иной, если, например, заряд будет находиться в центре, а масса равномерно распределена по объему шарика.

5. Теперь следует классические представления заменить квантовыми. В квантовой механике формула (35.5) не может служить определением магнитного момента, поскольку не существует никакого состояния частицы, которое характеризовалось бы и ее точным положением $\mathbf{r}$, и ее точным импульсом $\mathbf{p}$. Как и в случае углового момента, от классической формулы (35.5) квантовая механика переходит к *операторному соотношению*

$$\widehat{\mathfrak{m}} = \frac{q}{2c}\left[\widehat{\mathbf{rv}}\right] = \frac{q}{2\mu c}\left[\widehat{\mathbf{rp}}\right], \tag{35.9}$$

или

$$\widehat{\mathfrak{m}} = \Gamma \widehat{\mathbf{L}}. \tag{35.10}$$

Изучение магнитного момента частицы тем самым сводится к изучению свойств оператора $\widehat{\mathfrak{m}}$. Поскольку операторы $\widehat{\mathfrak{m}}$ и $\widehat{\mathbf{L}}$ отличаются только постоянным множителем, их свойства совершенно аналогичны. В частности, оператор $\widehat{\mathfrak{m}}$, как и $\widehat{\mathbf{L}}$ совершенно *не зависит от выбора начала координат*. Магнитный и угловой моменты квантуются по одинаковым правилам. Составляющие магнитного момента на любые два различных направления *не могут одновременно иметь определенные значения*. В стационарном состоянии определенные значения могут иметь квадрат магнитного момента и одна из его проекций на координатные оси. За таковую обычно принято принимать ось Z. Из формул (35.8) и (35.10) для орбитального движения электрона непосредственно вытекает

$$\mathfrak{m}_z = -\frac{e}{2\mu_e c} L_z = -\mathfrak{m}_\text{Б} m, \tag{35.11}$$

где

$$\mathfrak{m}_\text{Б} = \frac{e\hbar}{2\mu_e c} = 9{,}274 \cdot 10^{-21} \text{ эрг/Гс}. \tag{35.12}$$

Постоянная $\mathfrak{m}_\text{Б}$ носит название *магнетона Бора*. Магнетон Бора можно рассматривать как *квант магнитного момента* (точнее, его проекции на избранное направление).

Возможен другой способ вывода формулы (35.11). Из временного уравнения Шредингера получают уравнение непрерывности $\partial\rho/\partial t + \operatorname{div}\mathbf{j} = 0$, где ρ и $\mathbf{j}$ — плотность вероятности и плотность тока вероятности. По значению последней и по волновой функции находят плотность вероятности электрического тока в стационарном состоянии атома, а затем непосредственным интегрированием находят и средний магнитный момент, создаваемый этим током. Этот прямой способ рассуждения обладает тем принципиальным недостатком, что плотность тока вероятности $\mathbf{j}$ определяется нерелятивистским уравнением Шредингера *неоднозначно*: к полученному выражению можно добавить любое слагаемое вида $\operatorname{rot}\mathbf{a}$ (поскольку $\operatorname{div}\operatorname{rot}\mathbf{a} = 0$), не меняя значения полного потока вероятности через любую замкнутую поверхность, который только и доступен наблюдению. Плотность самого электрического тока в атоме, в отличие от потока вероятности, конечно, — *величина наблюдаемая*, но для ее однозначного определения *одного нерелятивистского уравнения Шредингера недостаточно*. Неоднозначность можно устранить, но для этого надо перейти к *релятивистской теории*. В самом деле, величина ρ по своему смыслу есть величина *однозначная*. А в релятивистской теории скаляр ρ и три компоненты вектора $\mathbf{j}$ объединяются в один *релятивистски инвариантный четырехмерный вектор*, временной компонентой которого является ρ.

§ 36. Опыты Штерна и Герлаха. Спин электрона

1. Наличие у атомов магнитных моментов и их квантование было доказано прямыми опытами Штерна и Герлаха (1889–1979) в 1921 г. В сосуде с высоким вакуумом создавался с помощью диафрагм B и B' (рис. 64) резко ограниченный атомный пучок исследуемого элемента,

испаряющегося в печи K. Пучок проходил через сильное магнитное поле $\mathbf{H}$ между полюсными наконечниками N и S электромагнита. Один из наконечников (N) имел вид призмы с острым ребром, а вдоль другого (S) была выточена канавка. Благодаря такой конструкции полюсных наконечников магнитное поле получалось *сильно неоднородным.* После прохождения через магнитное поле пучок попадал на фотопластинку P и оставлял на ней след.

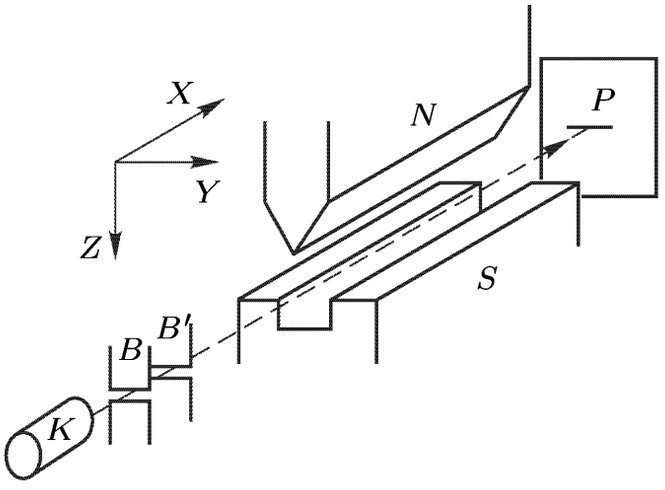


Рис. 64

Рассчитаем поведение атомного пучка сначала с классической точки зрения, предполагая, что никакого квантования магнитных моментов нет. Если $\mathfrak{m}$ — магнитный момент атома, то на атом в неоднородном магнитном поле действует сила

$$\mathbf{f} = (\mathfrak{m}\nabla)\mathbf{H}.$$

Направим ось Z вдоль магнитного поля (т. е. от N к S перпендикулярно к полюсным наконечникам). Тогда проекция силы в этом направлении будет

$$f_z = \mathfrak{m}_x \frac{\partial H_z}{\partial x} + \mathfrak{m}_y \frac{\partial H_z}{\partial y} + \mathfrak{m}_z \frac{\partial H_z}{\partial z}.$$

Первые два слагаемых в этом выражении не играют роли. В самом деле, по классическим представлениям атом в магнитном поле совершает прецессию вокруг оси Z, вращаясь с ларморовской частотой

$$\Omega = -\frac{eH}{2\mu c}$$

(заряд электрона обозначен через $-e$). Поэтому проекции $\mathfrak{m}_x$ и $\mathfrak{m}_y$ совершают колебания с той же частотой, становясь попеременно то положительными, то отрицательными. Если угловая скорость прецессии достаточно велика, то силу f_z можно усреднить по времени. При этом первые два члена в выражении для f_z обратятся в нуль, и можно написать

$$\overline{f}_z = \mathfrak{m}_z \frac{\partial H_z}{\partial z}. \tag{36.1}$$

Чтобы составить представление о степени допустимости такого усреднения, произведем численную оценку. Период ларморовской прецессии равен

$$\tau = \frac{2\pi}{|\Omega|} = \frac{4\pi\mu c}{eH} \approx \frac{7\cdot 10^{-7}}{H}\ \text{с},$$

где поле H измеряется в гауссах. Например, при $H = 1000$ Гс получаем $\tau \approx 7\cdot 10^{-10}$ с. Если скорость атомов в пучке равна $v = 100$ м/с $= 10^4$ см/с, то за это время атом пролетает расстояние $x \approx 7\cdot 10^{-6} \approx 10^{-5}$ см, пренебрежимо малое по сравнению со всеми характерными размерами установки. Это и доказывает применимость проведенного усреднения.

Но формула (36.1) может быть оправдана и с квантовой точки зрения. В самом деле, включение сильного магнитного поля вдоль оси Z приводит к состоянию атома только с одной определенной составляющей магнитного момента, а именно $\mathfrak{m}_z$. Остальные две составляющие $\mathfrak{m}_x$ и $\mathfrak{m}_y$ в этом состоянии не могут иметь определенных значений. При измерениях в этом состоянии $\mathfrak{m}_x$ и $\mathfrak{m}_y$ получили бы различные значения и притом их средние были бы равны нулю. Поэтому и при квантовом рассмотрении усреднение (36.1) оправдано.

2. Тем не менее следует ожидать различных результатов опыта с классической и с квантовой точек зрения. В опытах Штерна и Герлаха сначала получался след атомного пучка при выключенном магнитном поле, а затем при включенном. Если бы проекция $\mathfrak{m}_z$ могла принимать всевозможные непрерывные значения, как требует классическая теория, то сила $\overline{f}_z$ также принимала бы всевозможные непрерывные значения. Включение магнитного поля приводило бы только к *уширению пучка.* Не то следует ожидать по квантовой теории. В этом случае проекция $\mathfrak{m}_z$, а с ней и средняя сила $\overline{f}_z$ *квантованы*, т. е. могут принимать только ряд дискретных избранных значений. Если орбитальное квантовое число атома равно l, то по теории при расщеплении получится $2l+1$ пучков (т. е. оно равно числу возможных значений, которые может принимать квантовое число m). Таким образом, в зависимости от значения числа l следовало бы ожидать, что *пучок расщепится на* $1, 3, 5, \ldots$ *составляющих.* Ожидаемое число составляющих должно было бы быть *всегда нечетным.*

Опыты Штерна и Герлаха доказали квантование проекции $\mathfrak{m}_z$. Однако их результаты не всегда соответствовали теории, изложенной выше. В первоначальных опытах применялись пучки атомов серебра. В магнитном поле пучок расщеплялся на две составляющие. То же получалось для атомов водорода. Для атомов других химических элементов получалась и более сложная картина расщепления, однако число расщепленных пучков получалось *не только нечетным*, что требовалось теорией, *но и четным*, что противоречило ей. В теорию необходимо было внести коррективы.

3. К этому следует добавить результаты опытов Эйнштейна и де Газа (1878–1966), а также опытов Барнета (1873–1956) по определению гиромагнитного отношения, которые были изложены в т. III (§ 78). Для

железа, например, оказалось, что гиромагнитное отношение равно

$$\frac{\mathfrak{m}}{L} = \left| \frac{e}{\mu_e c} \right|, \tag{36.2}$$

т. е. вдвое больше, чем требуется по теории.

Наконец, оказалось, что спектральные термы щелочных металлов имеют так называемую *дублетную структуру*, т. е. состоят из двух близко расположенных уровней. Для описания этой структуры трех квантовых чисел n, l, m оказалось недостаточно — потребовалось *четвертое квантовое число*. Это явилось главным мотивом, послужившим Уленбеку (р. 1900) и Гаудсмиту (1902–1979) в 1925 г. для введения гипотезы о *спине электрона* [1]. Сущность этой гипотезы состоит в том, что у электрона есть не только момент количества движения и магнитный момент, связанные с перемещением этой частицы как целого. Электрон имеет также *собственный* или *внутренний механический момент количества движения*, напоминая в этом отношении классический волчок. Этот собственный момент количества движения и называется *спином* (от английского слова to spin — вертеться). Соответствующий ему магнитный момент называется *спиновым магнитным моментом*. Эти моменты обозначаются соответственно через $\mathbf{L}_s$ и $\mathfrak{m}_s$ в отличие от орбитальных моментов $\mathbf{L}_l$ и $\mathfrak{m}_l$. Спин чаще обозначают просто через $\mathbf{s}$.

В опытах Штерна и Герлаха атомы водорода находились в s-состоянии, т. е. не обладали орбитальными моментами. Магнитный момент ядра пренебрежимо мал. Поэтому Уленбек и Гаудсмит предположили, что расщепление пучка обусловлено не орбитальным, а *спиновым магнитным моментом*. То же самое относится к опытам с атомами серебра. Атом серебра имеет единственный наружный электрон. Атомный остов ввиду его симметрии спиновым и магнитным моментами не обладает. Весь магнитный момент атома серебра создается только одним наружным электроном. Когда атом находится в нормальном, т. е. s-состоянии, то орбитальный момент валентного электрона равен нулю — весь момент является спиновым.

4. Сами Уленбек и Гаудсмит предполагали, что спин возникает из-за вращения электрона вокруг собственной оси. Существовавшая в то время модель атома получила еще большее сходство с Солнечной системой. Электроны (планеты) не только вращаются вокруг ядра (Солнца), но и вокруг собственных осей. Однако сразу же выяснилась несостоятельность такого классического представления о спине (см. задачу 1 к этому параграфу). Паули систематически ввел спин в квантовую механику, но исключил всякую возможность классического истолкования этой величины. В 1928 г. Дирак показал, что спин электрона автоматически содержится в его теории электрона, основанной на релятивистском волновом уравнении. В теории Дирака содержится также и спиновый магнитный момент электрона, причем для гиромагнитного

[1] Гипотеза о спине электрона ранее вводилась Кронигом (р. 1904), но его работа не была опубликована.

отношения получается значение, согласующееся с опытом. При этом о внутренней структуре электрона ничего не говорилось — последний рассматривался как точечная частица, обладающая лишь зарядом и массой. Таким образом, *спин электрона оказался квантово-релятивистским эффектом, не имеющим классического истолкования.* Затем концепция спина, как внутреннего момента количества движения, была распространена на другие элементарные и сложные частицы и нашла подтверждение и широкие применения в современной физике.

5. Разумеется, в общем курсе физики нет возможности вдаваться в подробную и строгую теорию спина. Мы примем в качестве исходного положения, что спину $\mathbf{s}$ соответствует векторный оператор $\widehat{\mathbf{s}}$, проекции которого $\widehat{s}_x$, $\widehat{s}_y$, $\widehat{s}_z$ удовлетворяют таким же перестановочным соотношениям (31.2), что и проекции оператора орбитального момента, т. е.

$$\begin{aligned} \widehat{s}_y\widehat{s}_z - \widehat{s}_z\widehat{s}_y &= i\hbar\widehat{s}_x, \\ \widehat{s}_z\widehat{s}_x - \widehat{s}_x\widehat{s}_z &= i\hbar\widehat{s}_y, \\ \widehat{s}_x\widehat{s}_y - \widehat{s}_y\widehat{s}_x &= i\hbar\widehat{s}_z. \end{aligned} \tag{36.3}$$

Из них следует, что определенные значения в одном и том же состоянии могут иметь квадрат полного спина $\mathbf{s}^2$ и одна из его проекций на определенную ось (принимаемую обычно за ось Z). Если максимальное значение проекции s_z (в единицах $\hbar$) равно s, то число всех возможных проекций, соответствующих данному s, будет равно $2s+1$. Опыты Штерна и Герлаха показали, что для электрона это число равно 2, т. е. $2s+1=2$, откуда $s=1/2$. Максимальное значение, которое может принимать проекция спина на избранное направление (в единицах $\hbar$), т. е. число s, и принимается за значение спина частицы.

Спин частицы может быть либо целым, либо полуцелым. Для электрона, таким образом, спин равен $1/2$. Из перестановочных соотношений (36.3) следует, что квадрат спина частицы равен $\mathbf{s}^2 = s(s+1)$, а для электрона $\mathbf{s}^2 = (1/2)(1/2+1) = 3/4$ (в единицах $\hbar^2$).

Измерения проекции магнитного момента $\mathfrak{m}_z$ по методу Штерна и Герлаха показали, что для атомов водорода и серебра величина $\mathfrak{m}_z$ равна магнетону Бора $\mathfrak{m}_\text{Б}$, т. е. (35.12). Таким образом, гиромагнитное отношение для электрона

$$\left|\frac{\mathfrak{m}_z}{L_z}\right| = \frac{\mathfrak{m}_\text{Б}}{\hbar/2} = \frac{e}{\mu_e c}$$

в согласии с (36.2).

ЗАДАЧИ

1. С классической точки зрения спин электрона пытались истолковать как момент количества движения, возникающий из-за вращения электрона вокруг своего диаметра. Электрон считался шариком, его масса определялась по релятивистской формуле $\mathscr{E} = \mu c^2$, причем принималось, что собственная энергия электрона имеет чисто электростатическое происхождение. Проанализировать трудности, возникающие при таком классическом истолковании спина.

Смещения частицы $\Delta_1 z$ и $\Delta_2 z$, вызываемые силами (36.4) и (36.5), относятся как

$$\frac{\Delta_1 z}{\Delta_2 z} = \frac{\overline{f}_z}{(f_{\text{Лор}})_z} = \frac{\mathfrak{m}_z}{(ev/c)\Delta y}.$$

Считая, что атом — однозарядный ион, в качестве $\mathfrak{m}_z$ следует взять магнетон Бора (35.12). Тогда

$$\frac{\Delta_1 z}{\Delta_2 z} = \frac{\hbar}{2\mu_e v \Delta y}.$$

Если μ_a — масса атома, то

$$\frac{\Delta_1 z}{\Delta_2 z} = \frac{1}{4\pi}\,\frac{\mu_a}{\mu_e}\,\frac{\lambda}{\Delta y}, \tag{36.6}$$

где λ — длина волны де Бройля для атома: $\lambda = h/\mu_a v$.

В отсутствие магнитного поля щель, образуемая диафрагмами B и B', изобразится на пластинке P горизонтальной полоской. При включении неоднородного магнитного поля центр полоски сместится силой $\overline{f}_z$. Нецентральные точки полоски испытают дополнительные смещения под действием силы Лоренца. Максимальные дополнительные смещения, и притом в противоположные стороны, получат края полоски. В результате полоска на пластинке P перекосится. Для применимости метода Штерна и Герлаха необходимо, чтобы перекос был мал. Это значит, что должно выполняться условие $|\Delta_1 z| \gg |\Delta_2 z|_{\text{макс}}$, где $|\Delta_2 z|_{\text{макс}}$ — смещение края полоски, вызываемое силой Лоренца, т. е. ее значением при $|\Delta y| = |\Delta y|_{\text{макс}}$. Для свободного электрона $\mu_a = \mu_e$, и (36.6) дает $\Delta_1 z/\Delta_2 z = \lambda/(4\pi\Delta y)$. В этом случае условию $|\Delta_1 z| \gg |\Delta_2 z|_{\text{макс}}$ удовлетворить невозможно, поскольку должно быть $|\Delta y|_{\text{макс}} \gg \lambda$. Но удовлетворить ему оказывается возможным в случае атомов из-за того, что отношение μ_a/μ_e очень велико. Это и используется в методе Штерна и Герлаха.

§ 37. Эффект Садовского и спин фотона

1. В 1889 г. русский физик А. И. Садовский (1859–1921) теоретически предсказал, что свет, поляризованный по кругу или эллиптически, должен обладать моментом количества движения. Этот результат проще всего понять, исходя из закона сохранения момента количества движения. Согласно этому закону полный момент количества движения замкнутой системы должен оставаться постоянным. Рассмотрим эффект Садовского сначала *с классической точки зрения.*

Допустим, что электрический заряд e вращается по окружности радиуса r вокруг другого неподвижного заряда той же величины, но противоположного знака. Как известно, при вращении по окружности полная энергия, складывающаяся из кинетической и потенциальной, равна $\mathscr{E} = -e^2/2r$, т. е. половине потенциальной энергии заряда. Вращение по окружности есть движение ускоренное, а потому по законам классической электродинамики заряд e должен излучать электромагнитные волны. При наличии излучения заряд уже не может двигаться точно по окружности. Он будет непрерывно приближаться к ее центру. Предположим, что за время одного оборота уменьшение растояния r заряда до центра окружности очень мало по сравнению с r. Тогда

движение заряда e все еще можно охарактеризовать как вращение по окружности, радиус которой непрерывно уменьшается. Изменения энергии $\mathscr{E}$ и радиуса r при этом связаны соотношением

$$d\mathscr{E} = \frac{e^2}{2r^2}\, dr.$$

Вращающийся заряд обладает моментом количества движения $L = \mu r^2 \omega$, где μ — масса, а ω — круговая частота. При движении по окружности $\mu\omega^2 r = e^2/r^2$, откуда

$$\omega = \frac{e}{\sqrt{\mu r^3}},$$

и следовательно,

$$L = e\sqrt{\mu r}.$$

Значит,

$$dL = \frac{e\sqrt{\mu}}{2\sqrt{r}}\, dr, \quad \frac{d\mathscr{E}}{dL} = \frac{e}{\sqrt{\mu r^3}} = \omega.$$

Итак, при движении заряда по окружности его энергия и момент количества движения уменьшаются, причем их изменения связаны соотношением

$$\frac{d\mathscr{E}}{dL} = \omega. \tag{37.1}$$

Полная энергия и момент количества движения замкнутой системы должны оставаться постоянными. Система состоит из вещества и его излучения, которые могут обмениваться друг с другом и энергией, и моментом количества движения. Поэтому из постоянства этих величин для всей системы следует, что при изменении r на dr излучение уносит энергию $-d\mathscr{E}$ и момент количества движения $-dL$.

Структура излучения, конечно, определяется процессами, происходившими в излучателе. Но если излучение уже отделилось от излучателя, то теряется связь его с излучателем. Излучение продолжает существовать уже как самостоятельная система. Соотношение между его энергией и моментом количества движения поэтому есть *внутреннее свойство только самого отделившегося излучения.* Отсюда следует, что при рассмотренном нами способе возбуждения излучения его энергия $\mathscr{E}_{\text{изл}}$ и момент количества движения $L_{\text{изл}}$ должны быть связаны соотношением

$$\frac{\mathscr{E}_{\text{изл}}}{L_{\text{изл}}} = \omega = \frac{2\pi c}{\lambda}. \tag{37.2}$$

2. Излучение, отделившееся от излучателя, в нашем случае имеет довольно сложную структуру. Его интенсивность и поляризация по разным направлениям не одинаковы. В направлении, перпендикулярном к плоскости окружности, по которой вращается заряд e, излучение *поляризовано по кругу*, в плоскости окружности оно *поляризовано линейно*, а во всех остальных направлениях — *эллиптически.* Можно,

конечно, преобразовать все излучение в плоскую бегущую волну, поляризованную по кругу. Для этого можно, например, поместить центр окружности, по которой вращается излучающий заряд, в фокусе бесконечного идеально отражающего параболического зеркала, чтобы плоскость окружности была перпендикулярна к оси зеркала. Получится отраженная плоская волна, бегущая параллельно оси параболического зеркала. Она возникает в результате интерференции отраженных волн различной поляризации. Но ввиду цилиндрической симметрии результирующая волна будет поляризована по кругу. Однако нельзя заранее утверждать, что при отражении от зеркала общий момент количества движения излучения не изменится.

Чтобы преодолеть эту трудность, воспользуемся идеализированной моделью излучателя, аналогичной той, которая применялась в т. III, § 83, для получения плоских электромагнитных волн. Там было показано, что бесконечная заряженная плоскость, приведенная в ускоренное движение, является источником двух плоских электромагнитных волн, распространяющихся от нее в разные стороны с одной и той же энергией. Чтобы исключить статическое электрическое поле зарядов, мы помещали ранее рядом с рассматриваемой плоскостью вторую неподвижную бесконечную плоскость, заряженную электричеством противоположного знака. Теперь изменим слегка эту модель и возьмем снова бесконечную плоскость, но уже неподвижную и находящуюся в вакууме. Разместим на ней равномерно и достаточно густо электрические диполи с электрическими моментами, параллельными этой плоскости. Пусть каждый диполь вращается в этой плоскости вокруг своего центра с одной и той же угловой скоростью ω и одинаковой начальной фазой. Такая плоскость, покрытая вращающимися диполями, возбудит опять две плоские волны, распространяющиеся в разные стороны, но уже поляризованные по кругу. В силу симметрии энергия и момент количества движения распределятся поровну между обеими волнами. Поэтому для каждой из этих двух волн в отдельности соотношение (37.2) сохранится.

3. Направление вращения векторов $\mathbf{E}$ и $\mathbf{H}$, понятно, должно совпадать с направлением вращения диполя $\mathbf{p}$, т. е. в обеих волнах будет одно и то же. На рис. 65 направление вращения диполей $\mathbf{p}$ в плоскости указано стрелкой. В соответствии с указанным направлением моменты количества движения $\mathbf{L}_{\text{изл}}$ обеих волн направлены слева направо. Обе волны будут бегущими и уходящими, так что в обеих из них вектор Пойнтинга $\mathbf{S}$ направлен наружу. Значит, в волне, уходящей направо, вектор $\mathbf{S}$ направлен тоже направо, т. е. одинаково с вектором $\mathbf{L}_{\text{изл}}$. В волне же, уходящей налево, вектор $\mathbf{S}$ направлен налево, т. е. противоположно вектору $\mathbf{L}_{\text{изл}}$. Но волна, идущая вправо, поляризована по левому кругу (вращение векторов

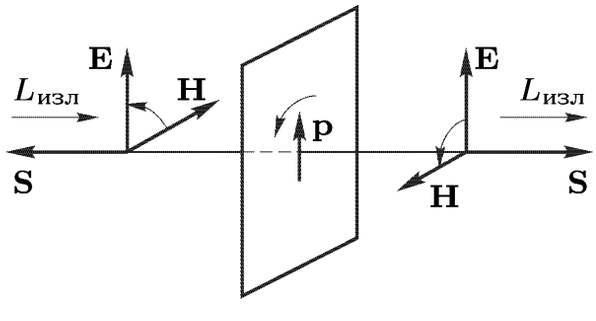


Рис. 65

E и **H** совершается против часовой стрелки, если волна идет к глазу наблюдателя); волна же, идущая влево, поляризована по правому кругу (вращение векторов **E** и **H** совершается по часовой стрелке, если волна идет также к глазу наблюдателя). Таким образом, в левополяризованной волне вектор $\mathbf{L}_{\text{изл}}$ направлен в сторону распространения волны, а в правополяризованной — в сторону, противоположную направлению распространения волны. То же заключение, разумеется, справедливо и для волн, поляризованных эллиптически.

Нелишне особо подчеркнуть, что взаимное расположение векторов **E** и **H** в бегущей волне однозначно определяет направление вектора Пойнтинга **S**, а с ним и направление распространения волны. Но этим расположением еще не определяется вид поляризации волны, поляризованной по кругу или эллиптически: поляризация может быть и левой, и правой. Например, на рисунках 66 *а* и 66 *б* векторы **E**, **H** и **S** имеют одну и ту же ориентацию и обе волны распространяются в одну и ту же сторону, хотя направления вращения векторов **E** и **H**, отмеченные круговыми стрелками, в них противоположны: рис. 66 *а* соответствует левой, а рис. 66 *б* — правой круговой поляризации.

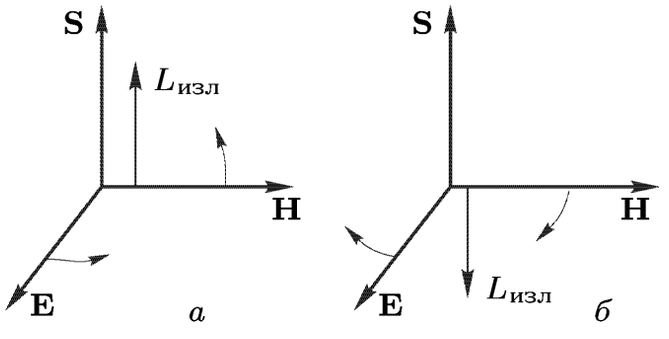


Рис. 66

Укажем теперь, насколько густо надо расположить диполи в излучающей плоскости, чтобы получились только плоские волны, распространяющиеся перпендикулярно к этой плоскости. Для этого надо учесть, что волны, излучаемые отдельными диполями, конечно, интерферируют между собой. Требуется, чтобы при интерференции они взаимно гасили друг друга во всех направлениях, за исключением направлений, перпендикулярных к излучающей плоскости. Для этого достаточно, чтобы *расстояние между диполями было меньше длины волны* λ. Тогда вдали от плоскости возникнут только плоские уходящие волны. Лишь вблизи самой плоскости на них наложатся *неоднородные волны*, не играющие роли в рассматриваемом нами вопросе, так как эти волны быстро затухают в тонком приграничном слое, толщина которого порядка расстояния между диполями.

В итоге получается, что *всякая плоская электромагнитная волна частоты* ω, *поляризованная по кругу, несет момент количества движения, связанный с энергией волны соотношением* (37.2). *Если поляризация левая, то вектор* $\mathbf{L}_{изл}$ *направлен в сторону распространения волны, если правая, то эти направления противоположны.* Это и есть основной результат, полученный А. И. Садовским.

Случай эллиптической поляризации сводится к случаю круговой поляризации. Действительно, волну, поляризованную по эллипсу, можно разложить на две волны, поляризованные по кругу: одну — по правому, другую — по левому.

4. Момент количества движения излучения можно найти и более непосредственно, исходя из свойств *только самого излучения*. Последнее, как известно, обладает количеством движения, объемная плотность которого дается выражением $\mathbf{g}_{\text{эл}} = \mathbf{S}/\omega$, где $\mathbf{S}$ — вектор Пойнтинга (см. т. III, § 81). Если взять момент вектора $\mathbf{g}_{\text{эл}}$ и проинтегрировать по всему пространству, занятому излучением, то и получится момент количества движения излучения. Это делается в задаче в конце этого параграфа, где указанная процедура проводится применительно к излучению электрического диполя Герца, дипольный момент которого, не меняясь по величине, равномерно вращается в одной плоскости. При этом, конечно, речь идет о моменте количества движения всего излучения, испускаемого источником в различных направлениях. Но в качестве источника излучения можно снова взять бесконечную плоскость с распределенными на ней достаточно густо диполями Герца, как это делалось в п. 2. Таким путем можно получить уже плоскую волну с круговой поляризацией. Для нее можно ввести и понятие вектора плотности потока момента количества движения излучения $\mathbf{M}$. На основании формулы (37.2) этот вектор определяется формулой

$$\mathbf{M} = \frac{\mathbf{S}}{\omega}. \tag{37.3}$$

5. При поглощении световой волны, поляризованной по кругу, на единицу площади тела будет действовать вращающий момент $\mathbf{M} = \mathbf{S}/\omega$, если только волна падает на поверхность тела нормально. Чтобы составить представление о величине эффекта, предположим, что плотность потока энергии в поляризованной по кругу плоской световой волне равна $S = 1$ Вт/см$^2 = 10^7$ эрг/(см$^2 \cdot$ с). (Это примерно в 7 раз больше плотности потока солнечного излучения вблизи Земли за пределами ее атмосферы.) Тогда для видимого излучения $\lambda = 500$ нм формула (37.3) дает

$$M = \frac{S}{\omega} = \frac{\lambda S}{2\pi c} = 2{,}7 \cdot 10^{-9} \text{ дин/см},$$

а для волны $\lambda = 1$ см той же интенсивности

$$M = 5{,}3 \cdot 10^{-5} \text{ дин/см}.$$

Если волна проходит через кристаллическую пластинку в полволны, вырезанную параллельно оптической оси, то она превращается из право- в левополяризованную и наоборот. В соответствии с этим величина M удваивается. При заданной мощности излучения эффект возрастает с увеличением длины волны. Но он все же очень мал и экспериментально был обнаружен только в 1935 г. американским физиком Бетом, и притом не только для радиоволн, но и для видимого света.

6. Перейдем теперь к рассмотрению эффекта Садовского с *квантовой точки зрения*. Одна из особенностей здесь состоит в том, что

испускание и последующее распространение света происходят не непрерывными порциями, а *неделимыми квантами — фотонами.* В соответствии с этим отпадает необходимость в искусственной концентрации излучения в определенном направлении, какая применялась при классическом рассмотрении. Многофотонные процессы, когда в одном акте излучения испускается не один, а несколько фотонов, как процессы маловероятные, рассматриваться не будут. Другая особенность заключается в том, что *у квантового вектора момента количества движения не могут одновременно иметь определенные значения все три проекции его на координатные оси.*

При переходе атома из одного стационарного состояния в другое испускается *один фотон* с энергией $\mathscr{E} = \hbar\omega$. Проекция момента количества движения атома на избранное направление (ось Z) при орбитальном движении электрона может принимать значения $m\hbar$. Пусть при излучении фотона эта проекция изменилась на $\hbar$. В таком случае в акте излучения атом потерял энергию $\hbar\omega$ и проекцию момента количества движения $\hbar$. В соответствии с законами сохранения энергия и момент количества движения, потерянные атомом, *перейдут к излучению.* Поэтому следует заключить, что проекция момента количества движения излученного фотона равна $\hbar$. Внутренний момент количества движения фотона, т. е. момент, не связанный с его орбитальным движением, называется *спином фотона.* Говорят, что спин фотона целочисленный и равен единице (т. е. на самом деле $\hbar$), хотя значение $\hbar$ относится *не к полному моменту*, а только *к его проекции на избранное направление.* Если проекция (в единицах $\hbar$) равна s, то, как для всякого квантового момента количества движения, квадрат вектора спина фотона определяется выражением $s(s+1)\hbar^2 = 2\hbar^2$. Отношение величин $\mathscr{E} = \hbar\omega$ и $L_z = \hbar$ дается формулой

$$\frac{\mathscr{E}}{L_z} = \frac{\hbar\omega}{\hbar} = \omega. \tag{37.4}$$

Это соотношение по форме совпадает с классическим (37.2), хотя между ними и есть существенное различие. В классической формуле (37.2) L означает *полный момент* количества движения излучения, тогда как в квантовой формуле (37.4) $L_z = \hbar$ дает только *проекцию момента* на избранное направление.

7. Масса покоя фотона равна нулю. Поэтому не существует системы отсчета, относительно которой фотон находился бы в покое. Его внутренний момент количества движения, или спин, поэтому нельзя определять как момент частицы, находящейся в состоянии покоя. *Фотон может существовать только в движении и притом со скоростью света c в любой системе отсчета.*

Строгое решение вопроса о моменте количества движения фотона может быть дано только в *релятивистской квантовой теории.* Нерелятивистская теория фотона принципиально невозможна, поскольку скорость фотона всегда равна скорости света c. В нашем курсе излагать релятивистскую теорию не представляется возможным. Ограничимся

только замечанием, что, как и всякая квантовомеханическая величина, момент количества движения фотона определяется через соответствующий *оператор.* Оказывается, что оператор момента количества движения фотона состоит из двух слагаемых. Одно слагаемое имеет вид $[\widehat{\mathbf{r}}\widehat{\mathbf{p}}]$, где $\widehat{\mathbf{p}}$ — *оператор импульса фотона.* Оно называется *орбитальным.* Дополнительное слагаемое называется *спиновым* или *оператором спина фотона.* Собственное значение проекции оператора $[\widehat{\mathbf{r}}\widehat{\mathbf{p}}]$ на избранное направление называется *орбитальным моментом количества движения фотона.* Собственное значение проекции оператора спина на то же направление есть *спиновый момент количества движения или просто спин фотона.*

Будем предполагать, что орбитального момента у фотона нет, так что весь его момент является спиновым. Наглядным оправданием этого может служить замечание, что обычно длина волны, излучаемой атомом, очень велика по сравнению с размерами последнего. Фотон же не может быть локализован в области пространства, линейные размеры которой меньше длины световой волны λ. С другой стороны, размеры излучающего атома очень малы по сравнению с λ. Поэтому фотон излучается атомом практически всегда «центрально». Фотон при этом не получает никакого орбитального момента количества движения, он уносит *только спиновый момент.* Чтобы у фотона появился дополнительный орбитальный момент, излучение должно произойти с далекой периферии атома — с расстояний порядка λ. Волновая функция атома на таких расстояниях, а с ней и вероятность излучения фотона ничтожны.

8. То обстоятельство, что фотон существует только в состоянии движения со скоростью c, проявляется еще в том, что в любой системе отсчета для него есть *только одно избранное направление — направление движения.* На это направление и проецируется вектор спина фотона. А так как спин фотона $s = 1$, то казалось бы, что относительно этого направления спин может ориентироваться $2s + 1 = 3$ способами: в первом проекция спина направлена по движению, во втором против движения, в третьем равна нулю. В действительности третья возможность не осуществляется.

Не вдаваясь в подробности, заметим, что к этому заключению приводит опыт. Из поперечности электромагнитных волн следует, что для получения любой поляризации волны достаточно наложения только двух, а не трех волн с различными поляризациями. В согласии с принципом соответствия следует ожидать, что в квантовой теории для получения любого состояния фотона достаточно суперпозиции только двух независимых состояний его. Какие же состояния фотона могут быть приняты в качестве независимых? Для этого рассмотрим связь между поляризацией и спином фотона.

9. Понятие *поляризации* (как и всякое другое понятие) в фотонной теории лишено того ясного наглядного смысла, которым оно отличается в классической теории. Поскольку, однако, поляризация света существует и проявляется на опыте, необходимо установить, что

соответствует ей в фотонной теории. *Единственной направленной величиной, характеризующей внутренние свойства фотона, является спин.* С другой стороны, в классической теории момент количества движения $\mathbf{L}$ волны, поляризованной по кругу, направлен по или против распространения волны. *Поэтому естественно считать фотон поляризованным по кругу, если он находится в состоянии с определенным значением проекции спина на направление распространения.* Если спин направлен в сторону распространения света, то поляризация фотона называется *левой*; в противном случае ее называют *правой* [1].

В классической оптике любая поляризация (линейная или эллиптическая) бегущей плоской волны может быть получена путем суперпозиции двух (когерентных) поляризованных по кругу плоских волн, распространяющихся в том же направлении, поляризация одной из которых правая, а другой левая. Так и состояние фотона с круговой поляризацией, распространяющегося в определенном направлении, следует рассматривать как его собственное состояние, которому соответствуют собственные значения проекции спина $s_z = +1, 0, -1$. Путем линейной суперпозиции таких состояний может быть получен фотон любой поляризации. Но состояние с $s_z = 0$ не осуществляется. Поэтому *состояние фотона с любой поляризацией, распространяющегося в определенном направлении, может быть получено линейной суперпозицией только двух состояний*: состояния с $s_z = +1$ и состояния с $s_z = -1$.

Суперпозиция таких состояний, конечно, *не будет классической.* Она понимается в том же смысле, как и суперпозиция квантовомеханических состояний частицы, характеризуемых волновыми функциями. А так как состояния фотона с $s_z = +1$ и с $s_z = -1$ являются собственными, то квадраты модулей коэффициентов при этих состояниях в суперпозиции определяют относительные вероятности самих состояний. Это проявляется, например, в том, что при измерении проекции s_z (скажем, по величине вращающего момента, сообщаемого телу при поглощении фотона) может с соответствующей вероятностью получиться либо $s_z = +1$, либо $s_z = -1$. Никакой промежуточный результат получиться не может.

ЗАДАЧА

Твердый диполь с электрическим моментом $\mathbf{p}$ равномерно вращается вокруг своего центра с постоянной угловой скоростью $\boldsymbol{\omega}$, причем вектор $\mathbf{p}$ все время расположен в одной плоскости. Согласно классической электродинамике он излучает, как диполь Герца. Излучение обладает количеством движения, объемная плотность которого определяется выражением

$$\mathbf{g}_{\text{эл}} = \frac{1}{4\pi c}\,[\mathbf{EH}] \tag{37.5}$$

[1] Такое определение правой и левой поляризаций соответствует тому, что принято в классической оптике. В квантовой электродинамике применяется противоположное соглашение: правой поляризации соответствует спин, направленный в сторону распространения фотона, левой — спин, направленный противоположно.

(см. т. III, § 84). Излучение уносит и момент количества движения. Для вычисления полного момента, уносимого излучением, достаточно знать $\mathbf{g}_{эл}$ на бесконечно удаленной сфере с центром в месте нахождения диполя. Откуда может взяться такой момент, если вдали от диполя (в волновой зоне) поля $\mathbf{E}$ и $\mathbf{H}$ связаны между собой, как в плоской волне, и оба перпендикулярны к радиусу-вектору $\mathbf{r}$, проведенному от диполя к точке наблюдения, так что вектор $\mathbf{g}_{эл}$ направлен вдоль $\mathbf{r}$, а потому момент $[\mathbf{r}\mathbf{g}_{эл}]$ равен нулю?

Р е ш е н и е. Утверждение, что в волновой зоне поля $\mathbf{E}$ и $\mathbf{H}$ убывают с расстоянием как $1/r$ — приближенное и выполняется только асимптотически при $r \to \infty$. Такое приближение достаточно для вычисления полной энергии или полного импульса, уносимых излучением, так как тогда вектор $[\mathbf{EH}]$ будет меняться как $1/r^2$. Высшие степени величины $1/r$ учитывать не надо, поскольку при интегрировании по сфере они ничего не внесут, если выполнить предельный переход $r \to \infty$. Но плотность момента количества движения $\mathbf{l}_{эл} = [\mathbf{r}\mathbf{g}_{эл}]$ получается из $\mathbf{g}_{эл}$ векторным умножением на $\mathbf{r}$. Величину $\mathbf{g}_{эл}$ на удаленной сфере, понятно, достаточно вычислить также с точностью до членов $1/r^2$ включительно, а для этого надо учесть в выражении для $\mathbf{g}_{эл}$ и члены третьей степени по $(1/r)$. Чтобы это сделать, достаточно пользоваться следующими формулами для поля излучения диполя Герца в волновой зоне в вакууме:

$$\begin{aligned} \mathbf{E} &= \left[\frac{3(\dot{\mathbf{p}}\mathbf{r})}{cr^4}\,\mathbf{r} - \frac{\dot{\mathbf{p}}}{cr^2}\right]_{t-r/c} - \left[\frac{(\ddot{\mathbf{p}}\mathbf{r})}{c^2r^3}\,\mathbf{r} - \frac{\ddot{\mathbf{p}}}{c^2r}\right]_{t-r/c}, \\ \mathbf{H} &= \frac{1}{cr^3}\,[\dot{\mathbf{p}}\mathbf{r}]_{t-r/c} + \frac{1}{c^2r^2}\,[\ddot{\mathbf{p}}\mathbf{r}]_{t-r/c}. \end{aligned} \tag{37.6}$$

Они получаются из формул (141.10) тома III, если их написать для вакуума и в соответствии с этим положить $\mathbf{D} = \mathbf{E}$, $v = c$. При этом в первой формуле (141.10) отброшен первый член, пропорциональный $1/r^3$, так как на $\mathbf{g}_{эл}$ он может повлиять только в члене порядка $1/r^4$. Из формул (37.6) надо найти $[\mathbf{EH}]$ в нужном нам приближении, опуская при этом члены, коллинеарные с $\mathbf{r}$, поскольку они не играют роли при вычислении $[\mathbf{r}\mathbf{g}_{эл}]$. Таким путем, опуская значок $t - r/c$, получаем

$$\mathbf{g}_{эл} = \frac{1}{2\pi c^4 r^4}\,(\dot{\mathbf{p}}\mathbf{r})\ddot{\mathbf{p}} + (\ldots)\mathbf{r}, \tag{37.7}$$

$$\mathbf{l}_{эл} = \frac{1}{2\pi c^4 r^4}\,(\dot{\mathbf{p}}\mathbf{r})[\mathbf{r}\ddot{\mathbf{p}}]. \tag{37.8}$$

Преобразуем эту формулу, воспользовавшись тем, что вектор $\mathbf{p}$ не меняет своей длины, а изменяется только из-за вращения. В таком случае $\dot{\mathbf{p}} = [\boldsymbol{\omega}\mathbf{p}]$. То же относится и к $\dot{\mathbf{p}}$, т. е. $\ddot{\mathbf{p}} = [\boldsymbol{\omega}\dot{\mathbf{p}}]$. В результате формула (37.8) преобразуется:

$$\mathbf{l}_{эл} = \frac{1}{2\pi c^4 r^4}\,(\dot{\mathbf{p}}\mathbf{r})[\mathbf{r}[\boldsymbol{\omega}\dot{\mathbf{p}}]] = \frac{1}{2\pi c^4 r^4}\,(\dot{\mathbf{p}}\mathbf{r})\{(\dot{\mathbf{p}}\mathbf{r})\boldsymbol{\omega} - (\boldsymbol{\omega}\mathbf{r})\dot{\mathbf{p}}\}. \tag{37.9}$$

Чтобы найти полный момент излучения, испускаемого диполем в единицу времени, надо это выражение умножить на c и результат проинтегрировать по всей поверхности бесконечно удаленной сферы. Ясно, что из-за симметрии вращения вокруг $\boldsymbol{\omega}$ при таком интегрировании получится вектор, направленный вдоль $\boldsymbol{\omega}$. А так как вектор $\dot{\mathbf{p}} = [\boldsymbol{\omega}\mathbf{p}]$ перпендикулярен к $\boldsymbol{\omega}$, то последний

член в (37.9) можно опустить. Тогда

$$\mathbf{l}_{\text{эл}} = \frac{1}{2\pi c^4 r^4}(\dot{\mathbf{p}}\mathbf{r})^2\boldsymbol{\omega} = \frac{1}{2\pi c^4 r^4}\dot{\mathbf{p}}^2\cos^2\varphi\cdot\boldsymbol{\omega}. \tag{37.10}$$

При интегрировании можно поступать так, как если бы вектор $\dot{\mathbf{p}}$ оставался неподвижным, и выбрать сферическую систему координат, указанную на рис. 67. В этом случае элемент поверхности сферы будет $dS = r^2\sin\theta\,d\theta\,d\varphi$. В результате для момента импульса излучения получим

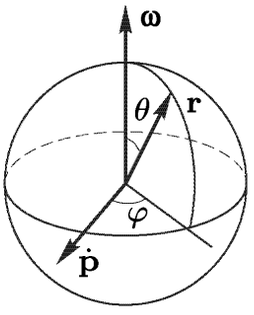


Рис. 67

$$\mathbf{L}_{\text{изл}} = \frac{\omega\dot{\mathbf{p}}^2}{2\pi c^3}\int_0^{\pi}\int_0^{2\pi}\sin^3\theta\cos^2\varphi\,d\theta\,d\varphi = \frac{2}{3c^3}\dot{\mathbf{p}}^2\omega. \tag{37.11}$$

Энергия, излучаемая диполем в единицу времени, равна $\mathscr{E}_{\text{изл}} = (2/3c^3)\ddot{\mathbf{p}}^2$ (см. т. III, § 141). А так как $\ddot{\mathbf{p}}^2 = \omega^2\dot{\mathbf{p}}^2$, то получается

$$L_{\text{изл}} = \mathscr{E}_{\text{изл}}/\omega. \tag{37.12}$$

§ 38. Четыре квантовых числа электрона и тонкая структура спектральных термов

1. Наличие у электрона внутреннего момента количества движения (спина) означает, что для электрона (в отличие от точечной классической частицы) трех степеней свободы недостаточно для характеристики его состояния. Электрон в атоме обладает дополнительной — четвертой — степенью свободы, называемой *спиновой*. Заметим, что пока что мы имеем в виду водородоподобный атом, а также вообще многоэлектронный атом или ион с *одним* наружным (валентным или оптическим) электроном. Такой электрон сейчас и предполагается в нашем рассмотрении. В квантовой механике его состояние описывается четырьмя квантовыми числами: 1) *главным квантовым числом* n; 2) *орбитальным квантовым числом* l; 3) *орбитальным магнитным квантовым числом*, которое мы теперь будем обозначать через m_l, и 4) *спиновым квантовым числом* m_s.

Смысл первых трех квантовых чисел n, l, m_l уже был выяснен в § 33. Спиновое же число m_s определяет проекции вектора спина $\mathbf{s}$ на выделенное направление. Если атом уже находится в состоянии с определенным значением орбитального момента $\mathbf{l}$ (т. е. с определенными $\mathbf{l}^2$ и l_z), то выделенное направление (ось Z) при $\mathbf{l}^2 \neq 0$ определяется вектором $\mathbf{l}$. Спин $\mathbf{s}$ может быть ориентирован либо по $\mathbf{l}$, либо против $\mathbf{l}$. Это означает, что проекция вектора $\mathbf{s}$ на это выделенное направление может принимать только два значения: $+\hbar/2$ и $-\hbar/2$, или $m_s\hbar$, где $m_s = \pm 1/2$. При $\mathbf{l} = 0$ (т. е. когда атом находится в s-состоянии) весь момент количества движения атома чисто спиновый: $\mathbf{s}$. Если состояние атома таково, что одна из проекций s_x, s_y, s_z имеет определенное значение (равное $\pm\hbar/2$), то соответствующая ось и определяет выделенное направление в атоме.

2. Орбитальный момент количества движения $\mathbf{l}$ и спиновый момент $\mathbf{s}$ складываются в полный момент количества движения $\mathbf{j} = \mathbf{l} + \mathbf{s}$ по правилам векторного сложения (см. § 32). Проекция полного момента на избранное направление может принимать значения $m_j\hbar$, где $m_j = m_l + m_s = m_l \pm 1/2$ называется *квантовым числом проекции полного момента.* Ясно, что операторы проекций полного момента на координатные оси удовлетворяют тем же перестановочным соотношениям (31.6), что и операторы проекций орбитального момента. Отсюда следует, что определенные значения в одном и том же состоянии могут иметь квадрат полного момента $\mathbf{j}^2$ и одна из его проекций на координатные оси. Отсюда же следует, что

$$\mathbf{j}^2 = \hbar^2 j(j+1),$$

где j — максимальное значение, которое может принимать квантовое число m_j. Иногда j называют *внутренним квантовым числом.*

Поскольку j есть максимальное значение числа m_j, а l — максимальное значение числа m_l, то из соотношения $m_j = m_l \pm 1/2$ следует

$$j = l \pm 1/2. \tag{38.1}$$

Знак «плюс» соответствует случаю, когда спин электрона ориентирован в направлении орбитального момента, а «минус» — когда он ориентирован противоположно. В обоих случаях число j полуцелое, поскольку l всегда целое.

При заданном j возможно $2j + 1$ квантовых состояний, отличающихся одно от другого значениями квантового числа m_j:

$$m_j = -j,\ -(j-1),\ \ldots,\ +(j-1),\ +j. \tag{38.2}$$

Число этих состояний у атомов с одним валентным электроном всегда четное, поскольку j полуцелое.

3. Вместо квантовых чисел n, l, m_l, m_s для характеристики состояния одноэлектронного атома можно применять и другие четверки квантовых чисел, например n, l, j, m_s. В спектроскопии принято пользоваться числами $n, l, j, 2s+1$ и заменять число l соответствующей буквой латинского алфавита в соответствии с таблицей, приведенной ранее в § 34 (п. 2). Сначала пишут числовое значение главного квантового числа n, за ним букву, заменяющую число l, число j пишут справа от этой буквы в виде нижнего индекса, а в качестве верхнего индекса слева от той же буквы пишут число $2s + 1$, называемое *мультиплетностью уровня.* Оно показывает, сколькими способами спин может ориентироваться относительно направления орбитального момента $\mathbf{l}$. В случае атома с одним валентным электроном значок $2s+1$ по существу излишен, так как для электрона $s = 1/2$, а потому всегда $2s+1 = 2$. Но если бы спин частицы имел другое значение, то указание мультиплетности $2s + 1$ приобрело бы существенное значение. То же самое можно сказать относительно аналогичного числа в случае атома с несколькими валентными электронами (см. п. 10).

Рассмотрим, например, состояние

$$3^2s_{1/2}$$

(называемое «три, дублет $s_{1/2}$». Смысл этого названия раскрывается в п. 4). В этом состоянии $n = 3$, $l = 0$, $j = 1/2$. Полный момент количества движения — чисто спиновый. В формуле $j = l \pm 1/2$ знак минус надо исключить, поскольку $j > 0$, так что в рассматриваемом случае $j = l + 1/2$. Состояние чисто формально называется дублетом, так как при $l = 0$ все направления для ориентации спина равноправны. По существу это есть синглет. Это, разумеется, относится ко всем s-состояниям (т. е. состояниям с $l = 0$).

В качестве второго примера возьмем состояние «четыре, дублет $d_{3/2}$», т. е.

$$4^2d_{3/2}.$$

В этом состоянии $n = 4, l = 2, j = 3/2$, причем $j = l - 1/2$, т. е. спиновый момент ориентирован против направления орбитального момента. Но в состоянии $4^2d_{5/2}$ $j = l + 1/2$, т. е. ориентации спина и орбитального момента одинаковы. Таким образом, состояние d действительно является дублетом. То же справедливо для всех остальных состояний: $p, f, g, \ldots$ (за исключением только состояния s).

4. Основное взаимодействие между электроном атома и ядром есть *электростатическое взаимодействие* их зарядов. Но так как электрон движется относительно атомного ядра, то возникает дополнительное взаимодействие, обусловленное спином электрона и зарядом ядра. Его называют *спин-орбитальным взаимодействием.* В существовании спин-орбитального взаимодействия можно убедиться наглядно, воспользовавшись представлениями полуклассической теории Бора. Простейшей является модель атома водорода, в которой электрон вращается по круговой орбите. Перейдем в ней к системе отсчета, в которой электрон покоится, т. е. сама система движется вместе с электроном. В такой системе отсчета ядро движется и создает магнитное поле $\mathbf{H}$, воздействующее на спиновый магнитный момент $\mathfrak{m}_s$ покоящегося в этой системе электрона. Поскольку заряды протона и электрона численно равны и противоположны по знаку, движущееся ядро в движущейся системе отсчета создает в месте нахождения электрона такое же магнитное поле, как и вращающийся электрон в покоящейся системе отсчета в месте нахождения ядра. Поэтому спин-орбитальное взаимодействие можно формально рассматривать как взаимодействие между спиновым и орбитальным магнитными моментами электрона.

Спиновый магнитный момент электрона $\mathfrak{m}_s$ может ориентироваться либо вдоль орбитального магнитного поля, либо противоположно. В первом случае потенциальная энергия взаимодействия электрона и ядра атома уменьшается, во втором увеличивается. Поэтому из-за спин-орбитального взаимодействия каждый энергетический уровень

атома расщепляется на два подуровня. Исключением является случай, когда атом находится в s-состоянии, поскольку в этом состоянии у атома нет орбитального магнитного момента, так что спин-орбитальное взаимодействие пропадает. Расщепление энергетического уровня в результате спин-орбитального взаимодействия называется *тонкой структурой уровня*. Совокупность подуровней, на которые расщепился рассматриваемый уровень, называется *мультиплетом*. В зависимости от числа подуровней, из которых состоит мультиплет, различают *дублеты*, *триплеты*, *квартеты*, *квинтеты*,... Простые уровни, не расщепляющиеся на подуровни, называются *синглетами*. Такие же термины употребляются и для совокупностей спектральных линий, получающихся путем расщепления из одной линии (см. § 40).

Таким образом, в случае атомов или ионов с одним валентным электроном спин-орбитальное взаимодействие приводит к тому, что все энергетические уровни, за исключением s-уровней, становятся дублетами, s-уровень остается синглетным. Теперь понятен смысл названий, которые употреблялись выше в п. 3. Например, уровень $4^2d_{3/2}$ был назван «четыре, дублет $d_{3/2}$». Употребление термина «дублет» для синглетных уровней s, как уже подчеркивалось, чисто условное. Оно применяется для того, чтобы не выделять эти уровни среди действительно дублетных уровней p, d, f, ... Впрочем, синглетные уровни s формально можно рассматривать как дублеты, состоящие из двух слившихся подуровней. Понятен также физический смысл мультиплетности $2s+1$: она определяет число подуровней в мультиплете, возникающем из-за спин-орбитального взаимодействия.

5. Легко оценить по порядку величины дополнительную потенциальную энергию, возникающую из-за спин-орбитального взаимодействия. Возьмем для этого атом водорода в основном состоянии и воспользуемся тем механизмом возникновения спин-орбитального взаимодействия, который был описан в п. 4. Перейдем снова к системе отсчета, движущейся вместе с электроном. Магнитное поле в месте нахождения электрона, создаваемое в этой системе протоном, движущимся со скоростью $\mathbf{v}$, определяется формулой $\mathbf{H} = e[\mathbf{vr}]/cr^3$, где $\mathbf{r}$ — радиус-вектор электрона относительно протона. По абсолютной величине $H = \alpha e/r^2$, где $\alpha = v/c$. Согласно (13.19) последняя величина есть *постоянная тонкой структуры*, определяемая формулой (13.18), т. е. $\alpha = e^2/\hbar c$. В магнитном поле $\mathbf{H}$ электрон обладает потенциальной энергией $-(\mathfrak{m}_s\mathbf{H})$, причем вектор $\mathfrak{m}_s$ может быть направлен либо по $\mathbf{H}$, либо против. По абсолютной величине эта энергия равна $\mathfrak{m}_s H = \mathfrak{m}_Б H$, где $\mathfrak{m}_Б = e\hbar/(2\mu_e c)$ — магнетон Бора. Сравним ее с полной энергией атома водорода в основном состоянии. Согласно формуле (13.20) она дается выражением $\mathscr{E}_1 = -\alpha^2\mu_e c^2/2$. В качестве r следует взять боровский радиус, определяемый формулой (13.16), т. е. $r_Б = \hbar^2/\mu_e e^2$. В результате получим

$$\frac{\mathfrak{m}_Б H}{\mathscr{E}_1} = \alpha^2 = 5{,}325 \cdot 10^{-5}. \tag{38.3}$$

6. Поскольку $\alpha = v/c$ (где v — скорость электрона на первой боровской орбите), *спин-орбитальное взаимодействие есть эффект, квадратичный относительно параметра* α. Поэтому его теория должна быть *релятивистской.* Этого и следовало ожидать, так как сам *спин есть квантово-релятивистский эффект, исчезающий в нерелятивистском приближении. Зависимость массы от скорости также приводит к тонкому расщеплению энергетических уровней* уже в рамках полуклассической теории Бора, как это впервые показал Зоммерфельд. Дело в том, что в боровской нерелятивистской теории всем эллиптическим орбитам электрона (включая и круговую) с одной и той же большой осью соответствует *одна и та же энергия.* Учет зависимости массы от скорости *снимает такое вырождение* — величина энергии начинает зависеть и от эксцентриситета эллипса. Это и приводит к тонкому расщеплению энергетического уровня. Таким образом, уточняя приведенное выше определение тонкой структуры, следует сказать, что *она вызывается не только спин-орбитальным взаимодействием, но и зависимостью массы электрона от скорости.* Оба расщепления — второго порядка по параметру α, а потому должны рассматриваться одновременно.

Наиболее последовательно тонкая структура может быть рассчитана и исследована на основе *релятивистской квантовой теории Дирака,* в которой автоматически учитывается и спин электрона, и зависимость массы от скорости.

В случае водородоподобного атома решение волнового релятивистского уравнения Дирака приводит к следующей формуле для энергии в стационарном состоянии:

$$\mathscr{E} = -\frac{(Ze^2)^2\mu_e}{2\hbar^2 n^2}\left[1 + \frac{\alpha^2 Z^2}{n}\left(\frac{1}{j+1/2} - \frac{3}{4n}\right)\right]. \tag{38.4}$$

В квадратных скобках опущены члены четвертой и высших степеней по α. Благодаря малости постоянной α^2 поправка к нерелятивистской формуле (13.8) получается очень малой, так что рассматриваемое расщепление уровней оправдывает название «тонкой структуры».

Заметим, что энергии уровней в водородоподобных атомах по теории Дирака вырождены по l, т. е. они зависят (и притом в любом приближении) только от главного квантового числа n и квантового числа полного момента j, но не зависят от орбитального числа l (об отступлениях от этого результата говорится в § 44). Иначе говоря, в водороде и водородоподобных атомах уровни с одинаковыми квантовыми числами n и j, но различными l совпадают. Такое совпадение имеет место только у водорода и водородоподобных атомов. Для остальных одноэлектронных атомов, например атомов щелочных металлов, совпадения нет.

7. Величина тонкого расщепления энергетических уровней для легких атомов не превышает 10^{-5} эВ и сильно возрастает с увеличением заряда ядра. Для тяжелых атомов она может достигать десятых долей

эВ, так что в этих случаях нет смысла называть расщепление «тонким». (Напомним, что энергия ионизации атома водорода из основного состояния составляет 13,6 эВ.)

Для полноты заметим, что, помимо тонкой структуры, в спектре водорода и многих других атомов наблюдается еще так называемая *сверхтонкая структура.* Она возникает из-за взаимодействия магнитных моментов электронов со слабыми магнитными полями атомных ядер. Формула, аналогичная (38.4), к сверхтонкому расщеплению неприменима. Сверхтонкая структура будет рассмотрена ниже.

8. Чтобы не возвращаться к вопросу о квантовых числах и не излагать дважды правил отбора при излучении света, рассматриваемых в следующем параграфе, остановимся кратко на *сложных*, т. е. *многоэлектронных*, *атомах.* Подробный разбор затрагиваемых здесь вопросов относится к специальным курсам спектроскопии. В общем курсе физики об этих вопросах можно дать лишь общее предварительное представление, совсем не претендуя при этом на полноту и достаточную убедительность изложения.

В случае многоэлектронных атомов каждый (i-й) электрон электронной оболочки атома можно было бы характеризовать орбитальным $\mathbf{l}_i$ и спиновым $\mathbf{s}_i$ векторами момента количества движения. Однако опыт показывает, что при рассмотрении наиболее важных вопросов можно обойтись значительно менее подробной характеристикой, объединяя (связывая) определенным образом по правилу векторного сложения орбитальные и спиновые моменты отдельных электронов. Если бы нас интересовал только полный момент количества движения атома $\mathbf{J}$, то порядок сложения векторов $\mathbf{l}_i$ и $\mathbf{s}_i$ не имел бы значения, так как окончательный результат не зависит от порядка расположения слагаемых. В действительности наряду с $\mathbf{J}$ существенны также другие моменты и соответствующие им квантовые числа. Такие моменты получаются из $\mathbf{l}_i$ и $\mathbf{s}_i$ путем выделения соответствующих групп слагаемых. Какие группы надо выделить и произвести в них сложение $\mathbf{l}_i$ и $\mathbf{s}_i$ — это зависит от относительной величины различных взаимодействий между электронами атома. Наиболее важной и распространенной является так называемая *нормальная связь*, или *связь Рассела–Саундерса*, предложенная этими американскими астрофизиками в 1925 г. Она осуществляется, когда электростатическое взаимодействие электронов — их отталкивание по закону Кулона — велико по сравнению со спин-орбитальным взаимодействием, т. е. взаимодействием между орбитальными и спиновыми магнитными моментами электронов. Это, как правило, имеет место в легких и не слишком тяжелых атомах.

9. Нормальная связь заключается в том, что орбитальные и спиновые моменты электронов электронной оболочки в отдельности складываются по правилам векторного сложения в *общие орбитальный и спиновый моменты атома*, обозначаемые соответствующими прописными (большими) буквами $\mathbf{L}$ и $\mathbf{S}$, т. е.

$$\begin{aligned} \mathbf{L} &= \mathbf{l}_1 + \mathbf{l}_2 + \mathbf{l}_3 + \ldots, \\ \mathbf{S} &= \mathbf{s}_1 + \mathbf{s}_2 + \mathbf{s}_3 + \ldots \end{aligned} \tag{38.5}$$

Состояние электронной оболочки атома и характеризуется суммарными моментами $\mathbf{L}$ и $\mathbf{S}$, а также полным моментом количества движения атома, который, конечно, зависит от угла между векторами $\mathbf{L}$ и $\mathbf{S}$. Его можно получить по формуле

$$\mathbf{J} = \mathbf{L} + \mathbf{S}. \tag{38.6}$$

Векторам $\mathbf{J}$, $\mathbf{L}$, $\mathbf{S}$ соответствуют квантовые числа J, L, S, определяющие квадраты длин этих векторов по формулам (в единицах $\hbar$)

$$\mathbf{J}^2 = J(J+1), \quad \mathbf{L}^2 = L(L+1), \quad \mathbf{S}^2 = S(S+1). \tag{38.7}$$

Ясно, что при четном числе электронов в атоме квантовые числа S и J целые, а при нечетном — полуцелые. Квантовое число L всегда целое. Как всегда, квантовые числа J, L, S имеют смысл наибольших значений, которые могут принимать проекции векторов $\mathbf{J}$, $\mathbf{L}$, $\mathbf{S}$ на избранное направление. Соответствующие проекции, следовательно, могут принимать значения (в единицах $\hbar$):

$$\begin{aligned} m_J &= -J,\ -(J-1),\ \ldots,\ +(J-1),\ +J, \\ m_L &= -L,\ -(L-1),\ \ldots,\ +(L-1),\ +L, \\ m_S &= -S,\ -(S-1),\ \ldots,\ +(S-1),\ +S. \end{aligned} \tag{38.8}$$

В частности, при заданных L и S квантовое число J может принимать следующие значения:

$$J = |L+S|,\ |L+S-1|,\ \ldots,\ |L-S|. \tag{38.9}$$

Конечно, при определении векторов $\mathbf{L}$, $\mathbf{S}$, $\mathbf{J}$ достаточно ограничиться только наружными, валентными электронами, если внутренние оболочки атома *полностью заполнены электронами*, так как в этом случае моменты количества движения внутренних электронов, как орбитальные, так и спиновые, полностью *скомпенсированы*, т. е. полные моменты внутренних оболочек равны нулю.

Электроны в атоме подвергаются действию электрического поля ядра, обладающего центральной симметрией. Благодаря этому вектор полного момента $\mathbf{J}$ точно сохраняется. Но векторы $\mathbf{L}$ и $\mathbf{S}$ в отдельности не сохраняются, а изменяются из-за спин-орбитального взаимодействия. При этом, однако, длины векторов $\mathbf{L}$ и $\mathbf{S}$, а значит, и квантовые числа L и S остаются практически неизменными. Практически сохраняются также проекции векторов $\mathbf{L}$ и $\mathbf{S}$ на направление вектора $\mathbf{J}$. Благодаря этому картину временного изменения $\mathbf{L}$ и $\mathbf{S}$ можно наглядно представить как прецессию (вращение) этих векторов вокруг неизменного направления вектора $\mathbf{J}$, и притом с общей угловой скоростью. Аналогом этого может служить свободная прецессия оси фигуры и угловой скорости $\boldsymbol{\omega}$ симметричного гироскопа вокруг неизменного направления вектора момента количества движения (см. т. I, § 49). Различие состоит в том, что в случае гироскопа направления оси фигуры и вектора $\boldsymbol{\omega}$ могут меняться непрерывно, тогда как в случае атома они квантуются. Это происходит из-за того, что проекции векторов $\mathbf{L}$ и $\mathbf{S}$ на направление

вектора **J** могут принимать только квантованные значения $m_L\hbar$ и $m_S\hbar$, где m_L и m_S — соответствующие квантовые числа, которые могут принимать значения в соответствии с формулами (38.8).

10. В спектроскопии состояние наружных (валентных) электронов атома суммарно характеризуют квантовым числом L, причем вместо числового значения L применяют соответствующую букву латинского алфавита. Именно, поступают так же, как в случае одного электрона (см. § 34, п. 2). Только вместо строчных букв применяют такие же, но *прописные* (большие) буквы латинского алфавита. Иначе говоря, пользуются следующей схемой:

Квантовое число L	0	1	2	3	4	5	6	7	8
Обозначение состояния	S	P	D	F	G	H	I	K	L

и далее по алфавиту с пропусками букв P и S.

В качестве нижнего индекса справа от соответствующей буквы ставят квантовое число полного момента J, а в качестве верхнего индекса слева — число $2S+1$, называемое *мультиплетностью уровня*. По этому числу можно вычислить не только спин S, но и число уровней, на которые расщепляется рассматриваемый уровень из-за спин-орбитального взаимодействия. Впрочем, число $2S+1$ дает число компонент в расщепившемся уровне только в случае, когда $S \leqslant L$. В противоположном случае, когда $S \geqslant L$, число компонент в расщепившемся уровне определяется числом возможных проекций вектора **L** на более длинный вектор **S**, т. е. оно равно $2L+1$. Правда, и в этом случае, хотя и чисто формально, число $2S+1$ называют мультиплетностью уровня.

Например, когда наружная оболочка атома состоит из двух электронов, то возможны два случая: 1) спины электронов направлены противоположно, а потому $S=0$; 2) спины электронов параллельны, тогда $S=1$.

В первом случае $J=L, 2S+1=1$, т. е. все уровни *синглетны*. Соответственно различным значениям L получаются следующие уровни:

$L=J$	0	1	2	3	4	5	6
Уровни	$^1S_0$	$^1P_1$	$^1D_2$	$^1F_3$	$^1G_4$	$^1H_5$	$^1I_6$

Во втором случае $2S+1=3$, т. е. все уровни *триплетны*, за исключением, конечно, уровней s, которые всегда *синглетны*. Здесь возможны три случая: $J=L-1$, $J=L$, $J=L+1$. В соответствии с этим получается следующая схема:

L	0		1			2			3			4		
J	0	1	0	1	2	1	2	3	2	3	4	3	4	5
Уровни	$^3S_0$	$^3S_1$	$^3P_0$	$^3P_1$	$^3P_2$	$^3D_1$	$^3D_2$	$^3D_3$	$^3F_2$	$^3F_3$	$^3F_4$	$^3G_3$	$^3G_4$	$^3G_5$

Читателю рекомендуется разобрать аналогичный вопрос, когда наружная оболочка атома содержит три электрона.

можно записать в виде

$$\mathbf{J} = \mathbf{J}' + \mathbf{s}_\text{ф}, \tag{39.1}$$

где $\mathbf{J}$ — момент количества движения атома до излучения фотона (в единицах $\hbar$), $\mathbf{J}'$ — после излучения, а $\mathbf{s}_\text{ф}$ — вектор спина фотона. В дальнейшем индекс «ф» для краткости будем опускать. Закон (39.1) записан в символической форме, поскольку в одном и том же состоянии все три компоненты квантовомеханического вектора $\mathbf{J}$ не могут иметь определенные значения. Однако это не вносит никаких неопределенностей в дальнейшие рассуждения, поскольку в них речь идет не о самих векторах $\mathbf{J}$, $\mathbf{J}'$, $\mathbf{s}$, а о соответствующих им квантовых числах J, J', s. Разумеется, квантовые числа в обеих частях равенства (39.1) должны быть одинаковы. Это и используется в дальнейшем, причем квантовые числа правой части (39.1) получаются по правилу векторного сложения (см. § 32).

Впрочем, есть частный случай, когда и в квантовой механике вектор $\mathbf{J}$ определен однозначно. Это — случай, когда квантовое число полного момента $J = 0$. Тогда $\mathbf{J}^2 = J(J+1) = 0$, т. е. сам вектор $\mathbf{J}$, а с ним и все его проекции имеют определенные значения. В этом отношении вектор $\mathbf{J}$ ведет себя так же, как и в классическом случае. Поэтому переходы из квантового состояния с $J = 0$ в другое состояние также с $J = 0$ (так называемые $0-0$-переходы) *абсолютно запрещены.* В противном случае из-за наличия спина у фотона момент количества движения атома, по крайней мере в одном из этих состояний, был бы отличен от нуля, а этого по предположению не должно быть.

3. Строгий квантовомеханический вывод правил отбора потребовал бы введения понятий и математических методов, выходящих за пределы нашего курса. Поэтому мы поступим не вполне последовательно и применим модельный *метод векторных диаграмм*, условный смысл которых уже отмечался ранее в § 32 (п. 5). Такой прием не является настоящим выводом — его скорее следует рассматривать как способ запоминания и осмысливания правил отбора. Оправданием метода может служить только то, что он приводит к правильным результатам. В рассматриваемом методе символы $\mathbf{J}$ и $\mathbf{s}$ рассматриваются как обычные *классические векторы.* Только длины этих векторов считаются равными не J и s, а $\sqrt{J(J+1)}$ и $\sqrt{s(s+1)}$. (Впрочем, если принять $|\mathbf{J}| = J$ и $|\mathbf{s}| = s$, то получатся те же правила отбора.) Рисунок 68 *а* выражает закон сохранения момента импульса при излучении фотона в рассматриваемой векторной модели: $\mathbf{J} = \mathbf{J}' + s$.

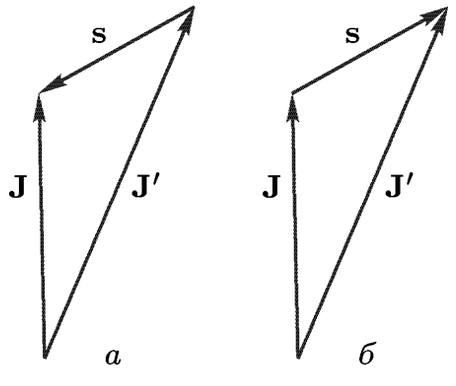


Рис. 68

Рассмотрим сначала случай излучения фотона, когда ни один из векторов $\mathbf{J}$ и $\mathbf{J}'$ не обращается в нуль, причем $|\mathbf{J}'| \geqslant |\mathbf{J}|$. Всякая сторона треугольника короче суммы длин остальных двух сторон. Возьмем из двух сторон $\mathbf{J}$ и $\mathbf{J}'$ более длинную, т. е. воспользуемся неравенством

$|\mathbf{J}'| \leqslant |\mathbf{J}| + |\mathbf{s}|$ или

$$\sqrt{J'(J'+1)} \leqslant \sqrt{J(J+1)} + \sqrt{s(s+1)}. \tag{39.2}$$

Так как для фотона $s = 1$, то последнее слагаемое равно $\sqrt{2}$. Квантовые числа J и J' целые, когда число электронов в атоме четное, и полуцелые, когда оно нечетное. Приращение $\Delta J \equiv J' - J$ может поэтому равняться только положительному целому числу или нулю, так как при излучении фотона число электронов в атоме не меняется. Заменяя в неравенстве (39.2) J' на $J + \Delta J$ и возводя его в квадрат, получим

$$\Delta J^2 + (2J+1)\Delta J - 2 \leqslant 2\sqrt{2J(J+1)}. \tag{39.3}$$

При фиксированном J и при $\Delta J \geqslant 0$ левая часть этого неравенства возрастает с возрастанием ΔJ, так как ее производная по ΔJ существенно положительна. При $\Delta J = 0$ неравенство (39.3) выполняется. Неравенство (39.3) выполняется и при $\Delta J = 1$, так как в этом случае оно переходит в очевидное неравенство $J \leqslant \sqrt{2J(J+1)}$. Но уже при $\Delta J = 2$ неравенство (39.3) не выполняется. В этом случае оно переходит в $2(J+1) \leqslant \sqrt{2J(J+1)}$, а такое неравенство неверно, в чем легко убедиться, возводя его в квадрат. Неравенство (39.3) тем более не выполняется при больших значениях ΔJ.

Случай $J' \leqslant J$ сводится к предыдущему заменой J на J' и наоборот.

Таким образом, когда ни одно из квантовых чисел J и J' не равно нулю, получается правило отбора при излучении фотона

$$\Delta J \equiv J' - J = \pm 1 \quad \text{или} \quad 0. \tag{39.4}$$

Когда одно из квантовых чисел J или J' обращается в нуль, треугольник на рис. 68 вырождается в два равных отрезка прямых, направленных одинаково или противоположно. Тогда в (39.4) случай $\Delta J = 0$ исключается. Возможны только переходы с $\Delta J = \pm 1$.

Случай, когда оба числа J и J' равны нулю, невозможен, на что было указано уже выше.

Правила отбора при *поглощении* фотона получаются так же, как и при излучении. В этом случае $\mathbf{J} + \mathbf{s} = \mathbf{J}'$, а вместо рис. 68 *а* надо пользоваться рис. 68 *б*.

Сформулируем теперь правила отбора, которым должны удовлетворять квантовые числа m_J и m'_J проекций полного момента импульса атома до и после излучения или поглощения фотона. При этом нет необходимости переходить к векторной модели, а можно написать сразу

$$\Delta m_J \equiv m'_J - m_J = \pm 1 \quad \text{или} \quad 0. \tag{39.5}$$

Эти правила, конечно, должны выполняться при одновременном выполнении предыдущих правил отбора. В частном случае, когда проекции m_J и m'_J максимальны, они совпадают с J и J', а правила (39.5) переходят в (39.4). Однако возможны и такие случаи, когда по крайней мере одна из этих проекций меньше соответствующего квантового числа J.

4. В связи с изложенным отметим следующее. В § 37 указывалось, что спин фотона может ориентироваться вдоль направления его распространения только двумя способами. Это означает, что любое состояние поляризации фотона может быть осуществлено путем линейной комбинации двух состояний, в одном из которых поляризация правая, а в другом левая. Между тем при спине s число состояний с различными проекциями вектора $\mathbf{s}$ на избранное направление должно было бы равняться $2s + 1$. Поэтому казалось бы, что спин фотона должен быть $1/2$. Но в таком случае при излучении и поглощении фотона квантовое число J полного момента количества движения атомной оболочки должно было бы меняться на $\pm 1/2$, т. е. из целого переходить в полуцелое и наоборот. Это находится в противоречии с уже отмеченным фактом, что при излучении и поглощении фотона число электронов в атоме не меняется, а *квантовое число J всегда целое при четном числе электронов и полуцелое — при нечетном.* В п. 8 § 37 уже указывалось, что из трех возможных проекций спина при $s = 1$ в случае фотона одна не осуществляется из-за поперечности электромагнитных волн.

5. Выведенные правила отбора для однофотонных процессов основаны на строгом законе сохранения момента количества движения. Посмотрим теперь, какие правила отбора связаны с поведением векторов $\mathbf{L}$ и $\mathbf{S}$. Излучение электромагнитных волн обусловлено *электромагнитными свойствами электрона*, т. е. его зарядом и магнитным моментом. Излучение фотона возникает либо в результате изменения движения заряда (изменение вектора $\mathbf{L}$), либо в результате поворота спинового магнитного момента, либо в результате обоих этих процессов сразу. Излучение, вызванное поворотом спина, конечно, — существенно релятивистский эффект. Теория показывает, что при излучении света в оптическом диапазоне взаимодействие фотона с зарядом электрона на несколько порядков сильнее взаимодействия его с магнитным моментом. Это позволяет считать, что излучение фотона в рассматриваемом диапазоне не связано с изменением $\mathbf{S}$, т. е.

$$\Delta \mathbf{S} = 0. \tag{39.6}$$

Иными словами, излучение и поглощение света не слишком коротких волн происходит так, как если бы спина вообще не было, а весь магнитный момент атома был только орбитальным. Поэтому можно воспользоваться полученными выше результатами, заменив полный момент $\mathbf{J}$ на орбитальный момент $\mathbf{L}$. Таким образом, при однофотонных процессах излучения и поглощения не слишком коротких волн должны приближенно выполняться следующие правила отбора:

$$\Delta L \equiv L' - L = \pm 1 \quad \text{или} \quad 0, \tag{39.7}$$

причем когда одно из чисел L и L' обращается в нуль, значение $\Delta L = 0$ исключается. Значение $\Delta L = 0$ невозможно также для атомов с *одним валентным электроном*, например для атомов водорода и щелочных металлов. Однако этот запрет связан не с законом сохранения момента количества движения, а с *законом сохранения четности волновой*

функции. На этом вопросе мы остановимся ниже. Здесь же отметим только, что правило отбора $\Delta L = \pm 1$ уже было использовано нами в § 34 для объяснения спектральных серий щелочных металлов.

6. Когда $\Delta J = \pm 1$, то излучается фотон с *круговой поляризацией.* Когда же $\Delta J = 0$, то поляризация получается *линейной.* Казалось бы, что это не согласуется с тем фактом, что спин фотона равен 1. Квантовая механика находит оригинальный выход из этого затруднения. Она утверждает, что в рассматриваемом случае излучается фотон в состоянии с *неопределенным спином.* Однако это состояние является суперпозицией двух состояний с круговой поляризацией — правой и левой, представленных с равной вероятностью. При измерении момента импульса, который передает фотон телу при поглощении, с одинаковой вероятностью может получиться только либо $+1$, либо -1.

Наконец, особо подчеркнем, что все полученные здесь правила отбора связаны со свойствами фотона и относятся к квантовым переходам с излучением или поглощением только одного фотона. На многофотонные процессы излучения и поглощения они не распространяются. Они не распространяются и на такие квантовые переходы, которые осуществляются не с помощью электромагнитного излучения, а, например, вызываются электронными ударами в газовых разрядах, возникают при тепловом возбуждении атомов и пр.

Возможны и излучательные переходы с нарушением правил отбора, приведенных выше. Они называются *запрещенными переходами.* Их вероятность много меньше вероятности разрешенных переходов. Интенсивность *запрещенных спектральных линий*, как правило, много меньше интенсивности разрешенных.

§ 40. Тонкая структура спектральных линий водорода и щелочных металлов

1. В § 38 было показано, что из-за спин-орбитального взаимодействия и зависимости массы электрона от скорости спектральные термы расщепляются. Это расщепление называется *тонкой структурой* спектральных термов или энергетических уровней. В настоящем параграфе рассматриваются только атомы с *одним* валентным электроном. В соответствии с этим термы обозначаются строчными (малыми) буквами латинского алфавита. Энергия уровня зависит от квантовых чисел n, l, j, но в отсутствие магнитного и электрического полей не может зависеть от магнитного квантового числа m_j, так как в этом случае все направления в пространстве совершенно эквивалентны. Только в случае водорода и водородоподобных атомов имеет место *случайное вырождение* по l из-за того, что электрическое поле ядра, в котором движется единственный электрон атома, — *кулоновское.* В этом случае энергия уровня зависит только от квантовых чисел n и j, но не зависит от l. Она определяется формулой (38.4).

От *тонкой структуры термов* следует отличать *тонкую структуру спектральных линий*, т. е. расщепление спектральной линии на

несколько близко расположенных компонент. Это расщепление определяется разрешенными переходами между различными подуровнями расщепившихся энергетических уровней, т. е. правилами отбора (39.4) и (39.7). Для одноэлектронных атомов в отсутствие внешних полей они сводятся к следующим двум:

$$\Delta l = \pm 1, \tag{40.1}$$

$$\Delta j = \pm 1 \text{ или } 0. \tag{40.2}$$

2. В качестве примера рассмотрим тонкую структуру линии L_α серии Лаймана водорода:

$$\overline{\nu} = 1s - 2p,$$

где $\overline{\nu} = 1/\lambda$ обозначает спектроскопическое волновое число (см. § 11). Терм $1s$ синглетный, терм $2p$ — дублет и состоит из двух подтермов $2p_{1/2}$ и $2p_{3/2}$ (рис. 69). Правила отбора (40.1) и (40.2) допускают переход с обоих уровней $2p_{1/2}$ и $2p_{3/2}$ на уровень $1s_{1/2}$. Поэтому линия L_α должна быть дублетом, т. е. состоять из двух спектральных линий. Линия $1s_{1/2}-2p_{3/2}$ более интенсивна, чем линия $1s_{1/2}-2p_{1/2}$, и изображена на рис. 69 вместе с соответствующим переходом $2p_{3/2} \to 1s_{1/2}$ более жирно. Расстояние между обеими линиями, как нетрудно подсчитать с помощью формулы (38.4), составляет $\Delta\overline{\nu} = 0{,}365$ см$^{-1}$, или в длинах волн $\Delta\lambda = 5{,}3 \cdot 10^{-4}$ нм. Длина волны самой линии L_α составляет $\lambda = 121{,}6$ нм. Для разрешения тонкой структуры линии L_α спектральный прибор должен иметь разрешающую способность не менее

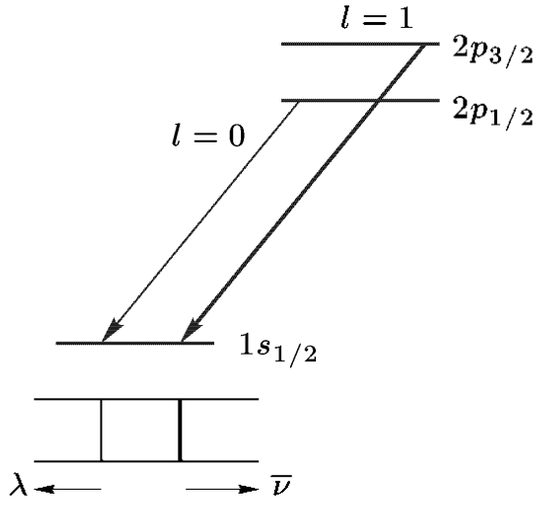


Рис. 69

$$\frac{\lambda}{\Delta\lambda} = 121{,}6 \cdot (5{,}3 \cdot 10^{-4}) \approx 2{,}23 \cdot 10^5.$$

Аналогичную дублетную структуру имеют и остальные линии серии Лаймана водорода: L_β, L_γ, L_δ.

В качестве второго примера рассмотрим тонкую структуру линии H_α серии Бальмера водорода (рис. 70). Главному квантовому числу $n = 2$ соответствуют состояния $2s_{1/2}$, $2p_{1/2}$ и $2p_{3/2}$. Поскольку энергия не зависит от орбитального числа l, первым двум состояниям соответствует одна и та же энергия, т. е. энергетический уровень, состоящий из двух слившихся уровней. Таким образом, при квантовом числе $n = 2$ получаются два уровня, из которых один слившийся. При $n = 3$ число различных уровней равно трем, причем один уровень, $3d_{5/2}$, простой, а два уровня, $(3p_{3/2}, 3d_{3/2})$ и $(3s_{1/2}, 3p_{1/2})$, слившиеся. Линия H_α возникает при квантовых переходах с верхних уровней ($n = 3$) на нижние ($n = 2$). Переходы, удовлетворяющие правилам отбора (40.1) и (40.2),

а также соответствующие им спектральные линии с указанием их примерной относительной интенсивности изображены на рис. 70. Таким образом, линия H_α, а также все остальные линии серии Бальмера H_β, H_γ, H_δ, ... состоят из пяти компонент.

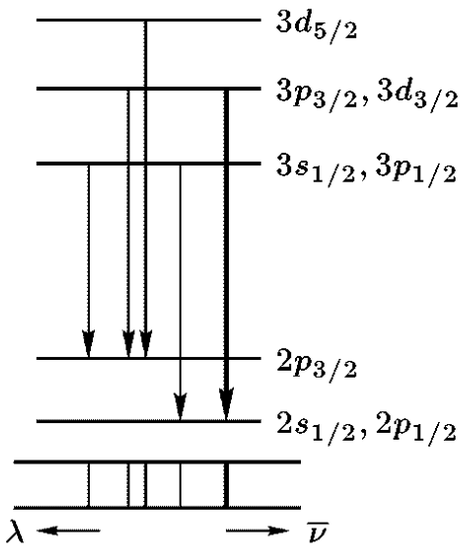


Рис. 70

3. Тонкая структура уровней и спектральных линий щелочных металлов и сходных с ними ионов в основном обусловлена *спин-орбитальным взаимодействием*, а не зависимостью массы электрона от скорости. Это связано с тем, что расщепление вследствие спин-орбитального взаимодействия быстро возрастает с увеличением Z. При больших Z расщепление уровней из-за зависимости массы электрона от скорости ничтожно по сравнению с расщеплением из-за спин-орбитального взаимодействия. Картина тонкого расщепления у щелочных металлов выглядит проще, чем у водорода. У водорода эта картина осложнена вырождением по l. У щелочных металлов это вырождение снимается, и закономерности тонкого расщепления предстают в чистом виде.

Происхождение тонкой структуры в спектральных сериях натрия пояснено на схематическом рис. 71. Главная серия возникает в результате переходов на наиболее глубокий уровень $3s_{1/2}$ с вышележащих

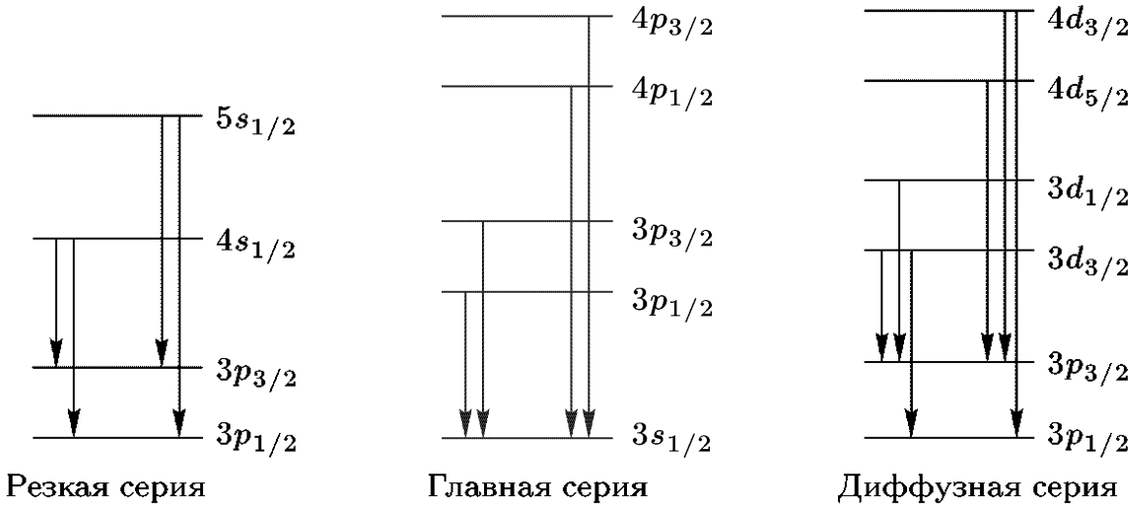


Рис. 71

p-уровней. Уровень $3s_{1/2}$ простой, а все p-уровни двойные, причем расстояние между компонентами этих уровней убывает с возрастанием главного квантового числа n. Поэтому и сами спектральные линии главной серии получаются двойными — *дублетами*. Расстояние между компонентами дублета уменьшается с возрастанием его номера (т. е. с увеличением частоты). Главная серия возникает и в спектре поглощения, так как в нормальном состоянии атомы натрия находятся на

наиболее низком уровне, т. е. на $3s_{1/2}$. Наиболее интенсивной является желтая резонансная линия натрия, возникающая при переходах $3p_{1/2} \to 3s_{1/2}$ и $3p_{3/2} \to 3s_{1/2}$. Этим переходам соответствуют длины волн $\lambda_1 = 589{,}6$ и $\lambda_2 = 589{,}0$ нм с разностью между ними $\Delta\lambda = 0{,}6$ нм.

Заметим, что с возрастанием Z расщепление спектральных линий щелочных металлов сильно увеличивается. Так, резонансная линия рубидия состоит из двух компонент $\lambda_1 = 794{,}8$ и $\lambda_2 = 780{,}0$ нм с расстоянием между ними $\Delta\lambda = 14{,}8$ нм. Для цезия соответствующие значения равны $\lambda_1 = 894{,}4$ и $\lambda_2 = 852{,}1$ нм, т. е. $\Delta\lambda = 42{,}3$ нм. Для таких расщеплений термин «тонкая структура» вряд ли является подходящим.

Линии резкой серии также являются *дублетами*, они возникают в результате переходов с простых s-уровней на лежащий ниже двойной $3p$-уровень, состоящий из подуровней $3p_{1/2}$ и $3p_{3/2}$. Поэтому расстояния между компонентами дублетов одни и те же для всей серии, причем сами компоненты являются *резкими* линиями. По этой причине и сама серия получила название *резкой*.

Диффузная серия возникает при переходах на двойной уровень $(3p_{1/2},\ 3p_{3/2})$ с вышележащих также двойных уровней $(3d_{3/2},\ 3d_{1/2})$, $(4d_{5/2},\ 4d_{3/2})$ и т. д. Ее спектральные линии являются *триплетами*, так как переходы типа $3d_{5/2} \to 3p_{1/2}$, в которых число j меняется на 2, запрещены правилами отбора (40.2). Расстояния между подуровнями различных d-уровней значительно меньше соответствующих расстояний для p-уровней. По этой причине при недостаточной разрешающей силе спектрального аппарата компоненты триплета не разрешаются, а сами линии получаются размытыми. Понятно поэтому, почему рассматриваемая серия получила название *диффузной*.

ЗАДАЧА

Рассчитать тонкую структуру спектральной линии H_α серии Бальмера водорода.

Р е ш е н и е. Энергетические уровни атома водорода определяются формулой (38.4), в которой следует положить $Z = 1$, т. е.

$$\mathscr{E}_{n,j} = -\frac{\mu e^4}{2\hbar^2 n^2}\left[1 + \frac{\alpha^2}{n}\left(\frac{1}{j+1/2} - \frac{3}{4n}\right)\right]. \tag{40.3}$$

Заменяя здесь массу электрона μ на приведенную массу, перепишем эту формулу так:

$$\mathscr{E}_{n,j} = -\frac{R_{\text{H}}}{n^2}\left[1 + \frac{\alpha^2}{n}\left(\frac{1}{j+1/2} - \frac{3}{4n}\right)\right], \tag{40.4}$$

где R_{H} — постоянная Ридберга для водорода, определяемая выражениями (13.10) и (13.9):

$$R_{\text{H}} = 109\,677{,}576 \text{ см}^{-1}.$$

Энергетические уровни здесь измеряются в обратных сантиметрах. Квадрат постоянной тонкой структуры равен $\alpha^2 = 5{,}3251 \cdot 10^{-5}$. Происхождение линии

H_α представлено на рис. 70. В соответствии с этим получаем

$$\mathscr{E}_{2,1/2} = -27\,419{,}395 \cdot 1{,}00001664 = -27\,419{,}850 \text{ см}^{-1},$$
$$\mathscr{E}_{2,3/2} = -27\,419{,}395 \cdot 1{,}00000333 = -27\,419{,}485 \text{ см}^{-1},$$
$$\mathscr{E}_{3,1/2} = -12\,186{,}397 \cdot 1{,}0000197 = -12\,186{,}639 \text{ см}^{-1},$$
$$\mathscr{E}_{3,3/2} = -12\,186{,}397 \cdot 1{,}0000067 = -12\,186{,}478 \text{ см}^{-1},$$
$$\mathscr{E}_{3,5/2} = -12\,186{,}397 \cdot 1{,}0000022 = -12\,186{,}423 \text{ см}^{-1}.$$

Взяв разности этих чисел в соответствии с рис. 70, находим спектроскопические волновые числа (в см^{-1}) и длины волн (в нм) для пяти компонент, на которые расщепляется линия H_α водорода:

$\overline{\nu}$, см^{-1}	λ, нм
15 233,372	656,4534
15 233,211	656,4604
15 233,067	656,4660
15 233,007	656,4692
15 232,846	656,4761

Заметим, что длины волн и спектроскопические волновые числа здесь отнесены к вакууму (а не к воздуху, как это обычно делается в спектроскопии).

§ 41. Простой и сложный эффект Зеемана

1. В кулоновском поле (в нерелятивистском приближении) все энергетические уровни электрона *вырождены* — энергия зависит только от главного квантового числа n, но не зависит от орбитального числа l. По этой причине все спектральные линии атома водорода *одиночные* (синглеты). Атомы щелочных металлов можно рассматривать как одноэлектронные атомы, в которых электрон движется в центрально-симметричном, но уже *не кулоновском* поле. Вырождение по l снимается — энергия уровня зависит не только от n, но и от l. С этим связано происхождение спектральных серий щелочных металлов. Наличие спин-орбитального взаимодействия приводит к *тонкой структуре* спектральных линий. Но в отсутствие внешних полей все направления в пространстве эквивалентны, а потому энергии уровней *не зависят от магнитного квантового числа* m, хотя при заданном J число m_J может принимать $2J+1$ значений. Кратность соответствующего вырождения равна, таким образом, $2J+1$. *Магнитное поле снимает и это вырождение*: каждый энергетический уровень расщепляется на $2J+1$ подуровней. Этим объясняется *эффект Зеемана*, подробно рассмотренный в т. IV (§ 92) с классической точки зрения. Однако до открытия спина электрона из-за наличия правил отбора квантовая теория, как и классическая, объясняла только *простой эффект Зеемана.*

В самом деле, если атом обладает магнитным моментом $\mathfrak{m}$, то его энергия в магнитном поле $\mathbf{B}$ равна $\mathscr{E} = \mathscr{E}_0 - (\mathfrak{m}\mathbf{B})$, где $\mathscr{E}_0$ — энергия

в отсутствие магнитного поля. Если нет спина, то магнитный момент обусловлен только орбитальным движением электронов. Его проекция на направление магнитного поля составляет целое число магнетонов Бора, т. е. $m_L \mathfrak{m}_Б = m_L(e\hbar/(2\mu_e c))$. Следовательно,

$$\mathscr{E} = \mathscr{E}_0 - m_L \frac{e\hbar B}{2\mu_e c} = \mathscr{E}_0 - \hbar\Omega m_L, \tag{41.1}$$

где

$$\Omega = \frac{eB}{2\mu_e c} \tag{41.2}$$

— *ларморовская частота.* Каждый уровень расщепляется в магнитном поле на $2L+1$ подуровней. В результате квантовых переходов между различными уровнями излучаются спектральные линии с частотами

$$\omega = \omega_0 - \Omega\Delta m_L,$$

где $\omega_0 \equiv \Delta\mathscr{E}_0/\hbar$ — частота линии, излучаемой в отсутствие магнитного поля. В силу правил отбора $\Delta m_L = 0$ или ± 1. Таким образом, частота излучаемой линии будет

$$\omega = \omega_0 \quad \text{или} \quad \omega_0 \pm \Omega, \tag{41.3}$$

т. е. получается *лоренцевский триплет.* В соответствии с классической теорией переходам $\Delta m_L = 0$ соответствуют колебания вдоль (π-компоненты), а переходам $\Delta m_L = \pm 1$ — поперек магнитного поля (σ-компоненты).

2. Учет *спина электрона* позволил объяснить и *сложный эффект Зеемана.* Будем исходить из векторной модели, предполагая, что осуществляется *нормальная связь* (связь Рассела–Саундерса, см. § 38, п. 3). В ней атом характеризуется орбитальным моментом количества движения $\mathbf{L}$, спиновым моментом $\mathbf{S}$ и общим моментом количества движения $\mathbf{J} = \mathbf{L} + \mathbf{S}$. В операторной форме $\widehat{\mathbf{J}} = \widehat{\mathbf{L}} + \widehat{\mathbf{S}}$, $\widehat{\mathbf{J}}^2 = \widehat{\mathbf{L}}^2 + \widehat{\mathbf{S}}^2 + 2(\widehat{\mathbf{L}}\widehat{\mathbf{S}})$. Будем рассматривать состояние, в котором квадраты моментов имеют определенные значения, т. е. равны соответственно $J(J+1)$, $L(L+1)$, $S(S+1)$. В том же состоянии имеет определенное значение и скалярное произведение $(\mathbf{LS})$, а именно

$$(\mathbf{LS}) = \frac{1}{2}[\mathbf{J}^2 - \mathbf{L}^2 - \mathbf{S}^2] = \frac{1}{2}[J(J+1) - L(L+1) - S(S+1)]. \tag{41.4}$$

Моментам количества движения $\mathbf{L}$ и $\mathbf{S}$ соответствуют орбитальный и спиновый магнитные моменты $\mathfrak{m}_L = -g_l\mathbf{L}$, $\mathfrak{m}_s = -g_s\mathbf{S}$. (Знак минус поставлен потому, что заряд электрона отрицателен, а потому векторы $\mathfrak{m}_L$ и $\mathbf{L}$, а также $\mathfrak{m}_s$ и $\mathbf{S}$ направлены противоположно.) Примем за единицу момента количества движения величину $\hbar$, а за единицу магнитного момента — магнетон Бора $\mathfrak{m}_Б$. В этих единицах для электрона $g_l = 1$, $g_s = 2$. То обстоятельство, что $g_l \neq g_s$, и обеспечивает сложный характер эффекта Зеемана. Однако мы временно не будем фиксировать числовые значения g_l и g_s, имея в виду, что величины g_s для протона и нейтрона не равны соответствующей величине для электрона.

В отсутствие внешнего поля общий момент количества движения $\mathbf{J}$ сохраняется как по величине, так и по направлению (применяется векторная модель). Моменты же $\mathbf{L}$ и $\mathbf{S}$ из-за спин-орбитального взаимодействия не сохраняются. Однако в рассматриваемом состоянии сохраняются их длины. В результате они совершают *регулярную прецессию* вокруг неизменного направления вектора $\mathbf{J}$ и притом с одной и той же угловой скоростью, так как векторы $\mathbf{L}$, $\mathbf{S}$, $\mathbf{J}$ должны все время лежать в одной плоскости. С той же угловой скоростью будут прецессировать и соответствующие им магнитные моменты $\mathfrak{m}_L = -g_l\mathbf{L}$ и $\mathfrak{m}_S = -g_s\mathbf{S}$, а также общий магнитный момент $\mathfrak{m} = -g_l\mathbf{L} - g_s\mathbf{S}$. Действительно, ввиду неравенства g_l и g_s вектор $\mathfrak{m}$ не коллинеарен вектору $\mathbf{J}$, а потому также должен менять направление (рис. 72, где принято во внимание, что ввиду отрицательного заряда электрона направления векторов $\mathbf{L}$ и $\mathfrak{m}_L$, а также векторов $\mathbf{S}$ и $\mathfrak{m}_S$ противоположны).

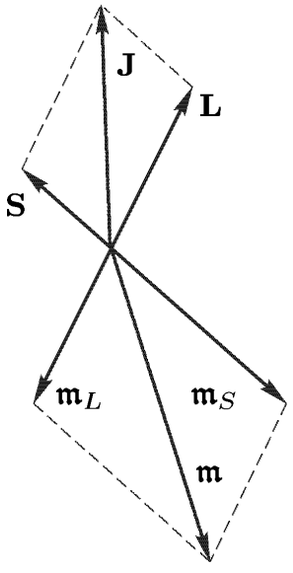


Рис. 72

Найдем теперь проекцию $\mathfrak{m}_{\|}$ вектора $\mathfrak{m}$ на направление вектора $\mathbf{J}$. Для этого прежде всего находим скалярное произведение

$$(\mathfrak{m}\mathbf{J}) = (-g_l\mathbf{L} - g_s\mathbf{S})(\mathbf{L}+\mathbf{S}) = -g_l\mathbf{L}^2 - g_s\mathbf{S}^2 - (g_l+g_s)(\mathbf{LS}),$$

или ввиду соотношения (41.4)

$$(\mathfrak{m}\mathbf{J}) = -g\mathbf{J}^2. \tag{41.5}$$

где через g обозначена величина

$$g = \frac{g_l+g_s}{2} + \frac{g_l-g_s}{2}\,\frac{\mathbf{L}^2-\mathbf{S}^2}{\mathbf{J}^2},$$

или в более подробной записи

$$g = \frac{g_l+g_s}{2} + \frac{g_s-g_l}{2}\,\frac{S(S+1)-L(L+1)}{J(J+1)}. \tag{41.6}$$

В частности, для электрона $g_l = 1$, $g_s = 2$, и выражение (41.6) переходит в

$$g = \frac{3}{2} + \frac{S(S+1)-L(L+1)}{2J(J+1)}. \tag{41.7}$$

В этом случае g является *рациональной дробью*. Величина g называется *множителем Ланде* (1888–1975).

Таким образом, на основании (41.5) можно написать $\mathfrak{m}_{\|}|\mathbf{J}| = -g|\mathbf{J}|^2$, откуда видно, что проекция вектора $\mathfrak{m}$ на направление вектора $\mathbf{J}$ имеет определенное значение, а именно $\mathfrak{m}_{\|} = -g|\mathbf{J}|$. Перпендикулярная проекция $m_\perp$, как и должно быть, определенного значения не имеет. В векторной модели она совершает прецессию вокруг вектора $\mathbf{J}$. При рассмотрении процессов, происходящих медленно по сравнению с этой

прецессией, от наличия перпендикулярной составляющей можно отвлечься, считая, что полный магнитный момент атома сводится к одной только его проекции $\mathfrak{m}_{\|}$. В этом приближении

$$\mathfrak{m} = \mathfrak{m}_{\|} = -g\mathbf{J}. \tag{41.8}$$

Именно такой магнитный момент атома проявляется, например, в опытах Штерна и Герлаха.

Формулы (41.6) и (41.7) могут быть получены и последовательно, квантовомеханически. Для этого надо только ввести *оператор проекции магнитного момента* $\widehat{\mathfrak{m}}_{\|}$ и определить его собственные значения. Но мы не будем этого делать, предпочитая наглядный, хотя и непоследовательный вывод, приведенный в тексте. Заметим только, что не представляет затруднений распространить этот наглядный метод на случай $(j-j)$-связи, а также на промежуточные виды связи. В этих случаях для g получаются формулы, отличающиеся от (41.6) и (41.7). Из-за этого характер расщепления спектральных линий в сложном эффекте Зеемана получается несколько иным, чем в случае нормальной связи. Интересующихся этим вопросом мы отсылаем к специальным курсам спектроскопии, например к книге М.А. Ельяшевича (р. 1908) «Атомная и молекулярная спектроскопия». — М.: Физматгиз, 1962.

3. До сих пор предполагалось, что магнитного поля нет. Допустим теперь, что атом помещен в однородное магнитное поле $\mathbf{B}$. Тогда не будет сохраняться и полный момент атома $\mathbf{J}$. Он под действием поля $\mathbf{B}$ начнет совершать *вынужденную прецессию*. Рассмотрим сначала случай слабого поля, когда эта прецессия совершается медленно по сравнению с прецессией, связанной со спин-орбитальным взаимодействием. Иными словами, магнитное поле $\mathbf{B}$ должно быть настолько слабым, чтобы вызываемое им зеемановское расщепление спектральных линий было мало по сравнению с расстояниями между компонентами тонкой структуры спектральных линий. В этом случае эффект Зеемана получается *сложным*. Для одиночных линий (синглетов) указанное условие никогда не может выполняться. Для таких линий всякое магнитное поле является *сильным*, и наблюдаемый на них эффект Зеемана всегда простой.

В случае слабого поля дополнительная энергия, приобретаемая атомом в магнитном поле, может быть вычислена в предположении, что полный магнитный момент атома дается выражением (41.8), или в обычных единицах

$$\mathfrak{m} = -g\frac{e\hbar}{2\mu c}\mathbf{J}. \tag{41.9}$$

Таким образом,

$$\mathscr{E} = \mathscr{E}_0 - (\mathfrak{m}\mathbf{B}) = \mathscr{E}_0 + gm_J\Omega\hbar, \tag{41.10}$$

где $\mathscr{E}_0$ — энергия атома в отсутствие магнитного поля. Эта формула показывает, на какие энергетические уровни расщепляется каждый уровень атома при внесении его в слабое магнитное поле. Она же

с учетом правил отбора определяет возможные радиационные переходы между уровнями, а следовательно, и длины волн соответствующих им спектральных линий. Очевидно, из правил отбора надо принять во внимание только следующее:

$$\Delta m_J = \pm 1 \text{ или } 0, \tag{41.11}$$

так как само собой разумеется, что спектральная линия, о зеемановском расщеплении которой идет речь, не запрещена, т. е. всем остальным правилам отбора удовлетворяет. Одним словом, формулы (41.10) и (41.11) составляют основу для понимания явления Зеемана, как сложного, так и простого.

4. Происхождение сложного эффекта Зеемана проследим на примере спектральной линии щелочного металла, соответствующей переходу с дублетного уровня $({}^2p_{1/2}, {}^2p_{3/2})$ на синглетный уровень ${}^2s_{1/2}$. В частности, при $n = 3$ получается всем известный желтый дублет натрия. Но мы не будем фиксировать главное квантовое число, чтобы подчеркнуть, что характер зеемановского расщепления от него не зависит. Рассчитаем сначала множитель Ланде по формуле (41.7). Для уровня $p_{3/2}$ $l = 1$, $j = 3/2$, $s = 1/2$ получаем $g = 4/3$. Аналогично для уровня $p_{1/2}$ $g = 2/3$, а для уровня $s_{1/2}$ $g = 2$. Согласно (41.10) смещения подтермов, на которые расщепляются соответствующие термы, равны gm_J, если за единицу принять лоренцевское расщепление, т. е. за единицу частоты принять ларморовскую частоту Ω. Соответствующие значения приведены в следующей схеме:

$p_{3/2}$	m_j	$-3/2$	$-1/2$	$+1/2$	$+3/2$
	gm_j	-2	$-2/3$	$+2/3$	$+2$
$p_{1/2}$	m_j		$-1/2$	$+1/2$	
	gm_j		$-1/3$	$+1/3$	
$s_{1/2}$	m_j		$-1/2$	$+1/2$	
	gm_j		-1	$+1$	

Расчет расщепления дублета $p_{1/2} \to s_{1/2}$; $p_{3/2} \to s_{1/2}$ в слабом магнитном поле приведен в табл. 3. Номера соответствующих линий приведены в порядке возрастания частоты. Таблице соответствует схема спектральных переходов, представленная на рис. 73. На этой схеме в левом столбце изображены энергетические уровни $s_{1/2}$, $p_{1/2}$, $p_{3/2}$ в отсутствие магнитного поля ($\mathbf{B} = 0$). Правее показаны те же уровни, расщепившиеся в слабом магнитном поле $\mathbf{B}$, и все разрешенные переходы между ними, создающие картину зеемановского эффекта в рассматриваемом случае. Мы видим, что линия $p_{1/2} \to s_{1/2}$ расщепилась на четыре компоненты, из которых крайние являются σ-компонентами, средние — π-компонентами. Вторая линия $p_{3/2} \to s_{1/2}$ расщепилась на шесть компонент: средние две являются π-компонентами, остальные четыре σ-компонентами. Расстояния между компонентами приведены в единицах лоренцевского расщепления, т. е. за единицу частоты принята ларморовская частота $eB/2\mu c$.

Таблица 3

Переходы $m_j^{(1)} \to m_j^{(2)}$	Поляризация	$(gm_j)^1 - (gm_j)^2$	Номер линии
Расщепление линии $p_{1/2} \to s_{1/2}$			
$-1/2 \to +1/2$	σ	$-1/3 - 1 = -4/3$	1
$-1/2 \to -1/2$	π	$-1/3 - (-1) = 2/3$	3
$+1/2 \to +1/2$	π	$1/3 - 1 = -2/3$	2
$+1/2 \to -1/2$	σ	$1/3 - (-1) = 4/3$	4
Расщепление линии $p_{3/2} \to s_{1/2}$			
$-3/2 \to +1/2$ — переход запрещен			
$-3/2 \to -1/2$	σ	$-2 - (-1) = -1$	6
$-1/2 \to +1/2$	σ	$-2/3 - 1 = -5/3$	5
$-1/2 \to -1/2$	π	$-2/3 + 1 = 1/3$	8
$+1/2 \to +1/2$	π	$2/3 - 1 = -1/3$	7
$+1/2 \to -1/2$	σ	$2/3 + 1 = 5/3$	10
$+3/2 \to +1/2$	σ	$2 - 1 = 1$	9
$+3/2 \to -1/2$ — переход запрещен			

5. Величины gm_J в формуле (41.10) являются небольшими рациональными числами. Поэтому из приведенной схемы происхождения сложного эффекта Зеемана непосредственно следует правило, эмпирически установленное Рунге (1856–1927) еще в то время, когда физическая природа сложного эффекта была не ясна. *Правило Рунге* утверждает, что расстояния между компонентами, получившимися в сложном эффекте Зеемана в результате расщепления в магнитном поле, являются *небольшими рациональными числами*, если за единицу расстояния принять нормальное лорентцевское расщепление Ω, получающееся в том же магнитном поле.

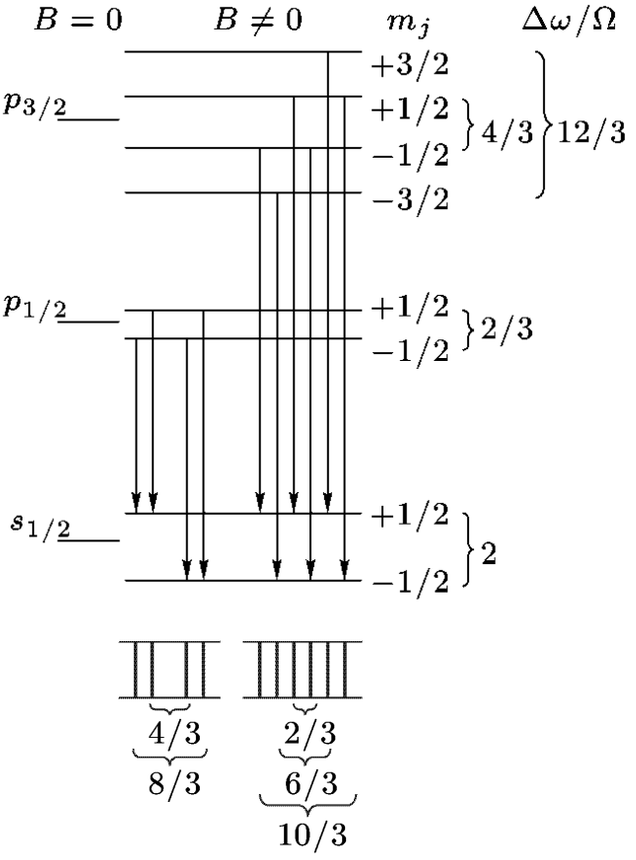


Рис. 73

Второе правило было установлено также эмпирически до создания теории сложного эффекта Зеемана. Это — *правило Престона*, состоящее в том, что спектральные линии, имеющие *один и тот же сериальный символ*, дают *одинаковый тип магнитного расщепления* независимо от значений главных квантовых чисел. Под се-

риальным символом понимают выражение вида ${}^2P_{3/2}{}^2D_{5/2}$, в котором ${}^2D_{5/2}$ относится к состоянию, из которого, а ${}^2P_{3/2}$ — на которое совершается радиационный переход. При этом значения главных квантовых чисел как в исходном, так и конечном состояниях атома могут быть какими угодно. По правилу Престона, например, компоненты ${}^2S_{1/2}{}^2P_{1/2}$ дублетов главной серии щелочных металлов расщепляются в слабом магнитном поле одинаково независимо от значений главных квантовых чисел и от химических элементов, к которым они принадлежат. То же относится ко вторым компонентам тех же дублетов, а также к мультиплетам резкой и диффузной серий щелочных металлов. Отступления от правила Престона связаны либо с узостью мультиплетной структуры, когда расстояния между компонентами мультиплета порядка или меньше величины расщепления, вызываемого магнитным полем, либо с нарушением нормальной связи между L и S.

6. С возрастанием напряженности магнитного поля, когда воздействие магнитного поля на атом становится сравнимым или больше спин-орбитального взаимодействия, должна меняться и картина расщепления в явлении Зеемана. В этом случае уже нельзя рассматривать поведение каждой компоненты мультиплета тонкой структуры в отдельности, а следует рассматривать *мультиплет в целом.* По мере усиления магнитного поля меняется и сам мультиплет. Особенно простых результатов следует ожидать, когда магнитное поле становится настолько сильным, что магнитное расщепление становится заметно больше ширины первоначального мультиплета. В этом случае можно пренебречь спин-орбитальным взаимодействием и рассматривать вынужденную прецессию векторов $\mathbf{L}$ и $\mathbf{S}$ в магнитном поле *независимо друг от друга.* Тогда

$$\mathfrak{m} = \mathfrak{m}_L + \mathfrak{m}_S = -\frac{e\hbar}{2\mu c}(\mathbf{L} + 2\mathbf{S}),$$

и, следовательно,

$$\mathscr{E} = \mathscr{E}_0 + \frac{e\hbar}{2\mu c}(\mathbf{L} + 2\mathbf{S})\mathbf{B} = \mathscr{E}_0 + \Omega\hbar(m_L + 2m_S).$$

Так как, согласно правилам отбора, $\Delta m_L = \pm 1$ или 0, а $\Delta m_S = 0$, то отсюда получаем

$$\Delta\omega = \frac{\Delta\mathscr{E}}{\hbar} = \pm\Omega \quad \text{или} \quad 0.$$

И, действительно, в 1912 г. Пашен и Бак (1881–1959) обнаружили, что с возрастанием магнитного поля происходят магнитооптические превращения, *конечным результатом которых в сильных полях является простой триплет Зеемана–Лоренца,* хотя в его трех компонентах все еще остаются малые по сравнению с расстояниями между ними и не зависящие от напряженности магнитного поля расщепления того же порядка, что и тонкая структура мультиплетов в отсутствие поля. Это явление получило название *явления Пашена–Бака.*

7. Рисунок 74 иллюстрирует, что́ получается на опыте при фотографировании поперек магнитного поля. В левом верхнем углу первого

ряда показана одиночная спектральная линия (синглет) в отсутствие магнитного поля, а под ней — ее расщепление в лорентцевский триплет

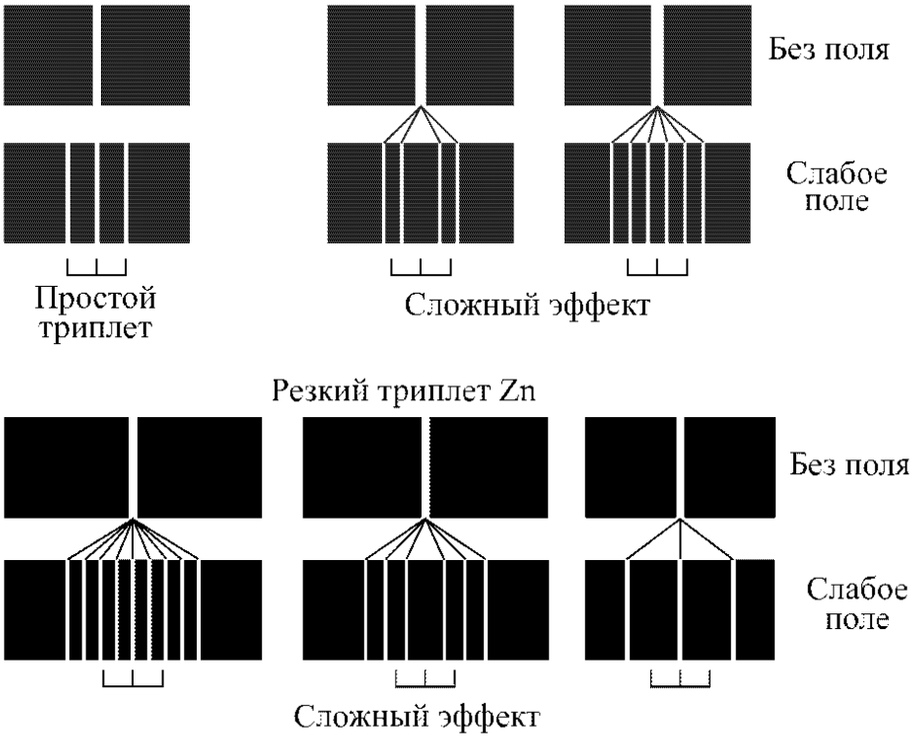


Рис. 74

в магнитном поле. Справа сверху сфотографирован желтый дублет натрия, под ним — расщепление в слабом магнитном поле. Этот случай был рассчитан выше в п. 4. Наконец, в нижней части рисунка приведен триплет цинка, а под линиями этого триплета — их расщепление в слабом магнитном поле.

8. Сформулируем теперь *количественно*, какие поля должны считаться *слабыми*, а какие *сильными*. Пусть $\Delta\omega$ означает ширину рассматриваемого мультиплета. Тогда поле будет слабым, если $\Omega \ll \Delta\omega$, и сильным в противоположном случае. Если перейти к длинам волн, то $\Delta\omega = |2\pi c\Delta(1/\lambda)| = 2\pi c\Delta\lambda/\lambda^2$. Подставляя значение $\Omega = eB/(2\mu c)$, найдем, что в случае слабых полей

$$B \ll \frac{4\pi\mu c^2\Delta\lambda}{e\lambda^2}. \tag{41.12}$$

Например, в случае резонансной желтой линии натрия $\lambda = 590$ нм, $\Delta\lambda = 0{,}6$ нм, и формула (41.12) дает для слабых полей $B \ll 3{,}7 \cdot 10^5$ Гс. Понятно поэтому, почему рассматриваемая линия в полях порядка 10^4 Гс дает сложный эффект. Для линии L_α лаймановской серии водорода $\lambda = 121{,}6$ нм, $\Delta\lambda = 5{,}3 \cdot 10^{-4}$ нм слабыми являются поля $B \ll 8000$ Гс. В случае линии H_α серии Бальмера $\lambda = 656$ нм, $\Delta\lambda = 0{,}0227$ нм (см. задачу к предыдущему параграфу), $B \ll 1{,}1 \cdot 10^4$ Гс.

Из этих данных видно, насколько неудачны были названия «нормальный» и «аномальный», которыми долго пользовались для обозначения соответственно простого и сложного эффектов Зеемана.

Подавляюще часто встречается сложный эффект, а потому именно его следовало бы считать нормальным, а не аномальным.

ЗАДАЧИ

1. На сколько компонент расщепится в слабом магнитном поле линия щелочного металла с сериальным символом $^2D^2_{5/2}F_{7/2}$? Сколько из них будет π- и сколько σ-компонент?

Ответ. На 18. Из них шесть π-компонент и двенадцать σ-компонент.

2. С помощью интерферометра Фабри–Перо наблюдается зеемановское расщепление коротковолновой компоненты D-линии натрия в магнитном поле $B = 5000$ Гс. Каково должно быть расстояние d между зеркальными плоскостями интерферометра, чтобы он был пригоден для исследования расщепления рассматриваемой линии?

Решение. Как показано в тексте (см. рис. 73), рассматриваемая линия расщепляется в магнитном поле на шесть равноотстоящих компонент, расстояние между которыми составляет $\Delta_1\omega = (2/3)\,\Omega$, а между крайними $\Delta_2\omega = (10/3)\,\Omega$. Спектральный прибор для исследования расщепления должен не только разрешать расщепленные линии, но и не должен давать перекрытия порядков. Значит, область дисперсии прибора должна быть не меньше $\Delta_2\omega$. Но область дисперсии $\Delta\omega$ в частотах дается соотношением $\Delta\omega = \omega/m$, где $m = 2d/\lambda = \omega d/\pi c$ — порядок спектра. Таким образом, должно быть

$$\frac{\omega}{m} \geqslant \Delta_2\omega,$$

т. е.

$$m \leqslant \frac{\omega}{\Delta_2\omega}. \tag{41.13}$$

В результате находим

$$d \leqslant \frac{3}{5}\,\frac{\pi\mu c^2}{eB} = 0{,}64 \text{ см}.$$

Разрешающая способность прибора дается выражением $N_{\text{эф}}m$, где $N_{\text{эф}}$ — эффективное число интерферирующих пучков. Для разрешения необходимо

$$N_{\text{эф}}m \geqslant \frac{\omega}{\Delta_1\omega}, \text{ т. е. } N_{\text{эф}}m \geqslant \frac{5\omega}{\Delta_2\omega}.$$

В сочетании с (41.13) это дает $N \geqslant 5$.

§ 42. Магнитный резонанс

1. В магнитном поле каждый уровень энергии атома расщепляется на $2J + 1$ подуровней. Подуровни одного и того же уровня отличаются квантовыми числами m_J, определяющими проекции вектора $\mathbf{J}$ на направление магнитного поля. Расщепление спектральных линий, наблюдаемое в эффекте Зеемана, возникает в результате квантовых переходов между подуровнями различных расщепившихся уровней. Спонтанные переходы между подуровнями одного и того же исходного уровня маловероятны: их вероятность пропорциональна кубу расстояния между подуровнями. Кроме того, в случае одного валентного электрона такие переходы запрещены правилом отбора $\Delta L \neq 0$.

Но правила отбора относятся к радиационным переходам *изолированных атомов.* Вынужденные переходы, т. е. переходы под действием внешних силовых полей, могут происходить и тогда, когда эти правила не выполняются. Точно так же могут существенно увеличиться и вероятности соответствующих квантовых переходов, если атом поместить в надлежащее внешнее силовое поле. Именно это происходит с квантовыми переходами между подуровнями одного и того же уровня, расщепившегося в постоянном магнитном поле $\mathbf{B}$. Они начинают осуществляться с заметными скоростями, если на поле $\mathbf{B}$ наложить поперечное слабое переменное магнитное поле. Цикл явлений и методов исследования, связанный с вынужденными переходами такого рода (по причинам, которые выяснятся в ходе изложения), называется *магнитным резонансом.*

2. Строго последовательное квантовомеханическое рассмотрение магнитного резонанса в общем курсе физики провести невозможно. Оно требует исследования вопроса о вероятностях вынужденных переходов атомных систем из одного состояния в другое. Однако для понимания сущности явления целесообразно воспользоваться простыми классическими соображениями. Это тем более оправдано, что явление магнитного резонанса было предсказано и понято именно на таком классическом пути. Результаты этого предсказания оказались в согласии с опытом, и классическими моделями пользуются и поныне, чтобы составить наглядную картину явления. Последовательное квантовомеханическое истолкование появилось несколько позже.

Итак, допустим, что частица с моментом количества движения $\mathbf{J}$ и магнитным моментом $\mathfrak{m}$ помещена в однородное постоянное магнитное поле $\mathbf{B}$. Пусть эти моменты связаны соотношением $\mathfrak{m} = g\mathbf{J}$, где g — *гиромагнитное отношение.* При этом предполагается, что за единицу $\mathbf{J}$ принимается постоянная Планка $\hbar$, а за единицу магнитного момента — магнетон Бора $e\hbar/(2\mu_e c) = 9{,}274 \cdot 10^{-21}$ эрг/Гс или ядерный магнетон $e\hbar/(2\mu_p c) = 5{,}050 \cdot 10^{-24}$ эрг/Гс, в зависимости от того, обусловлен ли магнитный момент частицы электронами или атомными ядрами. Здесь μ_e и μ_p — массы электрона и протона соответственно. Поэтому ядерный магнетон в 1836 раз меньше электронного магнетона Бора. Если же гиромагнитное отношение выражать в абсолютных единицах, то его мы будем обозначать через $g_{\text{абс}}$ и писать $\mathfrak{m} = g_{\text{абс}}\mathbf{J}$. Таким образом, величина g безразмерная, тогда как $g_{\text{абс}}$ имеет размерность величины $e/\mu c$. Например, для спина электрона $g_{\text{абс}} = e/\mu_e c$, тогда как $g = 2$. Спины ядер по порядку величины такие же, как и у атомов, и выражаются целыми или полуцелыми числами, тогда как ядерные магнитные моменты в тысячи раз меньше электронных. Для частиц разного рода величина g может быть различной, но это обстоятельство сейчас не имеет значения.

На частицу в магнитном поле действует вращающий момент $[\mathfrak{m}\mathbf{B}]$, так что

$$\mathbf{J} = [\mathfrak{m}\mathbf{B}] = g_{\text{абс}}[\mathbf{J}\mathbf{B}].$$

Это — *уравнение волчка*. В установившемся состоянии момент $\mathbf{J}$, а с ним и момент $\mathfrak{m}$ будут совершать вынужденную регулярную прецессию (рис. 75) с угловой скоростью

$$\mathbf{\Omega} = -g_{\text{абс}}\mathbf{B}. \tag{42.1}$$

Наложим теперь на поле $\mathbf{B}$ перпендикулярное к нему слабое магнитное поле $\mathbf{B}'$, вращающееся вокруг $\mathbf{B}$. Тогда прецессирующая частица подвергнется действию дополнительного переменного момента сил $[\mathfrak{m}\mathbf{B}']$. Этот момент, в зависимости от его направления, будет менять угол между векторами $\mathbf{J}$ и $\mathbf{B}$. Если скорость прецессии $\mathbf{\Omega}$ и угловая скорость вращения $\mathbf{\Omega}'$ поля $\mathbf{B}'$ значительно отличаются друг от друга, то фазы этих двух вращений будут непрерывно расходиться, а вращающий момент $[\mathfrak{m}\mathbf{B}']$ будет испытывать малые периодические изменения, быстро меняющиеся по величине и направлению. Такие изменения вызовут только малые изменения угла наклона между $\mathbf{J}$ и $\mathbf{B}$, периодически и быстро меняющиеся во времени. Они будут либо немного приближать по направлению вектор $\mathbf{J}$ к вектору $\mathbf{B}$, либо немного удалять, так что в среднем направление вектора $\mathbf{J}$ останется неизменным.

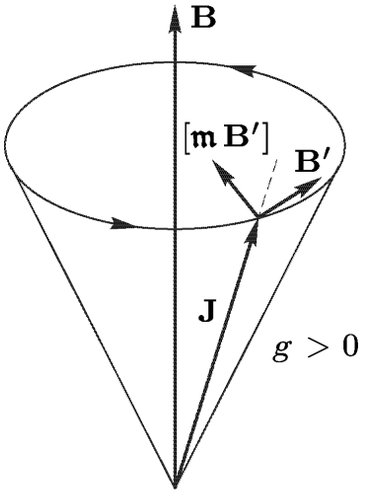


Рис. 75

Совсем иная картина будет наблюдаться, когда Ω' совпадает с Ω. В этом случае и возникает магнитный резонанс. Он состоит в том, что дополнительный момент сил действует все время в одну и ту же сторону, устанавливая вектор $\mathbf{J}$ либо параллельно, либо антипараллельно постоянному магнитному полю $\mathbf{B}$. При этом длина вектора $\mathbf{J}$ меняться не будет, но будет меняться его проекция J_z на направление поля $\mathbf{B}$.

3. Результат квантового рассмотрения отличается от классического только в том отношении, что в квантовом случае проекция J_z квантуется, и по этой причине ее изменения во времени происходят скачкообразно, тогда как в классическом случае они совершаются непрерывно. Классическое рассмотрение приводит к заключению, что резонанс должен наблюдаться только на одной частоте Ω. Если воспользоваться принципом соответствия, то это приводит к правилу отбора $\Delta m_J = \pm 1$, которое должно соблюдаться при квантовых переходах между подуровнями расщепившегося зеемановского уровня. Это означает, что разрешенными являются только переходы между соседними подуровнями. Действительно, расстояние между подуровнями составляет $\Delta\mathscr{E} = gB\Delta m_J\hbar$, так что частота перехода будет $\omega = \Delta\mathscr{E}/\hbar = gB\Delta m_J$, что совпадает с $\pm\Omega$ при условии $\Delta m_J = \pm 1$. Разумеется, это правило является также следствием сохранения момента импульса, так как вынужденный переход между подуровнями можно рассматривать как взаимодействие атомной системы с фотоном частоты Ω (см. § 39).

4. Для наблюдения магнитного резонанса удобнее пользоваться не вращающимся полем $\mathbf{B}'$, а полем, синусоидально колеблющимся в одном и том же направлении перпендикулярно к основному полю $\mathbf{B}$.

Принципиально это ничего не меняет, так как такое поле $\mathbf{B}'$ можно представить в виде суммы двух полей одинаковой напряженности, вращающихся с одной и той же угловой скоростью $\boldsymbol{\Omega}$, но в противоположных направлениях. Из них поле, вращающееся против прецессии частицы, как выяснено выше, оказывает на нее быстро осциллирующее воздействие, не играющее роли в рассматриваемом явлении. Существенно только поле, меняющее угол между $\mathbf{J}$ и $\mathbf{B}$ все время в одну и ту же сторону, т. е. поле, вращающееся в том же направлении, что и прецессирующий вектор $\mathbf{J}$. Круговая частота Ω, необходимая для получения магнитного резонанса, определяется формулой (42.1). При переходе к длинам волн из нее получаем

$$\lambda = \frac{2\pi c}{\Omega} = \frac{2\pi c}{|g_{\text{абс}}|B}. \tag{42.2}$$

Если магнитный и механический моменты частицы обусловлены электронами электронной оболочки атома, то магнитный резонанс называют *электронным парамагнитным резонансом* (ЭПР); если же атомными ядрами, то его называют *ядерным магнитным резонансом* (ЯМР). Различают также *ферромагнитный*, *антиферромагнитный*, *ферримагнитный* и *диамагнитный* (*циклотронный*) резонансы, о которых будет сказано в п. 10. Магнитный резонанс широко применяется для определения магнитных моментов атомов и атомных ядер, для изучения строения молекул и кристаллов и т. д.

Для электронов спин равен $1/2$, так что $g_{\text{абс}} = e/\mu c$. При напряженности магнитного поля $B \approx 3 \cdot 10^3$ Гс формула (42.2) в этом случае дает

$$\lambda = \frac{2\pi\mu c^2}{eB} \approx 3{,}5 \text{ см}.$$

Соответствующая частота

$$\nu = \frac{c}{\lambda} \approx 10^4 \text{ МГц}.$$

Это — частоты микроволнового диапазона ($\nu > 300$ МГц, $\lambda < 1$ м), называемые также СВЧ. Ядерный магнитный резонанс при тех же магнитных полях должен наблюдаться на волнах в тысячи раз более длинных, частоты которых порядка нескольких мегагерц.

5. Исторически сначала наблюдался ядерный магнитный резонанс с нейтральными атомными молекулярными пучками по методу, разработанному Раби (р. 1898) в 1938 г. Механический и магнитный моменты электронных оболочек атомов или молекул, из которых состоит пучок, должны быть скомпенсированы, так что оба момента являются чисто ядерными. В методе Раби атомы или молекулы пучка, пройдя через щель (рис. 76), сначала попадают в неоднородное поле магнита A с градиентом, направленным перпендикулярно к щели, например вниз. За магнитом A расположен магнит C, создающий сильное однородное постоянное поле $\mathbf{B}$, в котором расположена щель S. Затем идет магнит B, вполне аналогичный магниту A, но создающий неоднородное поле с градиентом, направленным противоположно градиенту поля

магнита A, т. е. в нашем случае вверх. Магниты A и B, создающие неоднородные поля, конструируются аналогично тому, как это делается в опытах Штерна и Герлаха (см. § 36).

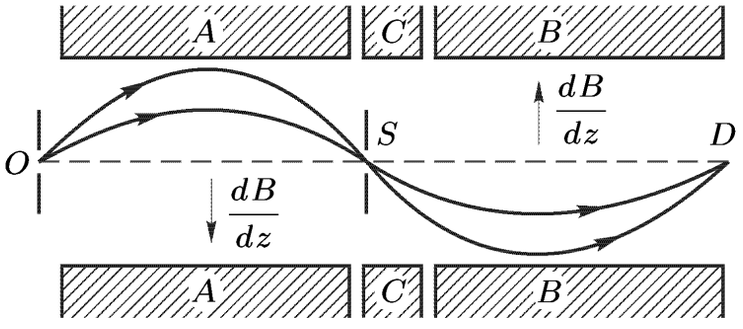


Рис. 76

Если бы магнитов A и B не было, то пучок частиц, направленный вдоль оси прибора, прошел бы через щель S и попал на детектор D. Но в неоднородных полях, благодаря наличию у частиц пучка магнитных моментов, на них действуют поперечные силы $F = \mathfrak{m}_z(dB/dz)$, направленные либо по, либо противоположно градиенту магнитного поля в зависимости от знака проекции $\mathfrak{m}_z$ на направление поля. Такие силы искривляют траектории частиц. Источник создает частицы различных скоростей и направлений. Из них представляют интерес только частицы, проходящие через щель S. Сила F действующая на частицу при ее движении по траектории, не меняется, так как при этом проекция $\mathfrak{m}_z$ остается постоянной. Поэтому в первом приближении траектории интересующих нас частиц в поле магнита A можно считать окружностями, радиус R которых определяется уравнением $v^2/R = f$, где f — нормальная сила, действующая на единицу массы частицы в неоднородном поле магнита A. Все частицы таких направлений, не задержанные краями диафрагм, пройдут через щель S.

В поле магнита B, поскольку проекции $\mathfrak{m}_z$ остаются прежними, траектории частиц будут такими же окружностями, но изогнутыми в противоположную сторону. При надлежащем поле магнита B отклонения частиц, вызванные магнитом A, компенсируются магнитом B, и частицы попадут на детектор D. Две траектории такого типа изображены на рис. 76. В этом случае детектор D зарегистрирует максимум тока частиц.

До сих пор предполагалось, что все магнитные поля постоянны. Наложим теперь на сильное постоянное поле $\mathbf{B}$ магнита C поперечное к нему слабое радиочастотное магнитное поле $\mathbf{B}'$, гармонически меняющееся во времени с частотой ω. Такое поле будет вызывать вынужденные квантовые переходы частиц, в результате которых проекции $\mathfrak{m}_z$ некоторых частиц будут изменяться, а с ними изменятся и силы, действующие на частицы в неоднородном поле магнита B. В пространстве, занятом полем магнита B, частицы будут сходить со своих прежних траекторий и перестанут попадать в детектор D.

Допустим сначала, что $\omega < \Omega$. Тогда с возрастанием ω интенсивность N пучка частиц, попадающих в детектор D, будет убывать. При

$\omega = \Omega$ наступает резонанс между колебаниями поля $\mathbf{B}'$ и ларморовской прецессией частицы вокруг сильного постоянного поля $\mathbf{B}$. Тогда интенсивность N пучка попадающих на детектор частиц обращается в минимум. При переходе через резонансную частоту Ω с возрастанием ω будет возрастать и указанная интенсивность. Из формулы (42.1) следует, что резонанс наступает при

$$\omega = |g_{\text{абс}}|B = g\frac{eB}{2\mu c}, \tag{42.3}$$

или

$$\nu = g\frac{eB}{4\pi\mu c}, \tag{42.4}$$

где под μ следует понимать массу протона (в случае ЯМР) или массу электрона (в случае ЭПР). Определив в минимуме частоту ν и напряженность поля $\mathbf{B}$, можно по формуле (42.4) вычислить и гиромагнитное отношение g для атомных ядер. Относительная точность таких измерений достигает 0,001 %. Зная спин ядра, можно найти и его магнитный момент $\mathfrak{m}$. (К этому вопросу мы вернемся ниже.)

На опыте частоту радиочастотного магнитного поля $\mathbf{B}'$ удобнее поддерживать постоянной, а напряженность сильного поля $\mathbf{B}$ плавно изменять в ту и другую сторону около некоторого среднего значения. С этой целью поле магнита C модулируют с низкой частотой (50 Гц) с помощью модулирующих катушек, питаемых от сети городского тока. Об остроте получающегося резонансного минимума можно судить по экспериментальной кривой рис. 77, полученной для ядер ^{7}Li (пучок состоял из молекул с компенсированными электронными спинами). По горизонтальной оси отложена напряженность «постоянного» поля B в гауссах, по оси ординат — относительная интенсивность N пучка частиц, регистрируемых детектором D. Частота радиочастотного поля, на которой получался резонансный минимум, равна $\nu = 5{,}585$ МГц.

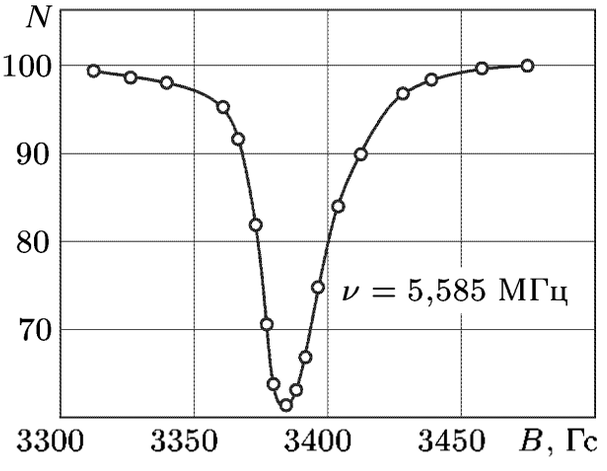


Рис. 77

6. Одним из важнейших применений магнитного резонанса с использованием атомных пучков является измерение магнитного момента электрона $\mathfrak{m}_{\text{эл}}$. Оказалось, что эта величина не совсем совпадает с магнетоном Бора. Если магнитный момент электрона представить в виде $\mathfrak{m}_{\text{эл}} = g\mathfrak{m}_{\text{Б}}/2$, то фактор $g/2$ не точно равен 1, как этого требует теория Дирака. Измерения с пучками атомов водорода в основном состоянии привели к результату

$$\frac{1}{2}\,g_{\text{эксп}} = 1{,}0011596524 \pm 20 \cdot 10^{-11}.$$

Этот результат находится в хорошем согласии с теоретическим значением

$$\frac{1}{2}\, g_{\text{теор}} = 1{,}0011596522091 \pm 31 \cdot 10^{-11}.$$

Таким образом, по квантовой электродинамике магнитный момент электрона

$$\mathfrak{m}_{\text{эл}} = (1{,}0011596522091 \pm 31 \cdot 10^{-11})\mathfrak{m}_{\text{Б}}.$$

Эта величина называется *аномальным магнитным моментом* электрона.

Метод молекулярных пучков применим только к *нейтральным* частицам, что сужает область его применимости. Действительно, на частицу с зарядом e, движущуюся со скоростью $\mathbf{v}$, действовала бы сила Лоренца $(e/c)[\mathbf{vB}]$, которая вызвала бы сильное боковое смещение ее, и притом различное для частиц различных скоростей.

Метод магнитного резонанса с использованием нейтральных молекулярных пучков отличается большой точностью. Важное достоинство этого метода состоит в том, что в нем воздействию радиочастотного поля подвергаются *свободные* частицы. Однако в экспериментальном отношении метод очень труден и требует, в частности, специальной вакуумной техники. Поэтому молекулярные пучки применяются сравнительно редко.

7. Широкое распространение получил не менее точный, но более простой в экспериментальном отношении метод магнитного резонанса, в котором используются *макроскопические количества* вещества в твердом, жидком или газообразном состоянии. Атомы, молекулы или ионы такого вещества должны обладать магнитными моментами — электронными или ядерными. При помещении исследуемого образца в сильное постоянное магнитное поле $\mathbf{B}$ в результате междуатомных столкновений спустя короткое время устанавливается состояние равновесия, в котором образец оказывается намагниченным. Это — *электронный* или *ядерный парамагнетизм*, а потому метод и получил название ЭПР или ЯМР.

Частица, магнитный момент которой ориентирован по полю, обладает меньшей энергией, чем такая же частица с магнитным моментом, ориентированным против поля. По формуле Больцмана в состоянии равновесия число первых частиц будет больше, чем вторых. Иными словами, нижние подуровни зеемановского расщепления окажутся заселенными больше, чем верхние.

Оценим разность заселенностей подуровней при комнатной температуре $T = 293$ К в предположении, что магнитный момент частицы равен одному магнетону Бора, $\mathfrak{m} = 9{,}27 \cdot 10^{-21}$ эрг/Гс, а $B = 5 \cdot 10^3$ Гс. Если n_2 — число частиц на верхнем подуровне $\mathscr{E}_2$, а n_1 — на нижнем подуровне $\mathscr{E}_1$, то по формуле Больцмана

$$\frac{n_1}{n_2} = \exp\frac{\mathscr{E}_2 - \mathscr{E}_1}{kT} \approx 1 + \frac{\mathscr{E}_2 - \mathscr{E}_1}{kT} = 1 + \frac{2\mathfrak{m}B}{kT},$$

так что

$$\frac{n_1 - n_2}{n} = \frac{2\mathfrak{m}B}{kT} = 2{,}3 \cdot 10^{-3}.$$

Значит, разность заселенностей двух соседних подуровней составляет всего около 0,2 % от заселенности одного из этих подуровней.

Несмотря на столь ничтожную величину этой разности, она может проявиться макроскопически, поскольку число частиц на каждом подуровне весьма велико. При вынужденном переходе частицы с верхнего подуровня на нижний испускается квант энергии, соответствующий разности энергий этих подуровней. При переходах с нижнего уровня на верхний расходуется энергия радиочастотного поля. Спонтанные переходы с излучением энергии из-за их относительной редкости могут не приниматься во внимание. В результате поглощение энергии будет превалировать над излучением, несмотря на то, что вероятности прямых и обратных переходов одинаковы (см. т. IV, § 119). Такой разностный эффект достигнет максимума при совпадении частоты радиочастотного поля с ларморовской частотой прецессирующей частицы, т. е. при резонансе. В максимуме поглощения энергии радиочастотного поля вблизи ларморовской частоты и проявится магнитный резонанс.

В результате преобладающих переходов частиц с нижних уровней на верхние энергии подуровней начнут выравниваться. Этому препятствуют релаксационные процессы, стремящиеся восстановить первоначальное равновесное состояние. Поэтому, чтобы магнитный резонанс был выражен достаточно резко, необходимо, чтобы период радиочастотных колебаний был мал по сравнению со временем релаксации, в течение которого восстанавливается равновесное состояние.

8. Электронный парамагнитный резонанс был открыт Е. К. Завойским (1907–1976) в 1944 г. Его первые наблюдения были произведены на солях группы железа. В дальнейшем круг изучаемых веществ значительно расширился. Завойский производил свои исследования с радиоволнами дециметрового диапазона, а потому в соответствии с формулой (42.2) пользовался магнитными полями $\mathbf{B}$ небольшой напряженности. В связи с развитием техники ультракоротких волн позднее в методе ЭПР стали применять радиоволны сантиметрового диапазона.

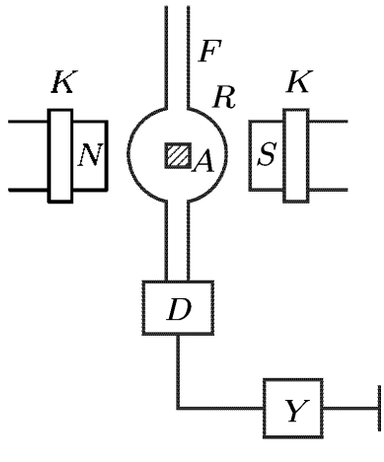


Рис. 78

В современных радиоспектроскопах частоту радиосигнала поддерживают постоянной, а магнитное поле $\mathbf{B}$ модулируют низкой частотой (50 Гц). Схема радиоспектроскопа показана на рис. 78. Электромагнит NS питается постоянным током и создает сильное постоянное магнитное поле. Это поле модулируется катушками KK, питаемыми переменным током с частотой 50 Гц. Исследуемый образец A объемом в несколько мм^3 помещают в объемный резонатор R, настроенный на длину волны $\lambda \sim 3$ см. Электромагнитные волны такой длины генерируются отражательным клистроном

и подводятся к резонатору R через волновод F. После частичного поглощения в образце A они, также через волновод, поступают к кристаллическому кремний-вольфрамовому детектору D, где детектируются и могут быть усилены. Парамагнитное поглощение может быть обнаружено, если детектор соединить с чувствительным гальванометром. Еще бóльшая чувствительность достигается, если сигнал от детектора, усиленный усилителем Y, развернуть на экране осциллографа (не указанного на рис. 78). Сигнал подается на вертикально отклоняющие пластины осциллографа. Горизонтальное отклонение фиксирует мгновенное значение магнитного поля $\mathbf{B}$. Это отклонение должно быть сфазировано с модуляцией магнитного поля $\mathbf{B}$. На экране осциллографа получается кривая парамагнитного поглощения вещества, аналогичная кривой рис. 77. Современные радиоспектроскопы при нормальной температуре позволяют обнаруживать ЭПР в образцах, содержащих до 10^{11}–10^{12} парамагнитных частиц.

9. Ядерный магнитный резонанс с поглощением коротких радиоволн макроскопическими количествами вещества впервые наблюдали в конце 1945 г. Парселл (р. 1912), Торри и Паунд (р. 1919) и независимо от них Блох (р. 1905), Хансен (1909–1949) и Пакард. Хотя в принципиальном отношении ядерный магнитный резонанс и не отличается от электронного, для его экспериментального исследования из-за большого различия в длинах волн высокочастотного электромагнитного поля требуется несколько изменить метод исследования. Идея одной из возможных экспериментальных схем понятна из схематического рис. 79. Главная особенность состоит в том, что исследуемый образец A помещается в катушку, последовательно соединенную с катушкой колебательного контура генератора высокой частоты. Электромагнит NS, питаемый постоянным током, создает сильное постоянное магнитное поле. Для создания перпендикулярного модулирующего магнитного поля применяется дополнительная катушка K, питаемая переменным током с частотой 50 Гц. При наступлении резонанса увеличивается поглощение высокочастотного электромагнитного поля. Это проявляется в уменьшении добротности колебательного контура генератора и даже в срыве генерации. Высокочастотный сигнал, появляющийся во внешней цепи, индуктивно связанной с колебательным контуром генератора, детектируется и усиливается усилителем Y. Затем получается его развертка, совершенно так же, как при исследовании ЭПР.

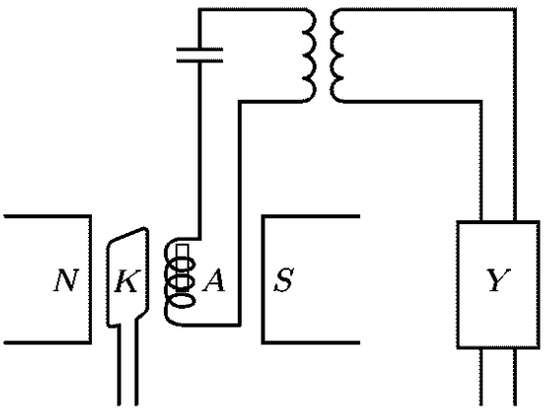


Рис. 79

10. В заключение заметим, что, помимо электронного и ядерного парамагнитных резонансов, различают также *ферромагнитный резонанс*, связанный с изменением ориентации электронных магнитных

моментов внутри доменов или между доменами ферромагнетика, а также *антиферромагнитный резонанс*, связанный с изменением ориентации спиновых магнитных моментов в антиферромагнетике. Частным случаем последнего резонанса является *ферримагнитный резонанс* — резонанс, происходящий в ферримагнетиках, т. е. антиферромагнетиках с неполностью компенсированными противоположно направленными магнитными моментами спинов решетки, а потому обладающими ферромагнитными свойствами. Ограничимся еще упоминанием, что к магнитному резонансу иногда формально причисляют *диамагнитный (циклотронный) резонанс*, имеющий, однако, совершенно иную физическую природу.

§ 43. Эффект Штарка

1. Явление Штарка (1874–1957) состоит в том, что при наложении электрического поля энергетические уровни атомов, молекул и кристаллов смещаются и расщепляются на подуровни. Это проявляется в расщеплении и смещении спектральных линий в спектрах испускания и поглощения указанных тел. Об этом явлении уже кратко говорилось в т. IV, § 93. Там указывалось, какие экспериментальные трудности возникают при наблюдении явления и как Штарку удалось их преодолеть. Штарк открыл явление, названное его именем, а затем подробно исследовал его на спектральных линиях серии Бальмера водорода. Впоследствии явление Штарка было обнаружено и на других атомах.

Уже с самого начала было выяснено, что классическая теория не в состоянии объяснить явление Штарка. Теория явления Штарка, основанная на полуклассической теории Бора, была независимо построена К. Шварцшильдом (1874–1916) и П. С. Эпштейном (1886–1966) в 1916 г. Их основные результаты были подтверждены в последовательно квантовомеханической теории, развитой Шредингером в 1926 г. В обеих теориях используются вычислительные методы теории возмущений, развитые в небесной механике Лагранжем (1736–1813), Лапласом (1749–1827) и др., а затем модернизированные применительно к задачам квантовой механики. Вычисления довольно сложны и не могут быть здесь воспроизведены. Можно ограничиться только некоторыми качественными соображениями и окончательными результатами При этом мы ограничимся штарк-эффектом только на атомах, а электрическое поле $\mathbf{E}$ будем предполагать однородным.

2. Уже из простых классических соображений легко понять, какую следует ожидать поляризацию компонент, на которые расщепляются спектральные линии при помещении источника света во внешнее электрическое поле $\mathbf{E}$. В электрическом поле частота колебаний элементарного источника света (электрона) зависит от того, совершаются ли колебания вдоль поля $\mathbf{E}$ или перпендикулярно к нему. Во всех случаях в наблюдаемом свете ввиду его поперечности возможны только колебания, перпендикулярные к линии наблюдения. Если линия наблюдения сама перпендикулярна к полю $\mathbf{E}$, то колебания, удовлетворяющие этому условию, могут происходить как по полю $\mathbf{E}$, так и перпендикулярно к нему. Они, вообще говоря, происходят с различными частотами, а потому в наблюдаемом спектре все линии окажутся поляризованными линейно: часть линий будет поляризована вдоль поля $\mathbf{E}$ (π-компоненты), а остальная часть перпендикулярно к нему (σ-компоненты).

Если же линия наблюдения направлена вдоль поля $\mathbf{E}$, то все колебания, сопровождающиеся излучением света, направлены только перпендикулярно

к $\mathbf{E}$. Поэтому в наблюдаемом спектре могут появиться только σ-компоненты. Все они будут неполяризованы, поскольку сила, действующая со стороны электрического поля $\mathbf{E}$ на колеблющийся электрон, не зависит от величины и направления скорости движения последнего. В этом существенное отличие электрического поля от магнитного. Сила, действующая на электрон со стороны магнитного поля, пропорциональна его скорости $\mathbf{v}$ и меняет свое направление на противоположное с изменением на противоположное направления скорости $\mathbf{v}$. Поэтому-то она и изменяет угловые скорости круговых вращений электрона, на которые можно разложить его колебательное движение. Это изменение зависит от направления вращения электрона, с чем и связан продольный эффект Зеемана. В случае электрического поля подобного изменения нет, а потому компоненты штарковского расщепления при продольном наблюдении оказываются *неполяризованными*. При наблюдении же под углом к полю $\mathbf{E}$ эти компоненты окажутся поляризованными *частично*.

3. Явление Штарка выглядит по-разному в зависимости от того, имеется у атома (в отсутствие электрического поля $\mathbf{E}$) дипольный электрический момент $\mathbf{p}$ или не имеется. В первом случае при наложении электрического поля $\mathbf{E}$, если ограничиться линейными по полю членами, атом получает дополнительную энергию $(-\mathbf{pE})$, пропорциональную первой степени электрического поля. Смещение и расщепление спектральных линий получатся также пропорциональными первой степени электрического поля. Такой эффект и был обнаружен Штарком.

Во втором случае у атома нет собственного электрического момента. В электрическом поле возбуждается лишь *индуцированный дипольный момент* $\mathbf{p} = \beta\mathbf{E}$, где β — поляризуемость атома, которая может быть вычислена методами квантовой механики. При увеличении электрического поля от 0 до $\mathbf{E}$ дипольный момент атома также увеличивается от 0 до $\mathbf{p}$. При этом над атомом совершается работа $(\mathbf{pE})/2 = \beta E^2/2$, которая идет на приращение потенциальной энергии атома в электрическом поле. (Коэффициент $1/2$ появляется по той же причине, что и в аналогичном случае при вычислении потенциальной энергии упруго деформированного тела, подчиняющегося закону Гука.) Смещение и расщепление спектральных линий окажутся пропорциональными E^2. Эффект Штарка в этом случае называется *квадратичным*. Он, разумеется, много меньше линейного эффекта, почему и был обнаружен позднее.

Конечно, атом с собственным дипольным моментом в электрическом поле получает и добавочный дипольный момент. В первом приближении этот добавочный момент можно считать пропорциональным полю. Тогда получится наложение *линейного* и *квадратичного* эффектов Штарка. Картина расщепления уровней окажется несимметричной: все подуровни будут смещены в сторону более низких энергий, и тем сильнее, чем выше они расположены. Сами линии окажутся смещенными в красную сторону спектра. Это смещение невелико. Например, для одной из штарковских компонент линии H_α оно составляет примерно 1 см$^{-1}$, тогда как расстояние между крайними штарковскими компонентами этой линии составляет 200 см$^{-1}$ [1]).

В полях, не превышающих 10^5 В/см, квадратичным эффектом Штарка в водороде можно полностью пренебречь. Квадратичный член $\sim E^2$ в водороде начинает сказываться только при более сильных полях. В полях, превышающих примерно $4 \cdot 10^5$ В/см, проявляется и член третьей степени

[1]) Так как $1/\lambda = \nu/c$, то 1 см$^{-1}$ $= 3 \cdot 10^{10}$ Гц $= 3 \cdot 10^4$ МГц.

$\sim E^3$, который также вычислен наряду с членом $\sim E^2$. С учетом этих членов теория хорошо согласуется с опытом в самых сильных электрических полях, вплоть до полей порядка 10^6 В/см, которых удалось достигнуть в настоящее время.

4. Причина, по которой в водороде, его изотопах (дейтерий и тритий) и водородоподобных ионах эффект Штарка *линейный*, состоит в том, что в этих случаях электрическое поле ядра, в котором движется электрон, *кулоновское*. В кулоновском же поле энергетические уровни электрона вырождены по l. Все состояния одноэлектронного атома с одним и тем же значением главного квантового числа n, отличающиеся значением l, в этом случае обладают *одной и той же энергией*. При этом состояния, суперпозицией которых получается любое состояние с заданным n, уже *в отсутствие внешнего электрического поля* обладают *собственными дипольными электрическими моментами*. При наложении внешнего электрического поля вырождение (частично) снимается, и энергетические уровни, соответствующие различным состояниям, испытывают разные смещения. Но все эти смещения и связанное с ними расщепление спектральных линий пропорциональны полю E, почему эффект Штарка и получается *линейным*.

В случае более сложных атомов и ионов с одним валентным электроном атом может рассматриваться также как одноэлектроная система. Однако в этом случае поле ядра, в котором движется электрон, искажено внутренними электронными оболочками, а потому уже не является кулоновским. В таком поле вырождения по l нет. Более подробное исследование показывает, что в каждом из состояний, характеризуемых квантовыми числами n и l, средний собственный электрический момент атома равен нулю. Поэтому при наложении поля расщепление уровней начинается с членов, квадратичных по полю E. Эффект Штарка оказывается *квадратичным*.

5. Обращаемся к рассмотрению эффекта Штарка в водороде. При этом не будем учитывать спин электрона, т. е. будем пренебрегать спин-орбитальным взаимодействием. В этом приближении задача сводится к решению уравнения Шредингера с учетом потенциальной энергии атома во внешнем электрическом поле. В этом случае задача обладает *цилиндрической симметрией*, причем ось симметрии направлена параллельно электрическому полю. Сферические координаты r, θ, φ хорошо приспособлены для решения задач в полях, обладающих сферической симметрией, но неудобны в случае цилиндрической симметрии. В этом случае более удобны так называемые *параболические координаты*, обладающие нужной симметрией. Решение уравнения Шредингера для атома водорода приводит к результату, что в постоянном электрическом поле энергетический уровень с главным квантовым числом n распадается на $2n - 1$ подуровней. Переходы между этими подуровнями, подчиняющиеся правилам отбора, и определяют компоненты, на которые расщепляются спектральные линии водорода при наложении электрического поля.

При наличии внешнего электрического поля закон сохранения момента количества движения, вообще говоря, не имеет места. Однако в постоянном однородном электрическом поле должна сохраняться *проекция момента количества движения* на направление электрического поля. Поэтому в этом случае сохраняют силу и правила отбора по магнитному квантовому числу m_l, определяющему указанную проекцию (спин s, как сказано выше, не учитывается). При $\Delta m_l = 0$ возникает π-компонента, а при $\Delta m_l = \pm 1$ σ-компоненты. Эти правила отбора и определяют возможные переходы.

Простейшей является картина расщепления водородных линий серии Лаймана. Линии этой серии получаются при переходах с вышележащих

уровней на уровень $n = 1$, который не расщепляется ($2n - 1 = 1$). Уровень $n = 2$ расщепляется на $2n - 1 = 3$ подуровня. Переходы с этих подуровней на уровень $n = 1$ дают три компоненты, на которые расщепляется линия L_α водорода. Эти переходы изображены на рис. 80. Уровни $n = 3$ и $n = 4$ расщепляются соответственно на $2n - 1 = 5$ и $2n - 1 = 7$ подуровней. При переходах с них на уровень $n = 1$ возникают компоненты, на которые расщепляются линии L_β и L_γ. Картина расщепления представлена на схематическом рис. 81. Здесь π-компоненты изображены жирными линиями, отложенными вверх, а σ-компоненты — такими же линиями, отложенными вниз. Длины этих линий показывают относительные интенсивности спектральных компонент, возникающих при наложении электрического поля. Заметим, что в случае L_β центральная компонента отсутствует, так что линия L_β расщепляется на 4 компоненты. Линия L_γ расщепляется на 7 компонент, из которых четыре являются π-, а три σ-компонентами. Приведенные теоретические результаты подтверждаются

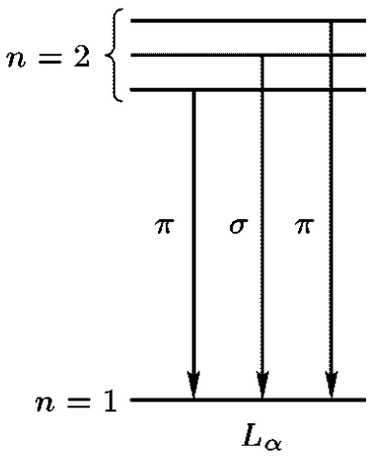


Рис. 80

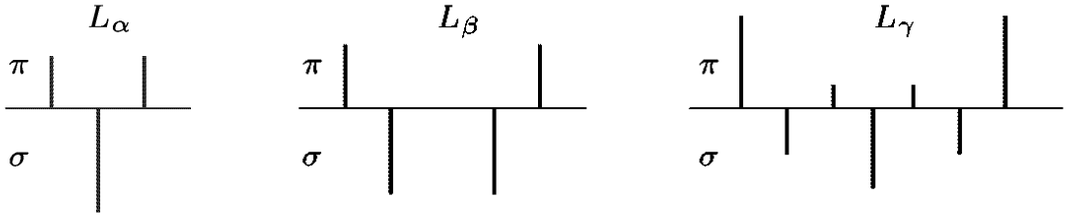


Рис. 81

опытами, которые, разумеется, должны выполняться с вакуумной спектральной аппаратурой (ультрафиолет!).

Несколько сложнее расщепляются спектральные линии серии Бальмера водорода. В этом случае переходы совершаются на три подуровня расщепившегося уровня $n = 2$. Ближайший уровень $n = 3$ расщепляется на 5 подуровней. В результате бальмеровская линия H_α, возникающая при переходах с уровня $n = 3$ на уровень $n = 2$, расщепляется на 15 компонент, как это видно из рис. 82. Линия H_β расщепляется на 20 компонент, линия H_γ — на 27, линия H_δ — на 32 и т. д. (Центральные компоненты при расщеплении H_β, H_δ не появляются, с чем и связано уменьшение числа компонент соответственно с 21 до 20 и с 33 до 32.) Расщепление линий H_α и H_β при эффекте Штарка, предсказываемое теорией (согласующейся с опытом), показано на рис. 83. Аналогично расщепляются линии H_γ, H_δ, ...

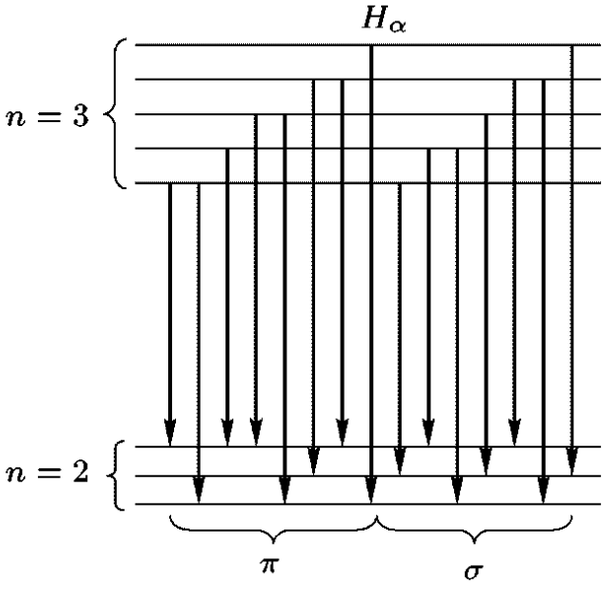


Рис. 82

6. Описанная картина штарковского расщепления получается, если *не учитывать спин электрона*, т. е. пренебречь *тонкой структурой* спектраль-

ных линий. Это можно делать, когда штарковское расщепление значительно превосходит ширину тонкой структуры спектральной линии. В полях порядка десятков тысяч В/см и выше тонкая структура практически не играет роли. Такие электрические поля (как и в случае магнитных полей в эффекте

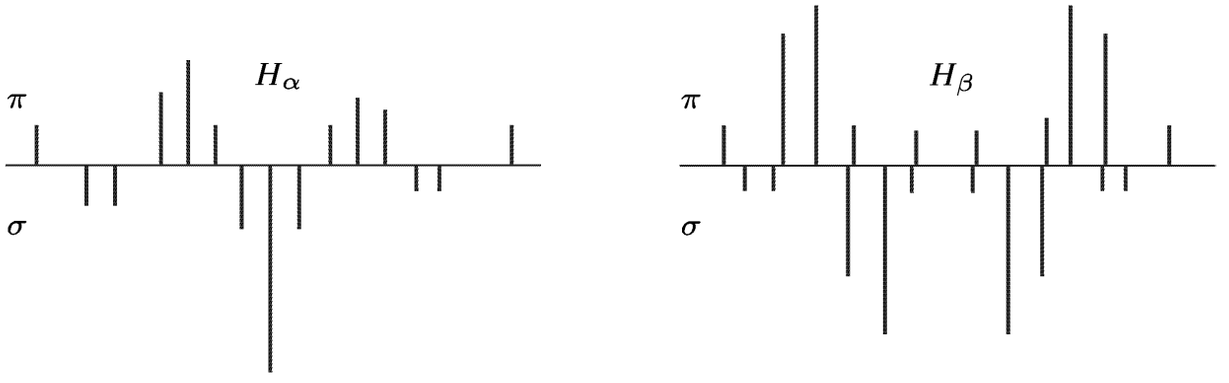


Рис. 83

Зеемана) можно назвать *сильными*. Когда же штарковское расщепление становится сравнимым или меньше ширины тонкой структуры, то электрическое поле называют *слабым*. Таким образом, приведенные выше результаты относятся к *сильным* (в указанном смысле) электрическим полям. В слабых полях эффект Штарка осложняется *тонкой структурой*.

§ 44. Лэмбовский сдвиг уровней атомных электронов

1. В § 38 указывалось, что, согласно релятивистской квантовой теории Дирака, энергии водородных и водородоподобных атомов и ионов с одинаковыми квантовыми числами n и j должны *точно совпадать* независимо от того, каково значение квантового числа l. При заданном j число l может принимать значения $j - 1/2$ и $j + 1/2$ в зависимости от того, как ориентирован спин электрона. Так, главному квантовому числу $n = 2$ (нижний терм водородной серии Бальмера) соответствуют три уровня $2s_{1/2}$, $2p_{1/2}$ и $2p_{3/2}$. Два первых уровня по теории Дирака должны совпадать, так как они имеют одно и то же квантовое число $j = 1/2$.

Действительно ли совпадают уровни $2s_{1/2}$ и $2p_{1/2}$ — этот вопрос пытались решить методами оптической спектроскопии, изучая тонкую структуру водородной линии H_α, возникающую при переходах с уровней с главным квантовым числом $n = 3$ на уровни с $n = 2$. Однако результаты получались противоречивыми. Одни исследователи находили полное согласие наблюдаемой тонкой структуры с теорией Дирака, тогда как другие приходили к результату, что уровни $2s_{1/2}$ и $2p_{1/2}$ смещены относительно друг друга приблизительно на величину порядка 0,03 $см^{-1}$ или около 1000 Гц. Эта величина примерно в 10 раз меньше расстояния между указанными уровнями и лежащим выше уровнем $2p_{3/2}$. Трудность исследования тонкой структуры состояла в исключительной близости спектральных линий, обладавших довольно большой шириной. Наблюдавшееся небольшое несовпадение уровней

$2s_{1/2}$ и $2p_{1/2}$ лежало в пределах экспериментальных ошибок. Вопрос был решен в 1947 г. Лэмбом (р. 1913) и Ризерфордом (р. 1912) *методами радиоспектроскопии.* Ожидаемые смещения уровней ~ 1000 Гц лежат в области сверхвысоких частот (СВЧ), где можно воспользоваться методами радиоспектроскопии, а последние позволяют обеспечить точность измерений до 1 Гц.

2. Идея опыта Лэмба и Ризерфорда основана на том, что возбужденный уровень $2p_{1/2}$ является нестабильным, а возбужденный уровень $2s_{1/2}$ — метастабильным. Время жизни на уровне $2s_{1/2}$ примерно в 10^8 раз больше, чем на уровне $2p_{1/2}$. Действительно, радиационный переход с уровня $2p_{1/2}$ на невозбужденный уровень $1s_{1/2}$ с испусканием одного фотона *разрешен правилом отбора* $\Delta l = \pm 1$. Переход же с уровня $2s_{1/2}$ на уровень $1s_{1/2}$ *запрещен*, поскольку при этом $\Delta l = 0$. Такой переход возможен только с испусканием *двух фотонов*, а потому он совершается примерно в 10^8 раз медленнее, чем переход $2p_{1/2} \to 1s_{1/2}$. По сравнению с переходом $2s_{1/2} \to 1s_{1/2}$ переход $2p_{1/2} \to 1s_{1/2}$ происходит примерно в 10^8 раз быстрее, т. е. практически мгновенно.

В опытах Лэмба и Ризерфорда молекулы водорода диссоциировались под действием высокой температуры в печи K (рис. 84). В результате получался пучок атомов водорода в основном состоянии $1s_{1/2}$, направлявшийся на металлическую пластинку (мишень) P, соединенную с гальванометром. Поскольку атомы пучка не были возбуждены, они не могли передавать энергию электронам мишени P. Вырывания электронов из мишени не происходило, и гальванометр не обнаруживал электрического тока.

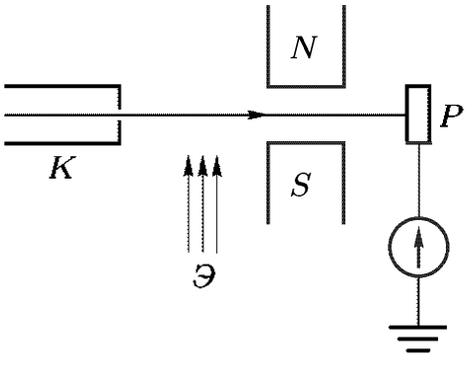


Рис. 84

Но часть атомов (примерно одну стомиллионную) можно было перевести в возбужденные состояния $2s_{1/2}$ и $2p_{1/2}$, заставив пучок атомов пересекать электронный пучок Э. Возбуждение атомов производится ударами электронов, так что правила отбора при излучении *не действуют.* Энергия возбуждения составляет 10,2 эВ. Атомы в состоянии $2p_{1/2}$ на мишень P не попадают, так как они практически мгновенно переходят в основное состояние $1s_{1/2}$. На мишень из возбужденных атомов могут попасть только атомы в состоянии $2s_{1/2}$. Энергию возбуждения они при столкновениях передают электронам мишени P, вырывая последние из мишени. Гальванометр начинает показывать ток, по величине которого можно судить о числе атомов в метастабильном состоянии $2s_{1/2}$, попадающих на мишень.

Если теперь на пути пучка атомов водорода включить циклически меняющееся магнитное поле надлежащей частоты, то в случае несовпадения уровней $2s_{1/2}$ и $2p_{1/2}$ начнутся вынужденные переходы между ними (см. § 42, п. 2). Они будут происходить с максимальной скоростью

при резонансе, когда частота внешнего магнитного поля будет соответствовать разности энергий между рассматриваемыми уровнями. Перейдя из состояния $2s_{1/2}$ в состояние $2p_{1/2}$, атом почти мгновенно возвращается в невозбужденное состояние $1s_{1/2}$. Поэтому число возбужденных атомов в состоянии $2p_{1/2}$, попадающих на мишень, начинает убывать. Вместе с тем начинает убывать и ток через гальванометр. По минимуму тока можно определить резонансную частоту поля, а с ней и разность уровней $2s_{1/2} - 2p_{1/2}$ [1]).

Опыты Лэмба и Ризерфорда доказали, что уровни $2s_{1/2}$ и $2p_{1/2}$ не совпадают между собой. Разность между ними по частотам составляет для водорода $1057{,}90 \pm 0{,}06$ МГц. Эта разность (как и вообще разность между уровнями тонкой структуры с одинаковыми n и j, но различными l) называется *лэмбовским сдвигом*. Взаимное расположение уровней $2s_{1/2}$, $2p_{1/2}$ и $2p_{3/2}$ с указанием расстояний между ними для водорода приведено на рис. 85. Лэмбовский сдвиг очень мал — он примерно в 10 раз меньше расстояния между уровнями тонкой структуры $2p_{3/2}$ и $2s_{1/2}$. Тем не менее это очень важное явление, теоретически истолкованное в квантовой электродинамике. Последняя дает для лэмбовского сдвига в случае водорода $1057{,}91 \pm 0{,}01$ МГц, что находится в превосходном согласии с экспериментальным значением, приведенным выше. Столь же превосходное согласие получается для дейтерия и однократно ионизованных атомов гелия.

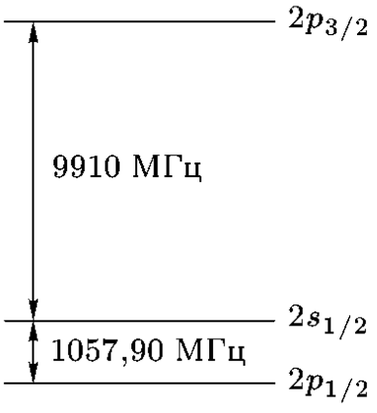


Рис. 85

3. Заметим, что лэмбовский сдвиг позволяет судить о точности, с которой оправдывается закон Кулона на расстояниях порядка атомных (10^{-8} см). Предположим, что истинный закон имеет вид $E \sim 1/r^{2\pm\gamma}$. Предположим далее, что $|\gamma| = 10^{-9}$. Тогда вычисления привели бы к изменению лэмбовского сдвига $2s_{1/2} - 2p_{1/2}$ на величину, бóльшую существующей ошибки измерения. На основании этого можно заключить, что на атомных расстояниях $|\gamma|$ не может быть больше 10^{-9} (ср. т. III, § 6, п. 4). На самом деле отступления от закона Кулона при взаимодействии электрона с ядром начинаются тогда, когда включаются ядерные силы. Для случая же, когда у взаимодействующих частиц ядерные взаимодействия отсутствуют (электроны и позитроны), закон Кулона в настоящее время проверен на ускорителях со встречными пучками до расстояний $\sim 10^{-16}$ см.

[1]) Уровень $2p_{1/2}$ оказался лежащим ниже уровня $2s_{1/2}$. Поэтому возможны переходы $2s_{1/2} \to 2p_{1/2}$ и в отсутствие внешнего поля. Однако так как расщепление мало, а вероятность перехода пропорциональна кубу расстояния между уровнями, то эта вероятность крайне мала (соответствующее время жизни — порядка нескольких лет). Поэтому указанный переход можно не принимать во внимание.

§ 45. Физический вакуум и объяснение лэмбовского сдвига

1. Лэмбовский сдвиг был объяснен и рассчитан Бете (р. 1908) в квантовой электродинамике. Поскольку последняя нами не излагалась, в нашем курсе можно дать только качественное представление о теории лэмбовского сдвига.

Предварительно необходимо остановиться на вопросе о *физическом вакууме* и *виртуальных частицах* — важнейших понятиях квантовой теории поля. Согласно квантовой теории поля вакуум не есть абсолютная «пустота», в которой ничего нет. Вакууму присущи многие физические свойства, и он может находиться в различных физических состояниях. Поэтому-то он и получил почетное название «физического вакуума». В отличие от гипотетического эфира XIX века, которому приписывались механические свойства, принципиально не отличающиеся от механических свойств обычных материальных сред, современная физика пытается устанавливать свойства физического вакуума только на основе твердо установленных экспериментальных фактов и проверенных опытом физических теорий. Никаких механических свойств физическому вакууму не приписывается.

Строго говоря, следует различать не один вакуум, а несколько, в зависимости от того, с какими частицами и полями он связан. Так, электромагнитное поле, или поле фотонов, может отдавать свою энергию квантами величиной $h\nu$. При каждой такой отдаче число фотонов уменьшается на единицу. В результате последовательности таких процессов в конце концов возникает состояние, в котором число квантов в системе равно нулю. Однако, в отличие от классических представлений, электромагнитное поле при этом не исчезает, а переходит в состояние с наименьшей энергией, отнять которую от поля уже нельзя. Это заключение является следствием существования нулевой энергии, т. е. в конце концов принципа неопределенностей. Состояние электромагнитного поля с наименьшей возможной энергией, в котором фотонов нет, называется *вакуумным состоянием электромагнитного поля* или *фотонным вакуумом*. Электромагнитное поле в вакуумном состоянии не может быть поставщиком энергии, но из этого не следует, что оно вообще никак не может проявить себя. Оно может быть причиной различных наблюдаемых физических явлений, как об этом уже упоминалось выше и подробнее будет сказано дальше.

Аналогично и для других частиц вводится представление о вакууме как о низшем энергетическом состоянии поля соответствующих частиц. Так, различают *электронно-позитронный вакуум*, где такими частицами являются электрон и позитрон, отличающиеся друг от друга только знаками электрических зарядов. Существует *вакуум π-мезонов* и т. д. При рассмотрении взаимодействия полей вакуумом можно назвать низшее энергетическое состояние всей системы этих полей.

Если полю, находящемуся в вакуумном состоянии, сообщить достаточную энергию, то происходит его возбуждение, т. е. рождение частицы — кванта этого поля. Так, в вакууме происходит, например, рождение электронно-позитронных пар. Рождение частицы можно описать как переход из «ненаблюдаемого» вакуумного состояния в состояние реальное.

2. Согласно современным представлениям взаимодействие между частицами осуществляется посредством обмена какими-то другими частицами, которым соответствует свой вакуум. Так, электромагнитное взаимодействие между электрически заряженными частицами осуществляется посредством электромагнитного вакуума. Один электрический заряд испускает фотон, поглощаемый другим зарядом, который в свою очередь испускает фотон, поглощаемый первым зарядом. Таким образом, между зарядами происходит обмен фотонами. В результате этого изменяется нулевое (т. е. невозбужденное) состояние вакуума, что по современным представлениям проявляется в силе взаимодействия между частицами, обменивающимися фотонами. Аналогично, до недавнего времени считалось, что взаимодействие между нуклонами (протонами и нейтронами) осуществляется посредством вакуума π-мезонов. Один из нуклонов испускает π-мезон, а другой его поглощает, и наоборот. Со времени принятия кварковой модели ядерные силы сводят к взаимодействию кварков.

Изложенные представления о взаимодействии на первый взгляд находятся в противоречии с тем, что утверждалось в п. 5 § 1. Будем для конкретности иметь в виду взаимодействие электрических зарядов, хотя все сказанное ниже будет относиться и ко всем другим типам взаимодействий. Пока взаимодействие не началось, каждая из частиц является свободной, а свободная частица ни испустить, ни поглотить квант не может. В противном случае в процессах испускания и поглощения кванта нарушался бы либо закон сохранения энергии, либо закон сохранения импульса. Однако приведенное утверждение относится к испусканию истинных частиц. Взаимодействие же осуществляется не реальными, а *виртуальными частицами*, и приведенное выше противоречие устраняется. Поэтому в предыдущем абзаце точнее надо было бы везде говорить не просто о частицах, обменом которых осуществляется взаимодействие, а о виртуальных частицах.

Виртуальные частицы существуют только в промежуточных состояниях и притом очень короткое время, что препятствует экспериментальной регистрации их. Но время существования виртуальной частицы Δt связано с неопределенностью ее энергии соотношением $\Delta t\ \Delta\mathscr{E} \geqslant \hbar$. Аналогично $\Delta x \Delta p \geqslant \hbar$. При наличии этих соотношений неопределенности законы сохранения энергии уже не препятствуют более испусканию квантов истинными свободными частицами, если только эти кванты имеют энергию $\Delta\mathscr{E}$ и существуют в течение промежутка времени $\Delta t \sim \hbar/\Delta\mathscr{E}$. Можно сказать, что виртуальная частица есть такая частица, для которой не выполняется обычная классическая связь между энергией и импульсом $\mathscr{E}^2 = (pc)^2 + m_0^2c^4$.

3. За время Δt естественно принять продолжительность акта обмена между взаимодействующими частицами. Если принять далее, что взаимодействие распространяется с максимальной скоростью c, то $\Delta t = L/c$, где L — радиус переносимого взаимодействия. Поэтому энергия кванта $\Delta\mathscr{E} = m_0c^2 = \hbar c/L$, где m_0 — масса покоя кванта. Отсюда

$$L = \frac{\hbar}{m_0 c}. \tag{45.1}$$

Чем больше масса m_0 кванта, переносящего взаимодействие, тем короче радиус действия соответствующих сил. Значение этого факта мы рассмотрим ниже. Отметим только, что для фотона $m_0 = 0$. Поэтому радиус действия кулоновских сил бесконечен.

Таким образом, принцип неопределенности вынуждает принять, что каждая заряженная частица окружена облаком испускаемых и поглощаемых виртуальных фотонов. Другие частицы окружены также соответствующими квантами, переносящими взаимодействие. Такая картина соответствует воззрениям Фарадея и Максвелла о недопустимости непосредственного действия на расстоянии. Для взаимодействия необходимо наличие промежуточного агента. Однако конкретное представление, что таким агентом является классический непрерывный эфир, безнадежно устарело и может быть сохранено только ради исторического интереса.

4. Вакуум является суперпозицией нулевых колебаний поля, т. е. состояний с виртуально возникающими и исчезающими виртуальными фотонами, виртуальными электронно-позитронными парами, а также другими парами частиц и античастиц. Эти виртуальные частицы взаимодействуют между собой и с истинными частицами. Так, возникший виртуальный фотон может породить электронно-позитронную пару. При аннигиляции последней возникнут новые виртуальные фотоны, и т. п. Во внешнем электрическом поле, например в поле атомного ядра, виртуальные электроны и позитроны располагаются неравномерно. Виртуальные позитроны смещаются преимущественно в направлении электрического поля, виртуальные электроны — против электрического поля. В результате во внешнем электрическом поле возникает явление, называемое *поляризацией вакуума*. Это явление аналогично обычной поляризации диэлектриков во внешнем поле. Только в диэлектриках речь идет о смещении истинных электрических зарядов, а в вакууме — о смещении виртуальных.

5. Теперь можно обратиться к объяснению лэмбовского сдвига. Основной вклад ($\sim \alpha^3 R$, где α — постоянная тонкой структуры, R — постоянная Ридберга) вносят два вакуумных эффекта (называемых иначе *радиационными поправками*). Во-первых, испускание и поглощение связанным электроном виртуальных фотонов, что приводит у него к изменению эффективной массы и появлению аномального магнитного момента. Во-вторых, поляризация вакуума, т. е. рождение и аннигиляция в вакууме виртуальных электронно-позитронных пар, что искажает кулоновский потенциал ядра на расстояниях порядка

комптоновской длины волны электрона ($\sim 10^{-11}$ см). Поскольку комптоновская длина волны электрона много меньше среднего радиуса боровских орбит в водороде, лэмбовский сдвиг в водороде вызывается преимущественно первой причиной (изменением эффективной массы электрона). Поляризация вакуума приводит к одинаковому для всех уровней сдвигу. Вклад поляризации вакуума в величину лэмбовского смещения уровней незначителен (в водороде $\approx 3\,\%$ общего смещения для основного уровня).

По теории, согласующейся с опытом, лэмбовский сдвиг уровня пропорционален четвертой степени атомного номера и обратно пропорционален третьей степени главного квантового числа.

Следующее простое рассуждение поясняет влияние главного квантового числа на сдвиг уровней. Образно говоря, s-электрон проводит основную часть времени вблизи ядра, где электрическое поле сильное и резко неоднородное, а p-электрон в среднем находится на более далеких расстояниях, где поле слабее и менее неоднородно. С другой стороны, взаимодействие с вакуумом (испускание и поглощение виртуальных фотонов) как бы раскачивает (трясет) электрон. На классическом языке орбита электрона не плавная кривая (например, круговая), а извилистая. Электрон то удаляется от ядра, то приближается к нему, и притом хаотически. Потенциальная энергия в поле ядра $U \sim 1/r$. При увеличении r на Δr энергия U меняется на величину

$$\Delta U \sim \frac{1}{r+\Delta r} - \frac{1}{r} = -\frac{\Delta r}{r(r+\Delta r)},$$

при уменьшении r — на величину

$$\Delta U' \sim \frac{\Delta r}{r(r-\Delta r)},$$

что по абсолютной величине больше ΔU. Это значит, что вакуумное дрожание электрона меняет знак изменения его потенциальной энергии U. Особенно велико изменение потенциальной энергии вблизи ядра, где U велика и резко меняется с расстоянием. Таким образом, вакуумные добавки к полной энергии $\mathscr{E}$ больше для s-электронов, чем для p-электронов. Это в основном и раздвигает энергии s- и p-электронов, которые без этого совпадали бы (если не учитывать сдвиг уровней, вызываемый поляризацией вакуума).

Глава VI

АТОМНЫЕ СИСТЕМЫ СО МНОГИМИ ЭЛЕКТРОНАМИ

§ 46. Принцип тождественности одинаковых частиц. Принцип Паули

1. В отличие от макроскопических тел, однотипные частицы микромира (электроны, протоны, нейтроны, все так называемые элементарные частицы, атомы и пр.) обладают *совершенно одинаковыми* свойствами: у них одинаковы масса, электрический заряд, спин и пр. В связи с этим возникает вопрос, как отличить одну частицу от другой такой же частицы. Возьмем в качестве примера систему, состоящую из двух электронов. В некоторый момент времени (принимаемый за начальный) отметим положение обоих электронов, снабдив один из них номером 1, а другой номером 2. С классической точки зрения электрон движется по определенной траектории, так что принципиально возможно проследить за движением каждого из рассматриваемых электронов. Обнаружив электрон в какой-то последующий момент времени, можно в принципе сказать, будет ли это электрон 1 или электрон 2. Поменяв местами и скоростями оба электрона, мы получим *новое состояние* системы. Во всех отношениях оно обладает теми же свойствами, что и исходное состояние, но отличается от него нумерацией электронов. С изложенной точки зрения одинаковые частицы *принципиально различимы* или *индивидуализированы.*

Совсем иначе обстоит дело с точки зрения квантовой механики, отвергающей классическое представление о движении частицы по траектории. Состояние системы частиц описывается в квантовой механике волновой функцией, которой дается вероятностное истолкование. В нашем примере она является функцией времени и координат обоих электронов. Обнаружив в какой-то момент времени один из электронов, *принципиально невозможно* решить, будет ли это электрон 1 или электрон 2. Невозможность принципиального решения того или иного вопроса означает, что самый вопрос поставлен *неправильно.* К таким искусственно поставленным вопросам с точки зрения квантовой механики относится и разбираемый нами вопрос о различении двух одинаковых частиц. Если две одинаковые частицы поменять местами, то результат такого обмена *никак нельзя обнаружить экспериментально.* От всякой теории естественно требовать, чтобы два состояния, принципиально неразличимые никакими опытами, она рассматривала как *одно и то же состояние.* Именно так поступает квантовая механика, которая принимает, что при перестановке двух одинаковых частиц не возникает нового состояния системы: оно остается абсолютно тем же, что и до

перестановки. Одинаковые частицы *принципиально неразличимы* или *обезличены*. Можно говорить о состоянии системы одинаковых частиц *только в целом*, а не о состоянии каждой частицы в отдельности. Это положение можно формулировать в виде *принципа тождественности одинаковых частиц*:

в системе одинаковых частиц реализуются только такие состояния, которые не меняются при перестановке местами двух любых частиц.

Этот принцип в квантовой механике является существенно новым, т. е. логически не вытекает из остальных основных положений ее, но можно доказать, что он не противоречит им. Его следует принять, поскольку он подтверждается всей совокупностью опытных фактов.

2. Состояние системы частиц в квантовой механике характеризуется волновой функцией. Какие же волновые функции допустимы, т. е. удовлетворяют принципу тождественности одинаковых частиц? Для решения этого вопроса достаточно ограничиться рассмотрением системы из двух одинаковых частиц. От зависимости волновой функции от времени t можно отвлечься и писать ее в виде $\psi(q_1, q_2)$. В случае бесспиновых частиц под q_1 следует понимать совокупность трех пространственных координат первой частицы, а под q_2 второй. Если же частица обладает спином, то к тройке пространственных координат следует еще добавить *спиновые координаты*, которые могут принимать дискретный ряд значений. Например, в случае электронов проекция спина может принимать два значения: $+1/2$ и $-1/2$. Эти значения в рассматриваемом случае и являются спиновыми координатами.

Переставив теперь местами частицы 1 и 2, получим функцию $\psi(q_2, q_1)$. Эту операцию можно рассматривать как действие на функцию $\psi(q_1, q_2)$ линейного оператора $\widehat{P}$, называемого *оператором перестановки*:

$$\psi(q_2, q_1) = \widehat{P}\psi(q_1, q_2).$$

Переставив рассматриваемые частицы вторично, получим исходную функцию

$$\psi(q_1, q_2) = \widehat{P}\psi(q_2, q_1) = \widehat{P}^2\psi(q_1, q_2).$$

Отсюда $\widehat{P}^2 = 1$, а потому $\widehat{P} = \pm 1$. Значит, допустимы волновые функции двух типов:

$$\psi_s(q_1, q_2) = \psi_s(q_2, q_1) \tag{46.1}$$

и

$$\psi_a(q_1, q_2) = -\psi_a(q_2, q_1) \tag{46.2}$$

В первом случае волновая функция при перестановке частиц не изменяется. Она называется *симметричной*, а потому и снабжается индексом s. Во втором случае функция называется *антисимметричной* и снабжается индексом a. Антисимметричная функция при перестановке одинаковых частиц меняет знак.

Полученные результаты обобщаются и для систем, состоящих из какого угодно числа одинаковых частиц. В этом случае, как показывает опыт, симметрия или антисимметрия волновой функции имеет место при перестановке *любых двух* одинаковых частиц.

3. Частицы, состояние которых описывается симметричными волновыми функциями, называются *бозе-частицами* или *бозонами.* Частицы же, описываемые антисимметричными волновыми функциями, называются *ферми-частицами* или *фермионами.* Такие названия приняты потому, что системы, состоящие из бозонов, подчиняются *статистике Бозе–Эйнштейна,* а состоящие из фермионов — *статистике Ферми–Дирака* (см. т. II, § 82). К бозонам относятся фотоны, π- и K-мезоны — вообще все частицы с нулевым или целым спином. К фермионам относятся электроны, протоны, нейтроны, нейтрино и все элементарные частицы и античастицы с полуцелым спином. Связь между спином и статистикой, установленная сначала эмпирически для фотонов и электронов, была в 1940 г. распространена теоретически Паули на все элементарные частицы и античастицы. Паули установил эту связь, исходя из самых общих принципов квантовой теории — релятивистской инвариантности, неотрицательности полной энергии, причинности и т. п.

Указанная связь между спином и статистикой справедлива и для сложных частиц, построенных из элементарных, т. е. для атомных ядер, атомов и молекул, если только рассматриваются явления при достаточно низких энергиях, в которых каждая из сложных частиц ведет себя как целое. Принадлежность сложной частицы к бозонам или фермионам определяется ее спином.

Например, атом водорода состоит из двух ферми-частиц: протона и электрона, спин каждой из которых равен $1/2$. Суммарный механический момент, т. е. спин атома водорода в нормальном состоянии, может быть равен либо 0 (спины протона и электрона антипараллельны), либо 1 (спины параллельны). В обоих случаях атом водорода в нормальном состоянии будет бозоном.

Приведем второй пример. Ядро атома гелия $^4_2\mathrm{He}$, т. е. α-частица, состоит из двух протонов и двух нейтронов. Спин этого ядра равен нулю, т. е. оно является бозоном. Бозоном будет и сам атом $^4_2\mathrm{He}$ в нормальном состоянии. Но ядро $^3_2\mathrm{He}$ состоит из нечетного числа (3) частиц со спинами $1/2$: двух протонов и одного нейтрона. Спин этого ядра нечетный, т. е. оно является фермионом. Фермионом будет и сам атом $^3_2\mathrm{He}$. Ядра и атомы $^4_2\mathrm{He}$ подчиняются статистике Бозе–Эйнштейна, ядра и атомы $^{03}_2\mathrm{He}$ — статистике Ферми–Дирака. Это проявляется в том, что вблизи абсолютного нуля температуры $^4_2\mathrm{He}$ обладает сверхтекучестью, в то время как $^3_2\mathrm{He}$ не обладает (уточнение дается в § 61, п. 7).

4. Применим принципы симметрии и антисимметрии к системе одинаковых частиц, не взаимодействующих между собой. При этом сначала будем рассуждать так, как если бы эти частицы не обладали спинами (уравнения для волновых функций при наличии спина в этом

курсе не рассматриваются), а затем обобщим полученные результаты на случай наличия спина.

Начнем с простейшего случая, когда система состоит всего из двух одинаковых частиц, декартовы координаты которых обозначим соответственно через x_1, y_1, z_1 и x_2, y_2, z_2. Если $\widehat{H}$ — оператор Гамильтона всей системы, а $\widehat{H}_1$ и $\widehat{H}_2$ — каждой из частиц, то $\widehat{H} = \widehat{H}_1 + \widehat{H}_2$. Уравнение Шредингера для системы частиц в стационарных состояниях будет

$$\widehat{H}\psi \equiv (\widehat{H}_1 + \widehat{H}_2)\psi = \mathscr{E}\psi, \tag{46.3}$$

где оператор $\widehat{H}_1$, поскольку частицы не взаимодействуют между собой, зависит только от координат первой частицы, а $\widehat{H}_2$ — только от координат второй частицы, а именно:

$$\begin{aligned} \widehat{H}_1 &= -\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial y_1^2} + \frac{\partial^2}{\partial z_1^2}\right) + U(x_1, y_1, z_1), \\ \widehat{H}_2 &= -\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial y_2^2} + \frac{\partial^2}{\partial z_2^2}\right) + U(x_2, y_2, z_2). \end{aligned} \tag{46.4}$$

Оба оператора $\widehat{H}_1$ и $\widehat{H}_2$ совершенно одинаковы, так как одинаковы сами частицы, но эти операторы зависят от разных координат.

Для решения уравнения (46.3) применим тот же метод, какой был использован в § 26. Так как оператор $\widehat{H}_1$ действует только на координаты x_1, y_1, z_1, а оператор $\widehat{H}_2$ — только на координаты x_2, y_2, z_2, то уравнение (46.3) распадается на два:

$$\widehat{H}_1\psi = \mathscr{E}_1\psi, \quad \widehat{H}_2\psi = \mathscr{E}_2\psi, \tag{46.5}$$

где $\mathscr{E}_1$ и $\mathscr{E}_2$ — постоянные, удовлетворяющие условию $\mathscr{E}_1 + \mathscr{E}_2 = \mathscr{E}$. Решение первого уравнения (46.5) имеет вид $\varphi_\beta\varphi_\alpha(1)$, где для краткости совокупность координат x_1, y_1, z_1 первой частицы обозначена цифрой 1. Аналогично, совокупность координат x_2, y_2, z_2 второй частицы будем обозначать цифрой 2. В общем случае коэффициент φ_β может зависеть от координат 2, т. е. $\varphi_\beta = \varphi_\beta(2)$. Функцию $\varphi_\beta(2)$ следует выбрать так, чтобы удовлетворялось и второе уравнение (46.5). Для этого должно быть

$$\widehat{H}_1\varphi_\alpha(1) = \mathscr{E}_1\varphi_\alpha(1) \quad \text{и} \quad \widehat{H}_2\varphi_\beta(2) = \mathscr{E}_2\varphi_\beta(2). \tag{46.6}$$

Таким образом, функция $\varphi_\alpha(1)$ описывает состояние первой частицы с энергией $\mathscr{E}_1$, а функция $\varphi_\beta(2)$ — состояние второй частицы с энергией $\mathscr{E}_2$. Соотношение $\mathscr{E} = \mathscr{E}_1 + \mathscr{E}_2$ означает, что энергия системы двух невзаимодействующих частиц равна сумме энергий этих частиц, как этого и следовало ожидать. Решение уравнения (46.3) принимает вид $\psi = \varphi_\alpha(1)\varphi_\beta(2)$, т. е. оно является, как принято говорить, решением с *разделяющимися переменными.*

Уравнения (46.6) по существу одинаковы. Они отличаются одно от другого только обозначениями независимых переменных и значениями

постоянных $\mathscr{E}_1$ и $\mathscr{E}_2$. Конечно, собственному значению $\mathscr{E}_1$ могут соответствовать несколько независимых решений первого уравнения (46.6). В общем случае под $\varphi_\alpha(1)$ следует понимать их линейную суперпозицию с постоянными коэффициентами, которая также является решением уравнения (46.3). То же относится к функции $\varphi_\beta(2)$.

Прежде чем идти дальше, учтем наличие спина. Для этого достаточно понимать под 1 и 2 совокупность не только пространственных, но и *спиновых координат* частиц. По-прежнему находится функция $\varphi_\alpha(1)\varphi_\beta(2)$, являющаяся решением уравнения, заменяющего уравнение Шредингера при наличии спина (такое уравнение называется уравнением Паули). В силу тождественности частиц функция $\varphi_\alpha(2)\varphi_\beta(1)$ является решением того же уравнения. Она получается из функции $\varphi_\alpha(1)\varphi_\beta(2)$, если поменять местами частицы 1 и 2, т. е. произвести перестановку их пространственных и спиновых координат.

Однако ни одна из этих функций не удовлетворяет принципу симметрии или антисимметрии. Но из них можно составить линейные комбинации с постоянными коэффициентами, которые также являются решениями уравнения типа Шредингера. Среди этих комбинаций есть *симметричная функция*

$$\psi_s = \varphi_\alpha(1)\varphi_\beta(2) + \varphi_\alpha(2)\varphi_\beta(1) \tag{46.7}$$

и *антисимметричная*

$$\psi_a = \varphi_\alpha(1)\varphi_\beta(2) - \varphi_\alpha(2)\varphi_\beta(1). \tag{46.8}$$

Состояние, описываемое функцией ψ_s, может действительно реализоваться в природе в случае системы двух одинаковых бозонов, а состояние, описываемое функцией ψ_a, — в случае системы двух одинаковых фермионов.

5. Приведенное рассуждение без всяких затруднений распространяется на случай системы, состоящей из произвольного числа N тождественных частиц. Прежде всего получается волновая функция

$$\psi = \psi_\alpha(1)\psi_\beta(2)\psi_\gamma(3)\ldots\psi_\omega(N)$$

с разделяющимися переменными, в которой сохранены прежние обозначения. Отсюда, последовательно меняя местами каждые две частицы, получаем новые волновые функции с разделяющимися переменными:

$$\psi' = \psi_\alpha(2)\psi_\beta(1)\psi_\gamma(3)\ldots\psi_\omega(N),$$

$$\psi'' = \psi_\alpha(3)\psi_\beta(2)\psi_\gamma(1)\ldots\psi_\omega(N)$$

и т. д., которым соответствует то же значение параметра $\mathscr{E}$. Состояние с таким значением $\mathscr{E}$ N-кратно вырождено. Однако ни одно из этих состояний не реализуется в природе.

В случае бозонов реализуется только их симметричная комбинация

$$\psi_s = \sum \widehat{P}\psi, \tag{46.9}$$

где $\widehat{P}$ — оператор перестановки координат двух частиц, причем сумма распространяется на всевозможные перестановки. К числу таких перестановок относится и тождественная перестановка, не меняющая координаты всех частиц, при которой функция ψ не изменяется. Таким образом, сумма (46.9) содержит N слагаемых.

В случае фермионов реализуется антисимметричная комбинация

$$\psi_a = \sum \pm \widehat{P}\psi, \tag{46.10}$$

причем знаку плюс соответствует четное число транспозиций (т. е. перестановок двух частиц), а знаку минус — нечетное. Таким образом, антисимметричная функция может быть записана в форме определителя

$$\psi_a = \begin{vmatrix} \psi_\alpha(1) & \psi_\beta(1) & \dots & \psi_\omega(1) \\ \psi_\alpha(2) & \psi_\beta(2) & \dots & \psi_\omega(2) \\ \dots\dots\dots & \dots\dots\dots & \dots & \dots\dots\dots \\ \psi_\alpha(N) & \psi_\beta(N) & \dots & \psi_\omega(N) \end{vmatrix}. \tag{46.11}$$

Обычно волновые функции (46.10) и (46.11) нормируют, умножая их на определенные числовые коэффициенты. Однако в разбираемом нами вопросе нормировка не имеет значения и поэтому не производится.

6. В случае невзаимодействующих одинаковых частиц имеет смысл говорить о состоянии не только системы в целом, но и о *состоянии одной частицы*. Например, можно сказать, что одна какая-то частица находится в состоянии ψ_α, а другая в состоянии ψ_β. При этом волновые функции фермионов ведут себя существенно иначе, чем волновые функции бозонов.

В системе одинаковых фермионов не может быть двух частиц, находящихся в одном и том же состоянии. Для частного случая системы, состоящей из двух частиц, это непосредственно видно из формулы (46.8), так как в случае $\psi_\alpha \equiv \psi_\beta$ волновая функция ψ_a обращается в нуль, что физически не соответствует никакому состоянию. В общем случае N частиц это видно из формулы (46.10) и в особенности из формулы (46.11), так как, например, при $\psi_\alpha \equiv \psi_\beta$ получился бы определитель с двумя одинаковыми столбцами, а такой определитель обращается в нуль.

Таким образом, *в системе тождественных фермионов не может быть двух частиц, находящихся в одном и том же состоянии.* Это положение называется *принципом* или *запретом Паули*, который высказал его для электронов в атоме еще до создания квантовой механики. В первоначальной формулировке принцип Паули утверждал, что *в атоме не может быть двух электронов, характеризующихся одинаковыми четверками квантовых чисел.*

В смысле ясности и точности принцип Паули уступает *принципу антисимметрии волновых функций.* Последний принцип справедлив и при наличии взаимодействия частиц, тогда как в принципе Паули речь идет о состояниях отдельных частиц, о которых, строго говоря,

можно говорить лишь в отсутствие взаимодействия. Тем не менее принцип Паули, даже в первоначальной формулировке самого Паули, оказался очень плодотворным, например, для обоснования периодической системы Менделеева, а также некоторых закономерностей в спектрах. Принцип антисимметрии волновых функций фермионов иногда называют *обобщенным принципом Паули.*

Что касается бозонов, то на их состояния *принцип симметрии волновых функций* не накладывает никаких ограничений, аналогичных запрету Паули. В одном и том же состоянии может находиться *любое число* одинаковых бозонов. Это непосредственно видно из выражений (46.7) и (46.9) для волновых функций таких бозонов.

§ 47. Объяснение периодической системы химических элементов Д. И. Менделеева

1. Периодическая повторяемость свойств химических элементов, открытая Д. И. Менделеевым (1834–1907) в 1869 г. и отраженная в его *периодической системе*, была объяснена Бором в 1922 г. на основе созданной им теории строения атома. Оказалось, что не атомная масса (атомный вес по старой терминологии), а *заряд ядра* лежит в основе систематики химических элементов. Если за единицу принять элементарный заряд e, то заряд ядра будет выражаться целым числом, которое принято обозначать через Z. Число Z и определяет номер химического элемента в периодической системе. В этой связи его называют также *порядковым номером* элемента. Заряд ядра численно равен числу электронов в электронной оболочке, окружающей ядро нейтрального атома. Свойства же элемента зависят прежде всего от числа электронов в электронной оболочке и от ее строения. Что касается *химических свойств* элемента, то они определяются *наружными электронами* электронной оболочки.

Существенное усовершенствование объяснение периодической системы получило в 1925 г. после открытия *принципа Паули.* Этот принцип определил максимальное число электронов в электронной оболочке атома, которые могут находиться в определенном квантовом состоянии. После этого стало понятным распределение химических элементов по группам и периодам периодической системы, эмпирически угаданное Менделеевым и пополненное затем другими химиками.

В естественном состоянии в природе встречается 90 элементов. Наибольший атомный номер (92) имеет уран. Элементы $_{43}\mathrm{Tc}$ (технеций) и $_{61}\mathrm{Pm}$ (прометий) в естественных состояниях не встречаются. Они радиоактивны, причем период их полураспада (у наиболее важного изотопа $^{99}\mathrm{Tc}$ $2{,}12\cdot10^5$ лет, у изотопов Pm — от нескольких лет до десятков лет) много меньше возраста Земли. Со времени образования Земли эти элементы целиком распались. Образоваться же при радиоактивных превращениях соседних по заряду ядер (из числа встречающихся в природе) они не могли, так как все эти соседние ядра стабильны. Элементы с атомным номером больше 92 называются *трансурановыми.* Они все

радиоактивны и также получены искусственно. К настоящему времени получены элементы с порядковыми номерами до 109 включительно.

Способы получения трансурановых элементов будут кратко описаны ниже. Здесь же заметим, что в 1983 г. поступило сообщение о получении учеными ФРГ (Дармштадт) 109-го элемента в количестве одного атома. В 1984 г. в СССР (Дубна) было получено восемь атомов такого элемента. В том же году там же были синтезированы два изотопа 108-го элемента. Названия элементов 102 (нобелий) и 103 (лоуренсий) не являются общепринятыми, так как первоначальные сообщения об их открытии с присвоением им указанных названий не подтвердились. Ученые Дубны предложили 102-й элемент называть жолиотием (Jl) в честь Жолио-Кюри, а 103-й элемент — резерфордием (Rf) в честь Резерфорда. 105-й элемент ими же предложено называть нильсборием (Ns) в честь Нильса Бора. Остальные трансурановые элементы еще не получили названий. Их по примеру Д. И. Менделеева можно было бы назвать экавольфрам, экарений, экаосьмий, экаиридий. Приставка «эка» означает, что в периодической системе названные элементы лежат соответственно под вольфрамом, рением, осмием, иридием.

2. Периодическая повторяемость свойств химических элементов является внешним проявлением внутренней структуры электронных оболочек атомов. К выяснению указанной структуры и сводится построение теории периодической системы элементов.

В существующей ныне теории периодической системы характеризуются состояния *отдельных электронов* электронной оболочки атома, а не ее состояние в целом. Это, конечно, недостаток теории. Состояние электронной оболочки должно было бы характеризоваться волновой функцией всех координат ее электронов, причем при наличии взаимодействия между электронами эти координаты не разделяются, т. е. волновая функция не представляется в виде произведения волновых функций отдельных электронов. Это и значит, что точный смысл имеет состояние всей электронной оболочки атома, а не отдельных электронов, из которых она состоит. Но точное нахождение волновой функции системы многих взаимодействующих частиц еще более недоступно методам современной математики, чем решение знаменитой задачи многих (в частности, трех) тел в классической механике. Поэтому-то и приходится вводить приближенное представление о состояниях отдельных электронов.

Состояние отдельного электрона характеризуется четырьмя квантовыми числами, за которые удобно принять n, l, m_l, m_s. Принцип Паули, на котором основана теория периодической системы, тогда будет означать, что *в одном квантовом состоянии может находиться не более одного электрона.* Индивидуализация электронов при этом не используется — существенно только, сколько электронов находится в заданном состоянии (n, l, m_l, m_s), а не какие это электроны. Если два электрона обменяются своими координатами, то состояние электронной оболочки останется *абсолютно тем же.* Спиновое квантовое число m_s можно исключить из рассмотрения, если учесть, что оно может принимать только два значения: $m_s = \pm 1/2$. Если это сделать, то можно

сказать, что *в квантовом состоянии* (n,l,m_l) *может находиться не более двух электронов.*

3. Совокупность электронов атома с заданным значением главного квантового числа n образует *электронный слой* или просто *слой.* В соответствии с терминологией, принятой в рентгеновской спектроскопии, слои обозначаются прописными буквами латинского алфавита, а именно:

n	1	2	3	4	5
Слой	K	L	M	N	O
Максимальное число электронов в слое	2	8	18	32	50

Совокупность электронов с заданными значениями n и l образует *оболочку* [1]. Различные состояния в оболочке отличаются значениями квантового числа m_l. Поскольку последнее может принимать значения $m_l = -l, -(l-1), \ldots, 0, \ldots, +(l-1), +l$, в оболочке с квантовым числом l может находиться не более $2(2l+1)$ электронов Оболочки обозначаются строчными буквами латинского алфавита, а именно:

l	0	1	2	3	4
Оболочка	s	p	d	f	g
Максимальное число электронов в оболочке	2	6	10	14	18

Символ, указывающий слои, оболочки и числа электронов в каждой оболочке, называется *электронной конфигурацией* атома. Например, $1s^2 2s^2 2p^6$ означает, что в состоянии $n=1$, $l=0$ находятся два электрона, в состоянии $n=2$, $l=0$ — также два электрона, а в состоянии $n=2$, $l=1$ — шесть электронов. Из дальнейшего будет видно, что это — электронная конфигурация неона. Максимальное число электронов в слое с заданным значением n получится суммированием чисел электронов во всех оболочках с допустимыми значениями l, а именно:

$$\sum_{l=0}^{n-1} 2(2l+1) = 2n^2. \tag{47.1}$$

Соответствующие числа для различных n были приведены выше.

4. Посмотрим теперь, как меняются электронные конфигурации при переходе от одного атома к другому в порядке возрастания их атомных номеров Z. При возрастании Z на единицу увеличивается на единицу заряд ядра, а к электронной оболочке атома добавляется один электрон. Принцип, определяющий при этом изменение электронной конфигурации, состоит в том, что вновь получаемая конфигурация из $Z+1$ электронов должна обладать *наименьшей энергией* из всех

[1] Термин «оболочка», как это делалось выше, применяется также в смысле совокупности *всех электронов,* окружающих атомное ядро.

возможных значений ее, допускаемых квантовой механикой. Иначе говоря, энергия связи вновь присоединяемого электрона в атоме должна быть *максимально возможной*. Однако фактическое применение этого принципа требовало бы решения квантовомеханической задачи многих тел, что практически невыполнимо. Поэтому приходится пользоваться эмпирическими данными, в частности химическими и в особенности данными спектроскопии об ионизационных потенциалах атомов. Это придает теории периодической системы полуэмпирический, описательный характер. Лучше было бы говорить не о теории, а об *объяснении* периодической системы, что и отражено в заглавии настоящего параграфа.

Казалось бы, что слои должны заполняться последовательно друг за другом, а в пределах каждого слоя сначала должна заполняться целиком s-оболочка, затем также целиком p-, d- и f-оболочки. В действительности такой «идеальный» порядок заполнения не согласуется с только что сформулированным принципом. Нарушение этого порядка объясняется главным образом наличием у электронов орбитальных моментов количества движения $L = \sqrt{l(l+1)}\hbar$. Энергия связи электрона зависит не только от его потенциальной энергии в электрическом поле ядра и окружающей его электронной оболочки, но также от «центробежной энергии»

$$\frac{L^2}{2m_e r^2} = \frac{l(l+1)\hbar^2}{2m_e r^2},$$

где r — расстояние электрона от ядра. «Центробежная» и электрическая энергии имеют противоположные знаки и поэтому действуют в разные стороны. Образно говоря, центробежная сила, возникающая при орбитальном вращении электрона, стремится удалить электрон от ядра. По этой причине каждый из 10 электронов $3d$-оболочки обладает меньшей энергией связи, чем каждый из двух электронов $4s$-оболочки. Поэтому-то оболочка $4s$ заполняется электронами раньше, чем оболочка $3d$, хотя главное квантовое число во втором случае меньше, чем в первом. Особенно велика «центробежная энергия» в случае d- и f-оболочек, поскольку для этих оболочек соответственно $l(l+1) = 6$ и $l(l+1) = 12$. С наличием этих оболочек и связаны отступления от того «идеального» порядка заполнения, о котором говорилось выше.

В действительности слои и оболочки, если отвлечься от некоторых деталей, указанных в табл. 4 (см. ниже), заполняются в следующем порядке:

$1s^2$	2 электрона,
$2s^2 2p^6$	8 электронов,
$3s^2 3p^6$	8 электронов,
$4s^2 3d^{10} 4p^6$	18 электронов,
$5s^2 4d^{10} 5p^6$	18 электронов,
$6s^2 4f^{14} 5d^{10} 6p^6$	32 электрона,
$7s^2 5f^{14} 6d^{10} 7p^6$	32 электрона.

5. Слой $n = 1$ состоит всего из одной s-оболочки ($l = 0$). В водороде на этой оболочке (т. е. в состоянии $1s$) находится лишь один электрон. В атоме гелия к нему присоединяется второй электрон в том же состоянии $1s$. Средняя энергия связи одного электрона в атоме гелия приблизительно в два раза больше, чем средняя энергия связи электрона в атоме водорода. Это объясняется тем, что заряд ядра гелия вдвое больше заряда ядра водорода, и по этой причине электрон в нормальном состоянии гелия находится на меньшем расстоянии от ядра, чем в атоме водорода. Наличие же второго электрона уменьшает энергию связи первого. Два элемента — водород и гелий — образуют *первый период* периодической системы.

Присоединим теперь к атому третий электрон, увеличив одновременно на единицу заряд ядра. Третий электрон не может находиться в слое K ($n = 1$), так как этот слой уже целиком заполнен. Он начинает заполнять s-оболочку слоя L ($n = 2$), попадая в состояние $2s$. Получается щелочной металл ${}_3\mathrm{Li}$. Четвертый электрон также попадает в состояние $2s$ — получается элемент бериллий ${}_4\mathrm{Be}$. Пятый электрон уже не может быть присоединен к оболочке $2s$, так как она заполнена целиком. Поэтому с бора ${}_5\mathrm{B}$ начинает заполняться $2p$-оболочка. В результате последовательно получаются элементы ${}_6\mathrm{C}$, ${}_7\mathrm{N}$, ${}_8\mathrm{O}$, ${}_9\mathrm{F}$. Построение оболочки $2p$ заканчивается благородным газом ${}_{10}\mathrm{Ne}$. Так образуется *второй* (*короткий*) *период* периодической системы, состоящий из восьми элементов.

Затем, начиная с щелочного элемента ${}_{11}\mathrm{Na}$, идет заполнение слоя M ($n = 3$). Однако после заполнения s- и p-оболочек оно заканчивается благородным газом ${}_{18}\mathrm{Ar}$. Получается *третий* (опять *короткий*) *период* периодической системы, состоящий также из восьми элементов. С этого места появляются нарушения «идеального» порядка заполнения слоев и оболочек. Именно, сначала заполняется $4s$-оболочка и только после этого начинается заполнение пропущенной $3d$-оболочки, да и то при ее заполнении встречаются различные нерегулярности.

6. Порядок заполнения слоев и оболочек указан в табл. 4. В ней приведены электронные конфигурации для каждого элемента. Все элементы разбиты на две половины. Первая половина напечатана слева, а вторая смещена относительно нее вправо. Так же принято поступать в химии с символами тех же элементов при печатании периодической системы, с целью расположить друг под другом элементы со сходными химическими свойствами. В элементах первой половины заполняются s- и p-оболочки. Это заполнение происходит всегда закономерно. Сначала целиком заполняется s-оболочка, а затем также целиком p-оболочка того же слоя. К электронной конфигурации каждого элемента последовательно прибавляется по одному s- или p-электрону — тогда получается электронная конфигурация стоящего после него элемента.

В элементах второй половины идет заполнение d- и f-оболочек. При этом и происходит нарушение «идеального» порядка заполнения, о котором говорилось выше. Так, в *четвертом периоде* за электронной

Таблица 4

Электронные конфигурации элементов

Период	Элемент			Электронная конфигурация	Ионизационный потенциал, В
1	1	H	водород	$1s^1$	13,539
	2	He	гелий	$1s^2$	24,45
2				Конфигурация гелия +	
	3	Li	литий	$2s^1$	5,37
	4	Be	бериллий	$2s^2$	9,48
	5	B	бор	$2s^2 2p^1$	8,4
	6	C	углерод	$2s^2 2p^2$	11,217
	7	N	азот	$2s^2 2p^3$	14,47
	8	O	кислород	$2s^2 2p^4$	13,56
	9	F	фтор	$2s^2 2p^5$	18,6
	10	Ne	неон	$2s^2 2p^6$	21,48
3				Конфигурация неона +	
	11	Na	натрий	$3s^1$	2,12
	12	Mg	магний	$3s^2$	7,61
	13	Al	алюминий	$3s^2 3p^1$	5,96
	14	Si	кремний	$3s^2 3p^2$	7,39
	15	P	фосфор	$3s^2 3p^3$	10,3
	16	S	сера	$3s^2 3p^4$	10,31
	17	Cl	хлор	$3s^2 3p^5$	12,96
	18	Ar	аргон	$3s^2 3p^6$	15,69
4				Конфигурация аргона +	
	19	K	калий	$4s^1$	4,32
	20	Ca	кальций	$4s^2$	6,09
	21	Sc	скандий	$4s^2 3d^1$	6,57
	22	Ti	титан	$4s^2 3d^2$	6,80
	23	V	ванадий	$4s^2 3d^3$	6,76
	24	Cr	хром	$4s^1 3d^5$	6,74
	25	Mn	марганец	$4s^2 3d^5$	7,40
	26	Fe	железо	$4s^2 3d^6$	7,83
	27	Co	кобальт	$4s^2 3d^7$	7,81
	28	Ni	никель	$4s^2 3d^8$	7,61
	29	Cu	медь	$4s^1 3d^{10}$	7,62
	30	Zn	цинк	$4s^2 3d^{10}$	9,35
	31	Ga	галлий	$4s^2 4p^1 3d^{10}$	5,27
	32	Ge	германий	$4s^2 4p^2 3d^{10}$	7,85
	33	As	мышьяк	$4s^2 4p^3 3d^{10}$	9,4
	34	Se	селен	$4s^2 4p^4 3d^{10}$	
	35	Br	бром	$4s^2 4p^5 3d^{10}$	11,8
	36	Kr	криптон	$4s^2 4p^6 3d^{10}$	13,94

Переходные элементы

Продолжение табл. 4

Период	Элемент			Электронная конфигурация	Ионизационный потенциал, В
5				Конфигурация криптона +	
	37	Rb	рубидий	$5s^1$	4,16
	38	Sr	стронций	$5s^2$	5,67
	39	Y	иттрий	$5s^2 4d^1$	6,5
	40	Zr	цирконий	$5s^2 4d^2$	
	41	Nb	ниобий	$5s^1 4d^4$	
	42	Mo	молибден	$5s^1 4d^5$	7,65
	43	Tc	технеций	$5s^1 4d^6$	
Переходные элементы	44	Ru	рутений	$5s^1 4d^7$	7,7
	45	Rh	родий	$5s^1 4d^8$	7,7
	46	Pd	палладий	$4d^{10}$	8,5
	47	Ag	серебро	$5s^1 4d^{10}$	7,54
	48	Cd	кадмий	$5s^2 4d^{10}$	8,95
	49	In	индий	$5s^2 5p^1 4d^{10}$	5,76
	50	Sn	олово	$5s^2 5p^2 4d^{10}$	7,37
	51	Sb	сурьма	$5s^2 5p^3 4d^{10}$	8,5
	52	Te	теллур	$5s^2 5p^4 4d^{10}$	
	53	I	иод	$5s^2 5p^5 4d^{10}$	10,44
	54	Xe	ксенон	$5s^2 5p^6 4d^{10}$	12,08
6				Конфигурация ксенона +	
	55	Cs	цезий	$6s^1$	3,88
	56	Ba	барий	$6s^2$	5,19
	57	La	лантан	$6s^2 5d^1$	
Лантаниды	58	Ce	церий	$6s^2 4f^2$	
	59	Pr	празеодим	$6s^2 4f^3$	
	60	Nd	неодим	$6s^2 4f^4$	
	61	Pm	прометий	$6s^2 4f^5$	
	62	Sm	самарий	$6s^2 4f^6$	
	63	Eu	европий	$6s^2 4f^7$	
	64	Gd	гадолиний	$6s^2 5d^1 4f^7$	
	65	Tb	тербий	$6s^2 5d^1 4f^8$	
	66	Dy	диспрозий	$6s^2 4f^{10}$	
	67	Ho	гольмий	$6s^2 4f^{11}$	
	68	Er	эрбий	$6s^2 4f^{12}$	
	69	Tm	тулий	$6s^2 4f^{13}$	
	70	Yb	иттербий	$6s^2 4f^{14}$	
	71	Lu	лютеций	$6s^2 5d^1 4f^{14}$	
	72	Hf	гафний	$6s^2 5d^2 4f^{14}$	

Продолжение табл. 4

Период	Элемент			Электронная конфигурация	Ионизационный потенциал, В
6				Конфигурация ксенона +	
	Переходные элементы	73	Ta тантал	$6s^2 5d^3 4f^{14}$	
		74	W вольфрам	$6s^2 5d^4 4f^{14}$	
		75	Re рений	$6s^2 5d^5 4f^{14}$	
		76	Os осмий	$6s^1 5d^7 4f^{14}$	
		77	Ir иридий	$6s^2 5d^7 4f^{14}$	
		78	Pt платина	$6s^1 5d^9 4f^{14}$	
		79	Au золото	$6s^1 5d^{10} 4f^{14}$	
		80	Hg ртуть	$6s^2 5d^{10} 4f^{14}$	9,20
	81	Tl	таллий	$6s^2 6p^1 5d^{10} 4f^{14}$	10,59
	82	Pb	свинец	$6s^2 6p^2 5d^{10} 4f^{14}$	7,39
	83	Bi	висмут	$6s^2 6p^3 5d^{10} 4f^{14}$	8,0
	84	Po	полоний	$6s^2 6p^4 5d^{10} 4f^{14}$	
	85	At	астат	$6s^2 6p^5 5d^{10} 4f^{14}$	
	86	Rn	радон	$6s^2 6p^6 5d^{10} 4f^{14}$	10,69
7				Конфигурация радона +	
	87	Fr	франций	$7s^1$	
	88	Ra	радий	$7s^2$	
	Актиниды	89	Ac актиний	$7s^2 6d^1$	
		90	Th торий	$7s^2 6d^2$	
		91	Pa протактиний	$7s^2 6d^1 5f^2$	
		92	U уран	$7s^2 6d^1 5f^3$	
		93	Np нептуний	$7s^2 6d^1 5f^4$	
		94	Pu плутоний	$7s^2 5f^6$	
		95	Am америций	$7s^2 5f^7$	
		96	Cm кюрий	$7s^2 6d^1 5f^7$	
		97	Bk берклий	$7s^2 6d^1 5f^8$	
		98	Cf калифорний	$7s^2 5f^{10}$	
		99	Es эйнштейний	$7s^2 5f^{11}$	
		100	Fm фермий	$7s^2 5f^{12}$	
		101	Md менделевий	$7s^2 5f^{13}$	
		102	(No нобелий)	$7s^2 5f^{14}$	
		103	Lr лоуренсий	$7s^2 6d^1 5f^{14}$	
		104	(Ku курчатовий)	$7s^2 6d^2 5f^{14}$	
		105	(Ns нильсборий)	$7s^2 6d^3 5f^{14}$	
		106		$7s^2 6d^4 5f^{14}$	
		107		$7s^2 6d^5 5f^{14}$	

конфигурацией ванадия $4s^2 3d^3$ следует электронная конфигурация хрома $4s^1 3d^5$ (конфигурацию аргона для краткости опускаем), т. е. в $3d$-оболочке число электронов увеличивается сразу на два, причем один электрон заимствуется из $4s$-оболочки. В конфигурации следующего элемента марганца $4s^2 3d^5$ новый электрон присоединяется не к $3d$-, а к $4s$-оболочке, восстанавливая в ней прежнее число электронов. Аналогичные нарушения идеального порядка, как легко проследить по табл. 4, происходят и при заполнении остальных d-, а также f-оболочек.

Каждый период периодической системы начинается со щелочного металла, в электронной конфигурации которого имеется только один наружный s-электрон. Из всех элементов атомы щелочных металлов обладают наименьшими ионизационными потенциалами. Поэтому эти атомы легко могут отдавать наружные s-электроны и присоединять их к атомам других элементов. Этим объясняется большая химическая активность щелочных металлов.

Каждый период заканчивается атомом благородных газов He, Ne, Ar, Kr, Xe, Rn. У этих атомов (за исключением He) наружная $s-p$-оболочка состоит из восьми электронов, образующих особенно компактную, симметричную и прочно связанную систему. Поэтому у атомов благородных газов ионизационные потенциалы максимальны. Этим и объясняется их химическая пассивность — они не вступают (или практически не вступают) в химические соединения с другими атомами. Наоборот, элементы соседней седьмой группы периодической системы — фтор, хлор, бром, иод — химически очень активны. Это потому, что у них наружная $s-p$-оболочка состоит из семи электронов и может быть легко дополнена до замкнутой оболочки путем присоединения недостающего восьмого электрона.

Каждый полупериод периодической системы заканчивается так называемыми *переходными элементами*, обведенными в табл. 4 штриховыми рамками. Это триады: (железо, кобальт, никель), (рутений, родий, палладий), (осмий, иридий, платина), помещаемые обычно в одну клетку периодической системы.

Особого внимания заслуживают 14 элементов от церия до лютеция включительно, называемые *редкими землями* или *лантанидами*. В них происходит заполнение внутренней $4f$-оболочки. Поскольку при этом наружные оболочки практически остаются без изменения, все элементы редких земель обладают весьма близкими химическими свойствами, и поэтому их трудно отделить один от другого. Лантаниды в табл. 4 обведены сплошной рамкой.

Представляет интерес история открытия элемента 72 (гафния). Этот элемент до 1922 г. не был известен, но место для него ошибочно оставляли среди редких земель. Однако Бор указал, что по теоретическим соображениям элементы редких земель должны заканчиваться 71-м элементом (лютецием), а элемент 72 по химическим свойствам должен быть аналогом циркония ($Z = 40$). После этого указания его стали искать и действительно обнаружили в циркониевых рудах.

Подобно лантанидам ведут себя *актиниды*. Так называют 14 элементов, большинство из которых получено искусственно. Это — элементы от тория до лоуренсия включительно. В табл. 4 они также обведены сплошной рамкой. В них происходит заполнение внутренней $5f$-оболочки, в то время как наружные оболочки практически остаются незатронутыми. Поэтому, как и лантаниды, актиниды обладают весьма близкими химическими свойствами.

Мы оборвали периодическую систему на элементе 109, искусственно полученном к настоящему времени. Это, конечно, не означает, что в дальнейшем не могут быть получены элементы с бóльшими атомными номерами.

7. Успехи теории периодической системы химических элементов несомненны. Однако необходимо иметь в виду и принципиальные недостатки существующей теории. Один из них был уже указан в п. 2. Добавим к нему следующее. Электроны в каждой оболочке атома характеризуются орбитальными квантовыми числами l. Это предполагает сохранение числового значения орбитального момента количества движения для каждого электрона. Но закон сохранения момента количества движения справедлив для частицы, движущейся в *центрально-симметричном* силовом поле, и не имеет места в других случаях. Поле же, в котором движется электрон в атоме, создается ядром и остальными электронами. Оно не обладает сферической симметрией. Даже в случае одного наружного электрона, когда все внутренние оболочки заполнены, сферическая симметрия получается в результате квантово-механического усреднения. В уравнение же Шредингера, определяющее движение электронов, входит не усредненная потенциальная энергия, а потенциальная энергия, понимаемая классически, как функция координат всех (точечных) электронов (см. § 21, п. 3).

§ 48. Рентгеновские лучи

1. Рентгеновские лучи в рентгеновских трубках получаются при торможении ускоренных электронов на аноде (в старых трубках —на антикатоде; см. т. III, § 117). Волновые свойства рентгеновских лучей были установлены в 1912 г. Лауэ и его сотрудниками Фридрихом (1883–1968) и Книппингом (1883–1935), осуществившими дифракцию этих лучей в кристаллах (см. т. IV, § 61). Еще раньше, в 1905 г., Баркла (1877–1944) установил, что если рентгеновские лучи действительно являются волнами, то эти волны должны быть *поперечными*. В опыте Баркла пучок рентгеновских лучей S (рис. 86) рассеивался на теле A. Наличие рассеяния обнаруживалось с помощью ионизационной камеры — рентгеновские лучи, попадая в камеру, производили ионизацию воздуха, делая его проводящим. Рассеянное под углом 90° излучение попадало на тело B и вновь рассеивалось. Оказалось,

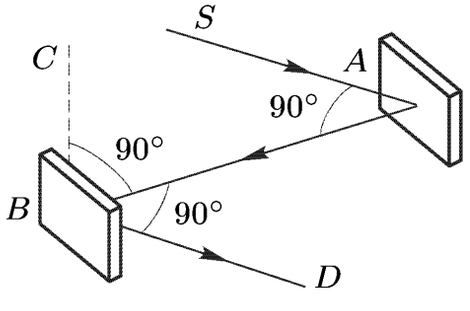


Рис. 86

что интенсивность вновь рассеянного излучения в направлении BC, перпендикулярном к плоскости SAB, равна нулю, тогда как в других направлениях, в частности в направлении BD, антипараллельном SA, рассеяние происходило. На основании этого Баркла и заключил, что рентгеновские волны, если рентгеновское излучение действительно состоит из них, должны быть поперечными.

В самом деле, если волны поперечны, то электрический вектор в падающем луче SA и возбужденные им колебания электронов в теле A можно разложить по направлениям AB и BC. Первые колебания, в направлении AB, излучения не дадут, так как колеблющийся электрон не излучает в направлении его колебаний. К телу B дойдет только волна с электрическим вектором, параллельным BC. Она возбудит колебания электронов в том же направлении. Следовательно, излучения в направлении BC не возникнет.

Разумеется, рентгеновские лучи в опытах Баркла должны были обладать достаточной жесткостью, чтобы на своем пути в воздухе не испытывать заметного поглощения. Для этого напряжение на трубке должно быть достаточно высоким. По той же причине тела A и B делались из материала с малым атомным номером Z (уголь в опытах Баркла). Иначе при высоких напряжениях на трубке в случае больших Z получилось бы собственное характеристическое излучение (см. ниже) и притом значительной интенсивности. Уголь, парафин и пр. дают слабое и мягкое собственное излучение, поглощающееся в воздухе на пути уже в несколько сантиметров, а потому не мешающее постановке опыта.

2. Баркла же установил, что рентгеновское излучение, исходящее из антикатода трубки, состоит из двух частей. Одна из них есть *тормозное излучение*, возникающее при торможении электронов в антикатоде. Его свойства совершенно не зависят от материала антикатода. Свойства другого существенно определяются материалом, из которого состоит антикатод. Рентгеновское излучение Баркла характеризовал его *жесткостью*, т. е. способностью проходить через различные вещества. Точная количественная характеристика стала возможной только после опыта Лауэ, упоминавшегося выше, когда были разработаны методы рентгеновской спектроскопии, позволившие измерять длины волн рентгеновского излучения (см. т. IV, § 61).

Тормозное излучение дает *сплошной спектр*. По аналогии с белым светом его называют также *белым рентгеновским излучением*. В сторону длинных волн интенсивность тормозного излучения спадает полого и асимптотически стремится к нулю. Со стороны коротких волн сплошной спектр обрывается резко (см. рис. 5). Как уже подробно указывалось в § 2, такая особенность сплошного рентгеновского излучения объясняется его *квантовой природой*. Коротковолновая граница сплошного рентгеновского спектра определяется формулой (2.6). Она совершенно не зависит от материала анода (антикатода), а определяется только напряжением на трубке. Вообще, относительное распределение энергии по длинам волн в спектре тормозного рентгеновского

излучения от материала анода не зависит. Последний влияет только на интегральную интенсивность излучения.

Характеристическое излучение, напротив, имеет *линейчатый спектр*, т. е. состоит из закономерно расположенных достаточно узких спектральных линий. Их длины волн зависят исключительно от материала анода. В этом отношении характеристическое излучение напоминает линейчатый спектр газов в оптической области. Электронная бомбардировка возбуждает как сплошное, так и характеристическое излучение, тогда как бомбардировка α-частицами или протонами возбуждает только характеристическое излучение. Рентгеновские лучи, испускаемые веществом при действии других рентгеновских лучей, состоят частично из рассеянного первичного пучка, частично из характеристического излучения самого вещества. Характеристическое излучение появляется только после того, как напряжение на трубке начинает превосходить некоторое определенное значение, зависящее лишь от материала анода.

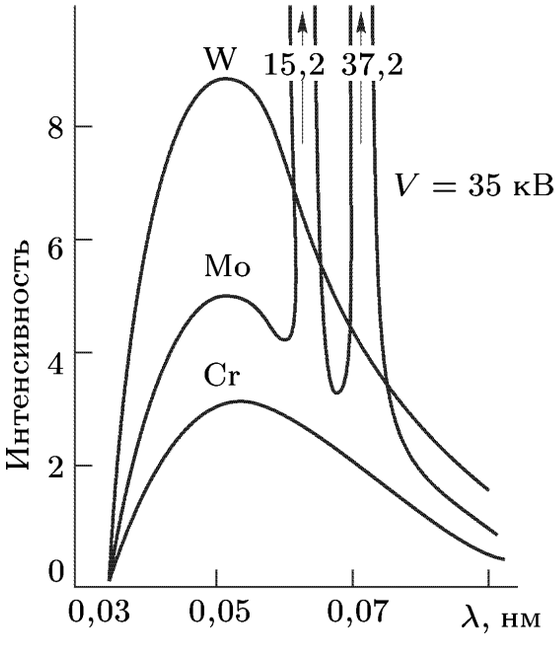


Рис. 87

На рис. 87 для напряжения на трубке $V = 35$ кВ приведены экспериментальные кривые спектрального распределения энергии рентгеновского излучения в случае антикатодов из вольфрама, молибдена и хрома. За меру интенсивности излучения принят ток в ионизационной камере. В случае молибдена, наряду со сплошным излучением, возбуждаются две линии характеристического излучения с длинами волн примерно $\lambda = 0{,}063$ и $\lambda = 0{,}071$ нм. Характеристическое излучение появляется и для хрома, но при более длинных волнах, не укладывающихся на рисунке. Но в случае вольфрама напряжения 35 кВ недостаточно для возбуждения характеристического излучения в диапазоне волн, приведенном на рисунке, — получается только сплошной спектр.

3. Каждая линия рентгеновского характеристического излучения, конечно, возникает в результате перехода атома с одного энергетического уровня на другой. Ее частота определяется правилом частот Бора:

$$\hbar\omega = \mathscr{E}_1 - \mathscr{E}_2. \tag{48.1}$$

Поскольку рентгеновское излучение коротковолновое, разность энергий в начальном и конечном состояниях атома очень велика и в случае тяжелых элементов превосходит в тысячи, десятки и сотни тысяч раз соответствующую разность в оптической области спектра. Это указывает на то, что квантовые переходы совершаются *внутренними*, а не наружными (валентными) электронами атома. Но для того, чтобы такие переходы были возможны, необходимо, чтобы внутри электронной

оболочки атома были *свободные места*, не заполненные электронами, на которые могут переходить электроны из других квантовых состояний атома. Такие свободные места образуются при воздействии на атом быстрых электронов, фотонов высоких энергий или других быстрых частиц.

Допустим, например, что с наиболее глубокого слоя K атома удален электрон. В результате этого энергия атома возрастает на энергию ионизации из K-слоя. Обозначим через $\mathscr{E}_1$ энергию атома после такого удаления электрона из K-слоя наружу. На освободившееся место перейдет электрон из соседнего L-слоя. K-слой окажется заполненным, а в L-слое не будет доставать одного электрона. Иными словами, получится атом, у которого из слоя L удален один электрон. Его энергия $\mathscr{E}_2$ меньше исходной энергии $\mathscr{E}_1$. Если в результате излучится рентгеновский квант, то его частота определится соотношением (48.1). На освободившееся место в слое L перейдет электрон из слоя M с излучением рентгеновского кванта более низкой частоты. То же произойдет с электроном из слоя N, и т. д. Конечно, переходы могут происходить не только между соседними слоями, но, например, между слоями M и K, N и K и т. д. Они возбуждают более жесткое излучение, чем переходы $L \to K$. В результате всех этих процессов возникнет *весь спектр* рентгеновского характеристического излучения атома.

Итак, частота линии характеристического рентгеновского излучения определяется разностью энергий энергетических уровней атома, найденной в предположении, что из каждого из них *удален один электрон*. При таком удалении атом переходит в возбужденное состояние — его энергия увеличивается на соответствующую энергию ионизации. Последняя тем больше, чем глубже лежит уровень невозбужденного атома. Из изложенного ясно, почему при рассмотрении рентгеновского излучения пользуются схемой уровней возбужденного атома, полученной в предположении, что с рассматриваемого уровня удален один электрон, тогда как число электронов на всех остальных уровнях остается неизменным. Из сказанного выше также следует, что схема уровней так возбужденного атома является *обращенной* по отношению к схеме энергетических уровней невозбужденного атома. Это значит, что в обращенной схеме энергетический уровень расположен тем выше, чем глубже он лежал до удаления с него электрона.

Обращенная схема энергетических уровней атома приведена на рис. 88. Электронные слои невозбужденного атома, соответствующие главным квантовым числам $n = 1, 2, 3, 4, \ldots$, как известно, обозначаются соответственно через K, L, M, ... Такие же обозначения применяются и к атому, возбужденному в указанном выше смысле. Точно так же к возбужденному атому применяют спектроскопические обозначения типа $2^2S_{1/2}$, $2^2P_{1/2}, \ldots$ Целесообразность этого мотивируется тем, что внутренние оболочки атома замкнуты; их момент количества движения равен нулю, а при удалении электрона уносимый им момент количества движения передается с противоположным знаком электронной оболочке, с которой он был удален.

В отсутствие магнитного поля энергия не может зависеть от магнитного квантового числа. Она практически зависит только от квантовых чисел n и j. Слой K состоит из одного энергетического уровня $1^2S_{1/2}$

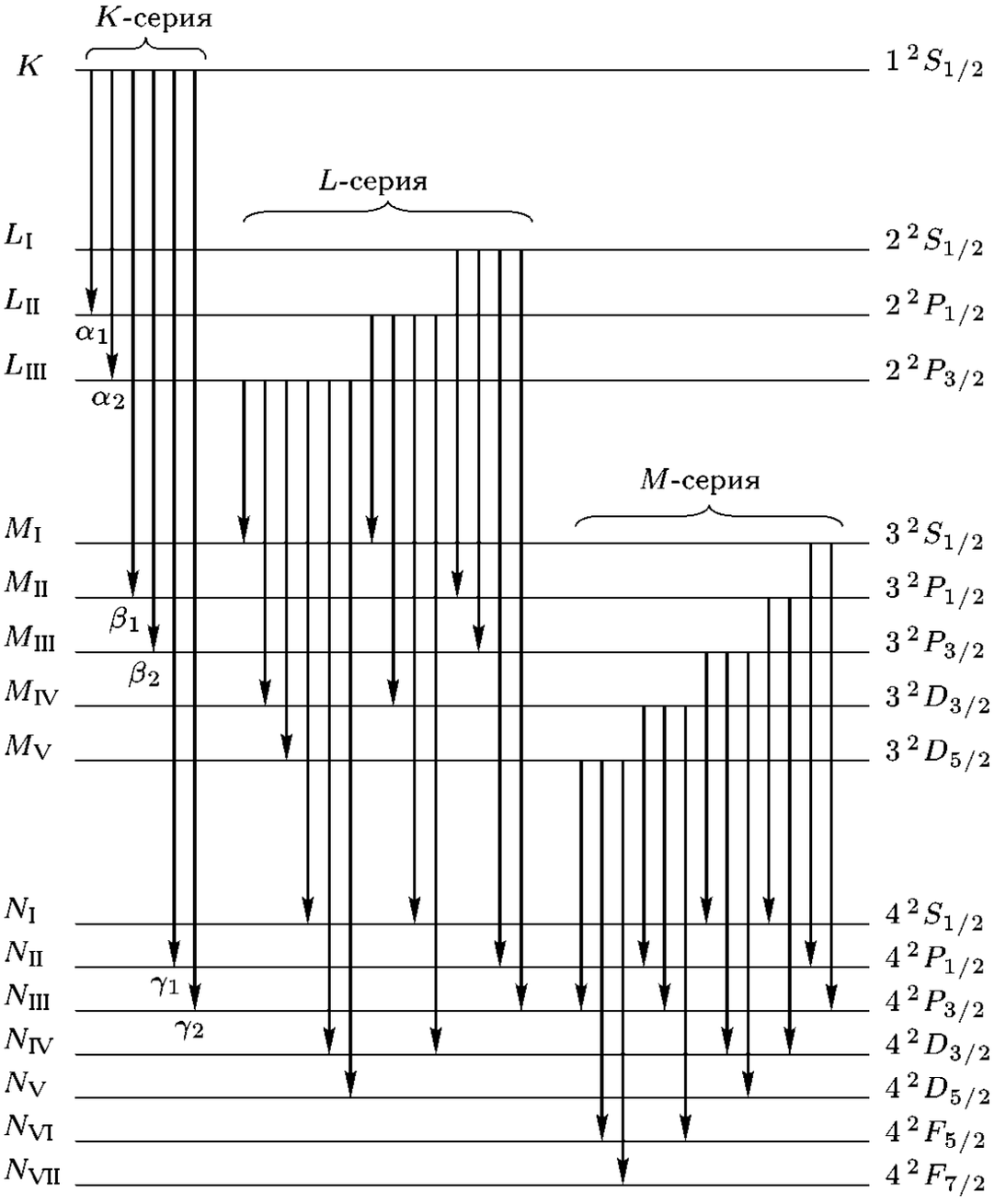


Рис. 88

($n = 1$, $l = 0$, $j = 1/2$). Слой L ($n = 2$) распадается на три подслоя, обозначаемые через $L_{\rm I}$, $L_{\rm II}$, $L_{\rm III}$. Подслой $L_{\rm I}$ состоит из одной $2S$-оболочки. Два остальных подслоя представляют $2P$-оболочку, расщепляющуюся на две подоболочки $2^2P_{1/2}$ и $2^2P_{3/2}$. Их энергии различаются из-за спин-орбитального взаимодействия. Аналогично слой M ($n = 3$) состоит из пяти подслоев, слой N ($n = 4$) — из семи подслоев, обозначенных на рис. 88, и т. д.

4. Недостаток электрона в замкнутой оболочке атома можно рассматривать как «дырку» в этой оболочке. Формально можно считать, что дырка может находиться в различных квантовых состояниях и пе-

реходить из одного состояния в другое. Когда в электронной оболочке атома есть дырка, эта оболочка не замкнута. При переходе дырки на другую замкнутую оболочку последняя перестает быть замкнутой, а исходная оболочка восстанавливается до замкнутой. С этой точки зрения обращенная схема рентгеновских уровней атома аналогична обычной схеме энергетических уровней атома с одним наружным (валентным) электроном. Понятно, что сохраняют силу и все правила отбора при излучении, когда атом (дырка) переходит из одного квантового состояния в другое. Такими правилами отбора являются

$$\Delta J = -1,\ 0,\ +1; \quad \Delta L = \pm 1. \tag{48.2}$$

Все переходы, удовлетворяющие этим правилам отбора, изображены стрелками на рис. 88. Стрелки означают переходы возбужденного атома с высших энергетических уровней на более низкие, или, что то же самое, переходы дырки с уровня на уровень. В целях сокращения мы ограничились наличием в атоме только слоев K, L, M, N. Рисунок нетрудно дополнить переходами на подуровни слоев O $(n = 5)$ и P $(n = 6)$, если таковые имеются.

Характеристический спектр рентгеновского излучения состоит из серий линий, которые обозначаются буквами K, L, M, N, O. Серия K возникает при переходах возбужденного атома с уровня K на лежащие ниже подуровни слоев L, M, N, . . .; серия L — при аналогичных переходах с подуровней слоя L; серия M — при переходах с подуровней слоя M и т. д. Как видно из рисунка, линии серии K имеют дублетную структуру. Компоненты дублетов обозначаются соответственно через α_1, α_2; β_1, β_2; γ_1, γ_2, как видно из рис. 88. Серии L, M, N имеют более сложную мультиплетную структуру. Для линий этих мультиплетов применяются также обозначения греческими буквами с индексами. Буква α указывает, что переход совершился с ближайшего слоя, буква β — со следующего после ближайшего и т. д. Цифровые индексы при буквах α, β, γ нумеруют линии в порядке убывания длин волн.

Из приведенного объяснения возникновения характеристического рентгеновского излучения следует, что при возбуждении наиболее глубоко лежащего слоя K возникает не только серия K, но и *весь рентгеновский спектр*. Вообще, при возбуждении какой-либо серии или линии рентгеновского излучения появляются и все серии и линии рентгеновского излучения с большими длинами волн.

При освобождении электрона из какой-либо внутренней оболочки электрон за пределами атома может обладать каким угодно запасом кинетической энергии. В этом случае его энергия в конечном состоянии не квантуется. При переходе из этого неквантованного состояния на одно из свободных мест в оболочках атома возникает *сплошное рентгеновское излучение*.

5. В отличие от оптических линейчатых спектров с их сложностью и разнообразием, рентгеновские характеристические спектры различных элементов характеризуются простотой и однообразием. Это связано с тем, что при переходе от одного элемента к следующему структура внутренних электронных оболочек атома изменяется очень

мало. При возрастании зарядового числа Z на единицу рентгеновский характеристический спектр элемента сохраняет свой вид; происходит лишь незначительное смещение всех рентгеновских линий в сторону более коротких волн. Эта особенность рентгеновских спектров впервые была обнаружена экспериментально Мозли в 1913 г. и истолкована им на основе теории Бора. Мозли систематически исследовал K- и L-серии рентгеновского излучения 38 различных элементов. Он пользовался кристаллическим спектрографом, работавшим по принципу брэгговских отражений (см. т. IV, § 61), в котором вместо ионизационной камеры была использована фотопластинка.

На рис. 89 воспроизведена фотография K-серии различных элементов, полученная Мозли. Спектры различных элементов расположены относительно друг друга так, что расстояние каждой линии от левого края рисунка приблизительно пропорционально длине волны этой линии. Сами элементы расположены в порядке возрастания атомных номеров от кальция ($Z = 20$) до цинка ($Z = 30$), входящего в состав латуни. Замечательна регулярность, с которой возрастает частота характеристического K-излучения с возрастанием атомного номера элемента. Такой же регулярностью отличаются и изменения частот L-, M- и N-линий характеристического излучения при переходе от одного элемента к следующему. Например, из рис. 89 сразу видно, что между кальцием и титаном пропущен элемент с атомным номером $Z = 21$. Это — скандий, предсказанный Менделеевым и открытый в 1879 г. Нильсоном и Клеве.

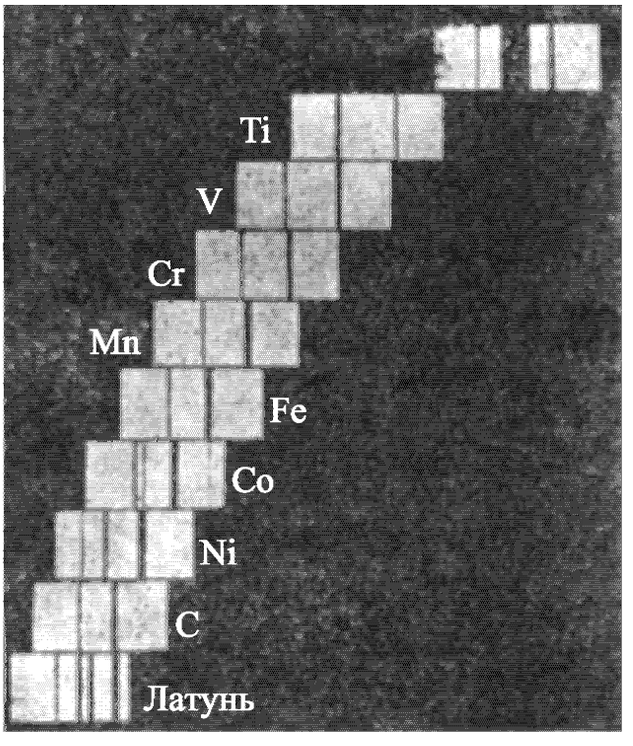


Рис. 89

Исследования Мозли впервые экспериментально показали, что основной величиной, определяющей место элемента в периодической таблице, является не атомная масса, а *атомный номер элемента.* Вместе с тем характеристические рентгеновские спектры позволяют однозначно определять атомные номера элементов и таким образом судить, заполнены ли в периодической таблице все места или должны существовать еще не открытые элементы. Уже сам Мозли оставил место под номером 43 для неизвестного в то время элемента, полученного позднее искусственно и названного технецием. До исследований Мозли не было выяснено, какой из элементов — кобальт с атомной массой 58,933 или никель с атомной массой 58,71 — надо поставить раньше в периодической таблице. Из рис. 89 ясно видно, что кобальт надо поставить между железом и никелем, хотя его атомная

масса и больше, чем у никеля. Участок периодической системы от $Z = 58$ до $Z = 71$ включительно занят элементами редких земель. Их химические свойства настолько близки, а атомные массы были известны настолько недостоверно, что правильность расположения этих элементов в периодической системе вызывала большие сомнения. Изучение рентгеновских спектров с использованием закона Мозли (см. ниже) устранило всякие сомнения. Оно показало, что от водорода до урана включительно должно существовать ровно 92 химических элемента.

Из изложенного выше о происхождении характеристических рентгеновских лучей следует, что длины волн таких лучей зависят исключительно от внутренней структуры электронных оболочек атома. Это есть *атомное свойство* элемента. Поэтому в сплавах и химических соединениях каждый элемент дает такой же характеристический спектр, какой получился бы в отсутствие других элементов. Это ясно проявляется на рис. 89. Спектр латуни получается простым наложением спектров меди и цинка, а в спектре кобальта отчетливо проявилось его загрязнение железом и никелем.

6. Мозли экспериментально установил, что квадратный корень из частоты колебаний $\sqrt{\nu}$ данной линии K-серии в зависимости от атомного номера элемента Z выражается плавной кривой, очень близкой к прямой линии:

$$\sqrt{\nu} = M(Z - a), \tag{48.3}$$

где M и a — постоянные. Та же формула, но с другими числовыми значениями M и a, справедлива и для L-серии, а также для последующих серий M, N, O. Эта эмпирическая формула называется *законом Мозли*. Она и определяет смещение характеристических рентгеновских спектров при переходе от одного элемента к следующему. Последующие более точные измерения обнаружили некоторые отступления от простой линейной зависимости (48.3). Впрочем, эти отступления для K- и L-серий не являются сколько-нибудь значительными, а становятся заметными лишь для M-, N- и O-серий.

Строгое доказательство и установление точности закона Мозли дать невозможно, так как этот вопрос сводится к проблеме многих тел. Можно дать лишь грубую интерпретацию закона Мозли, сведя проблему многих тел к одноэлектронной задаче. Это приводит к вполне удовлетворительным результатам, в особенности для K- и L-серий, а главное — устанавливает физический смысл постоянных, входящих в формулу (48.3).

Пусть в слое K, состоящем из двух электронов, образовалась дырка, т. е. один электрон из этого слоя был удален за пределы атома. Пусть эта дырка заполняется электроном из слоя L, в результате чего произойдет излучение K_α-линии. Поведение этого электрона и будет интересовать нас в дальнейшем, пока не произойдет заполнение дырки в слое K. Можно считать, что на этот электрон все внешние электроны не действуют, так как они как бы образуют внешнюю оболочку со сферически распределенным зарядом в ней. Электрическое поле, действующее на

рассматриваемый электрон, создается зарядом ядра и другим электроном, оставшимся в K-слое. Обе эти частицы можно заменить одной — ядром с «эффективным» зарядом $(Z-1)e$: другой электрон K-слоя как бы экранирует поле ядра атома.

Такое же рассуждение, хотя и менее четкое, применимо и в том случае, когда образовалась дырка в слое L, заполняемая в дальнейшем одним из электронов слоя M с излучением L_α-линии. В этом случае по-прежнему надо интересоваться поведением такого электрона, пока он не заполнит дырку в слое L. Теперь экранирование производится сначала двумя электронами слоя K и семью электронами слоя L. Однако в дальнейшем, пока рассматриваемый электрон не заполнил дырку в слое L, экранирующее действие некоторых электронов слоя L постепенно выпадает. Можно приближенно принять, что в экранирующем действии принимает участие как бы половина из оставшихся семи электронов из слоя L. Это значит, что эффективный заряд ядра можно считать равным $(Z-a)e$, где постоянная a называется *постоянной экранирования*. Можно ожидать, что для L_α-линии $a \approx 5{,}5$, что не так плохо согласуется с тем, что дает опыт. Впрочем, на приведенное обоснование значения $a \approx 5{,}5$ следует смотреть как на грубую оценку. В частности, в наших рассуждениях не учтено, что уровень L состоит из трех подуровней. Точные значения постоянных экранирования следует определять экспериментально.

В обоих случаях задача свелась к одноэлектронной, подобно тому, как это имело место при объяснении спектральных серий щелочных металлов. Частоты излучаемых линий будут определяться формулой

$$\nu = cR(Z-a)^2\left(\frac{1}{n_1^2}-\frac{1}{n_2^2}\right), \tag{48.4}$$

где R — постоянная Ридберга (см. § 13). Отсюда и получается формула (48.3), так как для рассматриваемой линии квантовые числа n_1 и n_2 фиксированы. Для K_α-линии $n=1$, $n_2=2$, а $a=1$, как это было обосновано выше. Поэтому

$$\nu_{K_\alpha} = \frac{3}{4}\,cR(Z-1)^2. \tag{48.5}$$

Для L_α-линии $n_1=2$, $n_2=3$, так что

$$\nu_{L_\alpha} = \frac{5}{16}\,cR(Z-a_{L_\alpha})^2, \text{ где } a_{L_\alpha} \approx 5{,}5. \tag{48.6}$$

7. Рентгеновские спектры поглощения отличаются той характерной особенностью, что в них *нет обращения линий*. Если через слой какого-либо элемента пропустить пучок тормозного рентгеновского излучения (т. е. излучения, разлагающегося в сплошной спектр), то на месте характеристических линий не появятся темные линии, как это было бы в случае оптических спектров. Это объясняется тем, что в веществе, через которое проходит рентгеновский пучок, *нет атомов в возбужденных состояниях*. Все внутренние электронные оболочки заполнены, в них

нет дырок, в которые при облучении рентгеновскими лучами могли бы перейти электроны из более глубоких энергетических уровней.

Поглощение рентгеновских лучей веществом совершенно не зависит от его оптических свойств. Например, белое прозрачное свинцовое стекло толщиной в несколько миллиметров практически полностью поглощает рентгеновские лучи, а потому и применяется для защиты персонала, обслуживающего рентгеновские установки. Тонкий же лист алюминия, абсолютно непрозрачный для видимого света, хорошо пропускает рентгеновские лучи; для рентгеновских лучей, получаемых в технических рентгеновских трубках (при напряжениях ~ 100 кВ), он почти совершенно прозрачен.

8. Параллельный пучок рентгеновских лучей при прохождении через вещество испытывает ослабление. Оно вызывается двумя причинами. Во-первых, *рассеянием*, когда часть лучей отклоняется в сторону и поэтому выходит из первоначального пучка. Это явление аналогично рассеянию видимого света в мутных средах. Но в случае видимого света мутность среды вызывается либо наличием в среде посторонних частиц с другим показателем преломления, либо (при отсутствии таких частиц) флуктуациями плотности самого вещества (см. т. IV, § 98). Рассеяние на флуктуациях плотности называется *молекулярным* или *релеевским*. Для рентгеновских же лучей ввиду малости их длин волн всякая среда является *сильно мутной*. Здесь роль неоднородностей, вызывающих рассеяние, играют не малые макроскопические кусочки вещества, а отдельные атомы и даже электроны в атомах. В случае мягкого рентгеновского излучения, когда длина волны больше или сравнима с размерами атома, электроны в атоме ведут себя как связанные — атом рассеивает как целое, а потому волны, рассеиваемые отдельными электронами атома, *когерентны*. В противоположном случае очень жестких рентгеновских лучей длина волны мала по сравнению с размерами атома. Тогда на электронах атома происходит *некогерентное* комптоновское рассеяние.

Количественное различие в рассеянии видимого света и рентгеновского излучения можно иллюстрировать следующим примером. Параллельный пучок видимого света вследствие молекулярного рассеяния ослабевает в e раз в слое совершенно чистой воды толщиной около 1 км, тогда как рентгеновский луч испытывает такое же относительное ослабление всего на расстоянии порядка 5 см.

Во-вторых, ослабление пучка рентгеновских лучей происходит из-за *истинного поглощения* или *абсорбции*, когда часть энергии пучка в конце концов переходит в тепло.

Если параллельный пучок рентгеновских лучей монохроматичен ($\lambda = \text{const}$), то ослабление его интенсивности I на отрезке dx однородного вещества определяется соотношением $dI = -\mu I\, dx$, где постоянная μ называется *коэффициентом ослабления* рентгеновских лучей. Отсюда следует

$$I(x) = I_0 e^{-\mu x}, \tag{48.7}$$

т. е. при прохождении расстояния $x = 1/\mu$ интенсивность убывает в e раз. Ослабление интенсивности пучка происходит как за счет истинного поглощения, так и за счет рассеяния рентгеновских лучей, так что можно написать

$$\mu = \tau + \sigma, \tag{48.8}$$

где τ называется *коэффициентом истинного поглощения*, а σ — *коэффициентом рассеяния* рентгеновских лучей.

Все коэффициенты μ, σ, τ пропорциональны плотности вещества ρ, а потому удобнее пользоваться так называемыми *массовыми коэффициентами*, т. е. величинами

$$\mu_m = \frac{\mu}{\rho}, \quad \tau_m = \frac{\tau}{\rho}, \quad \sigma_m = \frac{\sigma}{\rho}. \tag{48.9}$$

Введя μ_m в формулу (48.7), запишем ее в виде

$$I(x) = I_0 e^{-\mu_m \rho x}. \tag{48.10}$$

Если S — площадь поперечного сечения пучка, то величина $S\rho x$ означает массу вещества, проходимого рентгеновским пучком на расстоянии x. Следовательно, ρx есть масса вещества, проходимого пучком на расстоянии x, если площадь поперечного сечения пучка равна единице.

В теоретических расчетах еще удобнее пользоваться так называемыми *атомными коэффициентами*:

$$\begin{gathered} \mu_a = \frac{\mu}{\rho} m = \frac{\mu}{\rho} \frac{A}{N_\mathrm{A}}, \quad \tau_a = \frac{\tau}{\rho} m = \frac{\tau}{\rho} \frac{A}{N_\mathrm{A}}, \\ \sigma_a = \frac{\sigma}{\rho} m = \frac{\sigma}{\rho} \frac{A}{N_\mathrm{A}}, \end{gathered} \tag{48.11}$$

где m — масса атома, равная $m = A/N_\mathrm{A}$, A — масса моля, N_A — постоянная Авогадро. Эти коэффициенты имеют размерность площади, а потому могут быть истолкованы как *поперечные сечения атома* по отношению к ослаблению, поглощению и рассеянию рентгеновских лучей. Например, μ_a характеризует ослабление интенсивности рентгеновского пучка в слое, содержащем один атом на единице площади поперечного сечения пучка. Если в формулах (48.11) m означает массу не атома, а молекулы, то величины, определяемые этими формулами, следует назвать *молекулярными коэффициентами*.

9. Коэффициент поглощения рентгеновских лучей вообще возрастает с увеличением длины волны. Однако при некоторых значениях длины волны он резко падает, а затем начинает снова плавно возрастать. Такая особенность поглощения рентгеновских лучей имеет простое физическое объяснение. Пусть длина волны рентгеновских лучей настолько мала, что возбуждаются все рентгеновские уровни, включая K-уровень. На это возбуждение затрачивается энергия рентгеновских лучей. При увеличении длины волны, начиная с некоторого значения λ, энергии рентгеновского кванта становится недостаточно,

чтобы возбудить K-уровень. Это и сказывается на резком уменьшении поглощения, — появляется так называемый *край поглощения.* При дальнейшем увеличении длины волны перестают последовательно возбуждаться L_{I}-, L_{II}-, L_{III}-уровни (см. схему уровней на рис. 88), что ведет к появлению новых краев поглощения и т. д.

Довольно точно соблюдается эмпирически установленное соотношение

$$\tau_a = CZ^4\lambda^3, \tag{48.12}$$

где постоянная C испытывает всякий раз скачкообразное уменьшение при переходе через край поглощения. Из формулы (48.12), конечно, следует

$$\sigma_m = \frac{\tau_a}{\rho} = C'\frac{Z^4\lambda^3}{A}, \tag{48.13}$$

где $C' = CN_{\text{A}}$ — новая скачкообразно меняющаяся постоянная.

На схематическом рис. 90 изображен (в произвольных единицах) ход массового коэффициента поглощения для серебра и меди в зависимости от длины волны λ. На вертикальной оси отложена величина $\tau_m^{1/3}$. Зависимость всюду линейная, но при длинах волн $\lambda =$ $= 0{,}0485$ нм для серебра и $\lambda = 0{,}13785$ нм для меди наблюдается край поглощения, связанный с прекращением возбуждения K-уровня. Если бы кривую поглощения для серебра продолжить в сторону более длинных волн, то на ней появились бы три зубца при длинах волн 0,3245, 0,3506, 0,3993 нм, соответствующие новым краям поглощения (уровень L тройной, как видно из рис. 88).

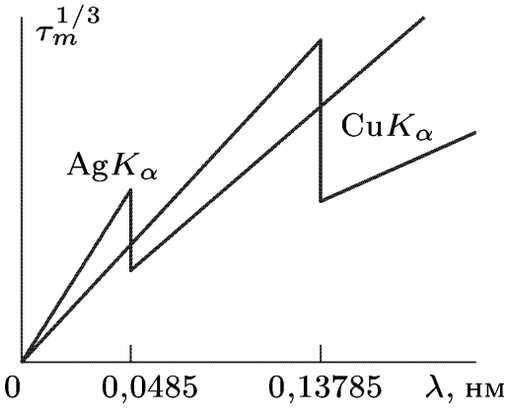


Рис. 90

Особенность поглощения рентгеновских лучей состоит в том, что оно является чисто *атомным свойством.* Молекулярный коэффициент поглощения *аддитивно складывается* из атомных коэффициентов элементов, входящих в состав молекулы. Пользуясь этим, молекулярные коэффициенты поглощения бесчисленного множества химических соединений можно вычислить, зная атомные коэффициенты поглощения элементов. Далее, как видно из формулы (48.12), атомные коэффициенты поглощения быстро возрастают с увеличением порядкового номера элемента — пропорционально Z^4. На этих особенностях поглощения рентгеновских лучей основано их применение для целей просвечивания (см. задачу 1 к этому параграфу).

10. В заключение этого параграфа рассмотрим еще одно явление, происходящее в электронной оболочке атома и вызывающее также поглощение рентгеновских лучей. Оно называется *эффектом Оже* (р. 1899) или *автоионизацией возбужденного атома.* Это явление связано с перераспределением энергии внутри возбужденного атома. В отличие от обычной фотоионизации, когда поглощенный рентгеновский

квант приводит непосредственно к вылету электрона за пределы атома, эффект Оже происходит в две стадии. На первой стадии происходит поглощение рентгеновского кванта, сопровождающееся возбуждением атома, а именно освобождением электрона из K-слоя, т. е. образованием дырки в этом слое. На второй стадии происходит заполнение этой дырки электроном, перескакивающим из L-слоя. Избыток энергии $\mathscr{E}_K - \mathscr{E}_L$ либо выделяется в виде кванта характеристического излучения, что было рассмотрено выше, либо приводит к выбрасыванию одного электрона из вышележащих оболочек атома. Этот процесс и называется эффектом Оже, а выброшенный электрон — *электроном Оже.* Таким образом, дырка в L-слое сохраняется, но возникает вторая дырка в одном из вышележащих слоев. Иными словами, атом становится дважды ионизованным.

Фотографируя в камере Вильсона фотоэлектроны, освобождаемые рентгеновскими лучами из атомов тяжелых инертных газов (например, криптона), Оже обнаружил, что в некоторых случаях в одной и той же точке берут начало *два электронных следа.* Один электрон образуется в результате обычной фотоионизации на K-слое, другой — вследствие внутреннего перераспределения энергии возбужденного атома, т. е. вследствие эффекта Оже.

Электроны Оже всегда уносят значительную долю энергии возбужденного атома. Особенно сильно эффект Оже проявляется в тех случаях, когда переходы с излучением фотонов запрещены (например, в $0 - 0$-переходах).

ЗАДАЧИ

1. Сравнить коэффициенты поглощения рентгеновских лучей для костей и тканей человеческого тела. Вещество кости в основном составляет фосфорнокислый кальций $Ca_3(PO_4)_2$. Поглощение ткани обусловлено главным образом входящей в нее водой H_2O.

Р е ш е н и е. На основании формулы (48.12) с учетом значений порядковых номеров элементов получим

$$\frac{(\tau_a)_{\text{кость}}}{(\tau_a)_{\text{ткань}}} = \frac{3 \cdot 20^4 + 2 \cdot 15^4 + 8 \cdot 8^4}{2 \cdot 1^4 + 8^4} \approx 150.$$

Для определения отношения массовых коэффициентов поглощения, согласно (48.13), надо учесть отношение молекулярных масс рассматриваемых химических соединений. Для $Ca_3(PO_4)_2$ $A = 3 \cdot 40 + 2 \cdot 30 + 8 \cdot 16 = 308$; для H_2O $A = 2 \cdot 1 + 16 = 18$. Поэтому

$$\frac{(\tau_m)_{\text{кость}}}{(\tau_m)_{\text{ткань}}} \approx 150 \cdot \frac{18}{308} \approx 9,$$

т. е. массовый коэффициент поглощения кости примерно в 10 раз превосходит такой же коэффициент ткани. Из-за различия плотностей отношение коэффициентов поглощения $(\tau)_{\text{кость}}$ и $(\tau)_{\text{ткань}}$ еще примерно в 2–3 раза больше. Этого достаточно, чтобы понять, почему на рентгеновских снимках тень от костей выделяется так резко.

2. Как известно, у редких земель внутренняя оболочка f заполняется после заполнения наружных оболочек. Таким образом, внутри электронной

атмосферы атома имеются незаполненные места — дырки. Может ли происходить спонтанное рентгеновское излучение, возникающее при заполнении таких дырок?

О т в е т. Не может, так как при переходе электрона из наружных оболочек в рассматриваемую дырку энергия атома не уменьшается, а увеличивается.

§ 49. Атом гелия

1. Простейшими после атома водорода являются атом гелия ($Z = 2$) и аналогичные ему ионы Li^{+} ($Z = 3$), Be^{++} ($Z = 4$), Be^{3+} ($Z = 5$), ..., электронные оболочки которых содержат по два электрона. В спектрах этих двухэлектронных атомов и ионов встречаются те же спектральные серии, что и у атомов щелочных металлов, но каждая из серий представлена в двух экземплярах: есть две главные серии, две резкие, две диффузные и т. д. В одном экземпляре серий все линии простые (синглеты), тогда как в другом — триплеты, т. е. каждая из линий состоит из трех близко расположенных линий.

Наиболее известной в спектре гелия является желтая линия D_3, благодаря которой гелий впервые был обнаружен на Солнце в 1867 г. Это — триплет с длинами волн 587,5963; 587,5643 и 587,5601 нм и отношениями интенсивностей $1 : 3 : 5$. Расстояние между двумя последними линиями составляет всего 0,0042 нм. Поэтому долгое время эти две линии принимали за одну, а триплет D_3 — за дублет. Триплет D_3 является первым членом первой побочной серии триплетов. Главная серия триплетов гелия лежит в инфракрасной части спектра. Соответствующие серии синглетов расположены преимущественно в области ультрафиолета.

Между синглетными и триплетными энергетическими уровнями гелия не осуществляются квантовые переходы (так называемый *запрет интеркомбинаций*). Этот факт послужил основанием для гипотезы, согласно которой гелий состоит из двух различных элементов: *ортогелия*, дающего триплетные линии, и *парагелия*, спектральные линии которого — синглеты. Гипотеза эта, как будет показано в дальнейшем, оказалась неправильной, а запрет интеркомбинаций — не абсолютно точным правилом. Это видно из того, что в спектре гелия имеется линия, правда единственная, с длиной волны 591,6 нм, получающаяся при переходе с триплетного уровня $^3P_{1/2}$ на синглетный уровень $^1S_0$.

2. Теория многоэлектронных атомов в квантовой механике не встречает принципиальных трудностей, хотя практические вычисления весьма сложны и громоздки и могут быть выполнены только на машинах. В простейшем случае атома гелия и аналогичных ему ионов, если не учитывать спины, задача сводится к решению уравнения Шредингера с двумя электронами. Ниже рассматривается этот случай, хотя все вычисления и опускаются. Для стационарного состояния уравнение Шредингера записывается в виде

$$\widehat{H}\psi \equiv (\widehat{H}_1 + \widehat{H}_2 + \widehat{U}_{12})\psi = \mathscr{E}\psi, \tag{49.1}$$

где волновая функция ψ зависит от координат обоих электронов, $\mathscr{E}$ — энергия стационарного состояния, $\widehat{H}_1$ и $\widehat{H}_2$ — гамильтонианы первого и второго электронов в отсутствие взаимодействия между ними:

$$\widehat{H}_1 = -\frac{\hbar^2}{2\mu}\nabla_1^2 - \frac{Ze^2}{r_1}, \tag{49.2}$$

$$\widehat{H}_2 = -\frac{\hbar^2}{2\mu}\nabla_2^2 - \frac{Ze^2}{r_2}, \tag{49.3}$$

а $\widehat{U}_{12}$ — часть полного гамильтониана, учитывающая взаимодействие между электронами:

$$\widehat{U}_{12} = \frac{e^2}{r_{12}}. \tag{49.4}$$

Здесь ∇_1^2 и ∇_2^2 — операторы Лапласа для первого и второго электронов:

$$\nabla_1^2 = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial y_1^2} + \frac{\partial^2}{\partial z_1^2}, \quad \nabla_2^2 = \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial y_2^2} + \frac{\partial^2}{\partial z_2^2}.$$

Через $\mathbf{r}_1(x_1, y_1, z_1)$, $\mathbf{r}_2(x_2, y_2, z_2)$ обозначены радиусы-векторы и декартовы координаты первого и второго электронов, а через r_{12} — расстояние между ними. Заряд ядра для общности принят равным Z (в случае гелия $Z = 2$). Ядро атома в рассматриваемом приближении считается неподвижным и принимается за начало координат.

3. Сформулированная задача аналогична классической задаче трех тел, рассматриваемой в небесной механике. В ней речь идет о движении двух планет в гравитационном поле Солнца с учетом гравитационного взаимодействия между самими планетами. Такая задача, хотя в принципе и допускает аналитическое решение в виде рядов, но эти ряды абсолютно непригодны для практических расчетов (см. подстрочное примечание к § 16 на с. 90). Но в небесной механике разработаны превосходные приближенные методы расчета, вполне удовлетворяющие высоким требованиям наблюдательной астрономии. В их основе лежит *теория возмущений*, использующая тот факт, что взаимодействие между планетами мало по сравнению с взаимодействием каждой планеты с Солнцем. В нулевом приближении взаимодействием между планетами можно совсем пренебречь. Используя решение, полученное в нулевом приближении, можно затем учесть взаимодействие между планетами в первом приближении, после этого найти второе приближение и т. д. По тому же пути идет и теория возмущений квантовой механики в случае атома с двумя электронами. Она также в нулевом приближении отбрасывает взаимодействие между электронами. Правда, здесь ситуация значительно менее благоприятна, чем в небесной механике, так как взаимодействие между электронами отнюдь не мало по сравнению с взаимодействием каждого из них с атомным ядром. Однако получающиеся результаты довольно удовлетворительны, чем и оправдывается использование методов теории возмущений.

Трудность задачи об атоме гелия обусловлена наличием в уравнении (49.1) члена $\widehat{U}_{12}\psi$, зависящего от координат *обоих* электронов. Благодаря этому уравнение (49.1) не имеет решений с разделяющимися переменными. В теории возмущений член $\widehat{U}_{12}\psi$ рассматривается как «малая» поправка и в нулевом приближении отбрасывается. Таким образом, уравнение нулевого приближения имеет вид

$$\widehat{H}^0\psi^0 = \mathscr{E}^0\psi^0, \tag{49.5}$$

где через $\widehat{H}^0$ обозначен гамильтониан в нулевом приближении, т. е. оператор $\widehat{H}^0 = \widehat{H}^1 + \widehat{H}^2$, в котором возмущающий член $\widehat{U}_{12}$ отброшен. Аналогично, через ψ^0 и $\mathscr{E}^0$ обозначены волновая функция и собственное значение энергии в нулевом приближении. Найдя величины ψ^0 и $\mathscr{E}^0$, ищем решение в первом приближении $\psi = \psi^0 + \psi^1$, $\mathscr{E} = \mathscr{E}^0 + \mathscr{E}^1$. Для этого служит уравнение

$$(\widehat{H}^0 + \widehat{U}_{12})(\psi^0 + \psi^1) = (\mathscr{E}^0 + \mathscr{E}^1)(\psi^0 + \psi^1),$$

или на основании (49.5)

$$\widehat{H}^0\psi^1 + \widehat{U}_{12}\psi^0 + \widehat{U}_{12}\psi^1 = \mathscr{E}^1\psi^0 + \mathscr{E}^0\psi^1 + \mathscr{E}^1\psi^1.$$

Здесь величины $\widehat{U}_{12}\psi^1$ и $\mathscr{E}^1\psi^1$ следует рассматривать как «малые» более высокого порядка и в первом приближении отбросить. Таким образом, в первом приближении

$$(\widehat{H}^0 - \mathscr{E}^0)\psi^1 = (\mathscr{E}^1 - \widehat{U}_{12})\psi^0. \tag{49.6}$$

Это — неоднородное уравнение, правая часть которого известна. Можно весьма просто доказать (на чем мы не останавливаемся), что уравнение (49.6) имеет решение только при таких значениях параметра $\mathscr{E}^1$, когда правая часть ортогональна к волновой функции ψ^0 нулевого приближения, т. е.

$$\int \psi^{0*}(\mathscr{E}^1 - \widehat{U}_{12})\psi^0\, d\tau = 0,$$

причем интегрирование производится по шестимерному объему $d\tau = dx_1\, dy_1\, dz_1\, dx_2\, dy_2\, dz_2$ обоих электронов. Таким образом, если функция ψ^0 нормирована к единице, то получается

$$\mathscr{E}^1 = \int \psi^{0*}\widehat{U}_{12}\psi^0\, d\tau, \tag{49.7}$$

т. е. поправка $\mathscr{E}^1$ к энергии в первом приближении равна потенциальной энергии U_{12} взаимодействия электронов, усредненной с помощью волновых функций нулевого приближения. Вычислив $\mathscr{E}^1$, можно затем решить уравнение (49.6) и найти ψ^1, на чем мы не останавливаемся.

Аналогичным путем можно найти второе приближение $\mathscr{E} = \mathscr{E}^0 + \mathscr{E}^1 + \mathscr{E}^2$, $\psi = \psi^0 + \psi^1 + \psi^2$, рассматривая поправки $\mathscr{E}^2$ и ψ^2 как величины еще более высокого порядка малости, и т. д.

4. Обратимся теперь к уравнению (49.5), т. е. к решению в нулевом приближении. В более подробной записи это уравнение гласит

$$(\hat{H}_1 + \hat{H}_2)\psi^0 = \mathscr{E}^0\psi^0, \tag{49.8}$$

где оператор $\hat{H}_1$ зависит только от координат первого электрона, а оператор $\hat{H}_2$ — только от координат второго. Поэтому решение ψ^0 будет решением с разделяющимися переменными:

$$\psi^0 = \psi^0(1)\psi^0(2), \tag{49.9}$$

где цифрой 1 обозначена совокупность координат первого, а цифрой 2 — второго электронов. После подстановки в предыдущее уравнение и деления на $\psi^0 = \psi^0(1)\psi^0(2)$ получится

$$\frac{\hat{H}_1\psi^0(1)}{\psi^0(1)} + \frac{\hat{H}_2\psi^0(2)}{\psi^0(2)} = \mathscr{E}^0.$$

Первое слагаемое в левой части зависит только от координат 1, а второе — только от координат 2. Так как сумма этих слагаемых $\mathscr{E}^0$ постоянна, то должно быть постоянно и каждое из слагаемых в отдельности. Иными словами, должны выполняться уравнения

$$\hat{H}_1\psi^0(1) = \mathscr{E}_1^0\psi^0(1), \quad \hat{H}_2\psi^0(2) = \mathscr{E}_2^0\psi^0(2), \tag{49.10}$$

где $\mathscr{E}_1^0$ и $\mathscr{E}_2^0$ — постоянные, удовлетворяющие условию

$$\mathscr{E}_1^0 + \mathscr{E}_2^0 = \mathscr{E}^0. \tag{49.11}$$

Оба уравнения (49.10) по существу тождественны. Они отличаются друг от друга только обозначениями координат первого и второго электронов, а также числовыми значениями постоянных $\mathscr{E}_1^0$ и $\mathscr{E}_2^0$ (поскольку электроны могут находиться в различных состояниях; если же эти состояния одинаковы, то $\mathscr{E}_1^0 = \mathscr{E}_2^0$. Каждое из уравнений (49.10) описывает стационарное состояние электрона в поле ядра в предположении, что взаимодействие между электронами не учитывается. В этом предположении

$$\hat{H}_1 = -\frac{\hbar^2}{2\mu}\nabla_1^2 - \frac{Ze^2}{r_1}, \quad \hat{H}_2 = -\frac{\hbar^2}{2\mu}\nabla_2^2 - \frac{Ze^2}{r_2}. \tag{49.12}$$

Тем самым задача в нулевом приближении сведена к задаче нахождения собственных функций и собственных значений одноэлектронного водородоподобного атома. Решение последней задачи хорошо известно.

5. Найдем прежде всего в нулевом приближении энергию полной ионизации нейтрального атома гелия, т. е. работу по удалению в бесконечность обоих его электронов, когда сам атом находится в основном состоянии. Работа по удалению электрона в бесконечность из атома водорода в основном состоянии равна $\mu e^4/2\hbar^2 \approx 13{,}539$ эВ. Для однократно ионизованного гелиеподобного атома эта работа в Z^2 раз больше (см. формулу (13.12)). Поэтому, согласно формуле (49.11), работа

по удалению обоих электронов гелиеподобного атома в бесконечность в нулевом приближении будет

$$\mathscr{E}^0_{\text{иониз}} = 2Z^2 \frac{\mu e^4}{2\hbar^2}. \tag{49.13}$$

Результат этот вполне очевиден, поскольку в нулевом приближении взаимодействие электронов не учитывается. В частности, для атома гелия формула (49.13) дает $\mathscr{E}_{\text{иониз}} = 108{,}3$ эВ.

Зная волновую функцию (49.9) в нулевом приближении, можно по формуле (49.7), в которой следует положить $U_{12} = e^2/r_{12}$, найти поправку к энергии полной ионизации гелиеподобного атома, которую дает первое приближение. Вычисления приводят к результату

$$(\mathscr{E}^0 + \mathscr{E}^1)_{\text{иониз}} = \left(2Z^2 - \frac{5}{4}\,Z\right)\frac{\mu e^4}{2\hbar^2}. \tag{49.14}$$

Числовые результаты приведены в табл. 5. В нулевом приближении, как и следовало ожидать, получается большое различие между вычислением и опытом: ошибка составляет примерно 40% для He и 10% для C^{4+}. Но уже в первом приближении получается хорошее согласие теории с опытом. Это тем более удивительно, что энергия взаимодействия между электронами отнюдь не мала по сравнению с энергией их взаимодействия с атомными ядрами. Многие авторы вычисляли в высших приближениях энергии ионизации и возбуждения атома гелия, а также разработали более совершенные методы, отличные от теории возмущений. Не имея возможности останавливаться на этих вопросах, приводим в последнем столбце таблицы результаты расчетов Хиллераса (1898–1965), полученные без учета конечности масс ядер и поправок на теорию относительности. Согласие с опытом очень хорошее.

Таблица 5

Атом или ион	Энергия полной ионизации, эВ			
	Наблюдено	Вычислено		
		в нулевом приближении	в первом приближении	по Хиллераасу
He	78,98	108,3	74,46	78,98
Li^+	198,04	243,7	192,9	198,03
Be^{2+}	371,51	433,2	365,5	371,49
B^{3+}	599,43	676,9	592,3	599,40
C^{4+}	881,83	974,8	873,3	881,82

6. Обратимся теперь к объяснению, почему в спектрах гелия и гелиеподобных ионов происходит удвоение спектральных серий, о котором говорилось в начале этого параграфа. Заметим прежде всего, что все серии получаются путем возбуждения одного электрона, а не обоих электронов сразу. Последний процесс значительно менее вероятен, чем

первый, и рассматриваться не будет. Поэтому рассмотрим в нулевом приближении такое состояние атома, когда один электрон не возбужден, а другой находится в возбужденном состоянии. Состояние первого электрона в нулевом приближении описывается волновой функцией $\psi_1^0(1)$, а второго — функцией $\psi_k^0(2)$. Нижний индекс означает совокупность трех квантовых чисел (n, l, m_l), характеризующих состояние электрона. В частности, нижнему индексу 1 соответствует нормальное (невозбужденное) состояние ($n = 1$, $l = 0$, $m_l = 0$). От наличия спина электрона мы временно отвлекаемся.

Волновая функция обоих электронов в нулевом приближении представится произведением $\psi_1^0(1)\psi_k^0(2)$. Однако такая функция не удовлетворяет требованию тождественности одинаковых частиц (электронов). Если в ней поменять местами первый и второй электроны, то получится функция $\psi_k^0(1)\psi_1^0(2)$, описывающая то же состояние атома: имеет место обменное вырождение (см. § 46). Линейной суперпозицией обоих квантовых состояний получается множество мыслимых состояний одной и той же энергии. Из них из-за тождественности электронов могут реализоваться только состояния, описываемые симметричными и антисимметричными волновыми функциями

$$\begin{aligned} \psi_s^0(1,2) &= \psi_1^0(1)\psi_k^0(2) + \psi_k^0(1)\psi_1^0(2), \\ \psi_a^0(1,2) &= \psi_1^0(1)\psi_k^0(2) - \psi_k^0(1)\psi_1^0(2). \end{aligned} \tag{49.15}$$

Симметрия и антисимметрия функций (49.15) относится к операции, в которой электроны обмениваются своими местами, т. е. аргументами в волновой функции, причем «место» характеризуется значениями только пространственных координат: спиновые координаты во внимание не принимаются. Между тем принцип антисимметрии волновых функций, которому должны подчиняться электроны, имеет в виду *полные* волновые функции, т. е. функции, зависящие не только от пространственных, но и от спиновых координат электронов. Здесь «место» электрона в волновой функции определяется не тремя, а *четырьмя* координатами. Поэтому обе неполные, т. е. пространственные функции (49.15), поскольку они удовлетворяют требованию неразличимости тождественных частиц, допустимы при квантовом описании состояния атома. Это утверждение содержится также в первоначальной упрощенной формулировке принципа Паули. Последний исключает такие состояния атома, в которых оба электрона имеют одинаковые четверки квантовых чисел n, l, m_l, m_s, но не запрещает, чтобы у электронов были одинаковыми три квантовых числа при различии четвертых. В частности, возможны состояния, у которых $k = 1$, когда оба электрона находятся в основном состоянии, которое характеризуется значениями квантовых чисел $n = 1$, $l = 0$, $m_l = 0$. Для такого состояния, понятно, из двух возможных функций (49.15) надо взять только симметричную, так как антисимметричная тождественно обращается в нуль В этом случае обменное вырождение отсутствует,

а потому основному состоянию соответствует только одна волновая функция и один уровень энергии.

Теперь каждое из выражений (49.15) можно принять в качестве нулевого приближения и по методу теории возмущений отыскать первое приближение. Таким путем будут найдены две волновые функции, одна из которых симметрична по координатам, а другая антисимметрична. Их можно обозначить соответственно через $\psi_s(1,2)$ и $\psi_a(1,2)$ без верхнего индекса 0. Состояние с антисимметричной волновой функцией $\psi_a(1,2)$ называется *ортосостоянием*, а с симметричной волновой функцией $\psi_s(1,2)$ — *парасостоянием*. Таким образом, двух гелиев, как различных химических элементов, не существует. Орто- и парагелий суть *один и тот же химический элемент, но находящийся в различных квантовых состояниях*. Тем самым объяснено происхождение двух систем термов и соответствующих спектральных переходов между ними. Понятно, что такие переходы, если они сопровождаются излучением света, должны подчиняться правилам отбора, о которых говорилось в § 39.

7. Знание волновых функций нулевого приближения позволяет вычислить по формуле (49.7) поправку к энергии атома в первом приближении. Эта формула предполагает, однако, что волновая функция нормирована к единице. Поэтому и обе функции (49.15), симметричную и антисимметричную, надо предварительно нормировать. Обозначим соответствующие постоянные нормирующие множители через C_s и C_a. Для фактического проведения расчета, конечно, эти нормирующие множители надо вычислить. Мы не будем этого делать, так как нашей целью является только выяснение *принципиальной физической стороны вопроса*, а для этого знания числовых значений C_s и C_a не требуется. Однако различие C_s и C_a должно быть учтено. (Если бы C_s и C_a были одинаковы, то нормирующие множители можно было бы опустить, предполагая, что комбинации (49.15) уже нормированы.)

Для состояний, описываемых симметричной волновой функцией, поправка к энергии равна

$$\mathscr{E}_s^1 = |C_s|^2 \int \psi_s^{0*} \frac{e^2}{r_{12}} \psi_s^0 \, d\tau = |C_s|^2 \int \frac{e^2}{r_{12}} |\psi_s^0|^2 \, d\tau_1 \, d\tau_2, \tag{49.16}$$

где $d\tau_1 = dx_1 \, dy_1 \, dz_1$, $d\tau_2 = dx_2 \, dy_2 \, dz_2$, а для состояний, описываемых антисимметричной волновой функцией,

$$\mathscr{E}_a^1 = |C_a|^2 \int \psi_a^{0*} \frac{e^2}{r_{12}} \psi_a^0 \, d\tau = |C_a|^2 \int \frac{e^2}{r_{12}} |\psi_a^0|^2 \, d\tau_1 \, d\tau_2. \tag{49.17}$$

Подставив сюда значения волновых функций из (49.15) и выполнив умножение с учетом тождественности электронов, получим

$$\mathscr{E}_s^1 = |C_s|^2 (I_к + I_об), \quad \mathscr{E}_a^1 = |C_a|^2 (I_к - I_об), \tag{49.18}$$

где введены обозначения

$$I_{\text{к}} = 2\int \frac{e^2}{r_{12}} \left|\psi_1^0(1)\right|^2 \left|\psi_k^0(2)\right|^2 d\tau,$$
$$I_{\text{об}} = 2\int \frac{e^2}{r_{12}} \psi_1^0(1)\psi_k^{0*}(1)\psi_k^0(2)\psi_1^{0*}(2)\, d\tau. \tag{49.19}$$

Из формул (49.16) и (49.17) видно, что энергии $\mathscr{E}_s^1$ и $\mathscr{E}_a^1$ существенно положительны. Величина $I_{\text{к}}$ также существенно положительна, как это ясно из первой формулы (49.19). Таким образом,

$$I_{\text{к}} + I_{\text{об}} > 0, \quad I_{\text{к}} - I_{\text{об}} > 0.$$

Величина $I_{\text{об}}$ также положительна, как показывает фактическое вычисление ее с использованием явных выражений волновых функций водородоподобных атомов. Таким образом, как видно из формул (49.18), *энергетические уровни парасостояния лежат ниже энергетических уровней ортосостояния.* Поэтому *нормальным состоянием гелия является парасостояние.*

Член $I_{\text{к}}$ имеет наглядное классическое истолкование. Заряд одного электрона как бы «размазан» в пространстве с объемной плотностью $\rho_1 = e|\psi_1^0(1)|^2$, а другого — с объемной плотностью $\rho_2 = e|\psi_k^0(2)|^2$. Подынтегральное выражение в $I_{\text{к}}$ имеет смысл потенциальной энергии отталкивания зарядов $\rho_1\, d\tau_1$ и $\rho_2\, d\tau_2$, а интеграл в целом — потенциальной энергии взаимодействия обоих размазанных по пространству зарядов. Член $I_{\text{об}}$ не имеет классического истолкования и является сугубо квантовым. Выражаясь образно, можно сказать, что $I_{\text{об}}$ появляется потому, что каждый электрон как бы одновременно находится и в состоянии ψ_1^0, и в состоянии ψ_k^0. Величина $I_{\text{об}}$ называется *обменной энергией*, а соответствующее ей взаимодействие — *обменным взаимодействием.* Наличием обменной энергии и объясняется понижение энергетических уровней парасостояния и повышение уровней ортосостояния.

Не следует думать, что обменная энергия есть какой-то особый вид энергии. Она появляется только в результате *принципиальной неразличимости* двух квантовых состояний, из которых одно получается из другого в результате перестановки двух тождественных взаимодействующих частиц. Поэтому обменное взаимодействие специфично не только для кулоновских сил, а имеет место в системе любых тождественных частиц независимо от природы сил взаимодействия между ними. Например, обменное взаимодействие существует между нуклонами атомного ядра и проявляется в так называемом *насыщении ядерных сил.* Обменным взаимодействием объясняется явление ферромагнетизма.

8. Объясним теперь разделение спектральных термов атома гелия и соответствующих им спектральных линий на синглетные и триплетные. Для этого надо учесть наличие спинов у электронов. В нулевом приближении пространственная и спиновая части полной волновой функции разделяются, так что она может быть представлена в виде

произведения пространственной и спиновой волновых функций. Конечно, как и пространственные, спиновые волновые функции должны удовлетворять требованию либо симметрии, либо антисимметрии. Поэтому спиновых функций может быть только четыре, а именно:

$$\begin{aligned}
\varphi_s^{(1)} &= \varphi_+(1)\varphi_+(2)\uparrow\uparrow,\\
\varphi_s^{(2)} &= \varphi_-(1)\varphi_-(2)\downarrow\downarrow,\\
\varphi_s^{(3)} &= \varphi_+(1)\varphi_-(2)+\varphi_+(2)\varphi_-(1)\uparrow\downarrow,\\
\varphi_a^{(4)} &= \varphi_+(1)\varphi_-(2)-\varphi_+(2)\varphi_-(1)\uparrow\downarrow .
\end{aligned} \tag{49.20}$$

Здесь спиновые функции, в отличие от пространственных, обозначены буквой φ. Значок $+$ у функции указывает, что проекция m_s спина электрона на направление, выделяемое орбитальным моментом, равна $+1/2$, а минус означает, что она равна $-1/2$. Это для наглядности отмечено направлениями стрелок, поставленных справа у волновых функций. Первая симметричная функция $\varphi_s^{(1)}$ означает, что проекции m_s спинов обоих электронов равны $+1/2$, вторая $\varphi_s^{(2)}$ — что обе они равны $-1/2$. В третьей симметричной функции $\varphi_s^{(3)}$ проекции спинов обоих электронов имеют противоположные знаки. Таким образом, в первых двух случаях полный спин обоих электронов будет 1, причем его проекции соответственно равны $+1$ и -1. Но если полный спин равен 1, то его проекции на избранное направление могут иметь три значения: $+1$, -1 и 0. Третий случай и осуществляется оставшейся симметричной волновой функцией $\varphi_s^{(3)}$.

Что касается антисимметричной спиновой функции $\varphi_a^{(4)}$, то ей соответствует проекция спина, равная нулю. А так как это единственная проекция, то полный спин в этом состоянии также равен нулю.

Четыре спиновые функции (49.20) можно комбинировать с пространственными функциями ψ_s и ψ_a, чтобы получить полную волновую функцию. Однако, поскольку речь идет о системах двух электронов, полная волновая функция их должна быть антисимметрична относительно операции перестановки частиц в полной волновой функции, характеризующейся пространственными и спиновыми координатами. Этому условию удовлетворяют только четыре комбинации:

$$\varphi_s^{(1)}\psi_a, \quad \varphi_s^{(2)}\psi_a, \quad \varphi_s^{(3)}\psi_a \quad \text{и} \quad \varphi_a^{(4)}\psi_s.$$

От их нормировки, не играющей роли в разбираемом вопросе, можно отвлечься. Пока что речь шла только о комбинациях в нулевом приближении. Но и во всех последующих приближениях из-за симметрии энергии взаимодействия Ze^2/r_{12} возникают волновые функции того же типа симметрии. Первые три функции образуют триплет с полным спином 1 и его проекциями $+1, -1, 0$. Этому триплету соответствует ортосостояние. Его энергетические уровни и спектральные линии, подчиняющиеся правилам отбора, расщепляются на три подуровня

и три компоненты из-за спин-орбитального взаимодействия. В этом состоянии атомы гелия имеют отличные от нуля магнитные моменты, а потому в магнитном поле дают зеемановское расщепление.

Волновая же функция $\varphi_a\psi_s$ описывает парасостояние. В этом случае проекция полного спина равна нулю. А так как это синглетное состояние, то и полный спин в парасостоянии будет нуль. Атомы гелия в синглетном состоянии не имеют магнитных моментов, а потому не дают зеемановского расщепления.

Что касается запрета интеркомбинаций, о котором было сказано в начале этого параграфа, то этот запрет связан с сохранением спина электронов при излучении, а потому не является абсолютно жестким и допускает исключения (см. § 39, п. 5).

ЗАДАЧИ

1. Зная энергию полной ионизации нейтрального атома гелия (см. табл. 5) в основном состоянии, найти энергию однократной ионизации того же атома, а затем энергию ионизации образовавшегося иона He^+.

Решение. После удаления одного электрона нейтральный атом гелия переходит в водородоподобный ион с зарядовым числом $Z = 2$. Энергия ионизации такого иона равна $13{,}539Z^2 = 54{,}156$ эВ. Поэтому для удаления первого электрона надо совершить работу $78{,}98 - 54{,}16 = 24{,}82$ эВ, а для удаления второго 54,156 эВ.

2. Оценить по порядку величины энергию однократной ионизации атома лития в основном состоянии.

Решение. Для оценки можно предположить, что два K-электрона атома лития экранируют поле ядра, как бы уменьшая зарядовое число до $Z = 3 - 2 = 1$. При таком рассмотрении сформулированная задача сводится к определению работы ионизации водородоподобного атома при удалении электрона из L-слоя. Она равна

$$\frac{13{,}539}{n^2} = \frac{13{,}539}{4} = 3{,}38 \text{ эВ}.$$

Экспериментальное значение равно 5,37 эВ (см. табл. 4).

3. Зная энергию однократной ионизации атома лития из основного состояния (см. предыдущую задачу), а также энергию полной ионизации гелиеподобного иона Li^+ (см. табл. 5), найти энергию ионизации нейтрального лития в основном состоянии при удалении из него двух электронов.

Ответ. 81,56 эВ.

§ 50. Химическая связь. Молекула водорода

1. Первоначально химия была вынуждена ввести предположение о существовании особых «химических сил», чтобы объяснить образование молекул из атомов, а также химические реакции между атомами или молекулами. На самом деле таких сил не существует — они сводятся к обычным электрическим (кулоновским) силам взаимодействия между заряженными частицами, т. е. электронами и ядрами, из которых состоят атомы. Но такое сведение возможно только на основе квантовой механики.

Различают два рода химических связей: *ионную* (*гетерополярную*) и *ковалентную* (*гомеополярную*). Ионная связь реализуется в тех случаях, когда молекулу можно представить как образование, состоящее из двух ионов: положительного и отрицательного (например, молекула NaCl состоит из ионов Na^+ и Cl^-). Если же это сделать невозможно, то связь называется гомеополярной.

Теория гетерополярной связи (носящая, впрочем, полуэмпирический характер) разрабатывалась, и не без успеха, еще до создания квантовой механики. Гомеополярная же связь не поддается классическому объяснению. Причина этого состоит в том, что силы, приводящие к такой связи, являются *обменными*, а потому требуют принципиально квантового рассмотрения. Уясним этот вопрос на примере простейшей гомеополярной молекулы — молекулы водорода. Впервые это было сделано в 1927 г. Гайтлером (1904–1981) и Ф. Лондоном (1900–1954), решившими задачу о молекуле водорода по методу теории возмущений. Впоследствии та же задача рассматривалась математически более совершенными методами, дающими значительно лучшую точность. Но прием Гайтлера и Лондона проще всего приводит к цели. Мы ограничимся принципиальным рассмотрением работы Гайтлера и Лондона, опуская фактическое проведение расчетов, так как нашей задачей является только уяснение физической природы гомеополярной связи, а не теоретический расчет числовых значений величин, характеризующих ее.

2. В квантовой механике задача о молекуле водорода сводится к решению уравнения Шредингера для стационарных состояний

$$\widehat{H}\psi(\mathbf{r}_1, \mathbf{r}_2) = \mathscr{E}\psi(\mathbf{r}_1, \mathbf{r}_2), \tag{50.1}$$

где через $\mathbf{r}_1(x_1, y_1, z_1)$ и $\mathbf{r}_2(x_2, y_2, z_2)$ обозначены координаты первого и второго электронов. Гамильтониан системы представляется в виде

$$\widehat{H} = \widehat{H}_1 + \widehat{H}_2 + \widehat{V}, \tag{50.2}$$

где

$$\begin{gathered} \widehat{H}_1 = -\frac{\hbar^2}{2\mu}\nabla_1^2 - \frac{e^2}{r_{1A}}, \quad \widehat{H}_2 = -\frac{\hbar^2}{2\mu}\nabla_2^2 - \frac{e^2}{r_{2B}}, \\ \widehat{V} \equiv V = -\frac{e^2}{r_{1B}} - \frac{e^2}{r_{2A}} + \frac{e^2}{r_{12}} + \frac{e^2}{R}. \end{gathered} \tag{50.3}$$

Гамильтониан $\widehat{H}_1$ помимо оператора кинетической энергии электрона 1, содержит в качестве второго слагаемого потенциальную энергию того же электрона в кулоновском поле ядра A. Аналогичный смысл имеет гамильтониан $\widehat{H}_2$. Что касается гамильтониана $\widehat{V}$, то он описывает четыре взаимодействия: электрона 1 с ядром B, электрона 2 с ядром A, электронов между собой и ядер также между собой. Здесь $r_{1A}, r_{1B}, r_{2A}, r_{2B}$ — расстояния электронов 1 и 2 до ядер A и B соответственно, r_{12} — расстояние между электронами, R — расстояние между ядрами A и B. Тяжелые ядра A и B могут считаться неподвижными

(такое приближение называется *адиабатическим*). Поэтому ψ рассматривается как функция координат только электронов, а не атомных ядер. Расстояние R между ядрами, существенное как раз в проблеме гомеополярной связи, в волновую функцию ψ входит как параметр.

Разумеется, полная волновая функция зависит не только от пространственных координат электронов, но и от их спинов. Такая полная функция, конечно, должна быть антисимметрична относительно перестановки электронов в пространстве, характеризуемом декартовыми и спиновыми координатами. Однако можно ограничиться только *пространственными волновыми функциями*, а влияние спинов учесть, расчленив задачу на две: 1) в одной *спиновая функция антисимметрична*, тогда пространственная функция, как было выяснено в предыдущем параграфе, должна быть симметричной; 2) в другой *спиновая функция симметрична*, тогда пространственная антисимметрична. В первом случае суммарный спин двух электронов равен нулю, и следовательно, возможно только одно (*синглетное по спину*) состояние с симметричной пространственной функцией. Во втором случае суммарный спин двух электронов равен 1 и может ориентироваться в пространстве тремя различными способами (*триплетное по спину* состояние).

3. Конечно, сформулированная задача не допускает аналитического решения. Необходимо использовать приближенные методы. Чтобы понять, как это можно сделать в рамках теории возмущений, допустим, что оба атома водорода разведены один от другого достаточно далеко. Тогда атомы можно считать независимыми и каждый из них характеризовать своей волновой функцией. А так как речь идет о химической связи невозбужденных атомов, следует принять, что оба атома находятся в *основном* состоянии. При достаточном разведении волновые функции электронов 1 и 2 почти не перекрываются, так что можно считать, что электрон 1 принадлежит «своему ядру» A, а «чужое ядро» B на его состояние влияет мало. Аналогично, можно считать, что электрон 2 принадлежит «своему ядру» B. Именно по этой причине мы выделили выше из полного гамильтониана $\widehat{H}$ рассматриваемой системы слагаемые $\widehat{H}_1$ и $\widehat{H}_2$, представляющие собой гамильтонианы отдельных атомов. Их мы будем рассматривать как *главные части* гамильтониана и только их учтем в нулевом приближении. Слагаемое же V представляет собой гамильтониан *взаимодействия атомов*. Его мы будем рассматривать как *малое возмущение*, учитываемое в первом и высших приближениях. Это мотивируется не только удаленностью атомов друг от друга, но также и тем, что в выражении (50.3) для V первые два слагаемых частично компенсируются остальными двумя, поскольку последние входят в гамильтониан с противоположными знаками.

В синглетном состоянии, когда спиновая функция антисимметрична, в нулевом приближении получается симметричная пространственная волновая функция с разделяющимися переменными:

$$\psi_s = C_s[\psi_A(1)\psi_B(2) + \psi_A(2)\psi_B(1)], \qquad (50.4)$$

где C_s — постоянная нормировки. Через ψ_A обозначена волновая функция электрона, когда он принадлежит ядру A, а через ψ_B — когда ядру B. Функции ψ_A и ψ_B считаются различными, хотя аналитически они и представляются одинаковыми выражениями. Это различие обусловлено тем, что они являются функциями различных переменных: в одном случае координаты отсчитываются от ядра A , в другом — от ядра B. В триплетном состоянии, когда спиновая функция симметрична, получается антисимметричная пространственная волновая функция

$$\psi_a = C_a[\psi_A(1)\psi_B(2) - \psi_A(2)\psi_B(1)], \tag{50.5}$$

где C_a — другая постоянная нормировки.

Постоянные C_s и C_a легко вычислить, хотя для наших целей существенно только то, что они различны. Определим для примера постоянную C_s. При этом мы будем считать, что сами функции ψ_A и ψ_B нормированы. Нормировка функции (50.4) означает, что $\int \psi_s^*\psi_s\, d\tau = 1$ или

$$\begin{aligned}|C_s|^2 \int [\psi_A^*(1)\psi_B^*(2) + \psi_A^*(2)\psi_B^*(1)]\,[\psi_A(1)\psi_B(2) + \\ + \psi_A(2)\psi_B(1)]\, d\tau_1\, d\tau_2 = |C_s|^2 \int \psi_A^*(1)\psi_A(1)\, d\tau_1 \int \psi_B^*(2)\psi_B(2)\, d\tau_2 + \\ + |C_s|^2 \int \psi_A^*(2)\psi_A(2)\, d\tau_2 \int \psi_B^*(1)\psi_B(1)\, d\tau_1 + \\ + |C_s|^2 \int \psi_A^*(2)\psi_A(1)\psi_B^*(1)\psi_B(2)\, d\tau_1\, d\tau_2 + \\ + |C_s|^2 \int \psi_A^*(1)\psi_A(2)\psi_B^*(2)\psi_B(1)\, d\tau_1\, d\tau_2 = 1,\end{aligned}$$

где $d\tau = d\tau_1\, d\tau_2$, $d\tau_1 = dx_1\, dy_1\, dz_1$, $d\tau_2 = dx_2\, dy_2\, dz_2$. Первые два интеграла в правой части равны единице в силу нормировки волновых функций ψ_A и ψ_B, последние два интеграла равны между собой в силу тождественности электронов. Обозначим один из этих интегралов, например первый, через S, т. е. введем обозначение

$$S = \int \psi_A^*(2)\psi_A(1)\psi_B^*(1)\psi_B(2)\, d\tau_1\, d\tau_2. \tag{50.6}$$

Тогда

$$|C_s|^2(2 + 2S) = 1,$$

откуда

$$C_s = \frac{1}{\sqrt{2}}\,\frac{1}{\sqrt{1+S}}, \tag{50.7}$$

если опустить несущественный фазовый множитель в выражении для C_s. Аналогично

$$C_a = \frac{1}{\sqrt{2}}\,\frac{1}{\sqrt{1-S}}, \tag{50.8}$$

Конечно, волновые функции ψ_s и ψ_a представляются выражениями (50.4) и (50.5) только в нулевом приближении (нуль в индексе

для простоты опущен). Однако характер симметрии пространственных волновых функций сохраняется в любом приближении.

4. Знания волновых функций нулевого приближения достаточно для нахождения энергии в первом приближении. В синглетном состоянии (т. е. состоянии с симметричной пространственной волновой функцией)

$$\mathscr{E}_s = \int \psi_s^*(\widehat{H}_1 + \widehat{H}_2 + V)\psi_s \, d\tau, \tag{50.9}$$

а в триплетном состоянии (т. е. состоянии с антисимметричной пространственной волновой функцией)

$$\mathscr{E}_a = \int \psi_a^*(\widehat{H}_1 + \widehat{H}_2 + V)\psi_a \, d\tau. \tag{50.10}$$

Так как ψ_s, рассматриваемая как функция координат $\mathbf{r}_1$, является собственной функцией оператора $\widehat{H}_1$, то $\widehat{H}_1\psi_s = \mathscr{E}_1\psi_s$. Аналогично $\widehat{H}_2\psi_s = \mathscr{E}_2\psi_s$, где $\mathscr{E}_1$ и $\mathscr{E}_2$ — энергии рассматриваемых изолированных атомов. А так как эти атомы тождественны и находятся в основном состоянии, то $\mathscr{E}_1 = \mathscr{E}_2 \equiv \mathscr{E}_0$, где $\mathscr{E}_0$ — энергия атома в основном состоянии. В точности то же самое имеет место и в случае функции ψ_a. Если учесть еще, что функции ψ_s и ψ_a нормированы, то ясно, что в обоих предыдущих выражениях гамильтониан $(\widehat{H}_1 + \widehat{H}_2)$ приводит к появлению слагаемого $2\mathscr{E}_0$. Далее, как видно из (50.3), величина V содержит член e^2/R, зависящий только от расстояния между ядрами, но не зависящий от координат электронов. Этот член в излагаемой теории играет роль параметра и добавляет к энергиям $\mathscr{E}_s$ и $\mathscr{E}_a$ величину e^2/R. Таким образом,

$$\mathscr{E}_s = 2\mathscr{E}_0 + \Delta\mathscr{E}_s, \quad \mathscr{E}_a = 2\mathscr{E}_0 + \Delta\mathscr{E}_a, \tag{50.11}$$

где

$$\Delta\mathscr{E}_s = \int \psi_s^* V \psi_s \, d\tau, \quad \Delta\mathscr{E}_a = \int \psi_a^* V \psi_a \, d\tau. \tag{50.12}$$

Подставляя в формулы (50.12) выражения (50.4) и (50.5) для ψ_s и ψ_a и учитывая тождественность электронов, получим

$$\Delta\mathscr{E}_s = \frac{K+S}{1+S}, \quad \Delta\mathscr{E}_a = \frac{K-S}{1-S}, \tag{50.13}$$

где введено обозначение

$$K = \int V\psi_A^*(1)\psi_A(1)\psi_B^*(2)\psi_B(2) \, d\tau_1 \, d\tau_2. \tag{50.14}$$

Величины $\Delta\mathscr{E}_s$ и $\Delta\mathscr{E}_a$ в первом приближении и представляют поправки к энергии электронов $2\mathscr{E}_0$, которые возникают из-за кулоновского взаимодействия атомов. В энергию этого взаимодействия ядра вносят член e^2/R. Кулоновское взаимодействие между электронами атомов носит двойственный характер. Оно выражается в одном случае через интеграл (50.14), в другом — через интеграл (50.6). В первом случае электрон 1 все время находится в одном и том же состоянии ψ_A,

а электрон 2 — в одном и том же состоянии ψ_B. В этом случае взаимодействие носит чисто классический характер, если только отвлечься от «размазывания» зарядов электронов в пространстве. Во втором случае каждый электрон находится одновременно и в состоянии ψ_A, и в состоянии ψ_B. Такое взаимодействие, как и в случае атома гелия, есть чисто квантовый эффект, не имеющий классического аналога. Оно называется *обменным взаимодействием* и возникает из-за тождественности электронов. В вычислениях тождественность проявляется в использовании симметричных и антисимметричных пространственных волновых функций.

5. Интегралы K и S, а с ними и энергии $\Delta\mathscr{E}_s$ и $\Delta\mathscr{E}_a$ являются функциями расстояния R между ядрами. На рис. 91 изображена зависимость этих энергий от расстояния R. По горизонтальной оси отложено отношение R/a, где a — боровский радиус. Из рисунка видно, что при сближении атомов в синглетном спиновом состоянии (спины антипараллельны) энергия взаимодействия сначала убывает, а затем быстро возрастает. Минимум достигается при расстоянии R, обозначаемом далее через R_0. В триплетном спиновом состоянии (спины параллельны) минимума не получается.

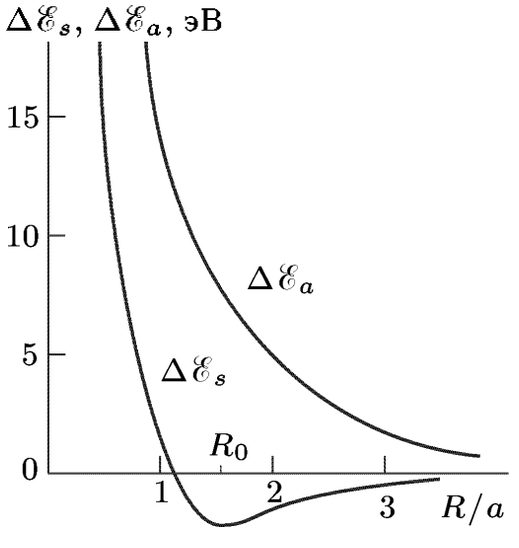


Рис. 91

Из рис. 91 видно, что *равновесное положение атомов существует только в синглетном спиновом состоянии*, а именно при $R = R_0$, когда энергия взаимодействия $\Delta\mathscr{E}_s$ обращается в минимум. *Только в синглетном состоянии и возможно образование молекулы*, причем R_0 имеет смысл равновесного расстояния между атомами в молекуле. *В триплетном состоянии молекула образоваться не может.* Гайтлер и Лондон в первом приближении теории возмущений нашли $R_0 = 1{,}51a \approx 0{,}080$ нм. Экспериментальное значение $R_0 = 0{,}07395$ нм. Лучшего согласия между теорией и опытом и не следовало ожидать, так как теория возмущений в первом приближении может считаться обоснованной только при $R \gg R_0$. Однако уже в первом приближении она приводит к качественно правильной картине взаимодействия атомов водорода в синглетном и триплетном состояниях. Более точные вариационные методы расчета привели в случае молекулы водорода к практически полному количественному согласию теории с опытом.

6. Различное поведение атомов в синглетном и триплетном состояниях легко понять на основе простых качественных соображений. В триплетном состоянии пространственная волновая функция (50.5) антисимметрична, а потому она имеет узел в плоскости симметрии, проходящей посередине между ядрами A и B перпендикулярно к соединяющей эти ядра прямой, так как в этой плоскости в силу тождественности электронов $\psi_A(1) = \psi_A(2)$, $\psi_B(1) = \psi_B(2)$. По

той же причине в синглетном состоянии пространственная волновая функция (50.4) в той же плоскости симметрии обращается в максимум. Поэтому в синглетном состоянии плотность вероятности нахождения электронов вблизи плоскости симметрии велика, тогда как в триплетном состоянии она близка к нулю. Таким образом, в синглетном состоянии при $R > R_0$ кулоновское отталкивание ядер компенсируется противоположно направленным действием электронов, так что при $R = R_0$ результирующая сила взаимодействия атомов меняет знак, т. е. отталкивание сменяется притяжением. Отрицательный заряд электронов, возникающий между ядрами A и B, и стягивает их, образуя связанное состояние. В триплетном же состоянии из-за малой вероятности нахождения электронов посередине между ядрами компенсирующее действие электронов недостаточно, чтобы получилось притяжение между атомами.

На малых расстояниях между ядрами атомы всегда отталкиваются независимо от того, находятся ли они в синглетном или триплетном состояниях. Силы отталкивания особенно велики, когда расстояние между ядрами становится меньше боровского радиуса a, так как в этом случае в средней точке между атомами, где волновые функции ψ_A и ψ_B совпадают, эти функции ничтожно малы. (Согласно полуклассической теории Бора, электрон вообще не может приблизиться к ядру на расстояние меньше a.) На малых расстояниях силы, с которыми отталкиваются атомы, обусловлены не столько кулоновским отталкиванием протонов, сколько отталкиванием электронов. Кулоновское отталкивание ядер меняется по степенному закону — обратно пропорционально квадрату расстояния между ядрами, тогда как кулоновское отталкивание электронов, как показывает расчет, приводит к более быстрому экспоненциальному возрастанию сил отталкивания с уменьшением расстояния между ядрами.

На расстояниях $R \gg a$, как также показывает расчет, силы притяжения (в синглетном состоянии) и отталкивания (в триплетном состоянии) экспоненциально убывают с расстоянием.

Как видно из формул (50.13), количественное различие во взаимодействии атомов в синглетном и триплетном состояниях обусловлено обменным интегралом (50.6), зависящим от расстояния R как от параметра. «Химические силы» могут проявиться только при таких R, когда подынтегральное выражение в обменном интеграле (50.6) заметно отлично от нуля, т. е. в области «перекрытия» электронных волновых функций $\psi_A(1)$ и $\psi_B(1)$ [или, что то же, $\psi_A(2)$ и $\psi_B(2)$] обоих атомов.

7. Хотя возможность образования нейтральной молекулы из атомов водорода в синглетном состоянии и обусловлена спином электрона, но само по себе силовое (магнитное) взаимодействие между спинами электронов при этом не играет никакой роли. Оно слишком слабо по сравнению с обменным взаимодействием. Влияние спина проявляется только в ограничении, накладываемом на полную электронную волновую функцию системы из двух атомов водорода, которое требует, чтобы она была антисимметрична относительно перестановки электронов.

Связанное состояние из двух ядер водорода (протонов) может образоваться и при наличии только одного электрона, когда ни о каком обменном взаимодействии между электронами говорить не приходится. Таким состоянием является ионизованная молекула водорода H_2^+. Связь осуществляется единственным электроном, в равной мере принадлежащим обоим ядрам. Возможность возникновения сил притяжения между ядрами можно пояснить на классической модели. Пусть отрицательный электрон находится посередине между положительными ядрами (рис. 92). Ядра отталкиваются одно от другого с силой $F_{++} = e^2/R$, тогда как электрон и ядро притягиваются с большей силой $F_{+-} = 4e^2/R$. Этот классический пример показывает только, почему могут появиться силы притяжения. Но классическая физика, конечно, ничего не может сказать, возможно ли устойчивое связанное состояние из рассматриваемых трех частиц. На этот вопрос положительный ответ дает только квантовая механика. Задача сводится к решению уравнения Шредингера для одного электрона в электрическом поле двух (неподвижных) положительно заряженных ядер. На этом вопросе мы останавливаться не можем.

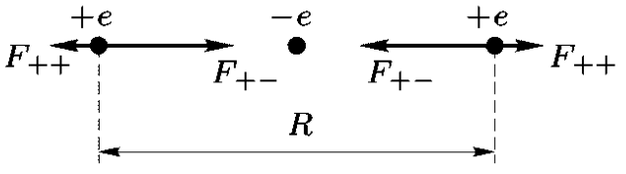


Рис. 92

8. Большой интерес представляет случай, когда химическая связь протонов осуществляется не электронами, а отрицательными мюонами. Тогда образуется *мюонная молекула*, состоящая из двух протонов и двух отрицательных мюонов. Один из протонов или оба вместе можно заменить дейтроном или тритоном, поскольку заряд этих частиц такой же, как у протона. Мюон — нестабильная частица, заряд которой равен заряду электрона. Он распадается на электрон и два нейтрино в среднем за время жизни $2 \cdot 10^{-6}$ с. В течение примерно такого же времени могут существовать и мюонные молекулы. Так как масса мюона в 207 раз больше массы электрона, то в такое же число раз меньше, чем у водорода, и среднее расстояние между ядрами мюонной молекулы, т. е. в мюонной молекуле оно составляет $0{,}07395 : 207 = 3{,}57 \cdot 10^{-4}$ нм $= 3{,}57 \cdot 10^{-11}$ см. Если ядрами служат протон и дейтрон, то на таком расстоянии между ними может произойти ядерная реакция с образованием ядра ^{3}He и выделением энергии $\approx 5{,}4$ МэВ. Следовательно, мюоны могут играть роль катализаторов в указанной ядерной реакции. Аналогичную роль они могут играть и в реакциях дейтерия с дейтерием или с тритием.

9. Построение удовлетворительной количественной теории гомеополярной химической связи сложных молекул, из-за большого количества электронов в них, встречает очень большие математические трудности. Такая теория еще не создана. Поэтому для объяснения качественных особенностей химической связи таких молекул в химии пользуются модельными представлениями, основанными на распространении теории молекулы водорода на случай сложных молекул. На этом вопросе мы останавливаться не будем.

§ 51. Параводород и ортоводород

1. Протоны, как и электроны, обладают спином $1/2$ (в единицах $\hbar$). Поэтому полная волновая функция системы, состоящей из этих частиц, должна быть антисимметрична относительно перестановки двух протонов. Спиновый магнитный момент протона примерно в 657,5 раза меньше, чем у электрона. Поэтому еще с большим основанием, чем для электронов, можно сказать, что в молекуле водорода полная протонная волновая функция может быть представлена в виде произведения пространственной волновой функции на спиновую. Такие же рассуждения, какие были проведены для двух электронов, показывают, что в молекуле водорода могут реализоваться только следующие два состояния.

1. Состояние, в котором спиновая волновая функция протонов антисимметрична, а следовательно, пространственная функция симметрична. В этом случае водород называется *параводородом*. Образно можно сказать, что в молекуле параводорода спины протонов антипараллельны.

2. Состояние, в котором спиновая волновая функция протонов симметрична, а следовательно, пространственная антисимметрична. На образном языке это означает, что в этом случае спины протонов параллельны. Такой водород называется *ортоводородом*.

Так как в молекуле водорода электронные спины всегда взаимно компенсируются, то ее спин I обусловлен только спинами ядер. Для молекулы параводорода $I = 0$, для молекулы ортоводорода $I = 1$.

В качестве координат, характеризующих пространственную волновую функцию двух протонов, удобно взять декартовы координаты центра масс молекулы, расстояние между ядрами (протонами) и два угла (угловые координаты), определяющие направление соединяющей их прямой. Координаты центра масс не меняются при перестановке протонов, от них можно отвлечься. Расстояние между ядрами также не меняется при их перестановке. Оно меняется из-за колебаний ядер относительно друг друга вдоль соединяющей их прямой. Энергия, потребная для возбуждения таких колебаний, довольно велика (в температурной шкале ей соответствуют тысячи Кельвинов). Мы будем предполагать, что относительные колебания ядер не возбуждены (нулевые колебания, конечно, остаются). Меняются только угловые координаты, т. е. возбуждены вращения молекулы. Перестановка протонов эквивалентна повороту молекулы на 180° вокруг ее центра масс. Следовательно, в случае параводорода при таком повороте пространственная волновая функция протонов не меняется, а в случае ортоводорода меняет знак. Из явного выражения угловой волновой функции протонов (которое мы не приводим) видно, что в случае параводорода это имеет место при $L = 0, 2, 4, \ldots$, а в случае ортоводорода при $L = 1, 3, 5, \ldots$, где L — квантовое число, определяющее квадрат орбитального углового момента молекулы при ее вращении. Таким образом, вращательные уровни параводорода характеризуются *четными*, а ортоводорода —

нечетными квантовыми уровнями. Отсюда видно, что наиболее глубокий энергетический уровень молекулы параводорода лежит ниже соответствующего уровня молекулы ортоводорода.

2. Для превращения параводорода в ортоводород или обратно требуется переориентировать спины ядер. Однако благодаря слабому взаимодействию спиновых магнитных моментов ядер такой процесс при столкновениях молекул водорода крайне маловероятен. При обычных условиях параводород практически не переходит в ортоводород и обратно. Это означает, что квантовые переходы в молекуле водорода подчиняются хотя и не абсолютно строгому, но практически строгому правилу отбора

$$\Delta I = 0, \tag{51.1}$$

из которого следует

$$\Delta L = \pm 2. \tag{51.2}$$

Значит, *переходы могут осуществляться только между четными или только между нечетными вращательными квантовыми состояниями.* Если содержание пара- и ортоводорода в газе разное, то это проявляется в закономерном чередовании интенсивностей спектральных линий во вращательном (полосатом) спектре молекулярного водорода.

Итак, обычный молекулярный водород ведет себя как смесь двух газов, обладающих несколько отличающимися свойствами. Так как наиболее глубокое вращательное состояние ($L = 0$) есть парасостояние, то устойчивой равновесной модификацией при низких температурах является параводород. Уже при 20 K водород в равновесном состоянии является практически чистым параводородом. Однако простое охлаждение до этой температуры не приводит к равновесию. Равновесного состояния все же можно достичь с помощью катализаторов (например, десорбцией с угля предварительно адсорбированного им водорода). Этот процесс, проводимый при 20 K, дает возможность получить практически чистый параводород. Параводород очень устойчив и может храниться в чистых стеклянных сосудах несколько недель.

С повышением температуры соотношение пара- и ортоводорода в равновесном состоянии изменяется. Так как параводороду соответствует только одно вращательное квантовое состояние ($I = 0$), а ортоводород может существовать в трех состояниях (со спинами -1, 0, $+1$), то статистические веса этих состояний относятся как $1 : 3$. При высоких температурах (практически уже при 300 K) влияние экспоненциального больцмановского множителя перестает сказываться, так что относительное содержание пара- и ортоводорода в равновесной смеси становится равным $1 : 3$. Параводород превращается в ортоводород в присутствии катализаторов (например, платинированного асбеста) при температуре в несколько сот градусов Цельсия или в электрическом разряде. Но так как не существует температуры, при которой равновесное состояние соответствовало бы чистому ортоводороду, то получить чистый ортоводород невозможно. Нижним пределом возможного содержания параводорода в равновесной смеси является 25 %, как это следует из отношения статистических весов пара- и ортоводорода.

3. При понижении температуры, начиная от комнатной, теплоемкость газообразного молекулярного водорода уменьшается. При комнатной температуре молярная теплоемкость водорода составляет классическое значение $C_V = (5/2)R$ и стремится при понижении температуры также к классическому значению $C_V = (3/2)R$ (см. т. II, § 85). Такое уменьшение теплоемкости связано с уменьшением вклада в теплоемкость, который вносят вращения молекул. Однако полного количественного согласия теории с опытом удалось достигнуть только после учета того обстоятельства, что водород является смесью пара- и ортоводорода. Теплоемкости компонентов этой смеси различны из-за того, что вращательные уровни ортоводорода сдвинуты вверх относительно вращательных уровней параводорода. Равновесное соотношение между пара- и ортоводородом достигается за длительное время — очень большое по сравнению со временем нагревания или охлаждения газа, которое требуется для измерения его теплоемкости. Поэтому в неравновесном состоянии теплоемкость газообразного водорода будет зависеть от относительного содержания в нем пара- и ортоводорода. Если же неравновесную смесь на длительное время предоставить саму себе, то ее состав со временем будет меняться. В соответствии с этим и теплоемкость смеси будет медленно меняться во времени.

§ 52. Молекулярные силы

1. Силы взаимодействия молекул и атомов сводятся к кулоновскому притяжению и отталкиванию положительно заряженных атомных ядер и отрицательных электронов. К ним относятся, во-первых, *химические* или *валентные силы*, рассмотренные в § 50. Эти силы действуют на близких расстояниях. Они связаны с противоположной ориентацией спинов атомных ядер и определяются степенью перекрытия электронных волновых функций взаимодействующих атомов. Так как электронная плотность убывает экспоненциально с увеличением расстояния от центра атома, то и химические силы убывают экспоненциально с возрастанием расстояния между атомами. В этом параграфе предполагается, что атомы и молекулы химически не реагируют, так что химические силы между ними не действуют.

Во-вторых, к молекулярным силам относятся *короткодействующие силы отталкивания*, возникающие при проникновении электронных оболочек взаимодействующих частиц при их сближении. Такие силы также экспоненциально убывают с расстоянием между взаимодействующими частицами и притом очень быстро. Это позволяет в кинетической теории газов пользоваться упрощенной моделью молекул и атомов, заменяя их твердыми упругими шарами, при столкновении которых развиваются бесконечно большие силы отталкивания.

В-третьих, между нейтральными атомами и молекулами действуют еще силы притяжения. Они убывают с расстоянием не по экспоненциальному, а по степенному закону, т. е. значительно более медленно. В отличие от короткодействующих сил, их можно было бы назвать

дальнодействующими. Эти силы возникают при взаимодействии электрических дипольных моментов атомов и молекул или вообще при деформации их электронных оболочек. Если молекулы полярные, т. е. обладают постоянными дипольными моментами, то наличие такого взаимодействия понятно. Однако взаимодействие существует, и даже, как правило, превосходит взаимодействие полярных молекул, когда постоянными дипольными моментами молекулы не обладают. В этом случае силы называются *силами Ван-дер-Ваальса* или *дисперсионными силами.*

Откуда же может возникнуть дипольное взаимодействие, если у молекул нет дипольных моментов? Дело в том, что у возбужденных молекул и атомов дипольные моменты появляются, хотя они и меняются во времени. Но если молекула или атом не возбуждены, то с точки зрения классических представлений дипольных моментов у них не должно быть. Не так обстоит дело с точки зрения квантовой механики. Она в согласии с опытом утверждает, что у невозбужденных атома или молекулы есть нулевая энергия, которой соответствуют весьма оживленные нулевые колебания. С наличием нулевых колебаний и связано происхождение сил Ван-дер-Ваальса. Нулевые колебания не связаны с тепловым движением, т. е. не зависят от температуры. Поэтому и силы Ван-дер-Ваальса не должны зависеть от температуры. Опыт показывает, что это действительно так.

2. Роль нулевых колебаний в происхождении сил Ван-дер-Ваальса и квантовую природу этих сил проще всего уяснить не на реальных атомах, а на идеализированных моделях их. Будем моделировать каждый из одинаковых взаимодействующих атомов одномерным гармоническим осциллятором с массой μ и собственной частотой колебаний ω_0. Будем предполагать, что колебания совершаются вдоль прямой, соединяющей осцилляторы. При отклонении из положений равновесия на x_1 и x_2 осцилляторы получат потенциальные энергии $(1/2)\mu\omega_0^2 x_1^2$ и $(1/2)\mu\omega_0^2 x_2^2$. При этом у них появятся дипольные моменты $p_1 = ex_1$ и $p_2 = ex_2$. Потенциальная энергия взаимодействия этих диполей равна $-p_1 p_2/R^3 = -e^2 x_1 x_2/R^3$, где R — расстояние между осцилляторами. В соответствии с этим гамильтониан системы рассматриваемых двух связанных осцилляторов представится выражением

$$\widehat{H} = \widehat{T}_1 + \widehat{T}_2 + \frac{1}{2}\,\mu\omega_0^2(x_1^2 + x_2^2) - \lambda x_1 x_2, \tag{52.1}$$

где $\lambda = e^2/R^3$, а $\widehat{T}_1$ и $\widehat{T}_2$ — операторы кинетических энергий осцилляторов, т. е.

$$\widehat{T}_1 = -\frac{\hbar^2}{2\mu}\,\frac{\partial^2}{\partial x_1^2}, \quad \widehat{T}_2 = -\frac{\hbar^2}{2\mu}\,\frac{\partial^2}{\partial x_2^2}. \tag{52.2}$$

Вместо x_1 и x_2 введем так называемые *нормальные координаты* q_1 и q_2 системы связанных осцилляторов, определяемые формулами

$$x_1 = \frac{1}{\sqrt{2}}\,(q_1 + q_2), \quad x_2 = \frac{1}{\sqrt{2}}\,(q_1 - q_2). \tag{52.3}$$

Тогда

$$\frac{\partial\psi}{\partial q_1} = \frac{\partial\psi}{\partial x_1}\frac{\partial x_1}{\partial q_1} + \frac{\partial\psi}{\partial x_2}\frac{\partial x_2}{\partial q_1} = \frac{1}{\sqrt{2}}\left(\frac{\partial\psi}{\partial x_1} + \frac{\partial\psi}{\partial x_2}\right)$$

и далее

$$\frac{\partial^2\psi}{\partial q_1^2} = \frac{1}{2}\left(\frac{\partial^2\psi}{\partial x_1^2} + 2\frac{\partial^2\psi}{\partial x_1\partial x_2} + \frac{\partial^2\psi}{\partial x_2^2}\right),$$

$$\frac{\partial^2\psi}{\partial q_2^2} = \frac{1}{2}\left(\frac{\partial^2\psi}{\partial x_1^2} - 2\frac{\partial^2\psi}{\partial x_1\partial x_2} + \frac{\partial^2\psi}{\partial x_2^2}\right).$$

Отсюда

$$\frac{\partial^2}{\partial q_1^2} + \frac{\partial^2}{\partial q_2^2} = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2}. \tag{52.4}$$

Для потенциальной энергии получаем

$$\frac{1}{2}\,\mu\omega_0^2(x_1^2 + x_2^2) = \frac{1}{2}\,\mu\omega_1^2 q_1^2 + \frac{1}{2}\,\mu\omega_2^2 q_2^2.$$

Здесь введены обозначения

$$\omega_1^2 = \omega_0^2 - \frac{\lambda}{\mu}, \quad \omega_2^2 = \omega_0^2 + \frac{\lambda}{\mu}. \tag{52.5}$$

Таким образом, гамильтониан системы связанных осцилляторов может быть представлен в виде

$$\widehat{H} = \widehat{H}_1 + \widehat{H}_2, \tag{52.6}$$

где

$$\widehat{H}_1 = -\frac{\hbar^2}{2\mu}\frac{\partial^2}{\partial q_1^2} + \frac{1}{2}\,\mu\omega_1^2 q_1^2, \quad \widehat{H}_2 = -\frac{\hbar^2}{2\mu}\frac{\partial^2}{\partial q_2^2} + \frac{1}{2}\,\mu\omega_2^2 q_2^2. \tag{52.7}$$

Это значит, что колебание системы в общем случае получается суперпозицией гармонических колебаний с частотами ω_1 и ω_2. Понятно, что в точности такое же утверждение получается и в классической механике. Колебания с частотами ω_1 и ω_2 называются *нормальными колебаниями* рассматриваемой системы. Целесообразность введения нормальных координат обусловлена тем, что в этих координатах задача о колебаниях связанной системы сводится к задачам о колебаниях двух *независимых* осцилляторов с разными частотами ω_1 и ω_2.

3. Наименьшая или нулевая энергия системы получится тогда, когда каждый из осцилляторов совершает нулевые колебания со своей частотой. Эта нулевая энергия системы равна

$$\mathscr{E}_0 = \frac{\hbar}{2}\,(\omega_1 + \omega_2). \tag{52.8}$$

Частоты ω_1 и ω_2 зависят от расстояния R между осцилляторами. Если R достаточно велико, то

$$\omega_1 = \omega_0\left(1 - \frac{\lambda}{\mu\omega_0^2}\right)^{1/2} \approx \omega_0\left(1 - \frac{\lambda}{2\mu\omega_0^2} - \frac{\lambda^2}{8\mu^2\omega_0^4}\right),$$

$$\omega_2 = \omega_0\left(1 + \frac{\lambda}{\mu\omega_0^2}\right)^{1/2} \approx \omega_0\left(1 + \frac{\lambda}{2\mu\omega_0^2} - \frac{\lambda^2}{8\mu^2\omega_0^4}\right).$$

Таким образом,

$$\mathscr{E}_0 = \hbar\omega_0 - \frac{\hbar\lambda^2}{8\mu^2\omega_0^3} = \hbar\omega_0 - \frac{\hbar}{8\mu^2\omega_0^3}\,\frac{e^4}{R^6}, \tag{52.9}$$

т. е. нулевая энергия системы связанных осцилляторов является функцией расстояния R между ними, а потому она играет роль потенциальной энергии взаимодействия их. Величина $\hbar\omega_0$ имеет смысл нулевой энергии осцилляторов при отсутствии связи между ними. Отбрасывая эту несущественную аддитивную постоянную, для потенциальной энергии можно написать

$$U(R) = -\frac{\hbar e^4}{8\mu^2\omega_0^3}\,\frac{1}{R^6}. \tag{52.10}$$

Она отрицательна, а потому взаимодействие между осцилляторами носит характер притяжения и притом независимо от того, в каких фазах колеблются осцилляторы. Соответствующие силы притяжения и будут силами Ван-дер-Ваальса для наших идеализированных атомов. Квантовая природа этих сил видна уже из того, что в формулу (52.10) в качестве множителя входит постоянная Планка $\hbar$ [1]).

Если дипольные моменты осцилляторов взаимно перпендикулярны, то потенциальная энергия дипольного взаимодействия равна нулю, и силы притяжения не возникают. Если же они направлены под углом друг к другу, то их можно разложить на составляющие вдоль прямой, соединяющей осцилляторы, и перпендикулярно к ней, так что и в этом случае возникает притяжение.

---

[1]) Притяжение является следствием того, что дипольное взаимодействие осцилляторов уменьшает их общую нулевую энергию. Это можно доказать и не прибегая к разложению частот ω_1 и ω_2 в степенные ряды. Действительно, из (52.5) получаем

$$\omega_1^2 + \omega_2^2 = 2\omega_0^2.$$

Поэтому, рассматривая квадраты частот как стороны прямоугольника и квадрата, на основании известной геометрической теоремы находим $\omega_1^2\omega_2^2 < \omega_0^4$ или $\omega_1\omega_2 < \omega_0^2$. Значит,

$$\omega_1^2 + 2\omega_1\omega_2 + \omega_2^2 < 4\omega_0^2, \quad \text{или} \quad (\omega_1 + \omega_2)^2 < 4\omega_0^2,$$

$$\omega_1 + \omega_2 < 2\omega_0.$$

А к установлению последнего неравенства и сводится доказываемое утверждение.

4. Формулу (52.10) можно преобразовать, введя в нее поляризуемость β атома в постоянном электрическом поле по формуле $p = ex = \beta E$. Так как в случае гармонического осциллятора сила, действующая на заряд e, равна $\mu\omega_0^2 x$, то при равновесии $\mu\omega_0^2 x = eE = ex/\beta$, откуда $e^2/(\mu\omega_0^2) = \beta$. Эта формула получена здесь классически, но она может быть найдена также из дисперсионной формулы квантовой механики. (Поэтому-то силы Ван-дер-Ваальса и называют также дисперсионными.) В результате получается

$$U(R) = -\frac{\varepsilon\beta^2}{8}\,\frac{1}{R^6}, \tag{52.11}$$

где $\varepsilon = \hbar\omega$ — разность между соседними энергетическими уровнями гармонического осциллятора.

Результат (52.11) существенно не меняется и для реальных атомов. В этом случае во втором приближении теории возмущений квантовая механика дает

$$U(R) = -k\frac{I\beta^2}{R^6}, \tag{52.12}$$

где I — энергия ионизации атома, а k — числовой коэффициент порядка единицы.

В заключение поясним с другой точки зрения, почему дипольное взаимодействие молекул приводит к притяжению между ними. Притяжение наблюдается тогда, когда молекулы обращены друг к другу разноименными зарядами (рис. 93 *а*), отталкивание — когда одноименными (рис. 93 *б*). В первом положении потенциальная энергия взаимодействия молекул, зависящая от их взаимной ориентации, минимальна, во втором максимальна. Первому положению соответствует устойчивое равновесие, второму — неустойчивое. Согласно формуле Больцмана при термодинамическом равновесии, например газа, первое положение более, а второе менее вероятно. В первом положении молекулы проводят большее время, чем во втором. Результирующим эффектом взаимодействия будет притяжение.

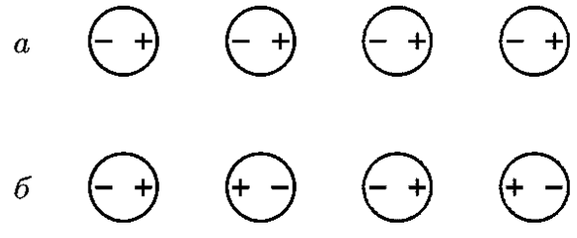


Рис. 93

Глава VII

НЕКОТОРЫЕ МАКРОСКОПИЧЕСКИЕ КВАНТОВЫЕ ЯВЛЕНИЯ

§ 53. Возможные состояния частицы в ограниченном объеме

1. В § 82 т. II были изложены основы квантовых статистик Бозе–Эйнштейна и Ферми–Дирака для идеального газа. Были получены выражения для среднего числа частиц в одном квантовом состоянии при заданной температуре T. Однако что́ это за состояния, которые может принимать частица, и каково их число в заданном интервале энергий — об этом до изложения основ квантовой механики, естественно, ничего не могло быть сказано. Теперь мы решим этот вопрос для произвольной нерелятивистской частицы, подобно тому как он был решен для фотона (см. т. IV, § 117).

Пусть частица находится внутри сосуда, который для простоты вычислений будем считать кубом со стороной L с непроницаемыми стенками. Стенка представляет собой потенциальный барьер: внутри сосуда потенциальная энергия частицы постоянна и принимается равной нулю, а при приближении к стенке неограниченно возрастает до бесконечности, оставаясь таковой при переходе через стенку. Такие предположения относительно потенциального барьера необходимо ввести, чтобы полностью исключить возможность выхода частицы из сосуда (см. § 24 и 28).

Сначала будем предполагать, что частица не обладает спином. Стационарное состояние такой частицы внутри сосуда описывается волновой функцией ψ, удовлетворяющей уравнению Шредингера

$$\nabla^2\psi + k^2\psi = 0, \tag{53.1}$$

где

$$k^2 = \frac{2\mu\mathscr{E}}{\hbar^2} = \frac{\mathbf{p}^2}{\hbar^2}. \tag{53.2}$$

На стенках сосуда функция ψ должна обращаться в нуль, чтобы частица не могла выйти из сосуда.

Частное решение уравнения (53.1) можно найти методом разделения переменных, полагая $\psi = \psi_x(x)\psi_y(y)\psi_z(z)$, так что

$$\nabla^2\psi = \psi_x''(x)\psi_y(y)\psi_z(z) + \psi_x(x)\psi_y''(y)\psi_z(z) + \psi_x(x)\psi_y(y)\psi_z''(z).$$

Подставив это значение в уравнение (53.1) и разделив его на ψ, получим

$$\frac{\psi_x''(x)}{\psi_x(x)} + \frac{\psi_y''(y)}{\psi_y(y)} + \frac{\psi_z''(z)}{\psi_z(z)} = -k^2.$$

Это уравнение должно выполняться, каковы бы ни были значения x, y, z. Первое слагаемое есть функция только x, а потому оно не зависит от того, какие значения имеют y и z. Если фиксировать y и z, то последние два слагаемых в левой части уравнения станут постоянными. Но тогда будет постоянным и первое слагаемое $\psi_x''(x)/\psi_x(x)$. Такое же рассуждение можно провести и в отношении остальных двух слагаемых. Таким образом, должны выполняться уравнения

$$\frac{\psi_x''(x)}{\psi_x(x)} = -k_x^2, \quad \frac{\psi_y''(y)}{\psi_y(y)} = -k_y^2, \quad \frac{\psi_z''(z)}{\psi_z(z)} = -k_z^2, \tag{53.3}$$

где k_x, k_y, k_z — постоянные, удовлетворяющие соотношению

$$k_x^2 + k_y^2 + k_z^2 = k^2. \tag{53.4}$$

Все эти постоянные должны быть положительными; в противном случае нельзя удовлетворить граничным условиям, как это будет ясно из дальнейшего. Общее решение первого уравнения (53.3) имеет вид

$$\psi_x(x) = a \sin k_x x + b \cos k_x x.$$

Постоянная b должна равняться нулю, так как в силу граничного условия $\psi_x(0) = b = 0$. Граничное же условие на стенке $x = L$ дает $\psi_x(L) = a \sin k_x L = 0$, так что $\sin k_x L = 0$. Аналогичные соотношения имеют место и для k_y, k_z. Следовательно,

$$k_x = \frac{\pi}{L} n_x, \quad k_y = \frac{\pi}{L} n_y, \quad k_x = \frac{\pi}{L} n_z, \tag{53.5}$$

где n_x, n_y, n_z — целые числа: n_x, n_y, $n_z = 1, 2, 3, \ldots$ (отрицательные значения их не приводят к новым линейно независимым решениям, а значения $n_x = n_y = n_z = 0$ дают тривиальные решения $\psi_x = 0$, $\psi_y = 0$, $\psi_z = 0$ и, следовательно, $\psi = 0$).

Таким образом, получается частное решение уравнения (53.1)

$$\psi = \sin k_x x \sin k_y y \sin k_z z, \tag{53.6}$$

обращающееся в нуль на стенках сосуда. Соответствующая волновая функция, зависящая от времени, представляет собой стоячую волну. Суперпозиция таких стоячих волн с постоянными амплитудами и будет общим выражением для волновой функции внутри сосуда. Каждой тройке целых чисел n_x, n_y, n_z соответствует одна стоячая волна, т. е. одно стационарное квантовое состояние частицы.

2. Чтобы найти число $dZ(k)$ стационарных состояний в интервале волновых чисел от k до $k + dk$, вообразим пространственную кубическую решетку, ячейки которой являются кубиками со стороной π/L и объемом π^3/L^3. Тогда число $dZ(k)$ будет равно числу узлов в зазоре положительного октанта такой решетки, который заключен между

сферами с радиусами k и $k+dk$, т. е. отношению объема такого зазора к объему ячейки:

$$dZ(k) = \frac{1}{8}\,\frac{4\pi k^2\,dk}{\pi^3/L^3} = \frac{Vk^2\,dk}{2\pi^2}, \tag{53.7}$$

где $V = L^3$ — объем сосуда.

Для электронов (и вообще частиц со спином $1/2$) выражение (53.7) следует удвоить, так как каждой пространственной волновой функции в этом случае соответствуют два спиновых состояния с противоположно ориентированными спинами. Для фотонов выражение (53.7) следует также удвоить, чтобы учесть возможность двух взаимно перпендикулярных поляризаций. В этих случаях

$$dZ_{\text{эл}}(k) = dZ_{\text{ф}}(k) = \frac{Vk^2}{\pi^2}\,dk. \tag{53.8}$$

Последняя формула, конечно, не может быть обоснована с помощью уравнения (53.1), так как уравнение Шредингера для фотонов неприменимо. Однако она уже была выведена нами с помощью уравнений Максвелла при рассмотрении вопросов теплового излучения (см. т. IV, § 117).

От волновых чисел можно перейти к импульсам, пользуясь формулой $p = \hbar k$ и, следовательно, $dp = \hbar\,dk$. В этих переменных

$$dZ_{\text{эл}}(p) = dZ_{\text{ф}}(p) = \frac{Vp^2}{\hbar^3\pi^2}\,dp. \tag{53.9}$$

Можно также в качестве переменной принять энергию частицы $\mathscr{E}$. Однако из-за различной связи энергии с импульсом в этом случае получаются различные выражения для электронов и для фотонов. Для электронов $\mathscr{E} = p^2/2\mu$,

$$dZ_{\text{эл}} = \frac{V\sqrt{2\mu^3\mathscr{E}}}{\pi^2\hbar^3}\,d\mathscr{E}. \tag{53.10}$$

Для фотонов $p = \mathscr{E}/c$,

$$dZ_{\text{ф}} = \frac{V\mathscr{E}^2}{\pi^2c^3\hbar^3}\,d\mathscr{E}. \tag{53.11}$$

§ 54. Теория Дебая теплоемкости твердых тел

1. Как было показано в т. II (см. § 69 и 85 указанного тома), применение квантовой теории позволило Эйнштейну уже в 1906 г. дать принципиальное объяснение падения теплоемкости твердых тел вблизи абсолютного нуля температур. Эйнштейн рассматривал твердое тело как совокупность N *независимых* частиц (гармонических осцилляторов), колеблющихся около положений равновесия с одной и той же частотой ω. Средняя энергия, приходящаяся на одну степень свободы,

в этом случае определяется формулой Планка

$$\overline{\varepsilon} = \frac{\hbar\omega}{e^{\hbar\omega/kT} - 1}, \tag{54.1}$$

в которой опущен член $\hbar\omega/2$, представляющий нулевую энергию осциллятора. Этот член надо учитывать в тех случаях, когда существенна амплитуда колебаний, например в случае зависимости рассеяния рентгеновских лучей от температуры. Но в вопросе о теплоемкости нулевая энергия роли не играет, поскольку она не зависит от температуры. По этой причине она и опускается в дальнейшем.

При высоких температурах формула (54.1) переходит в классическое выражение $\overline{\varepsilon} = kT$, а потому в вопросе о теплоемкости приводит к закону Дюлонга и Пти. При низких температурах формула, полученная Эйнштейном, дает убывание теплоемкости с температурой, причем теплоемкость стремится к нулю в согласии с тем, что требует эмпирически установленная теорема Нернста. Однако согласие теории с опытом получается только качественное. Так, по формуле Эйнштейна вблизи абсолютного нуля теплоемкость твердого тела должна убывать с температурой по *экспоненциальному* закону, тогда как опыт приводит к более медленному убыванию по *степенному* закону. Можно было думать, что такое расхождение теории с опытом связано не с принципиальными недостатками теории, а обусловлено грубостью примененной модели твердого тела. В теории Эйнштейна осцилляторы считаются независимыми. Но будет гораздо ближе к действительности, если их рассматривать *связанными*. В таком случае в теле возбудится не колебание с одной частотой, а получится целый *спектр частот* ω_i. Число этих частот равно $3N$, т. е. числу степеней свободы N частиц, из которых состоит тело (конечно, среди этих частот могут быть и совпадающие).

Если твердое тело рассматривать как систему N связанных частиц, совершающую нормальные гармонические колебания, то его средняя энергия определится по формуле

$$\overline{\mathscr{E}} = \sum_{i=1}^{3N} \overline{\varepsilon}_i = \sum_{i=1}^{3N} \frac{\hbar\omega_i}{e^{\hbar\omega_i/kT} - 1}.$$

Число нормальных колебаний с частотами меньше ω, конечно, дискретное, но оно очень велико и может быть аппроксимировано непрерывной функцией $Z(\omega)$. Число нормальных колебаний в интервале частот от ω до $\omega + d\omega$ тоже очень велико, но может рассматриваться как дифференциал $dZ(\omega)$ той же функции. В указанном приближении предыдущую формулу можно заменить на

$$\overline{\mathscr{E}} = \int_0^{\omega_{\text{макс}}} \frac{\hbar\omega}{e^{\hbar\omega/kT} - 1}\, dZ(\omega), \tag{54.2}$$

в которой $\omega_{\text{макс}}$ означает максимальную частоту нормальных колебаний. Она определяется из соотношения

$$Z(\omega_{\text{макс}}) = 3N, \tag{54.3}$$

так как общее число нормальных колебаний равно числу степеней свободы $3N$. Таким образом, в квантовой теории задача нахождения средней энергии твердого тела сводится к определению собственных частот нормальных колебаний, тогда как в классической теории этого делать не требуется, так как по этой теории средняя энергия зависит только от общего числа степеней свободы.

2. Вычислением собственных частот колебаний кристаллической решетки применительно к теории теплоемкости занимались Борн и Карман (1881–1963). Это — очень трудная задача. Однако в вопросе о теплоемкости твердых тел при низких температурах она может быть сильно упрощена, что и было сделано Дебаем. Рассмотрим среднюю энергию осциллятора $\overline{\varepsilon}$, определяемую планковской формулой (54.1) как функцию абсолютной температуры T. Для этого представим эту формулу в виде

$$\overline{\varepsilon} = kT\frac{x}{e^x - 1}, \tag{54.4}$$

где введена безразмерная переменная $x = \hbar\omega/kT$. График этой функции представлен на рис. 94. Из него видно, что в выражении (54.2) для средней энергии тела существенны члены, соответствующие только *низким частотам* нормальных колебаний. Им соответствуют длины волн, большие по сравнению с постоянной кристаллической решетки. Это позволяет отвлечься от атомистической структуры тела и рассматривать нормальные колебания в нем как стоячие инфразвуковые волны в упругой сплошной среде. Это — те же волны, которые вызывают тонкую структуру спектральных линий при молекулярном рассеянии света (эффект Мандельштама–Бриллюэна, см. т. IV, § 99). Таким образом, существенные низкие собственные частоты тела могут быть вычислены методами *теории упругости*, в которой среда считается *сплошной*.

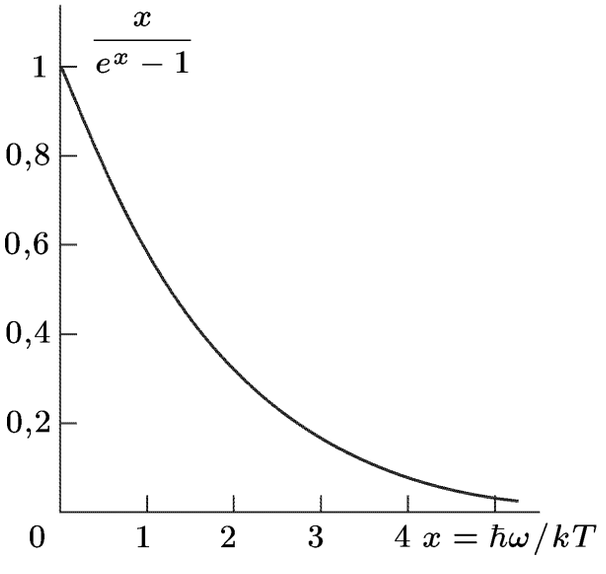


Рис. 94

Выражение для $dZ(\omega)$ может быть найдено из дифференциальных уравнений теории упругости совершенно так же, как была выведена формула (53.7) для такой же величины в случае волн де Бройля. При этом надо только принять во внимание, что в твердом теле могут распространяться как *продольные*, так и *поперечные* звуковые волны. В одном и том же направлении может распространяться только одна

продольная звуковая волна определенной частоты. Поэтому для продольных волн формула (53.7) может быть сохранена без изменений (разумеется, с заменой фазовой скорости волн де Бройля на скорость звука). Поперечных же волн, распространяющихся с той же частотой и в том же направлении, может быть две. Поэтому в этом случае выражение (53.7) надо удвоить. Таким образом,

$$dZ(\omega) = \frac{V\omega^2\,d\omega}{2\pi^2}\left(\frac{1}{c_{\parallel}^3} + \frac{2}{c_{\perp}^3}\right) = \frac{3V\omega^2\,d\omega}{2\pi^2 c^3}, \tag{54.5}$$

где V — объем тела, $c_{\parallel}$ — скорость продольных, а $c_{\perp}$ — скорость поперечных звуковых волн. Величина же c есть некоторая «средняя скорость», определяемая соотношением

$$\frac{3}{c^3} = \frac{1}{c_{\parallel}^3} + \frac{2}{c_{\perp}^3}. \tag{54.6}$$

В этом выводе не учтена анизотропия упругих свойств кристаллов, проявляющаяся даже для кристаллов кубической системы. Тело считалось изотропным, и его упругие свойства характеризовались двумя постоянными, за которые, в частности, можно принять обе скорости звука $c_{\parallel}$ и $c_{\perp}$. Но учет анизотропии малосуществен и вряд ли оправдан в рамках приближенного метода Дебая.

3. Средняя энергия кристалла, согласно (54.2), будет равна

$$\overline{\mathscr{E}} = \frac{3V\hbar}{2\pi^2 c^3}\int\limits_0^{\omega_{\text{макс}}} \frac{\omega^3\,d\omega}{e^{\hbar\omega/kT}-1}, \tag{54.7}$$

или, вводя прежнее обозначение $x = \dfrac{\hbar\omega}{kT}$, а также $x_{\text{макс}} = \dfrac{\hbar\omega_{\text{макс}}}{kT}$,

$$\overline{\mathscr{E}} = \frac{3Vk^4T^4}{2\pi^2 c^3\hbar^3}\int\limits_0^{x_{\text{макс}}} \frac{x^3\,dx}{e^x-1}. \tag{54.8}$$

Для низких температур подынтегральное выражение при высоких частотах ($x \gg 1$) очень мало. В этом случае точное определение верхнего предела $x_{\text{макс}}$ несущественно и его можно принять равным бесконечности, т. е.

$$\overline{\mathscr{E}} = \frac{3Vk^4T^4}{2\pi^2 c^3\hbar^3}\int\limits_0^{\infty} \frac{x^3\,dx}{e^x-1}. \tag{54.9}$$

Входящий сюда интеграл в точности совпадает с тем, который встречался при выводе закона Стефана–Больцмана из формулы Планка (см. т. IV, § 118). Он равен

$$\int\limits_0^{\infty} \frac{x^3\,dx}{e^x-1} = \frac{\pi^4}{15}. \tag{54.10}$$

Таким образом, при низких температурах ($x_{\text{макс}} \gg 1$)

$$\overline{\mathscr{E}} = DT^4, \tag{54.11}$$

где

$$D = \frac{V\pi^2 k^4}{10c^3\hbar^3}. \tag{54.12}$$

Для теплоемкости тела получаем

$$C_V = \left(\frac{\partial \overline{\mathscr{E}}}{\partial T}\right)_V = 4DT^3. \tag{54.13}$$

Таким образом, *теплоемкость кристаллической решетки вблизи абсолютного нуля температур меняется пропорционально третьей степени температуры.* Это — *закон кубов*, теоретически найденный Дебаем. Согласно этому закону при $T = 0$ теплоемкость обращается в нуль в согласии с теоремой Нернста. Закон кубов Дебая, вообще говоря, хорошо согласуется с опытом (см., впрочем, п. 5 настоящего параграфа). Конечно, надо помнить, что формула (54.11), как и закон кубов, относится только к теплоемкости кристаллической решетки. В случае металлов к теплоемкости решетки надо добавить теплоемкость свободных электронов, которая меняется пропорционально первой степени температуры (см. т. II, § 85).

4. Для высоких температур ($\hbar\omega/kT \ll 1$) формула (54.7) с учетом (54.3) приводит к правильному классическому выражению $\overline{\mathscr{E}} = 3NkT$. Этот результат, конечно, не следует из формулы (54.5), которая при высоких частотах неверна. Он является следствием того, что в классическом случае вид спектра частот нормальных колебаний вообще не играет никакой роли — важно только общее число таких колебаний, а оно правильно передается формулой (54.3).

Таким образом, формула (54.7) приводит к правильному результату как в области очень низких, так и в области очень высоких температур. Поэтому ее, как *интерполяционную формулу*, естественно распространить и на промежуточную область температур.

Определим теперь значение $\omega_{\text{макс}}$, пользуясь непрерывной моделью тела. На основании формулы (54.3)

$$Z(\omega_{\text{макс}}) = \int\limits_0^{\omega_{\text{макс}}} dZ(\omega) = \frac{3V}{2\pi^2 c^3} \int\limits_0^{\omega_{\text{макс}}} \omega^2\, d\omega = \frac{V}{2\pi^2 c^3}\omega_{\text{макс}}^3 = 3N,$$

откуда

$$\omega_{\text{макс}} = \left(\frac{6\pi^2 c^3 N}{V}\right)^{1/3} = \frac{2\pi c}{a}\left(\frac{3}{4\pi}\right)^{1/3}, \tag{54.14}$$

где a — постоянная решетки, определяемая соотношением $V = Na^3$. Минимальная длина упругой волны, соответствующая максимальной

частоте $\omega_{\text{макс}}$, равна

$$\lambda_{\text{мин}} = \frac{2\pi c}{\omega_{\text{макс}}} = a\left(\frac{4\pi}{3}\right)^{1/3} \approx 1{,}6a. \tag{54.15}$$

Конечно, этот расчет дает только порядок действительных величин $\omega_{\text{макс}}$ и $\lambda_{\text{мин}}$, так как для таких коротких волн непрерывная модель твердого тела уже не может дать точные результаты.

Введем «дебаевскую температуру», определяемую формулой

$$T_D = \frac{\hbar\omega_{\text{макс}}}{k} = \frac{2\pi c\hbar}{ka}\left(\frac{3}{4\pi}\right)^{1/3}. \tag{54.16}$$

Тогда формулу (54.8) для одного моля можно записать в виде

$$\overline{\mathscr{E}} = 9N_{\text{A}}kT\left(\frac{T}{T_D}\right)^3\int\limits_0^{T_D/T}\frac{x^3\,dx}{e^x-1} = 3RT\cdot 3\left(\frac{T}{T_D}\right)^3\int\limits_0^{T_D/T}\frac{x^3\,dx}{e^x-1}, \tag{54.17}$$

где $R = N_{\text{A}}k$ — универсальная газовая постоянная. Для высоких температур $T \gg T_D$ это выражение переходит в классическую формулу $\overline{\mathscr{E}} = 3RT$, т. е. получается закон Дюлонга и Пти. При температурах, много меньших дебаевской, получается формула (54.11) и закон кубов Дебая. Дифференцируя (54.17) по температуре, получаем интерполяционную формулу Дебая для молярной теплоемкости твердого тела:

$$C_V = 3R\left\{12\left(\frac{T}{T_D}\right)^3\int\limits_0^{T_D/T}\frac{x^3\,dx}{e^x-1} - \frac{3T_D/T}{e^{T_D/T}-1}\right\}. \tag{54.18}$$

В табл. 6 приведены значения дебаевских температур, вычисленные по формуле (54.16) и найденные из экспериментальной кривой теплоемкости. На рис. 95 представлена теплоемкость как функция от-

Таблица 6

Вещество	Дебаевская температура T_D, К	
	из экспериментальной кривой теплоемкости	вычислено по формуле (54.16)
Алюминий	398	402
Медь	315	332
Серебро	215	214
Свинец	88	73
Алмаз	1860	—

ношения T/T_D, вычисленная по формуле (54.18) (сплошная кривая); точки соответствуют экспериментальным данным для Pb, KCl и C (алмаз).

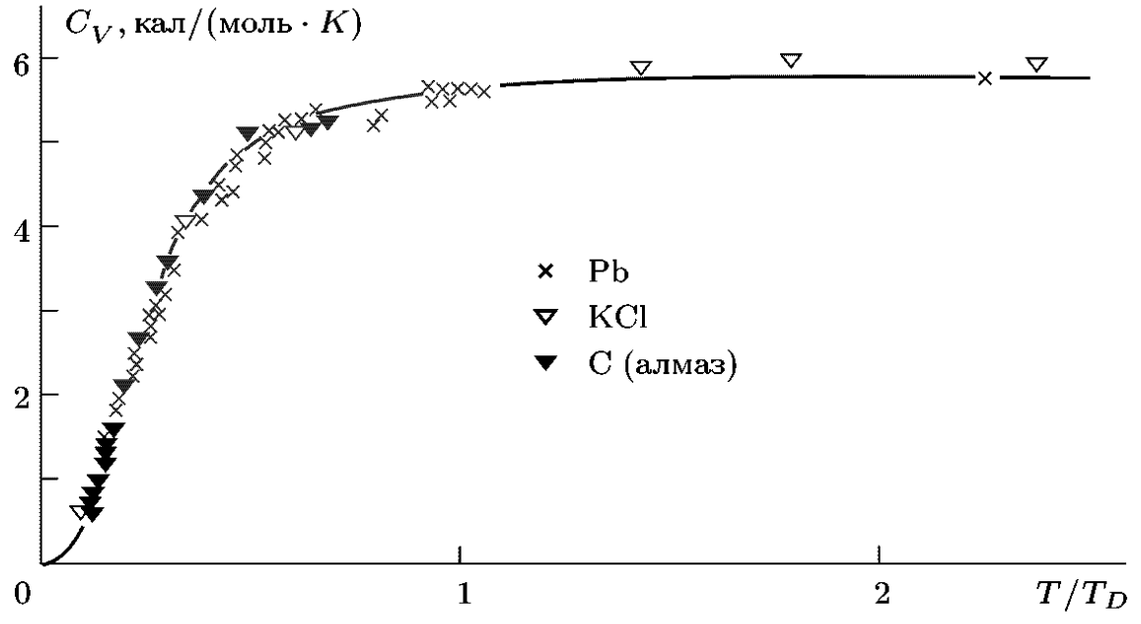


Рис. 95

5. Следует заметить, что изложенная теория Дебая справедлива только для кристаллов, построенных из частиц, на внутреннее состояние которых температура практически не оказывает никакого влияния. В этом случае внутреннее строение частиц не проявляется в теплоемкости. В большинстве случаев указанному условию удовлетворяют кристаллы, построенные из атомов или ионов, у которых расстояние между нормальным и первым возбужденным уровнями велико по сравнению с kT. Однако у некоторых атомов и ионов нижние энергетические уровни расположены очень близко друг к другу. Например, в кристаллическом сульфате гадолиния нижний энергетический уровень иона гадолиния состоит из восьми подуровней, расстояния между которыми соответствуют в температурной шкале дебаевской температуре 1,6 K. При очень низких температурах $T \approx 7$ K появляется добавочная теплоемкость, обусловленная возбуждением указанных подуровней. При столь низких температурах теплоемкость решетки весьма мала по сравнению с этой добавочной теплоемкостью. При $T = 1{,}6$ K теплоемкость кристалла почти в 500 раз превышает теплоемкость кристаллической решетки. При дальнейшем понижении температуры теплоемкость кристалла, естественно, падает до нуля.

В кристаллах, построенных из сложных молекул, может появиться теплоемкость, связанная с тепловым движением атомов или атомных групп внутри молекулы. В первом приближении можно считать, что колебания молекул внутри решетки не сказываются на их внутреннем состоянии. Тогда теплоемкость кристалла можно представить в виде

$$C = C_{\text{реш}} + C_{\text{внутр}}.$$

Вклад, вносимый в теплоемкость внутренним движением, в некоторых случаях может быть довольно значительным. Например, теплоемкость, связанная с внутренними колебаниями в молекулах бензола, при $T \approx 150$ K составляет около 20 % от теплоемкости решетки и достигает 80 % последней при $T \approx 270$ K.

§ 55. Типы связей атомов в твердых телах

1. В предыдущем параграфе при рассмотрении теплоемкости твердых тел мы отвлеклись от их внутренней атомистической структуры. Так поступать, вообще говоря, уже нельзя во многих других вопросах физики твердого тела, некоторые из которых рассматриваются ниже.

В твердых телах (кристаллах) атомы, молекулы или ионы совершают малые колебания около узлов соответствующей кристаллической решетки. Геометрические свойства кристаллических решеток — их симметрия — подробно изучаются в кристаллографии. Кратко эти свойства были рассмотрены в гл. XII второго тома. Содержание этой главы здесь повторяться не будет.

Правильное расположение структурных элементов кристалла (атомов, ионов, молекул) в узлах кристаллической решетки, конечно, требует объяснения с атомистической точки зрения. Можно было бы думать, что такое объяснение требует решения уравнения Шредингера

$$\widehat{H}\psi = \mathscr{E}\psi, \tag{55.1}$$

в котором волновая функция ψ зависит от координат всех ядер и электронов тела. Разумеется, это уравнение, как и всегда, имеет допустимые решения только для избранных значений энергии $\mathscr{E}$. Решение уравнения Шредингера определило бы значения $\mathscr{E}$, при которых тело должно быть кристаллом, а также структуру самой кристаллической решетки. Электроны, заполняющие пространство между атомными ядрами, необходимы для удержания структурных элементов кристалла в положениях равновесия — в узлах кристаллической решетки. В решении должны содержаться все возможные структуры кристаллических решеток. Оно должно охватывать все возможные фазовые переходы между ними, а также переходы в жидкое, а затем и газообразное состояние вещества.

Однако нахождение точного решения уравнения Шредингера макроскопического тела, содержащего громадное число переменных, — задача абсолютно невыполнимая. Она еще более сложна, чем аналогичная задача классической механики многих тел. Более того, явный вид функции ψ макроскопического тела сам по себе и не представляет интереса. Это не исключает возможности и целесообразности исследования функции ψ, не определяя ее явного вида, а непосредственно на основе самого уравнения Шредингера — нахождение таких ее свойств, которые определяли бы наблюдаемые особенности макроскопических тел, в частности существование и структуру кристаллической решетки. Но в такой постановке задача никогда не рассматривалась. Более того,

не доказано, получились ли бы на этом пути ожидаемые и необходимые результаты. При рассмотрении структуры кристаллической решетки и сил, удерживающих атомы, молекулы или ионы в положениях равновесия, приходится широко опираться на опытные данные.

2. Различают *ионные кристаллы*, *ковалентные кристаллы*, *молекулярные кристаллы*, *металлы*.

Кристаллическая решетка ионных кристаллов состоит не из нейтральных частиц, а из положительных и отрицательных ионов, притягивающихся друг к другу. Ионы образуются в результате перехода электронов от одного атома к другому. Электронные оболочки ионов, как правило, соответствуют электронным оболочкам инертных газов. Распределение электронов вблизи каждого иона близко к сферически симметричному и слабо нарушается в области соприкосновения соседних ионов. Типичными представителями ионных кристаллов могут служить NaCl, LiF, KI и т. д. Ионные кристаллы характеризуются большими *энергиями связи*. Под этим вообще понимают работу, которую надо совершить, чтобы при абсолютном нуле разложить систему на составные части. В случае ионных кристаллов такими составными частями являются положительные и отрицательные ионы, из которых построена решетка. Энергия связи решетки NaCl, например, составляет 764 кДж/моль (183 ккал/моль). Эта величина примерно в 1,7 раза больше отнесенной к одному молю энергии связи ионов Na^+ и Cl^- в молекуле NaCl, если предположить, что в молекуле они находятся на том же расстоянии, что и ближайшие ионы в кристаллической решетке. Такого же порядка энергии связи решеток аналогичных ионных кристаллов. Для ионных кристаллов характерна малая электрическая проводимость при низких температурах и хорошая проводимость при высоких температурах.

3. В ковалентных кристаллах (алмаз, кремний, германий и пр.), как и в молекуле водорода, связь *гомеополярная*, т. е. осуществляется *обменными силами*. Для таких кристаллов типично такое строение решетки, при котором число ближайших атомов, окружающих рассматриваемый атом, равно валентности элемента. Рассматриваемые кристаллы обычно образуются из легких элементов средних столбцов периодической системы. Валентные электроны атомов решетки *обобществлены*. Для кристаллов характерна высокая твердость, слабая электрическая проводимость у чистых образцов при низких температурах. Энергия связи кристаллической решетки у рассматриваемых кристаллов составляет примерно 700–1200 кДж/моль (170–280 ккал/моль).

4. Молекулярные кристаллы представляют собой слабо связанные агрегаты молекул. Связь обусловлена *силами Ван-дер-Ваальса*. К молекулярным кристаллам относятся почти все органические кристаллы и многие другие соединения. Молекулярная связь является единственной связью у кристаллов, образованных из атомов инертных газов. У таких веществ, как Ar, CH_4, энергия связи порядка 8 кДж/моль (2 ккал/моль). Для молекулярных кристаллов характерны низкие температуры плавления, кипения и возгонки, а также сильная сжимаемость.

5. Связь в металлах осуществляется посредством электронов, находящихся между ионами кристаллической решетки, подобно тому как электроны осуществляют гомеополярную связь между ядрами в молекуле водорода. Разумеется, невозможно провести детальный количественный расчет в такой многочастичной системе, какой является металл. Приходится довольствоваться качественными соображениями. К металлам относятся элементы, атомы которых содержат внешние недостроенные оболочки. Электроны таких оболочек сравнительно слабо связаны с атомными ядрами. Они могут переходить и действительно переходят от одного ядра к другому. Если даже на пути перехода электрона встречается потенциальный барьер, электрон может сравнительно легко преодолеть его туннельным способом. В результате ядра металла лишаются своих внешних оболочек. Их электроны не привязаны к индивидуальным атомам, а *обобществлены*, т. е. принадлежат всему кристаллу. Такие «свободные» электроны ведут себя подобно электронному газу. Из-за принципа Паули электроны не могут стоять на месте, а должны совершать оживленное движение — электронный газ в металле находится в *состоянии вырождения* (см. т. II, § 82).

Наличием свободных электронов объясняется высокая электрическая проводимость и теплопроводность металлов, специфический металлический блеск, особые механические свойства, позволяющие осуществлять ковку и штамповку.

Каждый ион кристаллической решетки металла заряжен положительно. Из-за этого между ионами действуют электрические силы отталкивания. Свободные электроны уравновешивают эти силы и удерживают ионы в положениях равновесия. Всякий раз, когда ион выходит из положения равновесия, легкие и быстрые свободные электроны перераспределяются в пространстве так, что возникают силы, возвращающие ион в положение равновесия. Этим и обеспечивается устойчивость кристаллической решетки и самого металла.

§ 56. Колебания атомов в одномерной прямолинейной цепочке

1. Для строгого обоснования допустимости и установления границ применимости сплошной модели твердого тела, использованной Дебаем в теории теплоемкости, надо, разумеется, рассмотреть задачу о колебаниях кристаллической решетки в последовательно атомистической постановке, как это сделали Борн и Карман. Рассмотрим этот вопрос на примере одномерной прямолинейной цепочки атомов. Таким путем, конечно, мы не получим точных количественных результатов, пригодных для трехмерного тела, но сильно упростим исследование и в то же время сохраним самые существенные качественные результаты.

2. Допустим сначала, что все атомы цепочки одинаковы и в положениях равновесия находятся на одинаковых расстояниях a друг от друга. Учтем только силы, действующие на каждый атом цепочки, которые исходят от двух соседних атомов. Действием всех остальных

атомов пренебрежем. Такое упрощение называется *приближением ближайших соседей.* Пусть атомы могут испытывать только продольные смещения из положений равновесия. Смещение n-го атома обозначим через ξ_n. Относительное смещение соседних атомов $(\xi_n - \xi_{n-1})$ будем считать малым по сравнению с «постоянной решетки» a. При смещении n-го атома относительно $(n-1)$-го возникает сила $F_{n,n-1}$, действующая на него и направленная противоположно относительному смещению. При малых относительных смещениях ее можно считать квазиупругой, т. е. положить равной

$$F_{n,n-1} = -\varkappa(\xi_n - \xi_{n-1}),$$

где «коэффициент упругости» $\varkappa$ для рассматриваемой цепочки есть величина постоянная. Полная сила, действующая на атом, будет

$$F_{n,n-1} + F_{n,n+1} = -\varkappa(\xi_n - \xi_{n-1}) - \varkappa(\xi_n - \xi_{n+1}) = \\ = \varkappa(\xi_{n-1} - 2\xi_n + \xi_{n+1}),$$

а уравнение движения имеет вид

$$m\ddot{\xi}_n = \varkappa(\xi_{n-1} - 2\xi_n + \xi_{n+1}), \tag{56.1}$$

где m — масса атома.

Нахождение общего решения уравнения (56.1) для очень большого числа атомов N — очень трудная задача. Для нахождения частного решения рассмотрим прежде всего случай, когда $N = \infty$, точнее, когда цепочка атомов бесконечно простирается в обе стороны. Цепочка обладает *трансляционной симметрией*, т. е. переходит сама в себя при сдвиге на любое целое число периодов a. Можно думать, что существует частное решение уравнения (56.1), отвечающее этому типу симметрии: все атомы совершают одинаковые гармонические колебания, но фазы этих колебаний сдвинуты на одну и ту же величину при переходе от каждого атома к соседнему с бóльшим номером. Такое решение представляется бегущей монохроматической волной постоянной амплитуды

$$\xi = \xi_0 e^{i(\omega t - kx)}. \tag{56.2}$$

Особенность этой волны состоит в том, что аргумент x может принимать только дискретные значения $x_n = na$ ($n = \ldots, -2, -1, 0, +1, +2, \ldots$). Если волновое число k заменить на $k = (2\pi/a)p$, где p — любое целое число, то колебания всех атомов цепочки не изменятся. Поэтому, не нарушая общности, можно ограничить изменения k одним интервалом длины $2\pi/a$, называемым *зоной Бриллюэна.* В частности, интервал

$$-\frac{\pi}{a} \leqslant k \leqslant \frac{\pi}{a} \tag{56.3}$$

называется *основной зоной Бриллюэна* (1889–1969). При положительных k волна бежит вперед (вправо), при отрицательных — назад (вле-

во). При таких k длина волны Λ (величина существенно положительная) может изменяться в пределах

$$\infty > \Lambda \geqslant 2a. \tag{56.4}$$

Таким образом, из-за дискретности структуры не имеет смысла говорить о распространении волн, длины которых меньше $2a$. Например, если положить $\Lambda = a$, то в этом случае смещения всех атомов в каждый момент времени были бы одинаковы, т. е. цепочка перемещалась бы как целое. А это эквивалентно длине волны $\Lambda = \infty$, входящей в интервал (56.4).

Найдем теперь условие, при котором волна (56.2) будет решением уравнения (56.1). Для этого замечаем, что $\ddot{\xi}_n = -\omega^2 \xi_n$. Подставляя это выражение в уравнение (56.1), найдем, что оно будет удовлетворено при условии

$$\omega^2 = \frac{\varkappa}{m}\left(2 - e^{ika} - e^{-ika}\right) = 2\frac{\varkappa}{m}\left(1 - \cos ka\right).$$

Ограничиваясь положительными значениями k, отсюда получаем

$$\omega = 2\sqrt{\frac{\varkappa}{m}}\sin\frac{ka}{2}, \tag{56.5}$$

так как, конечно, частота ω существенно положительна.

3. Фазовая скорость волны равна

$$c = \frac{\omega}{k} = a\sqrt{\frac{\varkappa}{m}}\,\frac{\sin\left(ka/2\right)}{ka/2}, \tag{56.6}$$

т. е. зависит от k, а значит, и от Λ. Следовательно, имеет место *дисперсия*, почему формула (56.5) и называется *дисперсионной*. Групповая скорость волны равна

$$u = \frac{d\omega}{dk} = a\sqrt{\frac{\varkappa}{m}}\cos\frac{ka}{2}. \tag{56.7}$$

При $ka = \pi$ (т. е. при $\Lambda = 2a$) она обращается в нуль. В этом случае *волна не переносит энергию*. Физическую причину этого легко уяснить при обращении к рис. 96. На нем стрелками представлены мгновенные

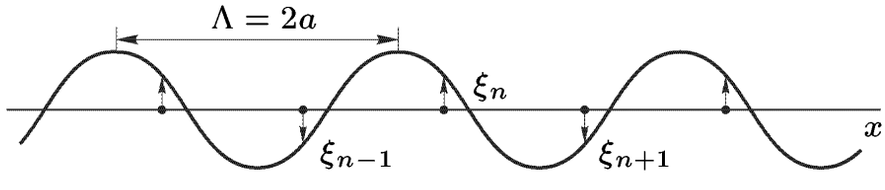


Рис. 96

смещения атомов для случая, когда $\Lambda = 2a$. В этом случае, как видно из рисунка, $\xi_{n-1} = \xi_{n+1} = -\xi_n$, так что уравнение (56.1) принимает вид

$$m\ddot{\xi}_n = -4\varkappa\xi_n.$$

Оно показывает, что в рассматриваемом случае атомы как бы не связаны, а изолированы и каждый из них совершает гармоническое колебание с частотой $\omega = 2\sqrt{\varkappa/m}$. Фазы колебаний соседних атомов сдвинуты на π. Поскольку силы взаимодействия любых двух соседних атомов, а также их смещения в любой момент времени равны и противоположны, работа атома 1 над атомом 2 в точности равна работе атома 2 над атомом 1. Это значит, что в приближении ближайших соседей передачи энергии от атома к атому не происходит. Что касается фазовой скорости c, то при $ka = \pi$ она будет

$$c = \frac{\omega}{k} = \frac{2\sqrt{\varkappa/m}}{\pi/a} = 2\,\frac{a}{\pi}\sqrt{\frac{\varkappa}{m}}.$$

Зависимость угловой частоты ω, а также фазовой и групповой скоростей от волнового числа k графически представлена на рисунках 97 и 98. При малых k, т. е. для длинных волн, формула (56.5) переходит

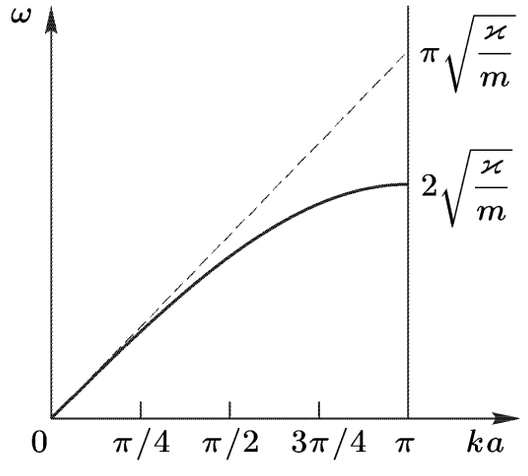


Рис. 97

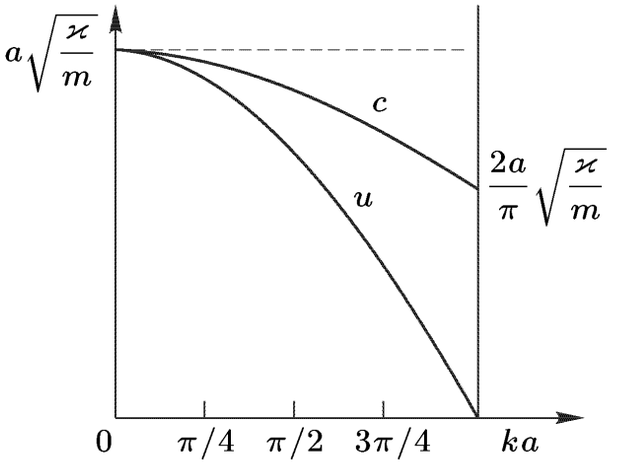


Рис. 98

в $\omega = ka\sqrt{\varkappa/m} = \text{const}$, а обе предельные скорости вырождаются в $c = u = a\sqrt{\varkappa/m} = \text{const}$. В этом случае дисперсия пропадает и цепочка ведет себя как сплошная среда. Поэтому значение скорости $c = u$ может быть получено из формулы, которую дает теория упругости для скорости звука в стержне (см. т. I, § 81). Действительно, модуль Юнга E для цепочки следует определить с помощью формулы $F_{n,n-1} = E(\xi_n - \xi_{n-1})/a$, поскольку $(\xi_n - \xi_{n-1})/a$ есть относительное растяжение цепочки. А так как сила натяжения равна $F_{n,n-1} = \varkappa(\xi_n - \xi_{n-1})$, то $E = \varkappa a$. Роль же плотности играет величина $\rho = m/a$. Поэтому из формулы для скорости звука в стержне получается

$$u = c = \sqrt{\frac{E}{\rho}} = a\sqrt{\frac{\varkappa}{m}},$$

как это и должно быть. Таким образом, в случае длинных волн ($\Lambda \gg a$) частота колебаний может быть вычислена по формулам, относящимся к непрерывной модели цепочки. Даже в случае самых коротких волн ($\Lambda = 2a$) ошибка, получаемая таким путем, не так уж велика: для ω

получается величина $\pi\sqrt{\varkappa/m}$, т. е. примерно в полтора раза больше правильного значения $2\sqrt{\varkappa/m}$.

4. Движение цепочки в общем случае может быть представлено как наложение волн различных частот, распространяющихся вперед и назад. Конечно, всякая реальная цепочка ограничена. Обозначим число атомов в ней через n, тогда ее длина будет $l = (n-1)a$. Если атомы могут совершать только продольные колебания, то число степеней свободы цепочки будет n. Закрепим неподвижно крайние атомы. Этим число степеней свободы уменьшится на две и станет равным $n-2$. Частным решением уравнения (56.1) будет волна (56.2), распространяющаяся вперед ($k \geqslant 0$). Волна той же частоты, распространяющаяся назад, также будет решением. Следовательно, частным решением будет и суперпозиция таких двух волн:

$$\xi = \xi_0 e^{i(\omega t - kx)} + \eta_0 e^{i(\omega t + kx)} = e^{i\omega t}(\xi_0 e^{-ikx} + \eta_0 e^{ikx}).$$

Но так как крайние атомы закреплены, то в любой момент времени их смещения должны быть равны нулю. Значит, должно быть

$$\xi_0 + \eta_0 = 0, \quad \xi_0 e^{-ikl} + \eta_0 e^{ikl} = 0.$$

Из первого соотношения следует, что $\eta_0 = -\xi_0$. С учетом этого второе соотношение дает

$$e^{ikl} - e^{-ikl} = 0,$$

или

$$\sin kl = 0 \tag{56.8}$$

Таким образом, получается стоячая волна

$$\xi = \xi_0 e^{i\omega t}(e^{-ikx} - e^{ikx}) = -2i\xi_0 \sin kx\, e^{i\omega t} = 2\xi_0 \sin kx \sin \omega t, \tag{56.9}$$

причем

$$kl = N\pi \quad (N = 0, 1, 2, \ldots, N_{\text{макс}}), \tag{56.10}$$

или

$$l = N\frac{\Lambda}{2},$$

т. е. *на длине цепочки должно укладываться целое число полуволн.* Каждой стоячей волне (56.9) соответствует собственное, или нормальное, колебание цепочки.

Максимальное значение N получается при максимальном значении $k = \pi/a$, т. е. $N_{\text{макс}} = l/a = n-1$. Однако это значение, как и значение $N = 0$, следует исключить, так как им соответствуют такие значения k, при которых все атомы получают одинаковые смещения. А поскольку крайние атомы неподвижны, то и все прочие атомы были бы неподвижны. Следовательно, в цепочке с закрепленными концами может возбудиться всего $n-2$ нормальных колебаний, что в точности равно числу степеней свободы цепочки. Общее движение цепочки с закрепленными концами может быть представлено наложением таких $n-2$ нормальных колебаний.

Итак, *для длинных волн действительно можно пользоваться непрерывной моделью твердого тела.* Это и делается в теории Дебая

для вычисления теплоемкости твердых тел при низких температурах, когда заметной энергией обладают только длинные волны. К тому же выводу можно прийти и в общем случае, рассматривая колебания трехмерной кристаллической решетки, состоящей из одинаковых частиц (одноатомное твердое тело). Только в таком трехмерном теле, если пренебречь анизотропией, к *продольным колебаниям*, рассмотренным выше для одномерной цепочки, добавляются еще две ветви *поперечных колебаний*, т. е. колебаний, перпендикулярных к волновому вектору $\mathbf{k}$ и совершающихся во взаимно перпендикулярных плоскостях. Число нормальных колебаний, как и число степеней свободы, утраивается.

5. До сих пор предполагалось, что кристалл состоит из *одинаковых* атомов. В случае кристаллов, элементарная ячейка которых содержит несколько *различных* атомов, добавляются колебания этих атомов относительно друг друга. Для качественного решения вопроса воспользуемся опять одномерной цепочкой, но состоящей из двух разных атомов, чередующихся друг с другом (рис. 99). Массы этих атомов обозначим через M и m $(M > m)$, а их смещения из положений

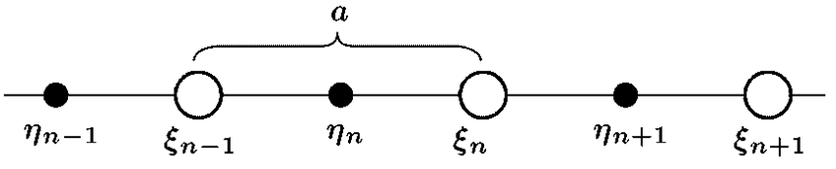


Рис. 99

равновесия — через ξ_n и η_n соответственно. Как и раньше, проведем расчет в приближении ближайших соседей, т. е. примем во внимание силы взаимодействия только соседних атомов. На рис. 99 показано расположение атомов цепочки в положении равновесия. Под каждым атомом обозначено смещение его из этого положения. Применительно к этому рисунку по аналогии с уравнением (56.1) получаем

$$\begin{aligned} M\ddot{\xi}_n &= \varkappa(\eta_{n-1} - 2\xi_n + \eta_n), \\ m\ddot{\eta}_n &= \varkappa(\xi_{n+1} - 2\eta_n + \xi_n). \end{aligned} \tag{56.11}$$

Затем ищем частное решение этой системы уравнений в виде монохроматической бегущей волны

$$\xi_n = \xi_0 e^{i(\omega t - kna)}, \quad \eta_n = \eta_0 e^{i(\omega t - kna)}, \tag{56.12}$$

предполагая, что a меняется в интервале (56.3), причем под a теперь понимается расстояние между двумя соседними одинаковыми атомами. После подстановки в (56.11) получится

$$\begin{aligned} (M\omega^2 - 2\varkappa)\xi_0 + \varkappa(1 + e^{ika})\eta_0 &= 0, \\ \varkappa(1 + e^{-ika})\xi_0 + (m\omega^2 - 2\varkappa)\eta_0 &= 0. \end{aligned} \tag{56.13}$$

Исключение ξ_0 и η_0 дает

$$Mm\omega^4 - 2\varkappa(M + m)\omega^2 + 2\varkappa^2(1 - \cos ka) = 0,$$

откуда

$$\omega^2 = \frac{\varkappa}{Mm}\left(M + m \pm \sqrt{M^2 + m^2 + 2Mm\cos ka}\right). \qquad (56.14)$$

Этой формулой и определяется спектр собственных частот колебаний цепочки. Из-за двойного знака перед квадратным корнем получаются две ветви частот. Знаку минус соответствует частота $\omega_1 = \omega_1(k)$, знаку плюс — частота $\omega_2 = \omega_2(k)$. Ветвь $\omega_1 = \omega_1(k)$ называется *акустической* или *дебаевской*, ветвь $\omega_2 = \omega_2(k)$ — *оптической* или *борновской*. Обе ветви представлены на рис. 100. При малых k (длинные волны) частота ω_1 также мала и меняется линейно в зависимости от k. В этом случае, как видно из (56.13), $\xi_0 = \eta_0$, а потому $\xi_n = \eta_n$. Это значит, что соседние атомы с массами M и m (и вообще все атомы, расположенные на отрезке, малом по сравнению с длиной волны) колеблются в *одинаковых фазах*. При таких колебаниях цепочка может быть аппроксимирована сплошной одномерной моделью.

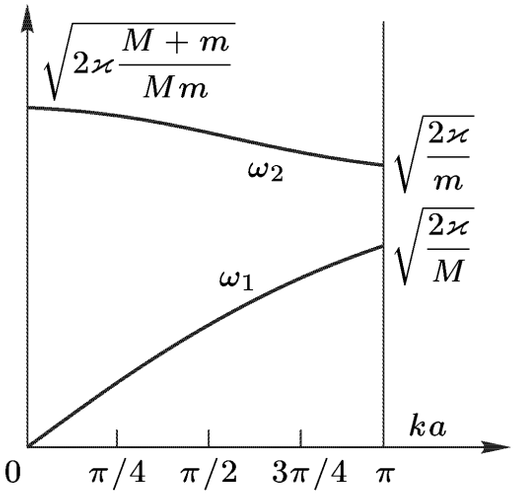


Рис. 100

Ветвь $\omega_2 = \omega_2(k)$ характеризуется тем, что для нее при $k \to 0$ ω_2 не стремится к нулю, а, наоборот, стремится к максимуму. При малых k каждое из уравнений (56.13) переходит в $M\xi_0 = -m\eta_0$, а потому $M\xi = -m\eta$. Это значит, что в этом случае соседние атомы с массами M и m колеблются в *противоположных фазах*, т. е. происходит колебание одного атома относительно другого.

6. В трехмерной кристаллической решетке, элементарная ячейка которой содержит s атомов, существуют $3s$ ветвей *нормальных колебаний*. Из них три ветви *акустические*: одной соответствуют *продольные* колебания, двум другим — *поперечные*. Остальные $3s - 3$ ветвей *оптические*. Частоты некоторых нормальных колебаний могут совпадать из-за симметрии решетки. Частоты акустических колебаний для длинных волн стремятся к нулю, для таких колебаний они пропорциональны волновому вектору. В этом случае соседние атомы элементарной ячейки движутся *синфазно*, и кристалл можно рассматривать как *сплошную среду*. Оптические колебания характеризуются высокими частотами, не обращающимися в нуль для бесконечно длинных волн. При оптических колебаниях происходят сильные смещения атомов элементарной ячейки относительно друг друга.

Рассеяние света на тепловых акустических волнах сопровождается изменением частоты, в этом состоит *явление Мандельштама–Бриллюэна* (см. т. IV, § 99, а также задачу 2 к § 57). Оптическая ветвь колебаний характеризуется частотами $\nu \equiv \omega/2\pi \sim 10^{12}$–$10^{13}$ Гц, они лежат в инфракрасной области спектра, почему и получили название оптических. Оптические колебания могут сопровождаться изменением

электрических моментов элементарной ячейки (например, в случае кристалла NaCl, когда имеет место относительное смещение ионов Na^+ и Cl^-). Тогда возникают инфракрасные полосы поглощения и соответствующая им аномальная дисперсия в оптическом спектре кристалла. С инфракрасными колебаниями связано явление *комбинационного рассеяния света* (см. т. IV, § 100).

Понятно, что ввиду дискретности пространственной решетки волновой вектор $\mathbf{k}$, как и в одномерной цепочке, определен не однозначно. К нему можно прибавить произвольный вектор $2\pi\mathbf{K}$, чтобы фаза колебаний всех атомов $\omega t - \mathbf{kr}$ изменилась на $2\pi n$, где n — произвольное целое число (положительное или отрицательное). Физически такое изменение ни в чем не проявляется. Если $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ — базисные векторы кристаллической решетки, то вектор $\mathbf{K}$ определяется выражением

$$\mathbf{K} = n_1\mathbf{a}_1^* + n_2\mathbf{a}_2^* + n_3\mathbf{a}_3^*, \tag{56.15}$$

где n_1, n_2, n_3 — произвольные целые числа (положительные и отрицательные), а $\mathbf{a}_1^*$, $\mathbf{a}_2^*$, $\mathbf{a}_3^*$ — базисные векторы обратной решетки, т. е.

$$\mathbf{a}_1^* = \frac{[\mathbf{a}_2\mathbf{a}_3]}{(\mathbf{a}_1[\mathbf{a}_2\mathbf{a}_3])}, \quad \mathbf{a}_2^* = \frac{[\mathbf{a}_3\mathbf{a}_1]}{(\mathbf{a}_2[\mathbf{a}_3\mathbf{a}_1])}, \quad \mathbf{a}_3^* = \frac{[\mathbf{a}_1\mathbf{a}_2]}{(\mathbf{a}_3[\mathbf{a}_1\mathbf{a}_2])} \tag{56.16}$$

(см. т. II, § 130, а также т. I, § 7). Минимальная область изменения вектора $\mathbf{k}$, которой можно ограничиться, чтобы представить любые колебания атомов решетки, называется *зоной Бриллюэна*. В частности, можно поступить так, чтобы точка $\mathbf{k} = 0$ в пространстве волновых векторов была точкой зеркальной симметрии. Тогда зону Бриллюэна называют *основной*. Вектор $\mathbf{K}$ (56.15) при любых целых числах n_1, n_2, n_3 принято называть *вектором обратной решетки*.

ЗАДАЧА

Почему формула (56.14) не переходит в формулу (56.5) при $M = m$?

§ 57. Фононы и квазичастицы

1. Внутреннее движение покоящегося тела может быть описано указанием движения каждой индивидуальной частицы, из которых состоит тело. Такой способ может быть назван *индивидуальным* описанием движения. Но возможен и *коллективный* способ, когда движение тела в целом рассматривается как результат наложения движений, в каждом из которых участвуют все частицы тела. Второй способ может обладать преимуществом в тех случаях, когда частицы тела взаимодействуют друг с другом. Тогда разложение полного движения тела на составляющие коллективные движения может быть произведено так, чтобы каждое составляющее коллективное движение могло быть возбуждено в отдельности. Если возбужденное движение тела не очень интенсивно, то оно всегда может быть разложено на плоские монохроматические волны различных частот, распространяющиеся в теле в различных направлениях практически независимо друг от друга. При увеличении интенсивности возбуждения наступают *нелинейные* явления. Однако

если отступления от линейности не очень значительны, то по-прежнему можно пользоваться разложением на плоские монохроматические волны, но между отдельными волнами возникает взаимодействие.

Оба способа описания движения в классической физике принципиально равноправны. Но в квантовой физике преимущество отдается второму способу. Причина этого заключается в квантовании. Уже Дебай в теории теплоемкости твердого тела (см. § 54) с успехом подверг квантованию энергию стоячих монохроматических волн, на которые может быть разложено движение тела. В вопросе о теплоемкости проводить дальнейшее разложение стоячих волн на бегущие не обязательно, поскольку в этом случае интерес представляет энергия тела в состоянии статистического равновесия, а, например, не его импульс, который для покоящегося тела равен нулю в любой момент времени. Но при рассмотрении различных процессов в телах, даже при наличии локального статистического равновесия, надо перейти к разложению движения на бегущие волны и к их квантованию.

В соответствии с гипотезой де Бройля, подтвержденной опытными фактами, с каждой бегущей монохроматической волной связаны энергия и импульс, определяемые соотношениями

$$\mathscr{E} = \hbar\omega, \quad \mathbf{p} = \hbar\mathbf{k}, \tag{57.1}$$

введенными по аналогии с теорией фотонов. Волна, несущая энергию и импульс, определяемые формулами (57.1), в каком-то отношении ведет себя как частица. Частица, уподобляемая звуковой волне в вышеуказанном смысле, называется *фононом*. Не надо вкладывать в представление о фононе нечто большее, чем то, что содержится в этом определении. Фонон несет энергию и импульс, связанные с частотой волны ω и волновым вектором $\mathbf{k}$ посредством постоянной Планка $\hbar$. Но бессмысленно, например, говорить о форме и размерах фонона, представляя его каким-то маленьким шариком. Поэтому фонон называют не «частицей», а «квазичастицей», а его импульс — «квазиимпульсом». В пп. 3 и 4 будут приведены дальнейшие соображения, оправдывающие эту терминологию.

Строго определенные значения величины ω и $\mathbf{k}$ имеют только в случае неограниченных плоских волн. Реальное же существование имеют только *пространственно ограниченные* волны. Реальным образом фонона является не бесконечная, а ограниченная волна, например волновой пакет.

2. Гипотеза фононов согласуется, например, с существованием *звукового давления*. Рассмотрим ради простоты изотропную сплошную среду (изотропное твердое тело, жидкость, газ), в которой распространяется монохроматическая плоская продольная звуковая волна, нормально падающая на плоскую границу твердого тела и поглощаемая им. Такая волна ежесекундно передает единице поверхности твердого тела импульс $cN\hbar k$, где c — скорость звука, а N — число фононов в единице объема среды. Этот импульс и есть давление $\mathscr{P}$, оказываемое звуком на тело. Поскольку сплошная среда недиспергирующая, скорость звука в ней c совпадает с фазовой скоростью $c_{\text{фаз}} = \omega/k$.

Поэтому

$$\mathscr{P} = \varepsilon, \tag{57.2}$$

где $\varepsilon = N\hbar\omega$ — объемная плотность звуковой энергии, падающей на тело. Формула (57.2) справедлива и в общем случае нормального падения волны при наличии отражения и прохождения. Только в этом случае плотность энергии дается выражением $\varepsilon = (1 + r)N\hbar\omega$, где r — коэффициент отражения. Полученные результаты согласуются с опытом и с тем, что дает классическая гидродинамика.

3. В изотропных твердых сплошных телах могут возбуждаться фононы двух типов — *продольные* и *поперечные*. В случае изотропных сред частоты поперечных фононов определяются только длиной волны и не зависят от их поляризации. В кристаллах, помимо продольных и поперечных, могут возбуждаться и другие фононы, соответствующие различным частотам и типам поляризации колебаний. (Как правило, в кристаллах «продольные» и «поперечные» волны не являются строго продольными и строго поперечными.) Для всех таких фононов справедливы соотношения (57.1). Только в этих случаях частота колебаний ω связана с волновым вектором $\mathbf{k}$ не обязательно линейной однородной зависимостью, как было при отсутствии дисперсии, а зависимостью более сложного вида $\omega = \omega(\mathbf{k})$. Такого рода зависимости называются *дисперсионными соотношениями*. Они различны для различных фононов. Примерами дисперсионных соотношений могут служить формулы (56.5) или (56.14), выведенные для одномерных цепочек атомов.

Волновой вектор $\mathbf{k}$ волны в кристаллической решетке определен *не однозначно*, а с точностью до слагаемого, равного вектору обратной решетки (см. § 56, п. 6). В соответствии с этим и вектор $\mathbf{p} = \hbar\mathbf{k}$ определен также *неоднозначно*. Но можно устранить эту неоднозначность, если ограничить область изменения $\mathbf{k}$ основной зоной Бриллюэна, что мы и будем делать. Так однозначно определенный вектор $\mathbf{p}$ называется *квазиимпульсом* фонона. Самый фонон, поскольку ему свойственны признаки частицы, называется *квазичастицей*, о чем уже было сказано раньше. Такой термин вводится, чтобы подчеркнуть, что квазичастица не является «настоящей» частицей.

Квазичастицы вводятся и в других разделах физики. Так, квазичастицами являются фотоны в вакууме и в особенности в среде. Представление о них согласуется с такими явлениями, как интерференция, эффект Доплера, эффект Вавилова–Черенкова, изменение частоты света при распространении в гравитационном поле. Мы уже рассматривали эти явления с точки зрения существования квазичастиц света, хотя и не пользовались самим термином «квазичастица».

4. В идеальной кристаллической решетке, свободной от посторонних примесей и лишенной различных дефектов (примесные атомы, атомы в междоузлиях, незаполненные узлы), плоская звуковая волна в линейном приближении должна распространяться *без затухания* и *рассеяния в стороны*. Линейное приближение означает, что разложение потенциальной энергии кристалла по степеням смещений атомов из положений равновесия обрывается на членах *второй* степени. Тогда

возникают волны с гармоническими колебаниями атомов, или фононы, *не взаимодействующие друг с другом*. При наличии членов высших степеней, если они достаточно малы (а это, как правило, имеет место всегда, пока решетка не разрушена, т. е. вплоть до температуры плавления), также можно говорить о распространении плоских волн, или фононов, в кристалле. Однако в этом случае наступает *взаимодействие различных волн* (*фононов*). Поскольку энергия и импульс фононов квантуются, такое взаимодействие носит характер *столкновений*, в которых происходит уничтожение старых и рождение новых фононов. Наличие в потенциальной энергии членов третьей степени приводит к столкновениям, в которых одновременно участвуют три фонона. При наличии членов четвертой степени появляются столкновения четырех фононов и т. д.

Фононы и вообще квазичастицы хорошо приспособлены для описания *слабых* коллективных возбуждений в телах. Между последовательными столкновениями фонон движется свободно, и если «длина свободного пробега» его достаточно велика по сравнению с постоянной кристаллической решетки, то возбужденное состояние кристалла можно в известном отношении рассматривать как *фононный газ*. При этом число фононов не сохраняется, что дает основание рассматривать их как *бозе-частицы* (*бозоны*).

На рис. 101 графически изображены примеры возможных взаимодействий фононов. Фононы изображены стрелками, а факты их взаимодействия — кружками. Рисунок 101 *а* соответствует распаду фонона

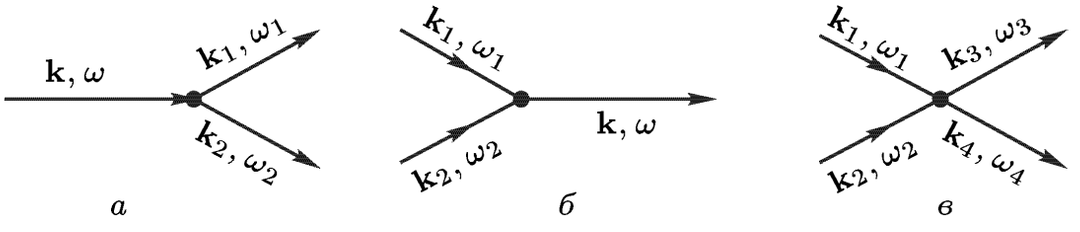


Рис. 101

$\mathbf{k}, \omega$ на два фонона $\mathbf{k}_1, \omega_1$ и $\mathbf{k}_2, \omega_2$. Рисунок 101 *б* изображает столкновение фононов $\mathbf{k}_1, \omega_1$ и $\mathbf{k}_2, \omega_2$, в результате которого возникает один фонон $\mathbf{k}, \omega$. На рис. 101 *в* изображено столкновение двух фононов $\mathbf{k}_1, \omega_1$ и $\mathbf{k}_2, \omega_2$, завершающееся возникновением двух новых фононов $\mathbf{k}_3, \omega_3$ и $\mathbf{k}_4, \omega_4$.

При взаимодействии фононов соблюдается *закон сохранения энергии*. В случае процесса, изображенного на рис. 101 *а*, он записывается в виде

$$\hbar\omega = \hbar\omega_1 + \hbar\omega_2 \tag{57.3}$$

и аналогично в других случаях. Однако закон сохранения квазиимпульса *может и не соблюдаться*. Причиной этого является неоднозначность волнового вектора $\mathbf{k}$, отмеченная выше. Действительно, разложим, например, вектор $\mathbf{k}$ на два вектора: $\mathbf{k} = \mathbf{k}_1' + \mathbf{k}_2'$ (рис. 102). Вектор $\mathbf{k}$ предполагается лежащим в основной зоне Бриллюэна, так что при нашем ограничении $\hbar\mathbf{k}$ является квазиимпульсом. Но пред-

положим, что составляющие векторы $\mathbf{k}_1'$ и $\mathbf{k}_2'$ (или по крайней мере один из них) настолько длинны, что они не умещаются в основной зоне Бриллюэна. Тогда, при нашем ограничении, векторы $\hbar\mathbf{k}_1'$ и $\hbar\mathbf{k}_2'$ не будут квазиимпульсами. Квазиимпульсы $\hbar\mathbf{k}_1$ и $\hbar\mathbf{k}_2$ получаются из них путем прибавления векторов вида $2\pi n\mathbf{K}$, где $\mathbf{K}$ — вектор обратной решетки (56.15), а $n = 0, \pm 1, \ldots$. Например, для процесса, соответствующего рис. 101 *а*, следует писать

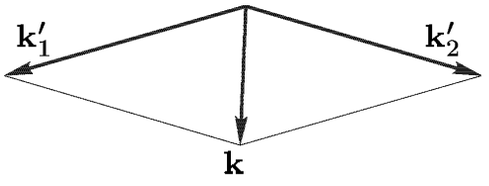


Рис. 102

$$\hbar\mathbf{k}_1 + \hbar\mathbf{k}_2 = \hbar\mathbf{k} + 2\pi\mathbf{K}. \qquad (57.4)$$

Если $n = 0$, то в процессе взаимодействия фононов квазиимпульс сохраняется. Такие процессы называются *нормальными*. Если же $n \neq 0$, то соответствующие процессы называются *процессами переброса*. Существование процессов переброса лишний раз оправдывает введение терминов «квазичастица» и «квазиимпульс» вместо простых терминов «частица» и «импульс».

Конечно, соотношения вида (57.3) и (57.4) справедливы не только при взаимодействии фононов между собой, но и при взаимодействии их с другими частицами и квазичастицами, например с фотонами. При переводе на классический язык эти соотношения выражают *законы интерференции волн*, *принцип Доплера* и вообще *законы изменения частоты волн при различных процессах*. Вот почему комбинационное рассеяние света, рассеяние Мандельштама–Бриллюэна, эффект Вавилова–Черенкова, изменение частоты света при распространении в гравитационном поле и другие явления могут быть истолкованы как с волновой точки зрения, так и с помощью представления о квазичастицах.

5. Воспользуемся теперь представлением о фононах для рассмотрения *теплопроводности твердых тел*. Мы имеем в виду диэлектрики, а не металлы. В диэлектриках перенос тепла осуществляется *фононами*, тогда как в металлах основную роль в этом процессе играют *электроны*. Само понятие теплопроводности относится к *локально равновесному* состоянию неподвижного тела, каждой точке которого можно приписать определенную температуру. Чтобы получить полностью равновесное состояние тела, можно, например, заключить его в жесткую оболочку, поддерживаемую при постоянной температуре. Тогда в результате теплового возбуждения фононов, их поглощения и рассеяния на других фононах, на примесях и дефектах решетки, на границах тела и окружающей оболочки, в конце концов установится *полностью термодинамически равновесное* состояние тела, однозначно определяемое только температурой оболочки. Оно характеризуется вполне определенным значением плотности энергии фононов в пространстве, вполне определенным распределением ее по спектру частот, хаотическим (в частности, изотропным) распределением направлений распространения фононов. В этом отношении равновесное состояние фононов в полости напоминает аналогичное состояние фотонов — черное излучение. Локально равновесное состояние тела отличается от

полностью равновесного тем, что температура тела меняется от точки к точке, а все прочие параметры, характеризующие состояние тела с фононами, успевают принять практически равновесные значения, соответствующие этой температуре. Локально равновесное состояние и имеется в виду при рассмотрении теплопроводности.

6. Будем сначала предполагать, что кристалл идеальный, т. е. не содержит примесей, а кристаллическая решетка лишена всех дефектов. Совокупность фононов в теле будем рассматривать как *фононный газ* и воспользуемся для его теплопроводности формулой

$$\varkappa = \frac{1}{3}\overline{v}C\lambda, \tag{57.5}$$

которую дает элементарная теория газов (см. т. II, § 89). Здесь C — теплоемкость единицы объема тела (в прежних обозначениях $C = nmc_v$), $\overline{v}$ — средняя скорость фонона в теле, λ — средняя длина свободного пробега фонона. Величина $\overline{v}$ имеет смысл средней скорости звука в теле, C определяется в квантовой теории теплоемкости твердого тела. Обе эти величины могут быть измерены экспериментально. Наибольшие трудности встречает определение величины λ. В гармоническом (линейном) приближении звуковые волны (фононы) распространяются в идеальном кристалле, не встречая никаких препятствий. В этом приближении нет столкновений между фононами. Если бы кристалл был безграничным, то λ, а с ней и теплопроводность $\varkappa$ были бы бесконечно большими. В следующих приближениях, когда в потенциальной энергии решетки учитываются члены третьей и высших степеней относительно смещений атомов из положений равновесия, появляются столкновения между фононами, ограничивающие их длины свободного пробега. Основное значение имеют члены третьей степени, приводящие к тройным столкновениям (см. рисунки 101 *а* и 101 *б*).

7. Однако *нормальные столкновения между фононами не влияют на теплопроводность кристалла.* Причина этого та же, что и в аналогичном случае прохождения электрического тока через металл: электрическая проводимость металлов связана с длиной свободного пробега электронов и дырок, но столкновения между самими электронами и самими дырками на нее не оказывают никакого влияния (т. III, § 42, п. 2). Действительно, при нормальных столкновениях квазиимпульс фонона совпадает с истинным импульсом, а последний при столкновениях сохраняется. Энергия при столкновениях также сохраняется. При распаде фонона прежний фонон исчезает, а вместо него появляются два новых фонона, которые и переносят через кристалл те же энергию и импульс. При столкновении двух фононов они исчезают, передавая импульс и энергию образовавшимся двум новым фононам, которые также продолжают нести их через кристалл. Так происходит при трехфононных процессах. Аналогичное имеет место в столкновениях с одновременным участием четырех, пяти и более фононов. Таким образом, нормальные столкновения между фононами не могут замедлить передачу энергии или импульса через кристалл. Если бы

все столкновения между фононами были нормальными, то теплопроводность бесконечной идеальной решетки была бы также бесконечной.

Возникает вопрос, почему приведенные рассуждения неприменимы к теплопроводности газа, состоящего из обычных частиц (атомов и молекул), хотя в этом случае при столкновениях также соблюдаются законы сохранения энергии и импульса? Дело в том, что при столкновениях частиц обычного газа они не уничтожаются и не рождаются. Налетающая частица, сама не уничтожаясь, при столкновении передает импульс и энергию уже существующим, а не рождающимся вновь частицам. При этом в газе нет переноса вещества, а передача энергии не полная. Энергия ударяющей частицы в результате столкновения может и уменьшаться, и увеличиваться Но если в газе есть градиент температуры, то энергия «горячих» частиц преимущественно уменьшается, а «холодных» увеличивается. Благодаря этому в газе и возникает поток тепла, направленный в сторону более низкой температуры.

Из приведенных рассуждений следует, что *теплопроводность идеального кристалла может быть связана только с такими столкновениями фотонов, которые сопровождаются процессами переброса*, так как при этих столкновениях не соблюдается закон сохранения квазиимпульса. Значит, только эти столкновения и должны быть приняты во внимание при вычислении средней длины свободного пробега фонона, входящей в формулу (57.5).

8. Впрочем, сами вычисления очень громоздки и не могут быть выполнены без подробных сведений о межатомных силах взаимодействия в кристалле, которыми для большинства кристаллов мы не располагаем. С этим связано неудовлетворительное состояние теории: есть четкие физические представления и методы расчета, которыми, однако, нельзя воспользоваться для получения окончательных количественных результатов. Мы ограничимся только краткими качественными соображениями.

Вблизи абсолютного нуля температур, когда тепловых фононов практически нет, средняя длина свободного пробега фонона ограничивается размерами кристалла. Здесь дело обстоит аналогично тому, что имеет место в случае ультраразреженных газов, когда длина свободного пробега молекулы велика по сравнению с размерами сосуда, в котором заключен газ (см. т. II, § 95). Полагая в формуле (57.5) $\lambda = l$, где l — размеры кристалла, мы получим величину $\varkappa$, которая будет характеризовать не только внутренние свойства кристалла, но будет зависеть и от его размеров. При низких температурах скорость $\overline{v}$ практически постоянна, а теплоемкость по теории Дебая пропорциональна T^3, поэтому и теплопроводность кристалла будет также пропорциональна T^3.

При повышении температуры влияние размеров кристалла отойдет на второй план. Определяющими будут *столкновения между фононами*, сопровождающиеся процессами переброса. За счет этого, а также за счет увеличения теплоемкости произойдет и быстрое увеличение теплопроводности. В этой области температур величина λ, а с ней и теплопроводность $\varkappa$ кристалла уже перестают зависеть от размеров кристалла, а становятся только его внутренними свойствами. В области

высоких температур можно ожидать зависимости $\varkappa \sim 1/T$. Действительно, в этой области справедлив классический закон равномерного распределения кинетической энергии по степеням свободы, в силу которого энергии всех фононов становятся одинаковыми (не зависящими от частоты ω). Поэтому плотность фононов N пропорциональна плотности энергии, т. е. T, а теплоемкость C достигает классического предела, который не зависит от T. Поэтому средняя длина свободного пробега фонона $\lambda \sim 1/N$, а с ней и теплопроводность $\varkappa$ становятся пропорциональными $1/T$.

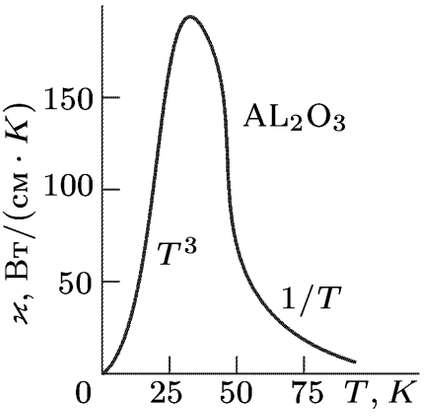


Рис. 103

Из изложенного ясно, что при повышении температуры теплопроводность диэлектрического кристалла должна проходить через максимум. Это отчетливо проявляется на рис. 103, где приведена экспериментальная кривая теплопроводности, полученная для искусственного сапфира (Al_2O_3). Максимум на кривых для различных веществ проявляется не всегда так резко. Причиной этого являются примеси и дефекты кристаллической решетки, вносящие дополнительное теплосопротивление и уменьшающие ее теплопроводность.

ЗАДАЧИ

1. Определить давление, оказываемое звуковой волной на границу раздела двух изотропных непоглощающих сплошных сред (рис. 104).

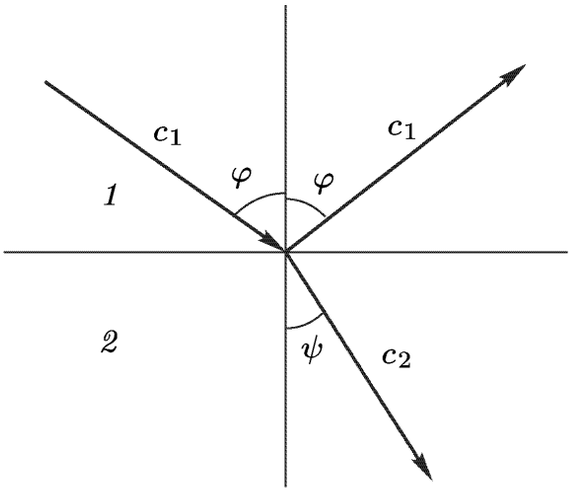


Рис. 104

Р е ш е н и е. Так как нормальные слагающие потока энергии по обе стороны границы раздела сред одинаковы, то

$$N_1 \varepsilon c_1 \cos\varphi - N_1' \varepsilon c_1 \cos\varphi = N_2 \varepsilon c_2 \cos\psi,$$

где N_1, N_1' и N_2 — числа падающих, отраженных и прошедших фононов в единице объема, c_1 и c_2 — скорости звука в средах *1* и *2*, $\varepsilon = \hbar\omega$ — энергия фонона (одинаковая в обеих средах). Введя коэффициент отражения $r = N_1'/N_1$, отсюда получим

$$N_2 = \frac{c_1 \cos\varphi}{c_2 \cos\psi}(1 - r)N_1.$$

Искомое давление $\mathscr{P}$ равно нормальной составляющей импульса, которую передает звук единице границы раздела сред:

$$\mathscr{P} = N_1 p_1 c_1 \cos^2\varphi - (N_2 p_2 c_2 \cos^2\psi - N_1' p_1 c_1 \cos^2\varphi).$$

Так как сплошная среда не обладает дисперсией, то $pc = \varepsilon$. Поэтому, используя значения N_1' и N_2, приведенные выше, и вводя плотность энергии падающей звуковой волны $\mathscr{E} = N_1\varepsilon$, получим

$$\mathscr{P} = \mathscr{E} \cos\varphi \operatorname{ctg}\psi\{(1 + r)\cos\varphi \operatorname{tg}\psi - (1 - r)\sin\varphi\}. \qquad (57.6)$$

2. Используя представление о фононах, получить формулу для тонкой структуры Мандельштама–Бриллюэна (см. т. IV, § 99).

Р е ш е н и е. Связь между энергией и импульсом для света в среде (фотона) и звука (фонона) имеет вид

$$p_{\text{фот}} = \frac{n\mathscr{E}_{\text{фот}}}{c}, \quad p_{\text{зв}} = \frac{\mathscr{E}_{\text{зв}}}{v_{\text{зв}}}, \tag{57.7}$$

где c — скорость света в вакууме, а $v_{\text{зв}}$ — скорость звука (фонона) в среде. К таким же соотношениям приводит и классическая теория. Уравнения сохранения энергии и импульса при излучении и поглощении фонона:

$$\mathscr{E}_{\text{зв}} = \pm(\mathscr{E}_{\text{фот}} - \mathscr{E}'_{\text{фот}}), \quad \mathbf{p}_{\text{зв}} = \pm(\mathbf{p}_{\text{фот}} - \mathbf{p}'_{\text{фот}}),$$

где плюс перед скобками относится к излучению, а минус — к поглощению фонона. Нештрихованными величинами обозначены энергия и импульс фотона до, а штрихованными — после излучения или поглощения фонона. Второе уравнение умножим на c/n, возведем оба уравнения в квадрат, а затем почленно вычтем. Тогда, используя связь между энергией и импульсом, получим

$$\left(\frac{c^2}{n^2 v_{\text{зв}}^2} - 1\right)\mathscr{E}_{\text{зв}}^2 = 4\mathscr{E}_{\text{фот}}\mathscr{E}'_{\text{фот}}\sin^2\frac{\theta}{2},$$

где θ — угол между направлениями падающего и рассеянного фотонов. В последнем уравнении слева единицей в скобках можно пренебречь, а справа $\mathscr{E}'_{\text{фот}}$ заменить на $\mathscr{E}_{\text{фот}}$, так как энергия фонона пренебрежимо мала. Сделав это и извлекая квадратный корень, получим

$$\mathscr{E}_{\text{зв}} = \pm 2n\frac{v_{\text{зв}}}{c}\,\mathscr{E}_{\text{фот}}\sin\frac{\theta}{2},$$

или

$$\mathscr{E}_{\text{фот}} - \mathscr{E}'_{\text{фот}} = \pm 2n\frac{v_{\text{зв}}}{c}\,\mathscr{E}_{\text{фот}}\sin\frac{\theta}{2}. \tag{57.8}$$

Это соотношение — чисто классическое. При его выводе были использованы только законы сохранения энергии и импульса, а также связь между энергией и импульсом для света и звука, которая также является классической. Переход от энергии к частоте производится уже с помощью квантовых соотношений $\mathscr{E}_{\text{фот}} = \hbar\omega$ и $\mathscr{E}_{\text{зв}} = \hbar\omega_{\text{фон}}$, причем существенно, что в обоих соотношениях постоянная $\hbar$ — одна и та же. В результате при квантовой интерпретации получается такая же формула

$$\omega - \omega' = \pm 2n\frac{v_{\text{зв}}}{c}\,\omega\sin\frac{\theta}{2}, \tag{57.9}$$

как и в классической теории. Однако окончательный результат совершенно не зависит от числового значения постоянной Планка.

§ 58. Энергетические зоны в твердых телах

1. В § 100 т. III было введено понятие об энергетических зонах в кристаллах. Это понятие существенно опирается на квантовые представления, о которых в т. III могли быть сообщены лишь предварительные сведения. Поэтому здесь мы опять вернемся к вопросу об энергетических зонах. Нас будет интересовать только принципиальная качественная сторона вопроса, а не точные количественные результаты,

требующие сложных и громоздких вычислений. К тому же для реальных кристаллов при настоящем состоянии теории провести точные вычисления невозможно. Мы вынуждены довольствоваться простейшими моделями и наглядными соображениями. Содержание этого параграфа полезно сопоставить с содержанием § 100 т. III.

Принципиально можно было бы поступить так, как это уже было намечено в т. III, § 100. Возьмем очень большое число тождественных атомов, удаленных друг от друга настолько далеко, что взаимодействие между ними очень мало и практически ни в чем не проявляется. В таком случае каждый из атомов ведет себя как изолированный. Ему свойственны определенные энергетические уровни. Система из N удаленных атомов будет иметь те же уровни энергии, но каждый уровень повторится N раз. Начнем теперь непрерывно сближать атомы. Появится взаимодействие между ними, в результате чего каждый первоначальный энергетический уровень станет как-то непрерывно смещаться. В конце концов из большого числа N атомов образуется кристалл. Проследив за эволюцией отдельных энергетических уровней, можно определить и энергетические уровни образовавшегося кристалла.

Конечно, при большом числе N атомов провести эту программу практически невозможно. Но это можно сделать для случая двух атомов. Полученные результаты можно будет использовать для выяснения вопроса, как качественно будет вести себя и система из большого числа атомов. Более того, для дальнейшего упрощения можно заменить реальный атом мысленным «одномерным атомом», проще всего — гармоническим осциллятором. Это фактически уже было сделано в § 52 при выяснении природы молекулярных сил (см. также т. III, § 137). Повторим еще раз полученные там результаты в форме, приспособленной для целей настоящего параграфа.

2. Пока осцилляторы раздвинуты достаточно далеко, гамильтониан системы обоих осцилляторов, если сохранять обозначения § 52, представится в виде

$$\widehat{H} = \widehat{H}_1 + \widehat{H}_2, \tag{58.1}$$

где

$$\widehat{H}_1 = -\frac{\hbar^2}{2\mu}\frac{\partial^2}{\partial x_1^2} + \frac{\mu}{2}\omega_0^2 x_1^2, \quad \widehat{H}_2 = -\frac{\hbar^2}{2\mu}\frac{\partial^2}{\partial x_2^2} + \frac{\mu}{2}\omega_0^2 x_2^2. \tag{58.2}$$

Осцилляторы ведут себя независимо друг от друга. Энергия каждого из них квантуется и равна $\hbar\omega_0(n + 1/2)$. Энергия системы в целом равна сумме энергий обоих осцилляторов. Ввиду тождественности осцилляторов одна и та же энергия $\mathscr{E}$ системы может быть представлена двумя способами: либо как $\mathscr{E} = \mathscr{E}_1^{\mathrm{I}} + \mathscr{E}_2^{\mathrm{II}}$, либо как $\mathscr{E} = \mathscr{E}_2^{\mathrm{I}} + \mathscr{E}_1^{\mathrm{II}}$. В первом случае осциллятор I имеет энергию $\mathscr{E}_1$, осциллятор II — энергию $\mathscr{E}_2$. Во втором случае полная энергия $\mathscr{E}$ та же, но энергией $\mathscr{E}_1$ обладает осциллятор II, а энергией $\mathscr{E}_2$ — осциллятор I. Это значит, что энергетический уровень $\mathscr{E}$ системы осцилляторов двукратно вырожден.

При сближении осцилляторов в гамильтониане системы

$$\widehat{H} = \widehat{H}_1 + \widehat{H}_2 - \lambda x_1 x_2 \tag{58.3}$$

появляется член $-\lambda x_1 x_2$, учитывающий взаимодействие между ними. Чтобы выяснить влияние этого члена, проще всего ввести нормальные координаты

$$q_1 = \frac{x_1 + x_2}{\sqrt{2}}, \quad q_2 = \frac{x_1 - x_2}{\sqrt{2}}. \tag{58.4}$$

Тогда гамильтониан системы преобразуется в

$$\widehat{H} = \left(-\frac{\hbar^2}{2\mu}\frac{\partial^2}{\partial q_1^2} + \frac{\mu}{2}\omega_1^2 q_1^2\right) + \left(-\frac{\hbar^2}{2\mu}\frac{\partial^2}{\partial q_2^2} + \frac{\mu}{2}\omega_2^2 q_2^2\right), \tag{58.5}$$

где

$$\omega_1^2 = \omega_0^2 - \frac{\lambda}{\mu}, \quad \omega_2^2 = \omega_0^2 + \frac{\lambda}{\mu}. \tag{58.6}$$

Вид гамильтониана (58.5) показывает, что в нормальных координатах q_1 и q_2 система совершает два независимых коллективных колебания с частотами ω_1 и ω_2. Можно сказать и иначе. Введение нормальных координат формально соответствует переходу к описанию движения системы осцилляторов посредством двух квазичастиц, гармонически колеблющихся с частотами ω_1 и ω_2. Движение каждой квазичастицы описывает не движение отдельного осциллятора, а обоих осцилляторов вместе. Общее движение всей системы складывается из наложения движений обеих квазичастиц. Энергия первой квазичастицы представляется выражением $\hbar\omega_1(n_1 + 1/2)$, второй $\hbar\omega_2(n_2 + 1/2)$, причем оба эти уровня не вырождены. Таким образом, в результате взаимодействия происходит расщепление двукратно вырожденного энергетического уровня на два уровня с частотами ω_1 и ω_2. Общее число уровней остается неизменным.

Все это можно было бы повторить без существенных затруднений для системы какого угодно числа тождественных гармонических осцилляторов. Задача сводится к введению нормальных координат, в которых кинетическая и потенциальная энергии системы одновременно приводятся к суммам квадратов. Такая процедура давно разработана в линейной алгебре и широко применяется, например, в теории малых колебаний колебательных систем со многими степенями свободы. Вся система по-прежнему формально сводится к совокупности независимых квазичастиц (гармонических осцилляторов) с различными собственными частотами. Первоначально вырожденный уровень энергии расщепляется на несколько уровней, причем общее количество уровней при сближении удаленных осцилляторов остается неизменным.

3. Теперь можно вернуться к вопросу, поставленному в начале этого параграфа: что следует ожидать при сближении тождественных бесконечно удаленных атомов, если число их N не вообразимо велико? В исходном положении взаимодействия между атомами нет. Поэтому кратность вырождения каждого энергетического уровня системы громадного числа атомов чудовищно велика. Сближение атомов включает

взаимодействие между ними, и вырождение снимается. По аналогии с системой осцилляторов следует ожидать, что энергетические уровни системы атомов начинают расщепляться. В конечном положении, когда из атомов образуется кристалл, первоначально вырожденный энергетический уровень расщепляется на громадное число уровней. Вместо одного вырожденного уровня возникает громадное количество их, образующих *полосу* или *энергетическую зону.*

Изложенное наглядно поясняет, как из отдельных уровней изолированных атомов возникает зонная структура энергии кристалла. Все сказанное без существенных изменений может быть распространено и на кристаллы, состоящие не из одинаковых, а из различных атомов (например, NaCl).

4. Взаимодействие атомов при их сближении проявляется прежде всего на *электронах наружных оболочек*, волновые функции которых сильно перекрываются. На внутренних электронах, волновые функции которых перекрываются незначительно, это взаимодействие сказывается гораздо слабее. Поэтому глубоко расположенные атомные уровни переходят в твердом теле в очень узкие полосы, а высоко расположенные уровни порождают значительно более широкие энергетические зоны. Самыми широкими оказываются зоны, соответствующие уровням *валентных электронов.* По этой причине при нагревании твердых тел испускаются не линейчатые, а непрерывные оптические спектры. Действительно, испускание линий оптического спектра ввиду их длинноволновости происходит при переходах между энергетическими уровнями крайней зоны, расстояния между которыми невелики, а сами уровни расположены непрерывно. В то же время в рентгеновских трубках с антикатодами из тяжелых металлов при достаточно высоких напряжениях наряду со сплошным излучением возбуждаются тонкие линии характеристического излучения, соответствующие переходам между узкими внутренними энергетическими зонами.

Таким образом, энергетический спектр кристалла состоит из нескольких полос, называемых *энергетическими зонами*, каждая из которых возникает в результате расщепления энергетических уровней отдельных атомов из-за их взаимодействия. В этом процессе расщепления участвуют главным образом валентные электроны, т. е. электроны наружных оболочек атомов. Из-за необычайно большого числа уровней, на которые расщепляются энергетические уровни изолированных атомов, энергия кристалла в пределах каждой зоны фактически меняется непрерывно. Соседние энергетические зоны могут быть разделены промежутками конечной ширины. Энергия кристалла не может принимать значения, лежащие в этих промежутках. Они образуют «щели» в энергетическом спектре кристалла и называются *запрещенными зонами.* Все остальные зоны называются *разрешенными.* Но могут быть и такие случаи, когда соседние зоны вплотную примыкают или перекрываются друг с другом. Крайняя зона, которой соответствуют наивысшие уровни энергии, называется *зоной проводимости.* Более глубоко лежащие зоны, получающиеся из-за расщепления уровней валентных электронов, называются *валентными зонами.*

5. Напомним и расширим теперь то, что было сказано в § 100 т. III относительно механизма прохождения электрического тока через кристалл. Все кристаллы разделяются на *металлы*, *диэлектрики* и *полупроводники*. Чтобы исключить влияние теплового движения, будем сначала предполагать, что температура кристалла равна абсолютному нулю. По принципу Паули на каждом энергетическом уровне может находиться не более двух электронов с противоположно направленными спинами. В равновесном состоянии будут заполнены электронами самые низкие энергетические уровни, а все вышележащие уровни окажутся свободными.

В диэлектрике валентная зона целиком заполнена электронами, а лежащая выше зона проводимости, отделенная от нее энергетической щелью значительной ширины, совсем не содержит электронов. Поэтому в электрических полях (слабее пробойного) диэлектрик не проводит электрического тока. В самом деле, при наложении постоянного внешнего электрического поля к гамильтониану кристалла добавляется малая поправка, соответствующая энергии кристалла во внешнем электрическом поле. Уравнение Шредингера всего кристалла слегка изменяется. Но так как число энергетических уровней зоны остается прежним, это приводит только к незначительному смещению энергетических уровней, что проявляется в *поляризации кристалла*. Существенно, однако, что и после наложения электрического поля зона остается по-прежнему *полностью заполненной электронами*. Никакие изменения состояния валентной зоны происходить не могут, поскольку с квантовой точки зрения все они сводятся к переходам электронов с одних энергетических уровней на другие. А это невозможно, если все уровни в валентной зоне заняты электронами. Электрический ток есть движение электронов, при котором они непрерывно переходят из одного состояния в другое. Следовательно, электроны, пока они находятся в целиком заполненной валентной зоне, *не могут участвовать в создании тока*. Для этого в валентной зоне необходимо создать незаполненные уровни энергии, т. е. перебросить некоторые электроны из валентной зоны в зону проводимости. Электрическое поле слабее пробойного этого сделать не может (если отвлечься от туннельных переходов из валентной зоны в зону проводимости). В более сильных полях возникает *электрический пробой* диэлектриков.

6. В металлах валентная зона заполнена электронами *целиком*, а зона проводимости лишь *частично*. Не имеет значения, существует ли щель между валентной зоной и зоной проводимости. Они могут вплотную примыкать и даже перекрываться между собой. Существенно только, чтобы в зоне проводимости были состояния, *не занятые электронами*. При наложении внешнего электрического поля $\mathbf{E}$ квантовые состояния внутри зоны проводимости слегка изменяются, однако состояния, не занятые электронами, в ней остаются. У электронов имеется возможность переходить в такие незанятые состояния. Переход, связанный с движением электрона в направлении поля $\mathbf{E}$, т. е. против действующей на него силы, сопровождается увеличением энергии кристалла. Переход, связанный с обратным направлением движения элек-

трона, наоборот, уменьшает эту энергию. Поэтому переходы электронов против поля **E** более вероятны и будут преобладать над переходами в обратном направлении. Через кристалл потечет электрический ток в направлении поля **E**, в соответствии с соглашением, принятым для определения направления тока.

Переносимые заряды, если бы они не убирались от границ кристалла, создали бы электрическое поле, противоположное по направлению приложенному полю **E**. В результате ток в конце концов прекратился бы. Чтобы этого не было, надо указанные заряды убирать и тем самым поддерживать приложенное поле **E**. Можно было бы также в кристалле создать вихревое электрическое поле, например, помещением его в переменное магнитное поле. В таком случае равновесие невозможно, и внутри кристалла все время будет циркулировать индукционный электрический ток.

7. В полупроводниках, как и в диэлектриках, валентная зона *полностью заполнена* электронами, а зона проводимости *полностью свободна*. Обе зоны отделены одна от другой щелью конечной ширины. Однако в полупроводниках эта щель значительно уже, чем в диэлектриках. При абсолютном нуле температуры это обстоятельство не играет роли, так как в этом случае переход электрона из валентной зоны в зону проводимости невозможен (если не учитывать туннельные переходы). Но если температура кристалла отлична от абсолютного нуля, то электрон в валентной зоне может получить от иона кристаллической решетки энергии порядка kT и перейти в зону проводимости. (Именно порядка kT, так как из-за малой концентрации электронов в зоне проводимости они практически подчиняются классической статистике Больцмана.) Такой переход может быть осуществлен и другим способом, например освещением кристалла. Независимо от способа перехода электронов в зону проводимости кристалл приобретает способность проводить электрический ток.

Как и в металлах, в полупроводниках проводимость создается электронами, перешедшими в зону проводимости. Но существует и другой механизм проводимости. Электрон, ушедший из валентной зоны, оставляет в ней незаполненное состояние, называемое *дыркой*. Другой электрон в валентной зоне получает возможность перейти в это незаполненное состояние. При этом в валентной зоне создается новая дырка, в которую может перейти третий электрон, и т. д. Разумеется, при наличии электрического поля **E** переходы, связанные с движением электронов против **E**, более вероятны, чем переходы, связанные с обратным движением. Это и создает ток в направлении **E**. Вместе с движением электрона происходит движение и соответствующей дырки, но в обратном направлении. Явление происходит так, как если бы ток вызывался не движением отрицательных электронов, а противоположно направленным движением положительно заряженных дырок.

Примеси, даже в ничтожных количествах, сильно повышают электрическую проводимость полупроводника. В запрещенной зоне (щели между валентной зоной и зоной проводимости) они создают добавочные энергетические уровни. Допустим, что такие уровни расположены

вблизи края зоны проводимости. Тогда создается возможность для перехода электронов с таких уровней в зону проводимости. В результате проводимость полупроводника возрастает. Соответствующие примеси называются *донорами*. Они поставляют электроны в зону проводимости и увеличивают проводимость полупроводника. Если же добавочные уровни находятся вблизи края валентной зоны, то электроны из валентной зоны получают возможность переходить на эти уровни. В валентной зоне образуются дырки. Проводимость полупроводника также увеличивается. Она называется *дырочной проводимостью*, а соответствующая примесь — *акцептором*.

8. Проводимость полупроводников сильно увеличивается с повышением температуры. В чистых полупроводниках увеличение интенсивности теплового движения электронов облегчает вероятность перехода их через энергетическую щель из валентной зоны в зону проводимости. При наличии примесей также возрастает вероятность переходов электронов с примесных уровней в зону проводимости или переходов электронов из валентной зоны на те же примесные уровни. Все эти процессы ведут к повышению в полупроводнике концентрации носителей тока: электронов и дырок. С этим и связано увеличение проводимости полупроводников с повышением температуры.

Иначе ведут себя металлы. Концентрация электронов в них почти не зависит от температуры. Проводимость определяется главным образом *длиной свободного пробега* электрона. Длина свободного пробега отчасти ограничивается примесями решетки. Соответствующая часть проводимости не зависит от температуры. В бесконечной идеальной кристаллической решетке, лишенной всяких примесей и дефектов, плоская волна де Бройля, соответствующая движению электрона, распространялась бы *без затухания и рассеяния*. Проводимость такого кристалла была бы *бесконечно велика*. В действительности свободный пробег электрона ограничен размерами кристалла (ср. с § 57, п. 4), а также тепловыми флуктуациями и дефектами кристаллической решетки. Они возрастают с температурой, уменьшая длину свободного пробега электрона. В результате *проводимость чистых металлов уменьшается с повышением температуры*. (Здесь мы исключили из рассмотрения сверхпроводники.)

§ 59. Зонная структура и волны Блоха

1. Зонную структуру энергетического спектра кристалла можно получить также из уравнения Шредингера для стационарных состояний. Однако точное уравнение Шредингера для кристалла в целом решить и исследовать невозможно из-за громадного числа частиц (электронов и атомных ядер) в кристалле. Это уравнение необходимо предварительно упростить. При таком упрощении считается, что состояние кристалла можно приближенно описать не полной волновой функцией, зависящей от координат всех электронов и атомных ядер, а *одночастичными волновыми функциями* каждого электрона. Каждая из таких волновых функций зависит уже только от координат *одного* электрона, который

находится в силовом поле, создаваемом атомными ядрами и остальными электронами кристалла. Ядра ввиду их массивности считаются при этом неподвижными, а электроны как бы «размазаны» по всему кристаллу. Таким путем многочастичное волновое уравнение Шредингера заменяется одночастичным для каждого электрона.

Необходимо отметить, что потенциальное силовое поле, в котором находится рассматриваемый электрон, не задано, а само зависит от состояний электронов. Такое поле называется *самосогласованным*. Согласование состоит в том, что, с одной стороны, одночастичные волновые функции отдельных электронов формируются самосогласованным полем; с другой стороны, самосогласованное поле само зависит от вида одночастичных волновых функций электронов. Метод самосогласованного поля находит свое оправдание в том, что большинство результатов, к которым он приводит, согласуются с опытом. Это в свою очередь связано не с явным видом самосогласованного поля, а главным образом с его *пространственной периодичностью*, определяемой периодичностью самой кристаллической решетки.

Необходимо заметить также, что волновые функции достаточно ввести не для всех, а только для внешних, т. е. *валентных электронов*, сравнительно слабо связанных с атомами решетки. Остальные электроны можно считать прочно связанными с атомными ядрами. Получается модель решетки из положительно заряженных ионов, которые рассматриваются неподвижными. Пространственно периодическое самосогласованное поле создается такими ионами и плавающими между ними валентными электронами.

2. В такой постановке рассмотрим задачу о виде стационарных одночастичных волновых функций электрона в кристалле и структуре энергетического спектра электрона, предполагая, что внешних силовых полей нет, а сама решетка простирается бесконечно во все стороны. Во избежание математических усложнений ограничимся рассмотрением *одномерного кристалла* (т. е. бесконечной прямолинейной цепочки одинаковых атомов, находящихся на постоянном расстоянии a друг от друга). Уравнение Шредингера для стационарных состояний в этом случае имеет вид

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2}+U(x)\psi=\mathscr{E}\psi, \tag{59.1}$$

или

$$\frac{d^2\psi}{dx^2}+\varkappa^2(x)\psi=0, \tag{59.2}$$

где введено обозначение

$$\varkappa^2=\frac{2m}{\hbar^2}(\mathscr{E}-U). \tag{59.3}$$

Это — линейное дифференциальное уравнение второго порядка с периодически меняющимися коэффициентами, поскольку из-за отсутствия внешнего поля $\varkappa^2(x+a)=\varkappa^2(x)$ для любого x. Оно называется *уравнением Хилла* (частный случай *уравнения Матьё*).

Исследуем общий вид решений уравнения Хилла, используя периодичность функции $\varkappa^2(x)$. В силу этой периодичности

$$\frac{d^2\psi(x+a)}{dx^2} + \varkappa^2(x+a)\psi(x+a) = \frac{d^2\psi(x+a)}{dx^2} + \varkappa^2(x)\psi(x+a).$$

Отсюда видно, что если функция $\psi(x)$ есть решение уравнения Хилла, то функция $\psi(x+a)$ будет также решением того же уравнения. Если $\psi_1(x)$ и $\psi_2(x)$ — какие-либо два произвольных линейно независимых решения уравнения Хилла, то общее решение его может быть представлено в виде

$$\psi(x) = c_1\psi_1(x) + c_2\psi_2(x),$$

где c_1 и c_2 — произвольные постоянные.

Докажем теперь, что среди решений уравнения Хилла существует такое решение $\Phi(x)$, что для любого $x\Phi(x+a) = \lambda\Phi(x)$, где λ — постоянная. Если такое решение существует, то его, разумеется, также можно представить в виде

$$\Phi(x) = c_1\psi_1(x) + c_2\psi_2(x).$$

Для упрощения вычислений выберем линейно независимые решения $\psi_1(x)$ и $\psi_2(x)$ так, чтобы они удовлетворяли условиям

$$\psi_1(0) = 1, \quad \psi_1'(0) = 0;$$

$$\psi_2(0) = 0, \quad \psi_2'(0) = 1.$$

(В этом случае говорят, что функции $\psi_1(x)$ и $\psi_2(x)$ образуют *фундаментальную систему решений.*) Тогда при $x = 0$ требования $\Phi(x+a) = \lambda\Phi(x)$ и $\Phi'(x+a) = \lambda\Phi'(x)$ запишутся в виде

$$c_1\psi_1(a) + c_2\psi_2(a) = \lambda c_1,$$

$$c_1\psi_1'(a) + c_2\psi_2'(a) = \lambda c_2,$$

или

$$[\psi_1(a) - \lambda]c_1 + \psi_2(a)c_2 = 0,$$

$$\psi_1'(a)c_1 + [\psi_2'(a) - \lambda]c_2 = 0.$$

Для совместности этих линейных однородных уравнений (относительно c_1 и c_2) необходимо и достаточно, чтобы выполнялось условие

$$\begin{vmatrix} \psi_1(a) - \lambda & \psi_2(a) \\ \psi_1'(a) & \psi_2'(a) - \lambda \end{vmatrix} = 0, \tag{59.4}$$

или

$$\lambda^2 - [\psi_1(a) + \psi_2'(a)]\lambda + [\psi_1(a)\psi_2'(a) - \psi_1'(a)\psi_2(a)] = 0.$$

Путем дифференцирования с учетом уравнения (59.2) нетрудно убедиться, что

$$\frac{d}{dx}\left[\psi_1(x)\psi_2'(x) - \psi_1'(x)\psi_2(x)\right] = 0.$$

Следовательно, функция, стоящая в квадратных скобках, постоянна. Ее значение найдется, если положить $x = 0$. Тогда она обратится в 1. Такое же значение эта функция будет иметь и при $x = a$. Введем, далее, обозначение

$$\psi_1(a) + \psi_2'(a) = 2L.$$

Величина L, разумеется, постоянна, поскольку используется однозначно определенное (фундаментальное) решение уравнения Хилла (59.2). Она называется *постоянной Ляпунова* (1857–1918). Ее значение определяется функцией $\varkappa$, т. е. в конце концов параметром $\mathscr{E}$. Теперь для определения λ получается квадратное уравнение

$$\lambda^2 - 2L\lambda + 1 = 0. \tag{59.5}$$

Из него найдем два значения λ:

$$\lambda_{1,2} = L \pm \sqrt{L^2 - 1}. \tag{59.6}$$

Тем самым определится не одно, а даже *два* решения уравнения (59.2): $\Phi_1(x)$ и $\Phi_2(x)$, обладающие требуемым свойством. (Случай равных корней мы исследовать не будем — он может быть рассмотрен предельным переходом $\lambda_1 \to \lambda_2$.) Решения $\Phi_1(x)$ и $\Phi_2(x)$ линейно независимы, и их удобно выбрать для представления общего решения в виде

$$\psi = c_1\Phi_1(x) + c_2\Phi_2(x). \tag{59.7}$$

Сами решения $\Phi_1(x)$ и $\Phi_2(x)$, конечно, не могут быть найдены, пока функция $\varkappa^2(x)$ (т. е. в конце концов самосогласованное поле U) неизвестна, а параметр $\mathscr{E}$ не фиксирован. Но это не мешает оперировать функциями $\Phi_1(x)$ и $\Phi_2(x)$ для установления общих свойств решений уравнения Хилла (59.2). А интересующие нас свойства таких решений существенно зависят от величины постоянной Ляпунова L.

3. Допустим сначала, что постоянная L по модулю больше единицы: $|L| > 1$. Тогда оба корня λ_1 и λ_2 вещественны и по модулю один больше, а другой меньше единицы, так как $\lambda_1\lambda_2 = 1$. Поэтому из уравнения $\Phi(x + a) = \lambda\Phi(x)$ следует, что при $x \to +\infty$ одна из функций $\Phi_1(x)$ или $\Phi_2(x)$ неограниченно возрастает, а при $x \to -\infty$ неограниченно возрастает другая. В силу этого при $|L| > 1$ ни одно из решений $\Phi_1(x)$ и $\Phi_2(x)$, а следовательно, и произвольная линейная комбинация их с постоянными коэффициентами (59.7), не может быть волновой функцией электрона в кристалле. Это означает, что в кристалле не существует состояний с энергией $\mathscr{E}$, для которых $|L| > 1$. Такие энергии образуют *запрещенные энергетические зоны.*

Если же $|L| < 1$, то можно положить $L = \cos ka$, где k — постоянная. Тогда

$$\lambda_{1,2} = \cos ka \pm i \sin ka = e^{\pm ika}.$$

Следовательно,

$$\Phi_{1,2}(x+a) = e^{\pm ika}\Phi_{1,2}(x), \tag{59.7а}$$

т. е. при изменении x на a модуль функции $\Phi(x)$ не изменяется. Меняется только ее фаза. Обе функции $\Phi_1(x)$ или $\Phi_2(x)$, а также их произвольная линейная комбинация с постоянными коэффициентами могут быть решениями уравнения Шредингера (59.1). Поэтому возможны только такие значения энергии $\mathscr{E}$, для которых модуль постоянной Ляпунова меньше единицы. Они образуют *разрешенные энергетические зоны кристалла.*

Как и в предыдущем параграфе, мы пришли к *зонной структуре энергетического спектра электрона в кристалле.* В пределах каждой зоны энергия электрона меняется непрерывно. Это, конечно, связано с предположением о *безграничности цепочки.* Если бы цепочка была ограничена, то на ее концах должны были бы выполняться определенные граничные условия, которые бы и превратили непрерывный спектр зоны в ряд более или менее тесно расположенных линий. Для пояснения изложенного полезно рассмотреть модельную задачу 1, приводимую в конце этого параграфа.

4. Волновые функции $\Phi_1(x)$ или $\Phi_2(x)$ могут быть представлены в виде

$$\Phi_{1,2}(x) = e^{\pm ikx}P_{1,2}(x).$$

Очевидно, с одной стороны,

$$\Phi_{1,2}(x+a) = e^{\pm ik(x+a)}P_{1,2}(x+a).$$

С другой стороны, ввиду (59.7а),

$$\Phi_{1,2}(x+a) = e^{\pm ika}\Phi_{1,2}(x) = e^{\pm ik(x+a)}P_{1,2}(x).$$

Следовательно,

$$P_{1,2}(x+a) = P_{1,2}(x), \tag{59.8}$$

т. е. обе функции $P_1(x)$ и $P_2(x)$ периодичны с периодом a.

Если еще учесть временной множитель $e^{-i\omega t}$, то в отсутствие внешних силовых полей возможные полные волновые функции в кристалле можно представить в виде

$$\begin{aligned}\Psi_1(x,t) &= P_1(x)e^{-i(\omega t-kx)},\\ \Psi_2(x,t) &= P_2(x)e^{-i(\omega t+kx)}.\end{aligned} \tag{59.9}$$

Эти волны описывают «свободное движение» электрона в кристалле, когда все действующие на него силы ограничены взаимодействиями с ионами кристаллической решетки и остальными электронами, а внешних силовых полей нет. Они называются *волнами Блоха* (р. 1905). В отличие от плоских волн де Бройля, распространяющихся в свободном пространстве, в волнах Блоха величины $P_1(x)$ и $P_2(x)$ не постоянны, а *пространственно модулированы*, т. е. периодически меняются вдоль цепочки с периодом a. Из-за такой пространственной модуляции при

свободном распространении Ψ-волн в кристалле величину $\hbar k$ называют *квазиимпульсом электрона*, тогда как при движении электрона в свободном пространстве такая величина есть просто импульс. В трехмерных кристаллах плоская волна Блоха имеет тот же вид, что и (59.9). Только $P(x)$ заменяется на функцию $P(\mathbf{r})$, обладающую той же пространственной периодичностью, что и сама решетка, а волновое число k — на волновой вектор $\mathbf{k}$, которому соответствует *квазиимпульс* $\hbar\mathbf{k}$.

5. Выбором постоянной Ляпунова, а с ней и квазиимпульса $\hbar\mathbf{k}$ определяется (с точностью до постоянного множителя) волновая функция Блоха. Следовательно, определяется и энергия электрона $\mathscr{E}$, которая входит в стационарное уравнение Шредингера в качестве постоянного параметра. Таким образом, в пределах каждой зоны допустимые значения энергии электрона могут быть представлены как функции квазиимпульса:

$$\mathscr{E} = \mathscr{E}(p). \tag{59.10}$$

Отсюда получается

$$\omega = \omega(k), \tag{59.11}$$

так как $\mathscr{E} = \hbar\omega$. Оба эти соотношения называются *законами дисперсии* электронных волн или электрона в кристалле. Законами дисперсии описывается взаимодействие рассматриваемого электрона с ионами кристаллической решетки и со всеми остальными электронами (заряд которых «размазан» в пространстве).

В плоской волне Блоха

$$\Psi(x,t) = P(x)e^{i(kx-\omega t)} \tag{59.12}$$

волновое число k определено не однозначно. Но причина этого иная, чем в случае акустической волны в кристаллической решетке, где неоднозначность связана с *дискретностью значений, которые может принимать координата* x (см. § 56). В волне же Блоха координата x меняется *непрерывно*. Зато «амплитуда» волны $P(x)$ является функцией x. Можно всегда изменить постоянную k и функцию $P(x)$ так, чтобы при этом волновая функция (59.12) осталась неизменной. С этой целью достаточно сделать замену

$$k = k' + 2\pi p/a, \quad P'(x) = P(x)e^{i2\pi px/a},$$

где p — произвольное целое число. Тогда (59.12) преобразуется в

$$\Psi(x,t) = P'(x)e^{i(k'x-\omega t)},$$

т. е. в волну с другой периодически меняющейся амплитудой $P'(x)$ и другим волновым числом k'. От такого преобразования физически ничего не изменится.

Пользуясь указанной неоднозначностью, при рассмотрении движения электрона в какой-либо зоне волновое число k (волновой вектор $\mathbf{k}$) можно выбрать так, чтобы его модуль принял *наименьшее значение* из всех возможных. Особый интерес представляет случай, когда

$k \ll 2\pi/a$, а «амплитуда» $P(x)$ содержит постоянное слагаемое, значительно превосходящее все остальные слагаемые, быстро меняющиеся в пространстве. (Это, как будет видно из дальнейшего, имеет место при движении электрона, энергия которого находится вблизи одной из границ зоны.) В этом случае функция Ψ станет волной постоянной амплитуды, на которую накладывается мелкая рябь, периодически повторяющаяся на каждом периоде решетки. При рассмотрении усредненного движения электрона от такой ряби можно отвлечься, т. е. оперировать с электронной волной уже постоянной амплитуды, получающейся из $P(x)$ путем пространственного усреднения. Длина такой усредненной волны $\Lambda \equiv 2\pi/k$ предполагается очень большой по сравнению с периодом решетки a. Можно построить пакет таких усредненных волн в узкой области Δk. Тогда скорость *усредненного движения электрона* будет равна *групповой скорости*

$$v = \frac{d\mathscr{E}}{dp}. \tag{59.13}$$

6. Выделим какую-либо разрешенную зону и рассмотрим соотношения вблизи ее границ, за которыми следуют запрещенные зоны. На нижней границе зоны энергия $\mathscr{E}$ минимальна, а на верхней максимальна. Значит, производная $d\mathscr{E}/dp$, а с ней и усредненная скорость электрона на обеих границах зоны, обращается в нуль. Поэтому вблизи границы каждой зоны разложение $\mathscr{E}$ по степеням $p = \hbar k$ может начаться только с квадратичного члена. Если энергию отсчитывать от края зоны, а квазиимпульсы в местах максимума или минимума считать равными нулю, то получится

$$\mathscr{E} = \frac{p^2}{2m_{\text{эф}}}, \tag{59.14}$$

где

$$\frac{1}{m_{\text{эф}}} = \left(\frac{d^2\mathscr{E}}{dp^2}\right)_{p=0}. \tag{59.15}$$

Величина $m_{\text{эф}}$ называется *эффективной массой электрона*. Она, как правило, не совпадает с настоящей массой электрона m_{e}. На нижней границе зоны (минимум $\mathscr{E}$) эффективная масса $m_{\text{эф}}$ *положительна*, тогда как на верхней границе (максимум $\mathscr{E}$) она *отрицательна* (рис. 105). В трехмерной кристаллической решетке роль величины $1/m_{\text{эф}}$ играет тензор $\partial^2\mathscr{E}/\partial p_i \partial p_j$, но сущность явлений, связанных с понятием эффективной массы, можно уяснить на одномерной модели, где масса $m_{\text{эф}}$ является скаляром.

Связь между энергией электрона и его квазиимпульсом вблизи границы зоны формально имеет такой же вид, что и связь между энергией и импульсом свободной частицы. Это и дает основание в рассматриваемом случае называть величину $m_{\text{эф}}$ эффективной массой, а квазиимпульс $\mathbf{p}$ во многих руководствах называют просто импульсом. Специфика усредненного движения электрона в периодической

кристаллической решетке вблизи границы зоны заключается, однако, в том, что эффективная масса электрона может быть и положительной, и отрицательной. В силу этого *вблизи нижней границы зоны векторы* $\mathbf{v}$ *и* $\mathbf{p}$ *направлены одинаково, а вблизи верхней — противоположно*, как это видно из формулы (59.13) и рис. 105.

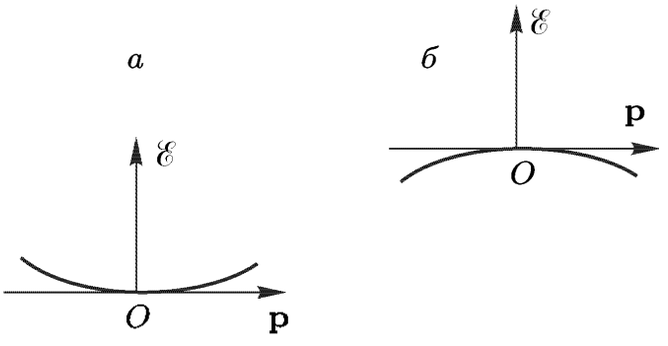


Рис. 105

7. До сих пор предполагалось, что внешнего силового поля нет. Наложим теперь на кристалл слабое постоянное электрическое поле. Оно в принципе несколько сместит энергетические уровни в кристалле. Но так как эти уровни практически непрерывны, то никакого изменения в положении уровней внутри зоны не произойдет. Могут только слегка сместиться границы самой зоны. Но в слабых полях этот эффект совсем не заметен и ни в чем не проявляется. Наложенное электрическое поле может, однако, вызвать изменения в заполнении энергетических уровней электронами. Это может произойти только тогда, когда по крайней мере некоторые энергетические уровни зоны *заполнены не целиком, а частично.* Согласно принципу Паули на одном уровне может находиться не более двух электронов, спины которых направлены противоположно. Если зона заполнена целиком, то электроны могут только обмениваться состояниями, что в силу их тождественности физически ни в чем не проявляется. Никакие движения электронов, связанные с их переходами из одного состояния в другое, невозможны. В этом случае при наложении электрического поля в кристалле и не возникнет электрического тока.

В состоянии равновесия при отсутствии внешнего поля в частично заполненной зоне будут заполнены электронами все *наиболее низкие* энергетические уровни. На каждом из них будут находиться два электрона с противоположно направленными спинами. При наложении электрического поля начнутся *переходы* из занятых состояний в свободные и *возникнет электрический ток.* Скорость усредненного движения электрона при этом определяется выражением (59.13). Влияние кристалла на движение электрона уже учтено дисперсионной формулой (59.10). Но на электрон в электрическом поле E действует еще внешняя сила $F = -eE$. Изменение энергии электрона за время dt под действием этой силы будет $d\mathscr{E} = Fv\,dt$. Но в силу (59.10) $d\mathscr{E} = (d\mathscr{E}/dp)dp = v\,dp$. Приравнивая оба выражения, получаем $dp = F\,dt$, т. е.

$$\frac{dp}{dt} = F. \tag{59.16}$$

Та же формула получается и в трехмерном случае. Только скаляры p и F следует заменить векторами $\mathbf{p}$ и $\mathbf{F}$. Получится формула, вполне соответствующая классической. Дифференцирование же соотношения

(59.13) по времени дает $\dot{v} = (d^2\mathscr{E}/dp^2)(dp/dt)$, или на основании формул (59.15) и (59.16)

$$m_{\text{эф}}\dot{v} = F. \tag{59.17}$$

8. Рассмотрим специально случай, когда электроны заполняют *почти всю зону*. В этом случае ток связан с наличием свободных состояний вблизи верхней границы зоны (рис. 105 *б*), так что эффективная масса электрона отрицательна. Тогда согласно (59.17) ускорение электрона $\dot{v}$ направлено против действующей силы $F = eE$, т. е. по полю E ($e < 0$). Электрон ведет себя как *отрицательно заряженная частица, но с отрицательной массой* $m_{\text{эф}}$. Но в точности *так же будет вести себя и воображаемая частица, у которой масса и заряд положительны*. Какие знаки приписать массе и заряду воображаемой частицы — не имеет значения. Существенно только, чтобы они были одинаковы. Но электрон с положительным зарядом и положительной массой ведет себя в точности так же, как «дырка», введенная в предыдущем параграфе. А так как число электронов, которые принимают участие в электрическом токе, в точности равно числу вакантных мест (дырок) в зоне, то *носителями тока формально могут считаться дырки*. Тем самым становится понятным, почему коэффициент Холла, например, может иметь не только отрицательный знак, но и положительный (см. т. III § 98).

9. Заметим в заключение, что в идеальной кристаллической решетке с неподвижными ионами плоская волна Блоха распространялась бы *без затухания*. Электрическое сопротивление кристалла в таком случае было бы равно нулю. Тепловые колебания, дефекты и примеси приводят к рассеянию электронных волн, т. е. ограничивают длины свободного пробега электрона, с чем и связано возникновение электрического сопротивления.

ЗАДАЧИ

1. Рассмотреть одномерную прямолинейную бесконечную кристаллическую решетку, моделируя потенциальную энергию $U(x)$ ступенчатой функцией, изображенной на рис. 106. Найти разрешенные и запрещенные зоны для такой цепочки, задав значения a и U, характерные для атомных размеров.

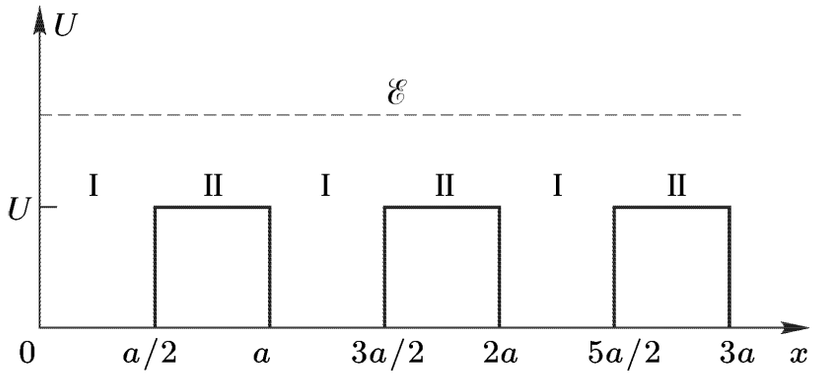


Рис. 106

Р е ш е н и е. На участках I примем потенциальную энергию равной нулю, а на участках II — постоянному значению U. Стационарное уравнение Шредингера с постоянной энергией $\mathscr{E}$ на участках I будет

$$\frac{d^2\psi}{dx^2} + \varkappa_1^2\psi = 0, \tag{59.18}$$

а на участках II

$$\frac{d^2\psi}{dx^2} + \varkappa_2^2\psi = 0, \tag{59.19}$$

где $\varkappa_1$ и $\varkappa_2$ — постоянные:

$$\varkappa_1^2 = \frac{2m}{\hbar^2}\,\mathscr{E}, \quad \varkappa_2^2 = \frac{2m}{\hbar^2}\,(\mathscr{E} - U).$$

Будем сначала предполагать, что $\mathscr{E} > 0$ и $\mathscr{E} - U > 0$. Тогда $\varkappa_1$ и $\varkappa_2$ будут вещественными. Без нарушения общности их можно считать положительными. В интервале $(0, a/2)$ система фундаментальных решений представится функциями

$$\psi_1 = \cos\varkappa_1 x, \quad \psi_2 = \frac{1}{\varkappa_1}\sin\varkappa_1 x.$$

Найдем теперь эти функции в интервале $(a/2, a)$. В этом интервале представим первую функцию в виде

$$\psi_1 = A\cos\varkappa_2\left(x - \frac{a}{2}\right) + B\sin\varkappa_2\left(x - \frac{a}{2}\right).$$

Неизвестные коэффициенты A и B найдутся из условий непрерывности функции $\psi_1(x)$ и ее производной $\psi_1'(x)$ на границе интервала $x = a/2$. Таким путем получаем, что в интервале $(a/2, a)$

$$\psi_1(x) = \cos\frac{\varkappa_1 a}{2}\cos\varkappa_2\left(x - \frac{a}{2}\right) - \frac{\varkappa_1}{\varkappa_2}\sin\frac{\varkappa_1 a}{2}\sin\varkappa_2\left(x - \frac{a}{2}\right).$$

Аналогично находим, что в том же интервале

$$\psi_2(x) = \frac{1}{\varkappa_1}\sin\frac{\varkappa_1 a}{2}\cos\varkappa_2\left(x - \frac{a}{2}\right) + \frac{1}{\varkappa_2}\cos\frac{\varkappa_1 a}{2}\sin\varkappa_2\left(x - \frac{a}{2}\right).$$

Для постоянной Ляпунова получается

$$L = \frac{1}{2}\left[\psi_1(a) + \psi_2'(a)\right] = \cos\frac{\varkappa_1 a}{2}\cos\frac{\varkappa_2 a}{2} - \frac{1}{2}\left(\frac{\varkappa_1}{\varkappa_2} + \frac{\varkappa_2}{\varkappa_1}\right)\sin\frac{\varkappa_1 a}{2}\sin\frac{\varkappa_2 a}{2}. \tag{59.20}$$

В случае, когда $\mathscr{E} > 0$, но $\mathscr{E} - U < 0$, изменим обозначения, заменив прежнее $\varkappa_2$ на мнимую величину $i\varkappa_2$, т. е. положим $\varkappa_2^2 = 2m(U - \mathscr{E})/\hbar^2$. Тригонометрические функции от мнимого аргумента следует заменить на гиперболические функции. Тогда формула (59.20) преобразуется:

$$L = \cos\frac{\varkappa_1 a}{2}\operatorname{ch}\frac{\varkappa_2 a}{2} - \frac{1}{2}\left(\frac{\varkappa_1}{\varkappa_2} - \frac{\varkappa_2}{\varkappa_1}\right)\sin\frac{\varkappa_1 a}{2}\operatorname{sh}\frac{\varkappa_2 a}{2}. \tag{59.21}$$

Наконец, когда $\mathscr{E} < 0$ и $\mathscr{E} - U < 0$, надо сделать вторую замену $\varkappa_1 \to i\varkappa_1$ (т. е. положить $\varkappa_1^2 = -2m\mathscr{E}/\hbar^2$). Тогда

$$L = \operatorname{ch}\frac{\varkappa_1 a}{2}\operatorname{ch}\frac{\varkappa_2 a}{2} + \frac{1}{2}\left(\frac{\varkappa_1}{\varkappa_2} + \frac{\varkappa_2}{\varkappa_1}\right)\operatorname{sh}\frac{\varkappa_1 a}{2}\operatorname{sh}\frac{\varkappa_2 a}{2}. \tag{59.22}$$

Формулы (59.20)–(59.22) имеют довольно сложный вид. Их исследование удобно проводить только графически на примерах. Приведем числовой пример, полагая ориентировочно $a = 2 \cdot 10^{-8}$ см, $U = 5$ эВ. Соответствующая кривая для $L = L(\mathscr{E})$ приведена на рис. 107. На заштрихованных участках величина $|L|$ меньше единицы. Эти участки в нашей модели являются разрешенными зонами. Светлые участки, где $|L| > 1$, соответствуют запрещенным зонам.

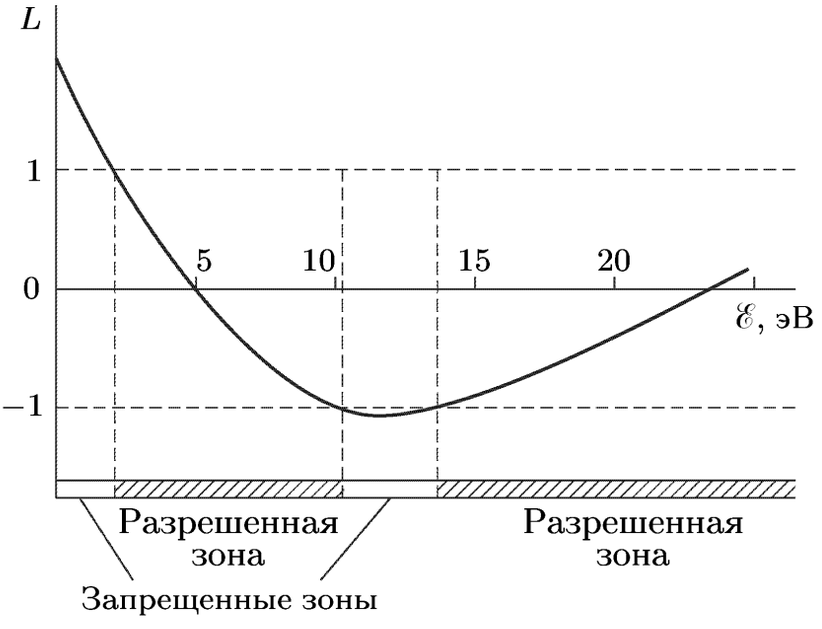


Рис. 107

2. Полюсы батареи соединены кристаллом, зона проводимости которого почти доверху заполнена электронами. В таком случае эффективная масса электрона отрицательна, и он движется через кристалл с ускорением в направлении электрического поля. Туда же будет направлена и средняя скорость электрона, так как до наложения поля средняя скорость была равна нулю. Иными словами, электрон движется через кристалл от анода к катоду. Следовательно, внутри батареи движение электрона происходит от катода к аноду, а ток внутри батареи в соответствии с принятым соглашением течет и противоположном направлении — от анода к катоду. Но такой ток заряжает батарею, а это противоречит закону сохранения энергии. Разрешить этот парадокс.

Решение. Все дело в том, что в этом рассуждении не учтено перераспределение электронов по возможным состояниям зоны при наложении электрического поля. На рис. 108 изображен участок дисперсионной кривой для рассматриваемого кристалла. При отсутствии внешнего поля штриховая горизонтальная прямая AB отсекает от этой кривой верхнюю часть, не заполненную электронами. В заполненных же частях столько же электронов движется направо, сколько и налево. Поэтому ток через кристалл не идет. При наложении электрического поля на электрон начинает действовать сила $F = eE$, направленная противоположно $\mathbf{E}$, так как заряд электрона отрицательный. Для конкретности примем, что поле направлено налево, а следова-

тельно, сила направо. В соответствии с (59.16) импульсы электронов справа по абсолютной величине должны увеличиваться, а слева уменьшаться. Это значит, что конец B незаполненного участка дисперсионной кривой должен опускаться, переходя в B', а конец A подниматься, переходя в A'. Скорость электронов направлена против $\mathbf{p}$, т. е. на нашем рисунке в левой половине направо, а в правой налево. Но в левой половине электронов больше, чем в правой. Поэтому через кристалл поток электронов будет идти слева направо, т. е. в направлении действующей силы $\mathbf{F}$(или против поля $\mathbf{E}$). Такой ток будет разряжать, а не заряжать батарею.

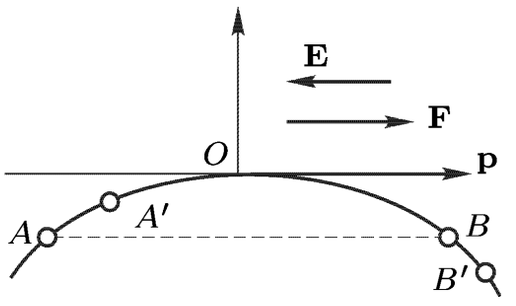


Рис. 108

Аналогичное рассуждение можно провести и в том случае, когда электронами заполнена небольшая нижняя часть зоны проводимости.

§ 60. Сверхтекучесть. Опытные факты

1. Кристаллизация (затвердевание) вещества происходит под действием сил притяжения между атомами (молекулами). Ей препятствует беспорядочное движение этих частиц. С классической точки зрения это движение только тепловое и полностью прекращается при абсолютном нуле температур. Поэтому с этой точки зрения при приближении к абсолютному нулю всякое вещество в конце концов должно перейти в твердое состояние. Не совсем так обстоит дело с точки зрения квантовой механики. Помимо теплового движения атомы совершают *нулевые колебания*, не прекращающиеся и при температуре абсолютного нуля. Их частота ω (а с ней и нулевая энергия $\hbar\omega/2$ тем больше, чем меньше масса атома ($\omega \sim 1/\sqrt{m}$). Если силы притяжения между атомами недостаточны, чтобы противостоять нулевым колебаниям, то даже при температуре абсолютного нуля тело не перейдет в твердое состояние, а останется жидким, если только давление не очень велико. Именно так обстоит дело с гелием. Причина этого двоякая. Во-первых, из всех инертных газов массы атомов гелия минимальны. Во-вторых, инертные газы имеют замкнутые (застроенные) наружные оболочки, характеризующиеся сферической симметрией. Такова и единственная K-оболочка атома гелия, состоящая всего из двух электронов. Из-за этого силы притяжения между атомами Не относительно слабы, а плотность жидкого гелия мала (при температуре 2,20 К и давлении насыщенных паров она равна около 0,146 г/см^3). Поэтому при нормальном давлении гелий остается жидким и при температуре абсолютного нуля. В этом прежде всего проявляются *квантовые свойства* жидкого гелия, благодаря чему его называют *квантовой жидкостью*. Из всех известных веществ свойством оставаться жидким вплоть до температуры абсолютного нуля обладает только гелий. Но гелий можно получить и в твердом состоянии. Для этого надо повысить внешнее давление до 25,3 атм или выше. Повышение давления как бы эквивалентно усилению сил притяжения между атомами гелия.

Существуют два стабильных изотопа гелия — $^4_2\mathrm{He}$ и $^3_2\mathrm{He}$. В воздухе содержится 0,0005 % He по объему. Природный гелий состоит почти исключительно из $^4_2\mathrm{He}$ и содержит только ничтожную часть $^3_2\mathrm{He}$. Например, в смеси с $^4_2\mathrm{He}$, полученной из воздуха, компонент $^3_2\mathrm{He}$ составляет всего 10^{-7} % общей массы газа. Детальное изучение свойств $^3_2\mathrm{He}$ стало возможным лишь после получения его в макроскопических количествах искусственными способами. Образуется $^3_2\mathrm{He}$ в результате β-распада трития:

$$^3_1\mathrm{H} \to {}^3_2\mathrm{He} + e^-.$$

Сам же тритий $^3_1\mathrm{H}$ получают с помощью другой ядерной реакции:

$$^6_3\mathrm{Li} + {}^1_0\mathrm{n} \to {}^3_1\mathrm{H} + {}^4_2\mathrm{He},$$

происходящей при бомбардировке ядер $^6_3\mathrm{Li}$ нейтронами. Существуют и другие ядерные реакции, в которых выделяется $^3_2\mathrm{He}$.

2. Ожижение природного гелия было произведено в 1908 г. Камерлинг-Оннесом (1853–1926) в Лейденской лаборатории. Только после этого открылась возможность исследования свойств веществ вблизи абсолютного нуля температур. Критическая температура $^4_2\mathrm{He}$ составляет $T_\mathrm{к} = 5{,}25$ K, так что в жидком состоянии он может существовать только ниже этой температуры. Под давлением в 1 атм $^4_2\mathrm{He}$ кипит при температуре 4,21 K. Еще более низкую критическую температуру (3,35 K) имеет более легкий изотоп $^3_2\mathrm{He}$. Под давлением в 1 атм жидкий $^3_2\mathrm{He}$ кипит при температуре 3,195 K. Естественный гелий был получен в твердом состоянии в 1926 г. Кеезомом (1876–1956), сменившим Камерлинг-Оннеса (после его смерти) на посту директора Лейденской лаборатории. Для получения твердого гелия при абсолютном нуле температур требуется давление не менее 25,3 атм, а при 60 K — порядка 10 000 атм. Твердый гелий — совершенно прозрачное и бесцветное вещество. В дальнейшем, где говорится о свойствах жидкого гелия, имеется в виду тяжелый изотоп $^4_2\mathrm{He}$. Только в следующем параграфе затрагиваются свойства легкого изотопа $^3_2\mathrm{He}$. Диаграмма состояния $^4_2\mathrm{He}$ приведена на рис. 109.

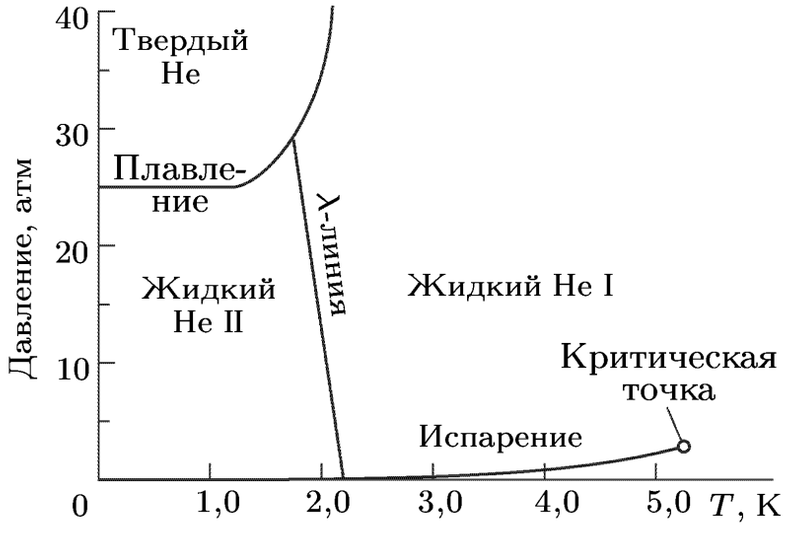


Рис. 109

При понижении температуры жидкий гелий претерпевает обратимое фазовое превращение второго рода, т. е. такое превращение, при котором не происходит выделения или поглощения тепла, а удельный объем жидкости остается неизменным (см. т. II, § 120). Температура перехода T_λ понижается с повышением давления по линейному закону от максимального значения 2,17 K при давлении насыщенных паров гелия (37,80 мм рт. ст.) до минимального значения 1,77 K при давлении 30 атм. На диаграмме состояния (T, p) точки, в которых совершается указанный фазовый переход, располагаются вдоль прямой, называемой *λ-линией*. Жидкость, температура которой (при одном и том же давлении) выше T_λ, называется *гелием-I*, а жидкость, температура которой ниже T_λ — *гелием-II*. Фазовый переход проявляется, во-первых, в *скачке теплоемкости* при температуре T_λ. По обе стороны от этой температуры теплоемкость гелия обращается в бесконечность по логарифмическому закону $C = \text{const} \cdot \ln\left(|T - T_\lambda|^{-1}\right)$, но с различными коэффициентами пропорциональности (рис. 110). Кривая теплоемкости напоминает греческую букву λ. Отсюда и произошли названия *λ-точка* и *λ-линия*. Во-вторых, при переходе He I $\rightleftarrows$ He II скачкообразно меняется *температурный коэффициент расширения*, который для He II оказывается даже отрицательным.

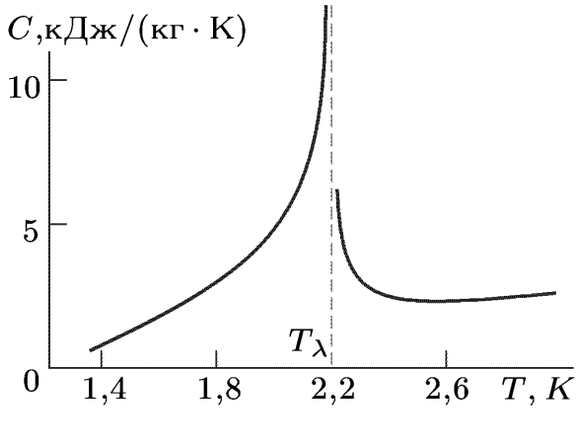


Рис. 110

Гелий-I ведет себя как обычная жидкость и обладает конечными значениями вязкости и теплопроводности. Гелий же II обладает рядом специфических свойств, совершенно непонятных с классической точки зрения и истолкованных только на основе квантовых представлений. В этом состоит вторая причина, почему его называют *квантовой жидкостью* (первая — уже упомянутое выше свойство гелия оставаться жидким при абсолютном нуле). Эти свойства являются проявлением одного — *сверхтекучести* гелия-II, открытой П. Л. Капицей (1894–1984) в 1937 г. Исследования течения жидкого гелия-II через узкие каналы и щели, производившиеся до Капицы, приводили к несогласованным результатам в отношении вязкости He II. Капица показал, что причина несогласованности состояла в том, что каналы и щели были *слишком широки*. Если оценить числа Рейнольдса, пользуясь полученными значениями вязкости He II, то они оказываются *очень большими*. При таких больших числах Рейнольдса течение He II было бы *турбулентным*, тогда как его считали ламинарным. Капица добился ламинарности, исследуя течение гелия-II через очень узкие капилляры (диаметром около 10^{-5} см и меньше) и щели. Оказалось, что при этих условиях жидкий гелий-II протекает через капилляры, *не испытывая сил вязкости*. Это явление исчезновения вязкости и было названо *сверхтекучестью*. С другой стороны, как показали измерения Кеезома и Мак Вуда в 1938 г., вязкость He II,

измеренная по методу затухания крутильных колебаний диска, погруженного в He II, оказалась конечной и вблизи λ-точки мало отличной от вязкости He I.

3. Ниже описываются некоторые явления, наблюдающиеся в He II. Однако для сокращения изложения мы сразу же даем и их толкование, основанное на так называемой *двухжидкостной модели* He II, предложенной в 1938 г. Л. Тиссой (р. 1907), а затем обоснованной и усовершенствованной Л. Д. Ландау (1908–1968) в 1941 г.

В основе этой модели лежит предположение, что He II состоит из двух компонентов: нормального (n) и сверхтекучего (s), каждый из которых характеризуется своей плотностью (так что полная плотность жидкости равна $\rho = \rho_n + \rho_s$) и своими гидродинамическими скоростями $\mathbf{v}_n$ и $\mathbf{v}_s$. Полная плотность потока импульса жидкости равна $\mathbf{j} = \rho_n \mathbf{v}_n + \rho_s \mathbf{v}_s$. Плотность ρ_n нормального компонента стремится к нулю при $T \to 0$, когда гелий становится полностью сверхтекучим, а плотность ρ_s сверхтекучего компонента обращается в нуль при $T > T_\lambda$, когда весь гелий превращается в He I.

Впрочем, двухжидкостная модель является не более как только *способом выражения*, удобным для описания явлений, происходящих в He II. Как и всякое описание квантовых явлений в классических терминах, двухжидкостная модель при буквальном ее понимании *не является адекватной*. Нельзя сказать, что гелий-II состоит из нормальных и сверхтекучих атомов. Разделить его на такие атомы невозможно уже потому, что атомы гелия совершенно тождественны. При перестановке местами «нормального» и «сверхтекучего» атомов, если бы гелий-II состоял из таковых, его состояние осталось бы в точности таким же, каким оно было до перестановки. В § 61 (п. 4) будет приведена другая аргументация, почему о реальном различении He II на нормальный и сверхтекучий компоненты и их фактическом отделении друг от друга не может быть речи. Кроме того, в покоящемся He II различие между нормальным и сверхтекучим компонентами не обнаруживается. Точнее, надо было бы говорить не о существовании двух компонентов в He II, а *об одновременном сосуществовании двух движений — нормального и сверхтекучего*. Плотности ρ_n и ρ_s играют при этом просто роль *коэффициентов*, показывающих вклад каждого из этих движений в тот или иной эффект. При использовании двухжидкостной модели надо всегда иметь в виду сделанную оговорку.

Нормальное движение обладает всеми свойствами вязкого течения жидкости, тогда как сверхтекучее движение приводит к явлению сверхтекучести. Двухжидкостная модель предполагает также, что сверхтекучее движение в гелии-II не только не сопровождается вязкостью, но и *не принимает участия в переносе тепла*.

4. С изложенной точки зрения понятно исчезновение вязкости He II при течении через очень узкие капилляры. Нормальное (вязкое) течение He II через узкий капилляр просто не происходит. Совершается только сверхтекучее течение, лишенное вязкости. Эффект выражен тем лучше, чем уже капилляр, так как в широких капиллярах течение становится турбулентным.

Прямой опыт Э.Л. Андроникашвили (р. 1910) в 1946 г. подтвердил справедливость двухжидкостной модели. Схема этого опыта представлена на рис. 111. Стопка очень тонких дисков из алюминиевой фольги (толщина 0,0015 см), расположенных на одной оси на равных расстояниях друг от друга (0,021 см), подвешивалась на упругой нити и совершала крутильные колебания в He II. При колебаниях дисков увлекался только нормальный компонент He II, тогда как сверхтекучий не оказывал никакого влияния на эти колебания. При $T > T_\lambda$ в колебаниях принимала участие вся жидкость. При переходе через λ-точку период колебаний резко уменьшался, что объясняется появлением в He II сверхтекучего компонента. Уменьшение периода колебаний непрерывно продолжалось и при дальнейшем понижении температуры. Измеряя периоды колебаний, можно было определить относительные содержания ρ_n/ρ и ρ_s/ρ нормального и сверхтекучего компонентов при различных температурах.

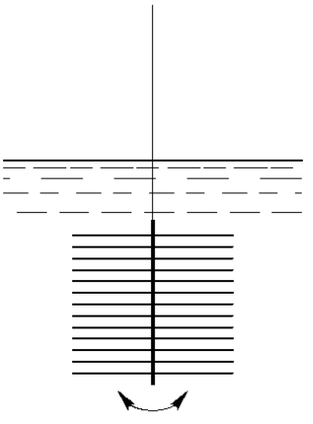


Рис. 111

Результаты графически представлены на рис. 112. Из него видно, что при абсолютном нуле содержится только сверхтекучий компонент. При повышении температуры его содержание уменьшается. Температура, при которой ρ_s/ρ обращается в нуль, и представляет собой точку перехода гелия-II в гелий-I. Таким образом, фазовый переход в жидком гелии связан с исчезновением (или появлением) сверхтекучей части жидкости. Это исчезновение (или появление) происходит постепенно, т. е. в λ-точке ρ_n/ρ принимает предельное значение $\rho_n/\rho = 1$ *непрерывным образом*, без скачка. Поэтому-то переход и является фазовым переходом второго рода.

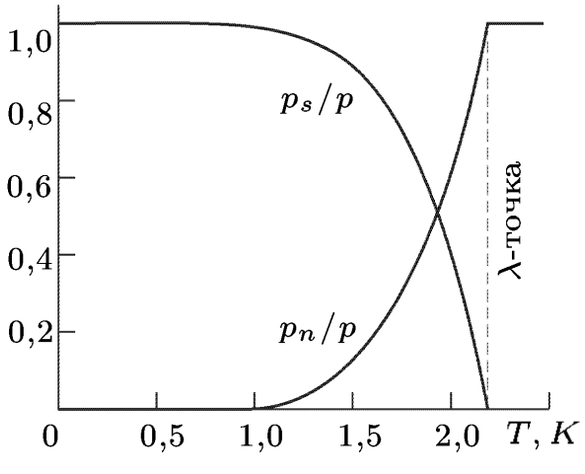


Рис. 112

С изложенной точки зрения понятно, почему при вытекании He II из сосуда через узкий капилляр или щель он не обнаруживает вязкости. Это связано с тем, что из сосуда вытекает сверхтекучая часть жидкости, не обнаруживающая трения, а нормальная часть задерживается в сосуде, протекая через капилляр несравненно медленнее, со скоростью, свойственной ее вязкости и толщине капилляра. Диск же, совершающий крутильные колебания в He II, тормозится трением о нормальную часть жидкости. Благодаря этому и получается конечное значение вязкости. Таким образом, в опытах с протеканием по капилляру обнаруживается наличие сверхтекучей части He II, а в опытах с колебаниями диска — нормальной.

5. Обратимся теперь к рассмотрению некоторых явлений, наблюдающихся в He II.

Пусть два сосуда, наполненные He II, соединены друг с другом узким капилляром. Практически роль капилляра может выполнять трубка, плотно забитая мелким порошком (например, наждачным). Между частицами порошка образуются извилистые каналы различной толщины (~ 100 нм), по которым и может течь сверхтекучая часть гелия. Если первоначально температуры He II в сосудах были одинаковы, а уровни жидкости различны, то начнется перетекание сверхтекучей части гелия из одного сосуда в другой. Оказывается, что при этом температура He II будет повышаться в сосуде, из которого гелий вытекает, и понижаться в сосуде, куда он втекает. Это явление называется *механотермическим* или *механокалорическим эффектом.* Оно объясняется тем, что сверхтекучая часть He II, с помощью которой только и осуществляется перетекание, не переносит никакого тепла. Поэтому внутренняя энергия в сосуде, откуда вытекает He II, не изменяется, а общее количество жидкости в нем уменьшается. На единицу массы приходится все большая и большая внутренняя энергия, что и приводит к повышению температуры. Напротив, из-за втекания сверхтекучей части во второй сосуд удельная внутренняя энергия He II в нем уменьшается, а поэтому температура понижается.

6. Существует обратный эффект, называемый *термомеханическим.* Допустим опять, что два сосуда с He II сообщаются между собой узким капилляром. Пусть теперь в начальный момент температуры и уровни жидкости в обоих сосудах одинаковы. Нагреем один из сосудов. Тогда в нем содержание сверхтекучего компонента уменьшится, а нормального увеличится. Термодинамическое равновесие нарушится. Механическое равновесие будет быстро восстанавливаться путем перетекания сверхтекучего компонента из более холодного в более теплый сосуд. Такой процесс из-за механотермического эффекта будет только увеличивать разность температур в обоих сосудах. Нормальный компонент практически не участвует в перетекании. Выравнивание температур происходит несравненно более медленно, чем установление механического равновесия. В результате, пока температуры не выравняются, уровень гелия в менее нагретом сосуде окажется ниже, чем в более нагретом. Таким образом, при наличии в He II градиента температуры возникает течение сверхтекучей части в направлении против этого градиента. Это и есть термомеханический эффект.

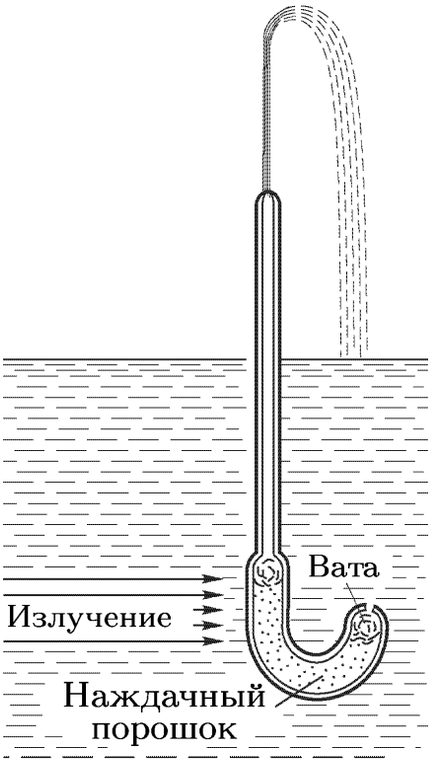


Рис. 113

Яркой демонстрацией термомеханического эффекта может служить так называемое *фонтанирование гелия*, впервые наблюдавшееся в 1938 г. Алленом (р. 1908) и Джонсом. Один конец широкой трубки (рис. 113), плотно заполненной наждачным порошком,

погружался в гелиевую ванну, а к другому был приделан вертикальный капилляр. При освещении наждачного порошка (достаточно было света от карманного фонарика) он нагревался, и сверхтекучий поток устремлялся из ванны в трубку с такой скоростью, что из капилляра била струя He II, достигавшая высоты 30–40 см.

7. На поверхности тел, погруженных в жидкий гелий, образуются тонкие пленки, толщина которых при давлении насыщенных паров составляет около 30 нм, или 100 атомных слоев. Образование аналогичных пленок происходит и в случае других жидкостей (например, керосина). Но из-за отсутствия вязкости He II его сверхтекучая часть течет внутри пленки, как по узкому каналу, — явление, наблюдавшееся Камерлинг-Оннесом еще в 1922 г. Благодаря этому пустая пробирка, опущенная в ванну с He II, начинает наполняться жидкостью, даже если ее верхний конец расположен выше уровня He II (рис. 114 *а*). Наполнение продолжается до тех пор, пока уровни жидкости в пробирке и в ванне не сравняются. Если после этого пробирку приподнять, то она опустошается (рис. 114 *б*). Опустошение продолжается и в том случае, когда пробирка приподнята над уровнем жидкости в ванне целиком. В последнем случае образуются капли, падающие в ванну с наружной

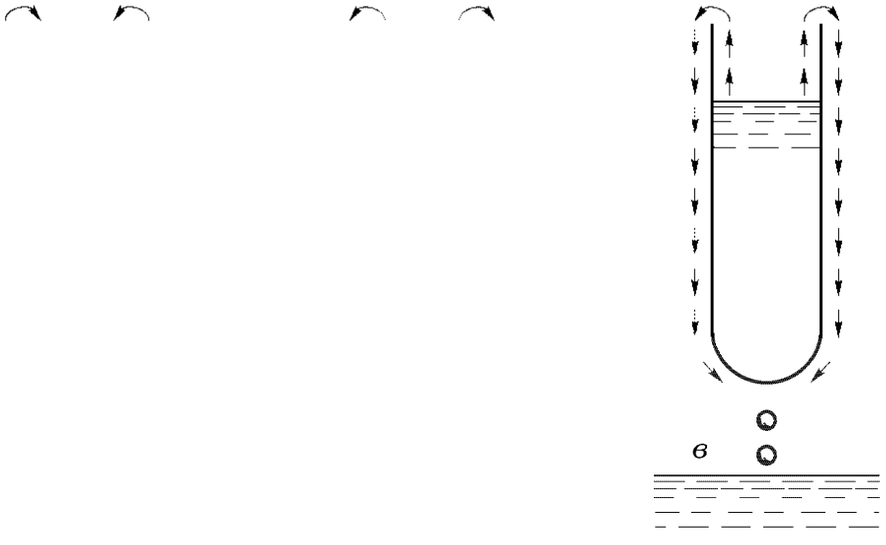


Рис. 114

поверхности дна пробирки (рис. 114 *в*). В этих опытах пленка He II работает как сифон, причем движущую силу для сверхтекучей части создает главным образом разность гравитационных потенциалов между концами пленки. Типичное значение для скорости сверхтекучей части He II составляет около 20 см/с. Нормальная часть жидкости, из-за своей вязкости, остается в пленке практически неподвижной.

8. Уже в ранних исследованиях (1935 г.) В. Кеезом и его дочь А. Кеезом обнаружили, что переход He I в He II сопровождается *резким возрастанием теплопроводности* (примерно в 10^6 раз). Теплопроводность He II в сотни раз больше, чем у лучших металлических проводников тепла. Кроме того, оказалось, что поток тепла в He II не пропорционален градиенту температуры, так что невозможно говорить о «коэффициенте теплопроводности» He II в обычном смысле слова.

Как доказал П.Л. Капица, аномалии в передаче тепла в He II являются следствием его сверхтекучести. Теплопередача в He II носит *конвективный* характер. Действительно, допустим, что He II как целое покоится и что между двумя точками его возникла разность температур. Тогда из-за термомеханического эффекта сверхтекучая часть He II начнет двигаться в сторону более высокой температуры. Возникнет и обратное движение нормальной части He II, так как полный импульс жидкости должен оставаться равным нулю. Сверхтекучая и нормальная части движутся «одна сквозь другую», не испытывая трения. Сверхтекучая часть не переносит тепла. Все тепло связано с движением нормальной части. Таким движением и объясняется аномально высокая теплопередача в He II.

Из-за аномально высокой теплопередачи фактически невозможно создать в толще He II температурный градиент. С этим связано *внезапное прекращение кипения* при охлаждении жидкого гелия. Выше температуры λ-точки He I при откачке ведет себя как всякая обыкновенная жидкость и бурно кипит. При переходе же через λ-точку кипение внезапно прекращается. В обыкновенной жидкости пузырьки, необходимые для парообразования внутри жидкости, образуются, когда локальная температура внутри жидкости существенно выше температуры свободной поверхности. Если бы в He II возникла подобная разность температур, то она исчезла бы столь быстро, что пузырьки не могли бы образоваться. Таким образом, в He II парообразование происходит только на *свободной поверхности*, а не внутри жидкости, с чем и связана невозможность кипения.

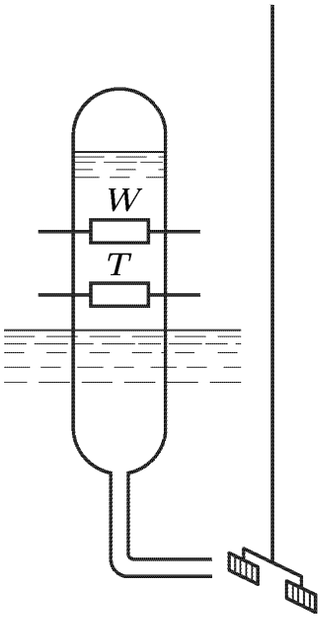


Рис. 115

9. Наличие сосуществующих сверхтекучего и нормального потоков в He II экспериментально наблюдал П.Л. Капица. В одном из его опытов (1941 г.), схематически показанном на рис. 115, небольшая стеклянная бульбочка (выполненная в виде миниатюрного сосуда Дьюара) оканчивалась в нижней части изогнутой стеклянной трубочкой (длиной 6 мм и внутренним диаметром 0,5 мм). Внутри бульбочки помещались константановый нагреватель W и бронзовый термометр T. Бульбочка погружалась в сосуд Дьюара с He II. Перед концом трубочки на коромысле, подвешенном на тонкой кварцевой нити, помещалось легкое крылышко, уравновешенное зеркальцем. Поворот коромысла можно было наблюдать по отклонению светового луча, отраженного от зеркальца. При включении нагревателя коромысло поворачивалось, а крылышко удалялось от конца трубочки. При этом, однако, уровень жидкости в бульбочке не изменялся. Явление объясняется тем, что при включении нагревателя сверхтекучая часть He II втекала внутрь трубочки, двигаясь в сторону повышения температуры. В противоположном направлении из трубочки устремлялся поток нормальной части жидкости. Происходило течение сверхтекучей и нормальной

частей друг через друга, так что жидкость в целом представлялась неподвижной. Нормальный поток, обладая вязкостью, оказывал давление на крылышко и вызывал его отклонение. Сверхтекучий же поток в силу парадокса Даламбера (см. т. I, § 100) на крылышко не действовал. Таким образом, наблюдалось своеобразное явление: на покоящееся тело, погруженное в He II, действовала сила, в то время как сам He II представлялся *неподвижным.*

10. Тисса на основе двухжидкостной модели предсказал существование в He II так называемого *второго звука.* Его теория была подробно разработана Ландау, а экспериментально явление впервые наблюдалось в 1946 г. В.П. Пешковым (1913–1980). В обычной жидкости звук представляет собой волны распространяющихся колебаний плотности. Такие волны могут распространяться и в He II. Они называются *первым звуком.* Скорость первого звука в He II при $T = 0$ составляет около 239 м/с, а с повышением температуры несколько уменьшается. В волне первого звука сверхтекучая и нормальная части в каждый момент времени смещаются в одну сторону, т. е. движутся в одинаковых фазах. Однако если в He II создать разность температур, то она начнет выравниваться не обычным механизмом теплопроводности, а быстрой конвекцией, при которой сверхтекучая и нормальная части движутся противоположно. Если в He II возбудить колебание температуры, то из-за движения сверхтекучей и нормальной частей в противоположных направлениях полная плотность He II останется неизменной, а полный поток импульса $\mathbf{j} = \rho_s \mathbf{v}_s + \rho_n \mathbf{v}_n$ будет равен нулю. Сверхтекучая и нормальная части колеблются *в противоположных фазах.* Однако нормальная часть несет тепло, а сверхтекучая не несет. Поэтому в рассматриваемом случае получается *температурная волна*, в которой меняется не плотность, а температура. Это и есть второй звук.

Скорость второго звука в He II сильно зависит от температуры. При $T = 0$ она составляет около 139 м/с, а при $T = T_\lambda$ обращается в нуль. По теории Ландау при $T = 0$ скорость второго звука, в согласии с опытом, в $\sqrt{3}$ раз меньше скорости первого звука.

ЗАДАЧА

Показать, что тепло, потребное для нагревания гелия от t до $t + dt$, стремится к нулю при $dt \to 0$ даже тогда, когда в рассматриваемом температурном интервале находится λ-точка, в которой теплоемкость гелия обращается в бесконечность.

§ 61. Понятие о теории сверхтекучести

1. Двухжидкостная модель He II, конечно, только *описывает*, но отнюдь *не объясняет* явление сверхтекучести. Природа этого явления в основном была понята, а двухжидкостная модель обоснована Л.Д. Ландау в 1941 г. Ниже дается качественное представление о теории Ландау. Прежде всего в этой теории принимается, что жидкий гелий-II при температуре абсолютного нуля находится *в основном*, т. е. наинизшем (невозбужденном), квантовом состоянии. При движении

тела в He II оно совсем не испытывает трения, если при этом жидкость остается в основном состоянии. Для возникновения трения необходимо, чтобы He II при движении тела переходил *в возбужденные* состояния. А так как между атомами жидкого гелия существует сильное взаимодействие, то возбужденные состояния должны быть *коллективными* (т. е. относиться ко всему гелию), а не возбуждениями изолированных атомов. Такие возбуждения носят квантовый характер. Кванты возбуждения в He II называются *элементарными возбуждениями* и характеризуются определенными значениями импульса и энергии, т. е. ведут себя как квазичастицы. Коллективный характер элементарных возмущений — существенный пункт в теории Ландау.

Теория прежде всего должна определить *закон дисперсии* или энергетический спектр элементарных возбуждений, т. е. зависимость энергии $\mathscr{E}$ от импульса $\mathbf{p}$ или, точнее, от модуля p импульса, поскольку жидкий гелий — изотропная среда. Эта часть задачи была решена Ландау *постулативно* с использованием эмпирических данных. Окончательная кривая $\mathscr{E} = \mathscr{E}(p)$, принятая Ландау (1947 г.), изображена на рис. 116. На начальном участке кривой $\mathscr{E}$ растет линейно с импульсом: $\mathscr{E} = c_{\text{зв}} p$, совершенно так же, как в случае возбуждения звуковых волн в кристаллической решетке. На этом участке в He II возбуждаются *фононы*, а постоянная $c_{\text{зв}}$ является фазовой скоростью звука в He II. Такой фононный спектр элементарных возбуждений предопределяет температурную зависимость теплоемкости He II вблизи абсолютного нуля: как и в случае твердых тел, теплоемкость пропорциональна кубу температуры T.

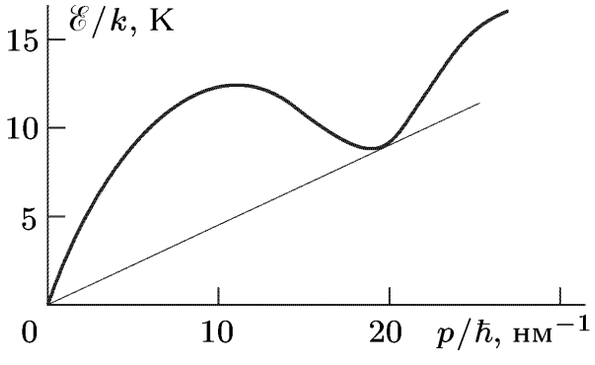


Рис. 116

При возрастании p кривая $\mathscr{E} = \mathscr{E}(p)$ загибается вниз, а в точке p_0 проходит через минимум. В окрестности самого минимума энергия может быть аппроксимирована формулой

$$\mathscr{E} = \Delta + \frac{(p - p_0)^2}{2m^*}, \tag{61.1}$$

причем постоянная m^* играет роль *эффективной массы*. Элементарные возбуждения, соответствующие этой части кривой, были названы *ротонами*.

Следует подчеркнуть, что в законе дисперсии $\mathscr{E} = \mathscr{E}(p)$ всюду имеется в виду *истинный импульс*, а не квазиимпульс, как это было в случае кристаллической решетки. Причина этого в том, что жидкий гелий — *изотропная среда*, не обладающая кристаллической структурой.

Позднее кривая $\mathscr{E} = \mathscr{E}(p)$, постулированная Ландау, была подтверждена экспериментально. Исследовалось рассеяние пучка монохроматических нейтронов в He II. Этот процесс можно рассматривать

как столкновение нейтронов с квазичастицами в He II, при котором соблюдаются законы сохранения энергии и импульса. Зная начальную энергию и направление падающих нейтронов и измеряя энергию рассеянных нейтронов при различных углах рассеяния (относительно исходного направления нейтронов), можно было рассчитать энергию и импульс квазичастиц, т. е. определить весь энергетический спектр элементарных возбуждений. Экспериментальная кривая $\mathscr{E} = \mathscr{E}(p)$, полученная таким образом, воспроизведена на рис. 117. Числовые значения постоянных (экстраполированные к нулевому давлению при плотности $\rho = 0{,}145$ г/см^3), полученные таким путем, а также из термодинамических данных, оказались равными

$$\begin{aligned} c_{\text{зв}} &= 2{,}4 \cdot 10^4 \text{ см/с}, \quad \Delta/k = 8{,}7 \text{ K}, \\ p_0/\hbar &= 1{,}9 \cdot 10^8 \text{ см}^{-1}, \quad m^* = 0{,}16 m_{^4\text{He}}. \end{aligned} \tag{61.2}$$

2. Несмотря на экспериментальное подтверждение, постулативный характер энергетического спектра в теории Ландау оставляет чувство неудовлетворенности. Однако установление такого спектра расчетным

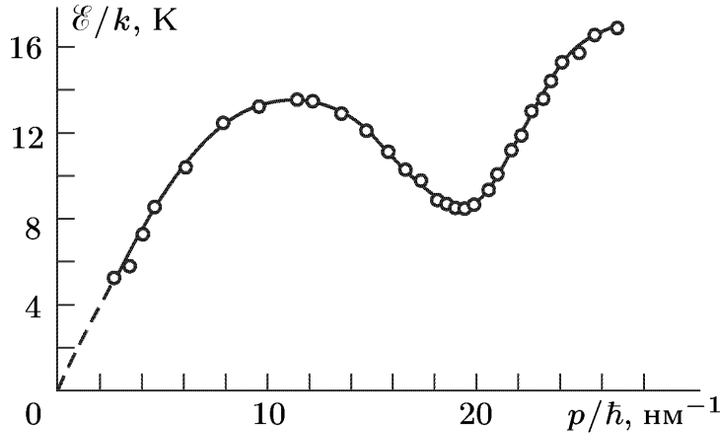


Рис. 117

путем совершенно безнадежно для жидкости с сильным взаимодействием между ее частицами, какой является He II. Позднее (1946 г.) Н. Н. Боголюбов (р. 1908) решил такую задачу для газа *слабо взаимодействующих* бозе-частиц (оказалось необходимым предположить, что силы отталкивания между такими частицами преобладают над силами притяжения). Спектр элементарных возбуждений при малых p оказался фононного типа, как и постулировал Ландау.

Нет никакой возможности излагать здесь теорию Боголюбова. Можно только отметить, что (как и во всех теориях сверхтекучести) необходимо принять, что газ состоит *из бозе-, а не из ферми-частиц*. В случае бозе-частиц имеет место конденсация Бозе–Эйнштейна (что в случае ферми-частиц невозможно из-за принципа Паули). Сущность этого явления (см. т. II, § 82) состоит в том, что, начиная с некоторой температуры T_0, называемой *температурой вырождения*, частицы бозе-газа начинают накапливаться на нижнем энергетическом уровне, а при температуре абсолютного нуля на этом уровне оказываются все частицы. Бозе-частицы, скопившиеся на нижнем энергетическом уровне,

образуют так называемый *конденсат*. Боголюбову удалось показать, что при определенных допущениях о силах взаимодействия между частицами образование конденсата в слабо неидеальном газе бозе-частиц можно рассматривать как фазовый переход второго рода, а движение конденсата обладает свойством сверхтекучести. Аналогия этого явления с наблюдаемой сверхтекучестью жидкого He II проявляется в том, что вычисленная температура вырождения бозе-газа оказалась равной $T_0 = 3{,}14$ K, что весьма близко к значению $T_\lambda = 2{,}17$ K — температуре фазового превращения He I в He II. Однако результаты Боголюбова относятся к *газу* со слабо взаимодействующими частицами, *а не к жидкому* He II, где взаимодействие велико. Заманчивой является перспектива построить на тех же основах теорию сверхтекучести и для жидкости, но до настоящего времени это сделать не удалось.

3. Покажем теперь, что при энергетическом спектре элементарных возбуждений, постулированном Ландау, жидкость должна обладать свойством сверхтекучести. Рассмотрим сначала He II при температуре абсолютного нуля, когда никаких возбуждений в нем нет. Пусть жидкость течет в капилляре с постоянной скоростью $\mathbf{v}$. Определим, при каких скоростях $\mathbf{v}$ не могут появляться элементарные возбуждения и, следовательно, гелий не будет испытывать трения, а при каких скоростях элементарные возбуждения появятся и возникнет трение жидкости о стенки капилляра. Для наших целей удобно перейти в систему отсчета, в которой жидкость покоится, а капилляр движется со скоростью $-\mathbf{v}$. Целесообразность этого обусловлена тем, что закон дисперсии (рис. 117) относится именно к системе отсчета, в которой *жидкость неподвижна.*

Допустим теперь, что в жидкости появилось элементарное возбуждение — квазичастица с энергией $\mathscr{E}$ и импульсом $\mathbf{p}$. Из-за отдачи скорость капилляра изменится и станет равной $-\mathbf{v}_1$. На основании закона сохранения энергии можно написать

$$\mathscr{E} + \frac{M}{2}\,\mathbf{v}_1^2 - \frac{M}{2}\,\mathbf{v}^2 = 0,$$

где M — масса капилляра. Так как она велика, то изменение скорости капилляра незначительно, в силу чего сумму $\mathbf{v} + \mathbf{v}_1$ можно заменить на $2\mathbf{v}$. Это дает

$$\mathscr{E} = \frac{M}{2}\,(\mathbf{v} - \mathbf{v}_1)(\mathbf{v} + \mathbf{v}_1) = M\mathbf{v}(\mathbf{v} - \mathbf{v}_1).$$

В силу закона сохранения импульса

$$\mathbf{p} + M\mathbf{v}_1 = M\mathbf{v},$$

так что окончательно

$$\mathscr{E} = \mathbf{pv}. \tag{61.3}$$

Отсюда во всяком случае следует, что $\mathscr{E} \leqslant pv$, т. е. $v \geqslant \mathscr{E}/p$. Таким образом, для появления элементарного возбуждения необходимо, чтобы

скорость жидкости v превышала минимальное значение $\mathscr{E}/p$:

$$v \geqslant \min \frac{\mathscr{E}}{p}. \tag{61.4}$$

Минимум $\mathscr{E}/p$ на кривой дисперсии получится в точке, где производная $\dfrac{d}{dp}\dfrac{\mathscr{E}}{p}$ обращается в нуль, т. е. в точке

$$\frac{d\mathscr{E}}{dp} = \frac{\mathscr{E}}{p}, \tag{61.5}$$

где прямая $\mathscr{E}/p = \text{const}$, проходящая через начало координат, касается дисперсионной кривой $\mathscr{E} = \mathscr{E}(p)$. На фононной ветви $\mathscr{E} = c_{\text{зв}}p$, так что для возбуждения фононов требуются скорости $v \geqslant c_{\text{зв}}$.

Этот результат не является неожиданным. При равномерном движении тела в жидкости, как известно, оно начинает генерировать *звуковые волны Маха* (1838–1916), когда скорость тела превышает фазовую скорость звука в жидкости. Аналогично ведет себя заряд, движущийся в среде со скоростью, превышающей фазовую скорость света в рассматриваемой фазе (*явление Вавилова–Черенкова*).Таким образом, скорость звука играет роль *критической скорости* $v_{\text{кр}}$, в том смысле, что генерация звука движущимся телом появляется, когда его скорость превышает критическую.

Критическая скорость для генерации ротонов в теории Ландау найдется, если из начала координат (рис. 116) провести касательную к ротонной части кривой $\mathscr{E} = \mathscr{E}(p)$. Она, очевидно, равна $v_{\text{кр}} = \mathscr{E}/p$, где $\mathscr{E}$ и p — энергия и импульс ротона в точке касания. Таким путем для ротонов получается $v_{\text{кр}} = 60$ м/с.

4. Допустим теперь, что температура жидкости, текущей через капилляр, отлична от абсолютного нуля, но близка к нему. Тогда даже в покоящейся жидкости появятся слабые возбуждения. Они ведут себя как *идеальный газ независимых квазичастиц*. Формула (61.4), полученная выше, остается в силе, поскольку она выражает условие рождения элементарных возбуждений из-за движения жидкости относительно стенок капилляра. При ее выводе предположение о равенстве абсолютному нулю температуры жидкости не использовалось.

В покоящейся жидкости квазичастицы, имеющиеся в ней при $T \neq 0$, движутся беспорядочно, так что в среднем газ квазичастиц не несет никакого импульса. Если же жидкость течет, то возникает преимущественное направление движения квазичастиц — в сторону течения жидкости. Средний импульс газа квазичастиц становится отличным от нуля и переносится в том же направлении. При столкновениях квазичастиц со стенками капилляра происходит передача этого импульса, по крайней мере частичная, из-за чего возникает трение. В результате часть жидкости, несущая газ элементарных возбуждений, течет через капилляр, как обычная нормальная жидкость, обладающая вязкостью. Остальная часть жидкости ведет себя как сверхтекучая, так как при скорости течения меньше критической рождения новых квазичастиц в результате движения жидкости относительно стенок капилляра не

происходит. Эта часть жидкости при течении не обменивается энергией и импульсом со стенками капилляра и с газом квазичастиц. Таким образом, в теории Ландау находят обоснование представления, введенные в двухжидкостной модели. Вместе с тем становится понятной невозможность реального отделения сверхтекучей части жидкости от нормальной. Причина этого в том, что элементарные возбуждения являются *коллективными*, а не возбуждениями индивидуальных частиц.

5. Критическая скорость течения ($v_{\text{кр}} \approx 60$ м/с), вычисленная в теории Ландау из кривой дисперсии, не согласуется с опытом. Опыт показывает, что критическая скорость сильно зависит от температуры и от радиуса капилляра или ширины щели, по которым течет He II. Она примерно на два-три порядка меньше того значения, которое дает теория Ландау. На этом основании следует заключить, что теория Ландау недостаточна. В ней не все принято во внимание. В действительности формула Ландау (61.4) определяет лишь *верхний предел* скорости течения, выше которого сверхтекучесть заведомо невозможна.

Прогресс в понимании указанного расхождения был достигнут при исследовании свойств вращающегося He II. Согласно двухжидкостной модели при вращении цилиндра с He II сверхтекучая часть жидкости не должна была бы вовлекаться во вращение. В действительности опыты показали, что при достаточно больших угловых скоростях цилиндра во вращении принимает участие не только нормальная, но и *сверхтекучая часть* жидкости. Тем не менее во вращающемся как целое He II наблюдаются такие специфические для сверхтекучести явления, как термомеханический эффект и распространение второго звука.

Онзагер (р. 1903) в 1949 г. для объяснения этого явления выдвинул гипотезу, позднее подтвержденную экспериментально, что ввиду *квантовой природы* сверхтекучего состояния He II вращения в ней должны возбуждаться, но они должны быть квантованы. Позднее (1955 г.) эта гипотеза была развита Фейнманом (р. 1918). Если атом гелия вращается вокруг оси цилиндра по окружности радиуса r, то его момент количества движения должен быть равен $n\hbar$, т. е.

$$m_{\text{He}} r v_s = n\hbar,$$

где v_s — скорость сверхтекучего движения, а n — целое число. Эта формула справедлива лишь для расстояний r, значительно превышающих среднее межатомное расстояние в ^{4}He, т. е. примерно $4 \cdot 10^{-8}$ см, так как только на таких расстояниях гелий можно рассматривать как сплошную жидкость. Радиус окружности r может принимать очень большие значения, вплоть до размеров цилиндра, так что речь идет о *квантовании макроскопических вращений.*

Линия, вокруг которой происходит вращение He II, называется *вихревой нитью.* При вращении вокруг оси цилиндра она, очевидно, совпадает с этой осью, заканчиваясь на дне цилиндра и на поверхности жидкости. Жидкость, вращающаяся вокруг вихревой нити, называется *квантовым вихрем.* Квантовый вихрь может появиться лишь при некоторой минимальной скорости вращения цилиндра $\Omega_{\text{кр}}$, так как для его образования требуется затрата энергии. При этом фактически

образуются только вихри с $n = 1$, так как кинетическая энергия пропорциональна $v_s^2 \sim n^2$, и поэтому энергетически более выгодно образование двух квантовых вихрей с $n = 1$, чем одного с $n = 2$. Таким образом, скорость движения сверхтекучей жидкости v_s при вращении вокруг вихревой нити определяется формулой

$$v_s = \frac{\hbar}{m_{\mathrm{He}} r}. \tag{61.6}$$

Когда скорость вращения цилиндра достигает критического значения $\Omega_{\mathrm{кр}}$, то образуется только один квантовый вихрь с вихревой нитью вдоль оси цилиндра. При увеличении Ω число квантовых вихрей растет, причем их вихревые нити параллельны оси цилиндра. При больших Ω вихревые нити практически равномерно пронизывают объем жидкости.

Элементарные возбуждения, соответствующие нормальной части He II, могут сталкиваться с квантовыми вихрями сверхтекучей части и обмениваться с ними энергией и импульсом. Это и вызывает возникновение вращения сверхтекучей части при вращении цилиндра.

6. При течении He II по капилляру с большими скоростями вихревые нити могут искривляться и даже замыкаться, образуя так называемые *вихревые кольца*. Последние обладают определенной энергией и импульсом и могут передвигаться по всему объему жидкого гелия. Можно сказать, что вихревые кольца представляют собой особого рода *элементарные возбуждения*, возникающие, когда скорость течения сверхтекучего гелия превосходит некоторое критическое значение $v_{\mathrm{кр}}$. Критическая скорость $v_{\mathrm{кр}}$ определяется энергией, необходимой для образования квантового вихревого кольца, и зависит от условий течения He II, в частности от радиуса капилляра, при увеличении которого критическая скорость быстро уменьшается. При радиусах капилляра порядка 10^{-5} см, когда и наблюдается сверхтекучесть, критическая скорость значительно меньше теоретического верхнего предела, даваемого формулой Ландау (61.4). Если скорость течения v меньше критической, то образование квантовых вихревых колец еще не происходит и наблюдается сверхтекучесть. Если же скорость течения превосходит критическую (хотя бы и остается меньше верхнего предела, даваемого теорией Ландау), то сверхтекучее течение He II становится неустойчивым к образованию квантовых вихревых колец. Квантовые вихревые кольца, возникающие в сверхтекучем компоненте He II, могут обмениваться энергией и импульсом с квазичастицами нормального компонента, что приводит к появлению трения между этими компонентами и к исчезновению сверхтекучести.

7. Рассмотрим коротко вопрос о сверхтекучести $^3_2$He. Атом $^3_2$He содержит нечетное число нуклонов (два протона и один нейтрон) и четное число (2) электронов. Каждая из этих частиц обладает спином 1/2. Поэтому спин атома $^3_2$He во всяком случае полуцелый, так что атомы $^3_2$He подчиняются статистике Ферми–Дирака. В этом отношении они отличаются от атомов $^{04}_{2}$He, для которых спин целый, а статистика Бозе–Эйнштейна. Образование конденсата и появление сверхтекучести

возможны только в системах частиц с бозе-эйнштейновской статистикой. Поэтому сначала казалось, что сверхтекучесть $^3_2$He невозможна. Это мнение изменилось в 1974 г. после создания теории сверхпроводимости (см. § 62). Дело в том, что между атомами $^3_2$He действуют силы притяжения Ван-дер-Ваальса, которые на больших расстояниях убывают обратно пропорционально седьмой степени расстояния. При наличии таких сил атомам $^3_2$He энергетически выгодно кооперироваться в пары, что и проявляется при достаточно низких температурах. Спин такой пары всегда целый. Поэтому пары подчиняются статистике Бозе–Эйнштейна, и можно ожидать появления сверхтекучести у $^3_2$He. Действительно, сверхтекучесть $^3_2$He была открыта в 1974 г. Она обнаруживается при температуре ниже 0,00265 K (при давлении около 34 атм). При уменьшении давления до 21 атм температура перехода жидкого $^3_2$He в сверхтекучее состояние понижается до 0,0024 K.

8. В заключение остановимся на возможности макроскопических квантовых явлений у водорода, аналогичных сверхтекучести. Квантовое поведение гелия, в частности свойство оставаться жидким вблизи абсолютного нуля, существенно обусловлено *нулевыми колебаниями.* Массы атомов и молекул водорода еще меньше, чем масса атома гелия, а потому у водорода нулевые колебания еще сильнее. С другой стороны, атом водорода состоит из двух частиц (протона и электрона), спин каждой из которых равен $1/2$, а потому он является *бозе-частицей.* Молекула водорода также *бозе-частица.* Казалось бы, что и водород вблизи абсолютного нуля температуры должен обладать сверхтекучестью. В действительности газ двухатомного водорода, хотя и обладает некоторыми квантовыми свойствами, но при охлаждении сжижается, а вблизи абсолютного нуля становится твердым.

Причина этого в том, что частицы водорода *взаимодействуют между собой.* В частности, гомеополярные химические силы, действующие между атомами, приводят к образованию связанного состояния — молекулы водорода H_2. А между молекулами водорода в свою очередь действуют силы, значительно превосходящие силы взаимодействия между симметричными инертными атомами гелия.

Однако следует принять во внимание, что силы взаимодействия нейтральных атомов водорода между собой на некотором расстоянии, превосходящем какое-то определенное значение, являются силами притяжения только в том случае, когда спины электронов атомов *антипараллельны* (синглетное состояние). Если же спины электронов атомов водорода *параллельны* (триплетное состояние), то силы взаимодействия всегда отталкивательные. В этом случае образование молекулы невозможно (см. § 50).

Однако можно получить атомарный водород, все атомы которого имеют параллельные спины. Тогда молекулы образовываться не будут. Для этого на диссоциированный водород надо воздействовать сильным ($\sim 10^5$ Гс) и резко неоднородным магнитным полем. Тогда произойдет пространственное разделение атомов с противоположно направленными спинами, подобно тому как это имеет место в классических опытах Штерна и Герлаха. При этом надо иметь в виду, что магнитный момент

ядра (протона) примерно в тысячу раз меньше магнитного момента электрона. Поэтому силой, действующей со стороны магнитного поля на ядро, можно полностью пренебречь по сравнению с силой, действующей на электрон.

Такой метод получения атомарного водорода с одинаково ориентированными спинами электронов был применен голландскими физиками Сильвера и Валравеном в Амстердаме (1980 г.). Экспериментальные трудности, которые удалось им преодолеть, очень значительны, но о них мы говорить не будем. Теоретически можно ожидать, что такой атомарный водород будет оставаться газообразным вблизи абсолютного нуля, не переходя даже в жидкое состояние. Указанным физикам удалось сохранить газ в атомарном состоянии в течение нескольких часов при предельно низкой температуре в сильном магнитном поле. Атомарный водород оставался газообразным вплоть до 0,08 К — низшей температуры, при которой производились исследования.

§ 62. Понятие о теории сверхпроводимости

1. Прежде чем читать настоящий параграф, необходимо обратиться к § 80 т. III, где были приведены основные опытные факты о сверхпроводимости и истории открытия этого явления. Здесь будет дано краткое представление о теории сверхпроводимости, в той мере, как это возможно сделать в рамках общего курса физики.

После создания теории сверхтекучести естественно было думать, что сверхпроводимость является сверхтекучестью «электронной жидкости», образованной валентными электронами в металле. Эта правильная мысль встретила, однако, серьезное затруднение. Электроны имеют спин 1/2. На их состояния наложен запрет Паули, а потому они подчиняются статистике Ферми–Дирака, т. е. являются фермионами. Напротив, на атомы $^4_2$He, с которыми имеет дело теория сверхтекучести, запрет Паули не распространяется, поскольку спин этих атомов целочисленный. В отличие от электронов, атомы $^4_2$He подчиняются статистике Бозе–Эйнштейна, т. е. являются бозонами. С этим связано глубокое различие между основными состояниями атомов $^4_2$He и электронов. В частности, для атомов $^4_2$He имеет место бозе-эйнштейновская конденсация, а для электронов, казалось бы, это явление невозможно.

2. Важным моментом на пути преодоления указанной трудности оказалось открытие *изотопического эффекта*, экспериментально обнаруженного в 1950 г. на изотопах ртути и олова. Этот эффект наблюдается у чистых сверхпроводников, имеющих несколько изотопов. Оказалось, что для кристаллических решеток, построенных из различных изотопов одного и того же элемента, выполняется соотношение $T_к\sqrt{M} = C = \text{const}$. Здесь $T_к$ — критическая температура перехода в сверхпроводящее состояние, а M — массовое число атомов, из которых построена кристаллическая решетка. Постоянная C одинакова для всех изотопов элемента, меняясь при переходе от одного элемента к другому. Но масса M определяет частоту колебаний решетки,

$\omega \sim \sqrt{1/M}$. Поэтому изотопический эффект заставляет ожидать, что сверхпроводимость обусловлена взаимодействием электронов с колебаниями кристаллической решетки.

На это указывает также то обстоятельство, что сверхпроводимость не обнаружена у чистых металлов с хорошей проводимостью (сюда относятся, например, элементы первой группы периодической системы Li, Na, K, Cu, Ag, Au). Дело в том, что у хороших проводников велика длина свободного пробега электронов, обусловленная их взаимодействием с кристаллической решеткой. Это значит, что взаимодействие электронов с колебаниями решетки слабое. У плохих проводников, наоборот, длина свободного пробега электронов мала, т. е. взаимодействие электронов с колебаниями решетки сильное, а это, согласно ожиданию, как раз и требуется для возникновения сверхпроводимости.

Открытие изотопического эффекта позволило выбрать правильное направление в построении микроскопической теории сверхпроводимости. Начало было положено в 1950 г. Фрёлихом (р. 1905), теоретически предсказавшим изотопический эффект практически одновременно с его экспериментальным обнаружением. После основополагающей работы Купера (р. 1930) последовательная микроскопическая теория сверхпроводимости была создана в 1957 г. Бардином (р. 1908), Купером и Шриффером (р. 1931). По имени авторов она сокращенно называется теорией БКШ. Более совершенный метод в теории сверхпроводимости был разработан Н. Н. Боголюбовым примерно в то же время. Предсказание Купера о спаривании электронов (см. ниже п. 4) позволило Боголюбову распространить и развить метод, предложенный им ранее в теории сверхтекучести. Другой метод в теории сверхпроводимости был разработан Л. П. Горьковым (р. 1929) и Намбу (р. 1921).

3. Колебания кристаллической решетки, как известно, квантованы и могут рассматриваться как квазичастицы — фононы. Электроны проводимости в металле могут непрерывно обмениваться фононами. Один электрон излучает, а другой поглощает фонон. Эти фононы называются *виртуальными*, так как они существуют в течение короткого времени жизни, а поэтому их энергия не фиксирована, а удовлетворяет принципу неопределенности время — энергия. Это позволяет не принимать во внимание закон сохранения энергии во время процесса взаимодействия виртуального фонона с электроном. Электрон, излучивший виртуальный фонон, испытывает отдачу, т. е. меняет свой импульс. Импульс другого электрона, поглотившего тот же фонон, также изменяется. Явление происходит так, как если бы между электронами, помимо кулоновского отталкивания, существовало дополнительное взаимодействие. Это дополнительное взаимодействие может проявиться в притяжении или отталкивании электронов.

Следующая классическая аналогия может служить иллюстрацией этого эффекта. Два конькобежца на льду непрерывно перекидывают друг другу мяч. Из-за отдачи между ними возникает отталкивание. Но оно может перейти в притяжение, если мяч заменить бумерангом. Для этого конькобежцы должны встать спиной друг к другу и каждый

из них должен бросать бумеранг в сторону, противоположную своему партнеру. Поведение бумеранга, конечно, определяется воздушной средой, в которой он летит. В случае фонона роль такой среды играет кристаллическая решетка.

4. В теории металлов, развитой Блохом (см. § 59), состояние электронов характеризуется *одночастичными* волновыми функциями. Это означает, что электроны рассматриваются независимыми. Явление сверхпроводимости не укладывается в эту схему. Благодаря электрон-фононному взаимодействию между электронами существуют корреляции, которые должны быть учтены. В теории БКШ учитываются только *парные корреляции*, что оказалось достаточным по крайней мере для качественного понимания основных свойств сверхпроводников. Силы электрон-фононного притяжения между парой скоррелированных электронов должны превышать силы кулоновского отталкивания, чтобы в целом между электронами действовало притяжение. Принято говорить, что такие спаренные электроны образуют *куперовскую пару* (в честь Купера, предсказавшего явление спаривания в 1956 г.). Куперовские пары являются бозонами — на них запрет Паули не распространяется. Поэтому-то они и могут совершать сверхтекучее движение.

Для выяснения свойств куперовских пар допустим сначала, что металл находится при температуре абсолютного нуля. Кроме того, допустим, что через металл не течет электрический ток. Так как электроны подчиняются принципу Паули, то в каждом квантовом состоянии может находиться не более одного электрона. Поэтому будут заполнены все состояния ниже некоторой определенной энергии $\mathscr{E}_F$, называемой *энергией Ферми*, а все состояния с большей энергией будут полностью свободны. Энергия $\mathscr{E}$ есть функция импульса: $\mathscr{E} = \mathscr{E}(\mathbf{p})$. В частности, уравнение $\mathscr{E}_F = \mathscr{E}(\mathbf{p})$ определяет в пространстве импульсов некоторую поверхность, называемую *поверхностью Ферми*. Для простоты будем считать металл изотропным. Тогда поверхность Ферми вырождается в сферу $\mathscr{E}_F = p^2/2m$, радиус которой мы обозначим через p_F. Когда температура металла повышается, но остается ниже $T_к$, электроны начинают выходить за пределы сферы Ферми, заполняя частично энергетически более высокие квантовые состояния. В результате по обе стороны сферы Ферми образуется тонкий слой, ограниченный сферами с радиусами $p_F - \Delta p/2$ и $p_F + \Delta p/2$, заполненный электронами не полностью, а только частично. Энергии электронов на границах этого слоя вблизи критической температуры $T_к$ отличаются друг от друга на величину порядка $kT_к$. Легко вычислить, что эта величина примерно в 10^4–10^5 раз меньше средней кинетической энергии, приходящейся на один валентный электрон в металле при фермиевском распределении (см. т. III, § 99).

Допустим теперь, что два электрона, образующие куперовскую пару, с импульсами $\mathbf{p}_1$ и $\mathbf{p}_2$ сталкиваются между собой и переходят в состояние с импульсами $\mathbf{p}'_1$ и $\mathbf{p}'_2$. Ясно, что перед столкновением состояния с импульсами $\mathbf{p}'_1$ и $\mathbf{p}'_2$ должны быть вакантными, т. е. находиться в пределах указанного выше тонкого сферического слоя. То

же самое относится к исходным векторам $\mathbf{p}_1$ и $\mathbf{p}_2$. В противном случае после столкновения состояние с импульсами $\mathbf{p}_1$ и $\mathbf{p}_2$ оказалось бы незаполненным, а это было бы невозможно, если бы это состояние лежало в импульсном пространстве, ограниченном внутренней сферой слоя. В силу закона сохранения импульса $\mathbf{p}_1 + \mathbf{p}_2 = \mathbf{p}'_1 + \mathbf{p}'_2 = \mathscr{P} = \text{const}$, где $\mathscr{P}$ — результирующий импульс обоих электронов, не меняющийся при столкновении. Процесс столкновения изображен на рис. 118. Конец вектора $\mathbf{p}_1$ и начало вектора $\mathbf{p}_2$ должны лежать в верхнем заштрихованном четырехугольнике, получающемся в результате пересечения двух сферических слоев со средними радиусами p_1 и p_2. Аналогично, в нижнем заштрихованном четырехугольнике должны лежать конец вектора $\mathbf{p}'_1$ и начало вектора $\mathbf{p}'_2$. Ясно, что столкновения будут происходить тем чаще, чем короче вектор $\mathscr{P}$. Практически будут происходить только такие столкновения, когда оба сферических слоя совпадают, т. е. когда $\mathscr{P} = 0$. Это значит, что должно быть $\mathbf{p}_1 = -\mathbf{p}_2$, $\mathbf{p}'_1 = -\mathbf{p}'_2$. Таким образом, в отсутствие тока импульсы электронов, образующих куперовскую пару, равны по величине и противоположны по направлению.

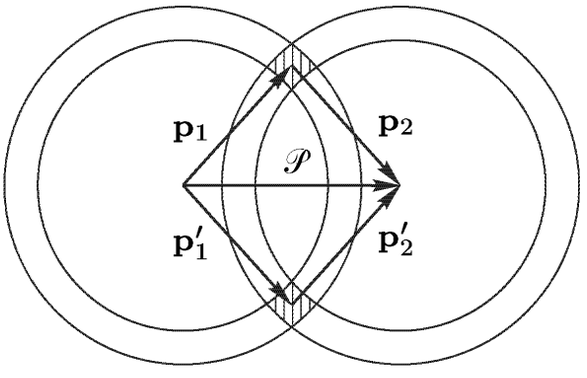


Рис. 118

Как уже было сказано выше, при температурах порядка $T_к$ энергия электронов в сферическом слое отличается от энергии Ферми на величину порядка $kT_к$ так что

$$kT_к \approx \frac{p^2}{2m} - \frac{p_F^2}{2m} \approx \frac{p + p_F}{2m}\,\Delta p,$$

где $\Delta p = p - p_F$. Таким образом, неопределенность в положении электронов $\Delta x \approx \hbar/\Delta p \approx \hbar p_F/(mkT_к)$. По порядку величины (при $T \sim 10$ К) неопределенность Δx обычно составляет 10^{-4} см. Величина Δx характеризует размеры области, в которой локализованы электроны, образующие куперовскую пару, т. е. «пространственные размеры» этой пары. Они, как видим, примерно в 10 000 раз превосходят среднее расстояние между электронами проводимости в металле (порядка 10^{-8} см). Значит, между электронами, связанными в пару, находится очень много других электронов. На этом основании говорят, что состояния электронов в куперовской паре *слабо коррелированы по координатам* обычного пространства. Сама эта корреляция на расстояниях такого порядка не может быть понята с классической точки зрения и является сугубо квантовым свойством. Напротив, *по импульсам корреляция сильная*, поскольку $\mathbf{p}_1 = -\mathbf{p}_2$.

5. Существование спаривания электронов в сверхпроводнике (при $T < T_к$) было доказано прямыми опытами по *квантованию магнитного потока*. Рассмотрим сверхпроводящее кольцо, по которому циркулирует сверхпроводящий ток. Пусть электроны движутся по окружности

радиуса r со скоростью v. Энергия тока представляется выражением $\mathscr{E} = (1/2c)I\Phi$, где I — сила тока, а Φ — магнитный поток через рассматриваемую окружность, создаваемый этим током. Если N — полное число электронов в кольце, а T — период обращения, то $I = Ne/T = Nev/(2\pi r)$. Таким образом, $\mathscr{E} = Nev\Phi/(4\pi rc)$. С другой стороны, та же энергия равна $\mathscr{E} = Nmv^2/2$. Приравнивая оба выражения, получим $\Phi = 2\pi rcmv/e$. Если электроны движутся куперовскими парами, то импульс каждой такой пары равен $p = 2mv$, так что $\Phi = \pi rcp/e$. Но импульс куперовской пары может принимать только квантованные значения согласно соотношению $pr = n\hbar = nh/2\pi$, где n — целое число. Следовательно,

$$\Phi = \frac{hc}{2e}\,n. \tag{62.1}$$

Эта формула выражает *квантование магнитного потока* в сверхпроводниках, причем *квант магнитного потока* определяется выражением

$$\Phi_0 = \frac{hc}{2e} = 2{,}07 \cdot 10^{-7}\ \text{Гс} \cdot \text{см}^2. \tag{62.2}$$

Формула такого вида была получена в 1950 г. Ф. Лондоном (1900–1954) еще до создания теории сверхпроводимости. Однако Лондон получил для Φ_0 вдвое большее значение по сравнению с тем, что дает формула (62.2). Это объясняется тем, что в 1950 г. явление спаривания электронов еще не было известно. Поэтому для импульса Лондон пользовался выражением $p = mv$, а не выражением $p = 2mv$, как делали мы. Опыт показал правильность формул (62.1) и (62.2) и тем самым подтвердил существование явления спаривания электронов. В одном из таких опытов сверхпроводящий ток возбуждался в оловянной трубочке с длиной 1 см и внутренним диаметром $2R = 1{,}5 \cdot 10^{-3}$ см. Благодаря малости диаметра при возбуждении всего одного кванта магнитного потока $\Phi_0 = \pi R^2 H$ получается уже макроскопически измеримое магнитное поле $H = 0{,}117$ Гс, которое слабее магнитного поля Земли примерно всего в 5 раз. Таким образом, квантование магнитного потока в подобных опытах проявляется уже *в макроскопических масштабах.*

В связи с изложенным отметим следующее обстоятельство. Известно, что в сверхпроводящем кольце можно возбудить незатухающий электрический ток. Например, один из опытов такого рода длился 2,5 года, и все же никакого затухания тока обнаружено не было. На первый взгляд в этом нет ничего удивительного, поскольку в сверхпроводнике не выделяется джоулево тепло, а потому и нет затухания. На самом деле вопрос сложнее. Электроны в сверхпроводящем кольце движутся ускоренно и *должны излучать*, а это должно привести к затуханию тока. Опыт же показывает, что затухания нет. Противоречие устраняется совершенно так же, как и соответствующее противоречие с излучением в классической теории атома. Чтобы не было излучения, Бор ввел *квантовый постулат* о стационарных состояниях атома. Так, и в сверхпроводящем кольце с током излучение не появляется из-за

квантования электрического тока. Но это квантование наблюдается уже в макроскопическом масштабе.

6. Электроны в куперовской паре обладают одинаковыми по величине, но противоположными по направлению импульсами. Их спины в принципе могут быть параллельны или антипараллельны. В обоих случаях полный спин куперовской пары получается целым, так что куперовская пара будет бозоном, а не фермионом. Более того, спин куперовской пары будет равен нулю, так как состояние с параллельными спинами электронов неустойчиво. Устойчивым является состояние с антипараллельными спинами электронов, которое и реализуется в действительности. Таким образом, *куперовские пары являются бозонами, а не фермионами.* На них запрет Паули не распространяется. Этим устраняется трудность, о которой говорилось в самом начале этого параграфа. Куперовские пары образуют «жидкость», которая может совершать сверхтекучее движение между ионами кристаллической решетки. Это и есть сверхпроводимость. Остановимся на этом вопросе несколько подробнее.

При $T = 0$ все куперовские пары находятся в основном состоянии с нулевыми импульсами. При $T < T_к$ в основном состоянии будет находиться большинство куперовских пар. Они образуют *связанный коллектив* и, как бозе-частицы, все находятся в *одном и том же квантовом состоянии.* Если разорвать куперовскую пару и удалить из этого коллектива образовавшийся электрон, то возникнет возбужденное состояние, энергия которого выше энергии исходного состояния коллектива. Существенно, что эти два состояния отделены одно от другого *энергетической щелью* конечной ширины, в которой нет квантовых состояний системы. Хотя ширина щели и очень мала (при $T = 0$ всего порядка 10^{-3}–10^{-4} эВ), она приводит к качественному изменению поведения всей системы электронов. В самом деле, из-за взаимодействия куперовских пар с колебаниями решетки хотя и происходят разрывы этих пар, но при $T < T_к$ образовавшиеся электроны не могут преодолеть энергетическую щель, так как их энергии недостаточно для этого. Освободившийся электрон «ищет» себе партнера среди других освободившихся электронов, чтобы образовать с ним новую куперовскую пару с нулевыми импульсом и спином. Но такой процесс смены партнеров» ввиду тождественности электронов не приводит к новому состоянию, т. е. к разрушению коллектива. При $T < T_к$ коллектив ведет себя и движется как целое. Если же $T > T_к$, то электроны, образующиеся при разрыве куперовских пар, преодолевают энергетическую щель и выбывают из коллектива, что приводит к разрушению последнего.

Наличие в энергетическом спектре сверхпроводящего состояния щели конечной ширины и объясняет исчезновение сопротивления электрическому току. В отсутствие электрического тока импульсы всех куперовских пар равны нулю: импульс одного электрона равен $+\mathbf{p}$, а другого $-\mathbf{p}$. При наложении электрического поля $\mathbf{E}$ оба электрона пары получают один и тот же дополнительный импульс $\mathbf{p}'$. Их полные импульсы становятся равными соответственно $\mathbf{p} + \mathbf{p}'$ и $-\mathbf{p} + \mathbf{p}'$, а импульс всей пары $2\mathbf{p}'$. Связанный коллектив куперовских пар начинает

двигаться как целое с определенной скоростью, соответствующей этому значению импульса, в результате чего возникает электрический ток. Пока последний недостаточно силен, связанный коллектив куперовских пар не разрушается из-за наличия энергетической щели, а потому ток не встречает сопротивления. Возникает *токовое состояние* коллектива куперовских пар, причем к моменту установления постоянного тока поле E в сверхпроводнике обращается в нуль. В движущемся коллективе куперовских пар, понятно, продолжают происходить процессы электрон-фононного взаимодействия, обеспечивающие устойчивость токового состояния коллектива куперовских пар. При токах достаточно большой силы коллектив распадается и сверхпроводимость исчезает.

При температуре абсолютного нуля все электроны сверхпроводника связаны в куперовские пары. При повышении температуры число куперовских пар уменьшается. Это связано с двумя обстоятельствами. Во-первых, с повышением температуры увеличивается вероятность отрыва электрона от куперовской пары и перескока его через щель. Во-вторых, уменьшается число процессов образования пар из-за уменьшения актов обмена фононами между электронами, приводящими к притяжению между ними. В результате энергетическая щель становится у́же, что еще более ускоряет процесс уменьшения числа куперовских пар. При $T = T_к$ ширина щели, а с ней и число куперовских пар сокращаются до нуля. Связанный коллектив таких пар распадается. Вместе с ним исчезает и сверхпроводимость — при $T > T_к$ металл становится нормальным проводником.

7. Родственность явлений сверхпроводимости и сверхтекучести проявляется и в предсказании сверхтекучести $^3_2$He. Поскольку атомы этого изотопа имеют полуцелый спин, сначала предполагали, что $^3_2$He не обладает сверхтекучестью. Лишь после создания теории сверхпроводимости обратили внимание на возможность объединения атомов $^3_2$He в пары, аналогичные куперовским. Такие пары, обладая целым спином, являются бозонами, на основании чего была предсказана сверхтекучесть жидкого $^3_2$He, которая и была обнаружена экспериментально (см. § 61).

8. Теория объясняет и эффект Мейсснера–Оксенфельда (см. т. III, § 80). Как известно, этот эффект состоит в том, что магнитное поле не проникает внутрь массивного сверхпроводника или вытесняется из последнего. Наиболее просто дать объяснение для сверхпроводника цилиндрической формы, предполагая его достаточно длинным, чтобы исключить влияние краевых эффектов. Если магнитное поле параллельно оси цилиндра, то достаточно принять во внимание, что сверхпроводящий ток не встречает сопротивления, даже если он течет по поверхности тела (точнее, вдоль тонкого поверхностного слоя его). Кроме того, надо учесть, что из всех мыслимых состояний сверхпроводника в действительности реализуется состояние наименьшей энергии.

Предположим, что $T < T_к$. Пусть сверхпроводник помещен в постоянное однородное магнитное поле $\mathbf{H}$. Логически допустимо, что это поле проникнет в сверхпроводник, заполняя весь объем его и оставаясь однородным. В силу граничных условий напряженности магнитного поля $\mathbf{H}$ внутри и вне цилиндра должны быть одинаковыми. Магнитная

энергия системы будет равна

$$\mathscr{E}_1 = \mathscr{E}_e + \mathscr{E}_i,$$

где $\mathscr{E}_e$ — магнитная энергия, локализованная вне, а $\mathscr{E}_i = (1/8\pi)\mu V\mathbf{H}^2$ — внутри цилиндра (V — объем цилиндра, μ — магнитная проницаемость сверхпроводника). Допустим теперь, что по поверхности цилиндра циркулируют круговые токи, создающие внутри цилиндра однородное поле $\mathbf{H}_i$. Поле, создаваемое ими вне цилиндра, как известно, равно нулю. Поэтому для магнитной энергии можно написать:

$$\mathscr{E}_2 = \mathscr{E}_e + \frac{\mu}{8\pi} V(\mathbf{H} + \mathbf{H}_i)^2,$$

причем энергия $\mathscr{E}_e$, локализованная во внешнем пространстве, будет такой же, что и в предыдущем случае. Энергия $\mathscr{E}_2$ обращается в минимум $\mathscr{E}_{2\,\text{мин}} = \mathscr{E}_e$, когда $\mathbf{H} + \mathbf{H}_i = 0$. Состояние с такой минимальной энергией и должно реализоваться в действительности, а не состояние с энергией $\mathscr{E}_1$, так как нет никаких препятствий для тока перераспределяться по объему сверхпроводника. Таким образом, при внесении цилиндра в однородное магнитное поле появляются сверхпроводящие круговые поверхностные токи, уничтожающие магнитное поле в объеме цилиндра. Это и есть эффект Мейсснера–Оксенфельда. Очевидно, он имеет место только при $T < T_\text{к}$, так как в противном случае сверхпроводящий поверхностный ток невозможен.

9. В приведенном доказательстве не учтена *поверхностная энергия*, существующая на границе раздела сверхпроводящей фазы с нормальной. Она определяется конечной глубиной проникновения магнитного поля из нормальной в сверхпроводящую фазу, притяжением между электронами куперовских пар, наличием энергетической щели между сверхпроводящей и нормальной фазами и пр. Эта энергия может быть как положительной, так и отрицательной. На это обстоятельство в 1957 г. обратил внимание А. А. Абрикосов (р. 1928), который ввел деление сверхпроводников на *сверхпроводники первого* и *второго рода.* Для первых поверхностная энергия положительна, для вторые отрицательна. К сверхпроводникам первого рода относится большинство чистых металлов, а второго рода — подавляющее число сплавов, а также многие чистые металлы с примесями. В сверхпроводниках первого рода наблюдается эффект Мейсснера–Оксенфельда, в сверхпроводниках второго рода — не всегда. Сверхпроводник второго рода может находиться в *сверхпроводящем* и *смешанном состояниях.* В сверхпроводящем состоянии имеет место эффект Мейсснера–Оксенфельда, в смешанном — нет.

На рис. 119 кривая $H = H_{\text{к}1}(T)$ определяет напряженность критического поля, при которой находятся в равновесии сверхпроводящая и смешанная фазы. Аналогично, кривая $H = H_{\text{к}2}(T)$ соответствует равновесию между сверхпроводящей и нормальной фазами. Область температур и магнитных полей, при которых металл находится в сверхпроводящем состоянии, обозначена двойной штриховкой, область смешанного состояния — простой штриховкой, а область нормального

состояния не заштрихована. Для сверхпроводников первого рода смешанного состояния не существует.

Понятно, что в сверхпроводнике должно реализоваться состояние *минимума* полной энергии, включающей поверхностную. По этой причине и возникает смешанное состояние. В сверхпроводник в смешанном состоянии внешнее магнитное поле проникает через *нити конечного поперечного сечения*. Конечное сечение получается потому, что из области, занятой магнитным полем, происходит его проникновение в окружающее пространство, находящееся в сверхпроводящем состоянии, причем этот процесс характеризуется конечной глубиной проникновения. Тело пронизано нитями, через которые проходят магнитные потоки, а сами нити отделены, одна от другой промежутками, сохраняющими сверхпроводимость, если только расстояние между соседними нитями превышает примерно удвоенную глубину проникновения магнитного поля в сверхпроводник.

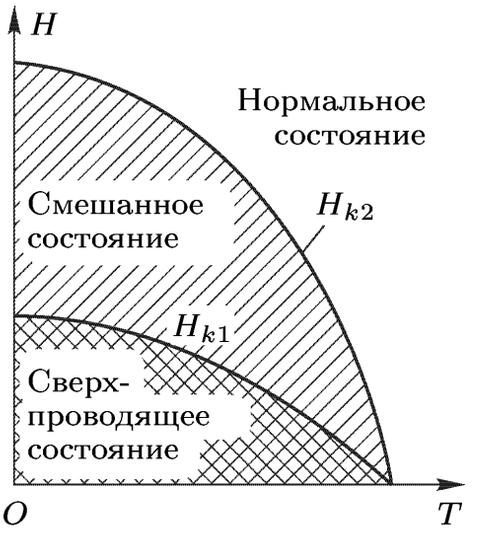


Рис. 119

Существенно, что магнитный поток через поперечное сечение нити *квантуется*. Энергетически выгодно, чтобы через каждую нить проходил *один квант* магнитного потока. Действительно, рассмотрим две нити радиуса r, через каждую из которых проходит один квант магнитного потока. Суммарный магнитный поток через обе нити равен $2\pi r^2 H$. Пусть обе нити сливаются в одну радиуса R. Тогда тот же магнитный поток будет $\pi R^2 H$. Сравнивая оба выражения, находим $R = r\sqrt{2}$. Поэтому длина окружности поперечного сечения нити, образовавшейся в результате слияния, будет $2\pi R = 2\pi r\sqrt{2}$, тогда как сумма длин окружностей поперечных сечений первоначальных двух нитей больше, так как она равна $2\pi r \cdot 2$. Таким образом, слияние двух нитей *уменьшает боковую поверхность*, по которой нити граничат с окружающим пространством. Это ведет к энергетически невыгодному увеличению поверхностной энергии, поскольку она отрицательна. Итак, через тело проходит магнитное поле, но оно сохраняет сверхпроводимость благодаря наличию сверхпроводящих промежутков между нитями. При усилении магнитного поля число нитей в теле увеличивается, а сверхпроводящие промежутки между ними сокращаются. В конце концов магнитное поле начинает пронизывать все тело, и сверхпроводимость исчезает.

Сверхпроводящие сплавы благодаря высоким значениям критических магнитных полей $H_{к2}$ нашли широкое применение при изготовлении обмоток соленоидов, предназначенных для получения сверхсильных магнитных полей (100 000 Гс и больше).

Глава VIII

СТАТИЧЕСКИЕ СВОЙСТВА АТОМНОГО ЯДРА

§ 63. Введение

1. Первое явление из области ядерной физики было открыто в 1896 г. Анри Беккерелем (1852–1908). Это — *естественная радиоактивность* солей урана, проявляющаяся в самопроизвольном испускании невидимых лучей, способных вызывать ионизацию воздуха и почернение фотоэмульсий. Через два года Пьер Кюри (1859–1906) и Мария Склодовская-Кюри (1867–1934) открыли радиоактивность тория и выделили из солей урана полоний и радий, радиоактивность которых оказалась в миллионы раз сильнее радиоактивности урана и тория.

Детальное экспериментальное изучение радиоактивных излучений было произведено Резерфордом. Он показал, что радиоактивные излучения состоят из трех типов лучей, названных соответственно α-, β- и γ-лучами. Бета-лучи состоят из отрицательных электронов, движущихся значительно быстрее, чем электроны в катодных лучах, α-лучи — из положительно заряженных частиц (α-частиц, масса которых равна массе ядра гелия), γ-лучи аналогичны лучам Рентгена, только значительно более жесткие. В соответствии с этим α- и β-лучи отклоняются магнитным полем, и притом в противоположные стороны, а на γ-лучи магнитное поле не действует.

Наименьшей проникающей способностью и наибольшим ионизующим действием обладают α-лучи. Они поглощаются слоем алюминия толщиной всего в несколько микрометров. Для поглощения β-лучей требуется слой алюминия в среднем толщиной 1 мм. Ионизующая способность β-лучей много меньше, чем α-лучей. Наибольшей проникающей способностью и наименьшим ионизующим действием обладают γ-лучи. Для защиты от них применяются свинцовые листы толщиной, зависящей от интенсивности излучения.

Ядерная природа радиоактивности была понята Резерфордом после того, как в 1911 г. он предложил ядерную модель атома (см. § 9) и установил, что радиоактивные излучения возникают в результате процессов, происходящих внутри атомного ядра. С этого момента и ведет свое начало ядерная физика.

2. Долгое время предполагалось, что само атомное ядро состоит из протонов и электронов. Однако такая гипотеза находилась в противоречии с экспериментальными фактами, относящимися к спинам и магнитным моментам ядер (см. § 68). Кроме того, присутствие электронов внутри ядра оказалось невозможным совместить с принципом неопределенности Гейзенберга (см. § 65, п. 10). В 1932 г. после открытия Чедвиком (1891–1974) нейтрона было установлено, что *ядро состоит*

из протонов и нейтронов (эти частицы получили общее наименование *нуклонов*). Такая модель ядра была предложена в том же году независимо друг от друга Д. Д. Иваненко (р. 1904) и Гейзенбергом. Свободный протон — стабильная частица. Масса нейтрона больше массы протона на 0,14 % или 2,5 электронных масс. В соответствии с этим в свободном состоянии нейтрон распадается на протон, электрон и электронное антинейтрино:

$$\mathrm{n} \to \mathrm{p} + \mathrm{e}^- + \overline{\nu}_\mathrm{e}. \tag{63.1}$$

Среднее время жизни нейтрона близко к 15,3 мин. Казалось бы, нет смысла говорить о нейтроне как об «элементарной» частице, а следует рассматривать его как «составную» частицу. Однако внутри ядра протон не свободен и ведет себя так же, как составная частица, распадающаяся на нейтрон, позитрон и электронное нейтрино:

$$\mathrm{p} \to \mathrm{n} + \mathrm{e}^+ + \overline{\nu}_\mathrm{e}. \tag{63.2}$$

Поэтому с тем же основанием протон можно считать «сложной» частицей, превращающейся в более «простую» — нейтрон. Это означает, что вопрос о том, какая частица — протон или нейтрон — более элементарна, лишен физического смысла. В этом отношении обе частицы равноправны. Какая из них распадается, зависит от энергетических соотношений. В свободном состоянии нейтрон радиоактивен, а протон стабилен, т. е. идет процесс (63.1). Внутри же ядра возможны оба процесса (63.1) и (63.2). Тип распада определяется массами рассматриваемого ядра и возможных продуктов распада. Это обстоятельство и дает возможность рассматривать обе частицы — протон и нейтрон — как *элементарные*, взаимно превращающиеся друг в друга.

Число протонов в ядре (*зарядовое число*) принято обозначать через Z, число нейтронов — через N. Их сумма $A = Z + N$ называется *массовым числом ядра*; число Z называют также *порядковым номером элемента*. Атомы с одинаковыми Z (т. е. атомы одного и того же элемента), но различными N называются *изотопами*, с одинаковыми A, но различными Z — *изобарами*, с одинаковыми N, но различными Z — *изотонами*. Наряду с термином *ядро атома* используется также термин *нуклид*.

3. Основное различие между протоном и нейтроном состоит в том, что протон — *заряженная* частица, заряд которой $e = 4{,}803$ СГСЭ $= 1{,}602 \cdot 10^{-19}$ Кл. Это *элементарный заряд*, численно равный заряду электрона. Нейтрон же, как показывает уже его название, *электрически нейтрален*. Спины протона и нейтрона одинаковы и равны спину электрона, т. е. $1/2$ (в единицах $\hbar$), — обе частицы являются *фермионами* и подчиняются статистике Ферми–Дирака. Массы протона и нейтрона почти равны: масса протона $m_\mathrm{p} = 1836{,}15 m_\mathrm{e}$, масса нейтрона $m_\mathrm{n} = 1838{,}68 m_\mathrm{e}$, где m_e — масса электрона; $m_\mathrm{e} = 9{,}1095 \cdot 10^{-28}$ г.

Масса частицы связана с ее полной энергией соотношением Эйнштейна $\mathscr{E}_\mathrm{пол} = mc^2$. Поэтому в ядерной физике и физике элементарных частиц массу принято измерять в единицах энергии, причем за единицу энергии принимается *мегаэлектронвольт* (МэВ). В этих

единицах $m_e = 0{,}511$ МэВ, $m_p = 938{,}3$ МэВ, $m_n = 939{,}6$ МэВ. При этом в указанных разделах физики под m понимается всегда *масса покоя* частицы (обозначение m_0 не применяется). Принято также массу частицы измерять в так называемых атомных единицах массы (а. е. м). За *атомную единицу массы* принимается $1/12$ часть массы нейтрального атома углерода ^{12}C. Нетрудно подсчитать, что 1 а. е. м. = 931,502 МэВ.

4. Малое различие в массах и совпадение спинов протона и нейтрона является проявлением общего свойства этих частиц — если отвлечься от различия, связанного с наличием у протона электрического заряда, и от различия в значениях магнитных моментов, то во всех остальных отношениях протон и нейтрон весьма сходны между собой. Это сходство, имеющее фундаментальное значение в ядерной физике, с особой отчетливостью проявляется в свойствах «зеркальных» ядер. Зеркальными называются два ядра с одинаковыми массовыми числами A каждое из которых получается из другого заменой всех протонов нейтронами, а всех нейтронов — протонами. Примером такой пары могут служить сами протон и нейтрон. Другими примерами являются ^3_1H и ^3_2He, ^7_4Be и ^7_3Li, $^{11}_5B$ и $^{11}_6C$, $^{13}_6C$ и $^{13}_7N$, $^{15}_7N$ и $^{15}_8O$, $^{17}_8O$ и $^{17}_9F$, $^{29}_{14}Si$ и $^{29}_{15}P$ и т. д. (более тяжелое ядро в каждой из этих пар радиоактивно). Как показывает опыт, оба зеркальных ядра имеют почти одинаковые энергии связи, сходное строение спектра возбужденных уровней энергии, одинаковые спины и пр. Это указывает на сходство сил, действующих между двумя протонами и двумя нейтронами.

На малых расстояниях (например, внутри ядра) между этими частицами действуют мощные ядерные силы, по сравнению с которыми электромагнитные силы в сотни раз слабее. В пренебрежении электромагнитными силами протон и нейтрон обладают одинаковыми свойствами: при прочих равных условиях ядерные силы, действующие между двумя протонами, равны ядерным силам, действующим между двумя нейтронами, а также между нейтроном и протоном. Это свойство называется *зарядовой симметрией ядерных сил.* Оно является проявлением еще более глубокой закономерности, называемой *изотопической инвариантностью.*

5. В настоящее время еще не создана последовательная законченная теория атомного ядра. Причина этого заключается прежде всего в том, что до сих пор неизвестна аналитическая зависимость ядерных сил от определяющих их параметров. Поэтому в отношении ядерных сил приходится вводить различные заведомо упрощенные и ограниченно применимые предположения, подобранные так, чтобы в определенной области явлений достигнуть согласия с экспериментальными данными.

В атомной физике, даже не зная закона Кулона, можно было бы в значительной степени восстановить его, изучая энергетический спектр атома водорода, поскольку последний содержит очень много уровней. В физике же ядра из-за короткодействующего характера ядерных сил такой возможности нет, так как имеется всего одно связанное состояние из двух нуклонов — дейтрон, состоящий из протона и нейтрона. Дейтрон же имеет всего *один* энергетический уровень (см. задачу

к § 26), а этого, конечно, мало для суждения о силах, действующих между протоном и нейтроном. Положение осложняется тем, что силы взаимодействия нуклонов зависят не только от *расстояния между ними*, но и от их *скоростей*, а также от *ориентации спинов*.

6. Если бы даже силы взаимодействия между нуклонами были полностью известны, то в теории ядра надо было бы еще решить квантовомеханическую *задачу многих тел*. Она много сложнее аналогичной классической задачи, с которой не может справиться современная математика (даже в случае трех взаимодействующих тел). Например, если ядро состоит из A нуклонов, то в пренебрежении их спинами его состояние определяется волновой функцией $\psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_A)$, зависящей от $3A$ пространственных координат. Эта функция должна быть определена из уравнения Шредингера с $3A$ независимыми переменными. Учет спинов нуклонов невообразимо усложняет эту задачу. С учетом спинов состояние ядра определяется волновой функцией $\psi(s_1, s_2, \ldots, s_A; \mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_A)$, зависящей не только от пространственных координат, но и от спиновых переменных (т. е. от квантовых чисел, определяющих проекции спина на выделенное направление), каждая из которых может принимать два значения: $-1/2$ и $+1/2$. Ее нахождение эквивалентно определению 2^A функций только от $3A$ пространственных координат. Например, при $A = 100$ число таких функций равно $2^{100} \approx 1{,}27 \cdot 10^{30}$. Ясно, что решение такой задачи абсолютно непосильно даже для самых мощных вычислительных машин.

В теории атома для преодоления аналогичной трудности разработаны превосходные приближенные методы — метод Хартри (1897–1958) и еще более точный метод Фока (1898–1974). Они используют то обстоятельство, что атом практически пуст — расстояния между электронами атомной оболочки, а также между последними и ядром атома очень велики по сравнению с размерами этих частиц. В случае ядер такой возможности нет, так как расстояния между нуклонами ядра того же порядка, что и размеры самих нуклонов. Поэтому указанные методы в случае ядра неприменимы.

Число нуклонов в ядре (за исключением только протона и дейтрона — ядер атомов водорода и дейтерия) всегда больше двух. Поэтому ядро следует рассматривать как систему многих частиц. Если бы число частиц в системе было очень велико, то надежное и даже единственно возможное описание ее состояния давал бы *статистический метод*. Но это условие совсем не выполняется для легких ядер. Для них статистический метод неприменим. Им можно пользоваться для средних и в особенности для тяжелых ядер. Но и здесь применимость статистического метода ограничена, так как число нуклонов даже в тяжелых ядрах все же недостаточно велико.

7. Сказанное делает понятным, почему мы не располагаем последовательной законченной теорией ядра, которая бы единым образом объясняла все его свойства. Вместо такой теории в ядерной физике используются различные *модели ядра*, каждая из которых охватывает

лишь ограниченный круг явлений [1]. Ядерные модели не дают истинной теории явлений, но позволяют систематизировать явления из различных областей ядерной физики, а главное предсказывать новые. В этом и заключается их польза. Достойно удивления, что на этом пути достигнуто так много в понимании известных из опыта ядерных явлений и предсказании новых.

8. Атомные ядра условно принято делить на *стабильные* и *радиоактивные*. Условность такого деления состоит в том, что в сущности все ядра подвергаются радиоактивному распаду, но с различной скоростью. Стабильными называют ядра, распадающиеся достаточно медленно. А понятие медленности зависит от конкретно поставленной задачи. Обычно ядра принято считать стабильными, если со времени образования видимой части Вселенной (порядка 10^{10} лет), т. е. со времени образования химических элементов, распалась ничтожная часть их. Ядра, распадающиеся более быстро, считаются радиоактивными.

Физические величины, характеризующие свойства атомных ядер, можно разделить на *статические* и *динамические*. Статические характеристики относятся к определенному, обычно невозбужденному состоянию ядра; динамические проявляются при возбуждениях и распадах ядра и в ядерных реакциях. Статические характеристики обычно называют свойствами стабильных ядер. Изменение терминологии, которому мы следуем, предложено Ю. М. Широковым и П. П. Юдиным в их известном учебнике «Ядерная физика» (М.: Наука, 1980). Оно оправдано тем, что между стабильными и радиоактивными ядрами, как уже указывалось выше, нет резкой границы. Статические же свойства присущи но только стабильным, но и радиоактивным ядрам и даже ядрам в возбужденных состояниях.

Важнейшими статическими характеристиками ядра являются: зарядовое число (атомный номер) Z, масса ядра M, энергия связи $\mathscr{E}_{\text{св}}$, спин $\mathbf{I}$, магнитный момент $\boldsymbol{\mu}$, электрический квадрупольный момент $\mathbf{Q}$, радиус R и несферичность ядра $\delta R/R$, четность P волновой функции ψ, изотопический спин T, спектр возбужденных состояний.

Радиоактивные ядра дополнительно характеризуются типом радиоактивного превращения (α- или β-распад, спонтанное деление и пр.), временем жизни τ (или периодом полураспада $T_{1/2}$), энергией испускаемых частиц и т. п.

9. В настоящее время в природе известны четыре вида фундаментальных взаимодействий: *сильное*, *электромагнитное*, *слабое*, *гравитационное*. Сильное взаимодействие удерживает нуклоны в атомных ядрах и присуще также большинству элементарных частиц, так называемых *адронов* (протон, нейтрон, гипероны, мезоны и др.). К электромагнитным взаимодействиям сводятся непосредственно воспринимаемые нами силы природы (за исключением тяготения): упругие, вязкие, молекулярные, химические и пр. Слабые взаимодействия вызывают

[1] Краткое описание некоторых моделей ядра дано в гл. X.

β-распад радиоактивных ядер (см. § 74) и наряду с электромагнитными силами управляют поведением *лептонов* — элементарных частиц, не участвующих в сильных взаимодействиях и обладающих спином $1/2$ (электрон, мюон, нейтрино и др.). Нейтральные лептоны (все нейтрино и антинейтрино) не участвуют и в электромагнитных взаимодействиях. Гравитационное взаимодействие присуще всем частицам.

Об интенсивности перечисленных взаимодействий можно судить по скорости процессов, вызываемых ими. Обычно для сравнения берут скорости процессов при кинетических энергиях: сталкивающихся частиц около 1 ГэВ; такие энергии характерны для физики элементарных частиц. При таких энергиях процессы, вызываемые сильными взаимодействиями, совершаются за времена порядка 10^{-23} с, электромагнитными — за времена порядка 10^{20} с, слабыми — за времена порядка 10^{9} с. Другой величиной, характеризующей интенсивность взаимодействия, является длина свободного пробега частицы в веществе. Сильновзаимодействующие частицы при энергии около 1 ГэВ можно задержать железной плитой толщиной в несколько десятков сантиметров. Нейтрино же с энергией 10 МэВ (такова энергия антинейтрино, образующихся в ядерных реакторах), которым свойственно только слабое взаимодействие, для своего задержания потребовали бы слоя железа толщиной не менее 10^{9} км.

Сильные и слабые взаимодействия проявляются только на *коротких расстояниях.* Радиус действии сильных взаимодействий составляет примерно 10^{-13} см (1 ферми), а слабых — $2 \cdot 10^{-16}$ см (0,002 ферми). Электромагнитные силы, напротив, являются *дальнодействующими*; они убывают обратно пропорционально квадрату расстояния между взаимодействующими частицами. По тому же закону убывают с расстоянием и гравитационные силы. Поэтому отношение электромагнитных и гравитационных сил не зависит от расстояния между взаимодействующими частицами. При одном и том же расстоянии оно равно $F_{\text{эл}}/F_{\text{гр}} = e_1 e_2/(G m_1 m_2)$, где G — гравитационная постоянная, m_1 и m_2 — массы частиц, e_1 и e_2 — их электрические заряды. Для взаимодействия двух протонов эта формула дает $F_{\text{эл}}/F_{\text{гр}} \approx 1{,}23 \cdot 10^{36}$.

Таким образом, в области, где проявляются слабые силы, гравитационное взаимодействие частиц на много порядков меньше даже слабого. Поэтому в физике микромира при современном ее состоянии гравитационное взаимодействие не учитывается. Но в макромире при рассмотрении движения больших масс: галактик, звезд, планет и пр., а также при рассмотрении движения небольших макроскопических тел в поле таких масс гравитационное взаимодействие становится определяющим. Особенно существенную, пожалуй, главную роль оно играет в процессах образования и эволюции звезд (см. гл. XV). Обусловлено это двумя обстоятельствами: во-первых, дальнодействующим характером гравитационных сил, во-вторых, положительностью масс всех тел. В силу второго обстоятельства гравитационные силы всегда являются силами притяжения. При увеличении масс взаимодействующих тел гравитационное взаимодействие между ними *усиливается* и притом

пропорционально массам обоих тел. Этого не происходит в случае электрического взаимодействия макроскопических тел, так как положительные и отрицательные заряды составляющих их частиц в высокой степени *нейтрализуют* друг друга. Вот почему электрические силы не оказывают существенного влияния на движение больших масс, хотя в случае взаимодействия тел, состоящих только из положительных или только отрицательных частиц, они значительно больше гравитационных и также являются дальнодействующими.

10. Классическая физика принимала, что взаимодействие между телами передается с конечной скоростью посредством силовых полей. Так, электрический заряд создает вокруг себя электрическое поле, которое в месте нахождения другого электрического заряда воздействует на него с некоторой силой (см. т. III, § 1). Так же, но уже посредством других силовых полей, осуществляются все взаимодействия в природе, например гравитационное. Квантовая физика не изменила такое представление, но учла *квантовые свойства самого поля.* Из-за корпускулярно-волнового дуализма всякому полю должна соответствовать определенная частица (*квант поля*), которая и является переносчиком взаимодействия. Одна из взаимодействующих частиц испускает квант поля, другая его поглощает. В этом и состоит механизм взаимодействия частиц. В случае электромагнитных взаимодействий квантами поля — переносчиками взаимодействия — являются *фотоны.* Сильные взаимодействия переносятся *глюонами*, слабые — $W^{\pm}$- и Z^0-*промежуточными векторными бозонами*, предсказанными в теории электрослабого взаимодействия и открытыми экспериментально в 1983 г., гравитационное — гипотетическими *гравитонами.* В настоящее время электромагнитное и слабое взаимодействия рассматриваются как различные проявления единого *электрослабого взаимодействия*, подобно тому как электрическое и магнитное поля объединены в единое *электромагнитное поле.* Слабые силы на малых расстояниях (порядка радиуса их действия) одного порядка с электромагнитными. Но они экспоненциально убывают с расстоянием, так что на больших расстояниях слабые силы исчезающе малы по сравнению с электромагнитными.

Против представления взаимодействия посредством обмена квантами поля можно выдвинуть следующее возражение. Пока частица свободна, она не может испустить или поглотить квант поля. Для испускания и поглощения фотона это утверждение уже было доказано в § 1. Повторим его еще раз в измененной форме, не предполагая, что масса покоя кванта поля равна нулю.

Рассмотрим сначала процесс испускания кванта и применим доказательство от противного. Пусть масса покоя частицы до испускания кванта равна m. После испускания кванта масса покоя может измениться, обозначим ее через m'. При испускании частица может получить импульс отдачи $\mathbf{p}$. В системе, в которой частица покоилась до испускания кванта поля, закон сохранения энергии записывается

в виде

$$mc^2 = \sqrt{(m'c^2)^2 + (pc)^2} + \mathscr{E}_{\text{кв}}, \qquad (63.3)$$

где $\mathscr{E}_{\text{кв}}$ — энергия испущенного кванта. Если импульс кванта $\mathbf{\Pi}$, то, очевидно, $p + \Pi = 0$. Из написанного уравнения видно, что $m'c^2 \leqslant mc^2$. Если частица стабильна, а это надо предполагать, то ее энергия минимальна и не может дальше уменьшаться. Поэтому должно быть $m = m'$, $p = \Pi = 0$, $\mathscr{E}_{\text{кв}} = 0$. Итак, испущенный квант поля не обладает ни энергией, ни импульсом. А это значит, что такого кванта вообще нет. Доказательство, по существу, предполагает, что при испускании кванта поля внутреннее квантовое состояние частицы *не изменяется*. Если то же самое имеет место и при поглощении, то не составит труда распространить доказательство и на поглощение. Это предлагается сделать читателю.

Квантовая теория снимает выдвинутое возражение, используя принцип неопределенности Гейзенберга. Если квант поля существует в течение короткого промежутка времени Δt, необходимого для переноса взаимодействия, то его энергия $\mathscr{E}$ не может быть точно определенной. Ее неопределенность $\Delta\mathscr{E}$ удовлетворяет соотношению Гейзенберга $\Delta\mathscr{E} \cdot \Delta t \approx \hbar$. Формально это означает, что для таких промежутков времени *закон сохранения энергии нарушается*. Иначе можно сказать, что *для частиц, переносящих взаимодействие, нарушается обычная связь между энергией и импульсом*. Поэтому их называют не просто частицами, а *виртуальными частицами* или *виртуальными квантами поля*. Из таких виртуальных квантов, испускаемых и поглощаемых действительными частицами, и состоит силовое поле, их окружающее. В частности, электромагнитное поле, окружающее электрический заряд, состоит из испускаемых и поглощаемых виртуальных фотонов. Процессы испускания и поглощения виртуальных частиц называют также *виртуальными*.

11. Определенная трудность возникает, если механизм взаимодействия частиц микромира пытаться представить наглядно. Вообразим, например, двух людей A и B, которые стоят в разных местах и перекидываются мячом. Бросив мяч в направлении к B, A испытывает толчок в обратном направлении. Дополнительный толчок в том же направлении он получает, когда ловит мяч, брошенный B. В результате между A и B возникает сила отталкивания. Можно ли придумать классический пример, когда в результате перекидывания каким-либо предметом между A и B возникает притяжение? Можно. Для этого мяч надо заменить бумерангом и поставить A и B спиной друг к другу. A бросает бумеранг в направлении от B и при этом испытывает силу, направленную к B. Бумеранг, изменив направление полета, пролетает за B а затем возвращается к B. B ловит его и посылает в сторону, противоположную A. В результате возникает сила, направленная к A, и т. д. Таким образом, осуществляется притяжение между A и B. Очевидно, в этом примере существенную роль играет среда (воздух), в которой летит бумеранг. Без такой среды бумеранг летел бы по инерции — прямолинейно. Все это является только наглядной иллюстрацией

и не претендует на истинное объяснение механизма взаимодействия. Взаимодействие посредством обмена виртуальными частицами не имеет наглядного объяснения.

12. Виртуальные частицы непосредственно не воспринимаются. Каждая частица является источником какого-то поля (например, электрон — источником электромагнитного поля). Согласно корпускулярно-волновому дуализму всякому полю соответствуют частицы, являющиеся квантами этого поля. Масса этих частиц m может быть равна нулю (как у фотонов), но может быть и отличной от нуля. Свободная частица, ввиду законов сохранения энергии и импульса, не может испустить реальный квант поля, так как для этого необходима энергия, не меньшая mc^2. С точки зрения классической физики такой процесс невозможен без поступления энергии извне. Но квантовая физика с ее принципом неопределенности $\Delta\mathscr{E}\cdot\Delta t\approx\hbar$ допускает нарушение закона сохранения энергии в течение малых промежутков времени Δt, необходимых для передачи взаимодействия. Чтобы получить требуемую энергию $\Delta\mathscr{E}=mc^2$, необходимо время $\Delta t\approx\hbar/mc^2$. За это время виртуальная частица не может уйти дальше чем на расстояние

$$R=c\Delta t=\frac{\hbar}{mc}, \tag{63.4}$$

т. е. на расстояние, равное комптоновской длине волны частицы массой m. Эта длина и может быть принята за радиус действия, переносимого виртуальной частицей. Чем меньше масса m, тем больше радиус действия, переносимого виртуальной частицей.

Электромагнитные взаимодействия осуществляются обменом фотонами. Масса фотона $m=0$, поэтому радиус действия электромагнитных сил $R=\infty$, т. е. эти силы являются дальнодействующими. Дальнодействующими являются и гравитационные силы, перенос которых осуществляется гипотетическими гравитонами. Гравитоны из-за исключительной слабости гравитационного взаимодействия не наблюдались на опыте и, по-видимому, еще очень долго не будут наблюдаться, но теоретики не сомневаются в их существовании. На основании дальнодействующего характера гравитационных сил следует заключить, что масса гравитона должна равняться нулю.

13. В 1934 г. Таммом (1895–1971) и Иваненко была развита теория ядерных сил, согласно которой переносчиками этих сил являются легкие частицы — электрон и позитрон, нейтрино и антинейтрино (частицы промежуточной массы — мюоны и пионы — в то время еще не были известны). Протон, например, испустив позитрон и нейтрино, превращается в нейтрон. Нейтрон же, поглотив те же частицы, превращается в протон. В результате этих процессов и возникают силы притяжения между протоном и нейтроном. В количественном отношении теория Тамма и Иваненко оказалась несостоятельной, так как она приводила к ядерному взаимодействию, которое было примерно в 10^{11}–10^{13} раз слабее требуемого. Но основные идеи этой теории сохранили свое значение и впервые были использованы в 1935 г. японским физиком Юкава (1907–1981).

Юкава в 1935 г. высказал предположение о существовании особого поля ядерных сил. Квантами этого поля являются частицы, которые, согласно гипотезе Юкавы, и осуществляют взаимодействие между нуклонами. По экспериментальным данным радиус действия ядерных сил равен $1{,}2 \cdot 10^{-13}$ см. Исходя из этого, можно оценить с помощью формулы (63.4) массу гипотетической частицы Юкавы. Она оказалась равной $270m_e$ (около 140 МэВ). Вскоре в 1937 г. Неддермайер (р. 1907) и Андерсон (р. 1905) открыли в космических лучах положительные и отрицательные частицы ($\mu^{\pm}$) массой $207m_e$ (106 МэВ), называемые теперь *мюонами* (первоначальное название — *мезоны*). Естественно было предположить, что мюоны и являются частицами Юкавы. Однако это предположение сразу же пришлось отвергнуть, так как мюоны очень слабо взаимодействуют с веществом и поэтому не могут быть переносчиками ядерного взаимодействия. (Время жизни мюона в системе отсчета, где он покоится, $\tau_\mu = 2{,}2 \cdot 10^{-6}$ с, так что релятивистские мюоны, образующиеся в верхних слоях атмосферы, могут доходить и действительно доходят до поверхности Земли.)

Лишь через 10 лет в 1947 г. Пауэлл (1903–1969), Оккиалини (р. 1907) и Латтес (р. 1924) в фотоэмульсиях, облученных космическими лучами в верхних слоях атмосферы, открыли следы заряженных частиц, названных *пионами* ($\pi^{\pm}$), которые распадаются на мюоны и нейтрино. В 1950 г. был открыт и нейтральный пион π^0. Пионы сильно взаимодействуют с веществом (время жизни $\tau_{\pi^{\pm}} = 2{,}6 \cdot 10^{-8}$ с, а $\tau_{\pi^0} = 0{,}83 \cdot 10^{-16}$ с). Поэтому пионы удовлетворяют требованиям, предъявляемым к частицам, переносящим ядерное взаимодействие. До недавнего времени считали, что пионы осуществляют сильные взаимодействия. Сейчас эта роль отводится глюонам.

Как было указано выше, слабые взаимодействия переносятся $W^{\pm}$- и Z^0-промежуточными бозонами. По экспериментальным данным массы $W^{\pm}$- и Z^0-бозонов соответственно приближенно равны 80 и 90 ГэВ. Эти данные с помощью формулы (63.4) позволяют оценить радиус действия R слабых сил. Для оценки возьмем $W^{\pm}$-бозоны, поскольку им соответствуют меньшая масса и, следовательно, больший радиус действия. Полагая $mc^2 = 80$ ГэВ, $\hbar c = 1{,}973 \cdot 10^{-14}$ ГэВ $\cdot$ см, получим

$$R = \frac{\hbar c}{mc^2} = 2{,}5 \cdot 10^{-16} \text{ см}.$$

14. Со времени возникновения *кварковой модели* принято считать, что основное взаимодействие между нуклонами осуществляется не мюонами, а сводится к взаимодействию кварков. Кварки — это гипотетические частицы, из которых состоят адроны, в том числе протоны и нейтроны. Предполагается, что кварки несут дробные заряды $\pm(2/3)e$ и $\pm(1/3)e$ и обладают спином $1/2$. Всего в настоящее время различают шесть кварков (и соответствующих им антикварков), которые группируются в три дублета. Взаимодействие между кварками осуществляется путем обмена безмассовыми частицами со спином 1 — *глюонами*. Кварки в свободном состоянии не удалось наблюдать в природе, в космических лучах и получить на ускорителях. Это привело к

гипотезе пленения кварков, согласно которой они существуют только внутри адронов и не могут существовать в свободном состоянии. Это объясняется тем, что силы взаимодействия между кварками убывают с уменьшением расстояния между ними и растут с увеличением расстояния. Благодаря этому на малых расстояниях кварки внутри адронов ведут себя как свободные частицы, называемые *партонами*. Гипотеза кварков, поскольку она объяснила множество фактов в поведении элементарных частиц и предсказала существование новых, в настоящее время считается общепринятой (см. гл. XVI).

§ 64. Энергия связи ядра

1. *Энергия связи ядра (относительно всех нуклонов) $\mathscr{E}_{\text{св}}$ есть мера его прочности, измеряемая минимальной работой, которую надо произвести, чтобы полностью расщепить ядро на составляющие его протоны и нейтроны.* Энергию связи ядра надо отличать от его *внутренней энергии*, т. е. от *энергии* образования ядра $\mathscr{E}_{\text{об}}$. Если энергию полностью расщепленного ядра принять за нуль, то, очевидно, $\mathscr{E}_{\text{об}} = -\mathscr{E}_{\text{св}}$. Через величину $\mathscr{E}_{\text{св}}$ определится и *энергия связи ядра по отношению к разделению его на любые две части, состоящие из протонов и нейтронов, т. е. минимальная работа, необходимая для разделения ядра на эти две части.* Например, энергия связи протона в ядре, иначе называемая *энергией отделения протона* от ядра, есть минимальная работа, которую надо произвести, чтобы удалить протон из ядра. Она определяется формулой

$$\mathscr{E}_{\text{р}} = \mathscr{E}_{\text{св}}(Z, A) - \mathscr{E}_{\text{св}}(Z-1, A-1), \tag{64.1}$$

т. е. равна разности энергий связи исходного и конечного ядра. Аналогично энергия связи нейтрона в ядре (иначе, *энергия отделения нейтрона от ядра*)

$$\mathscr{E}_{\text{n}} = \mathscr{E}_{\text{св}}(Z, A) - \mathscr{E}_{\text{св}}(Z, A-1), \tag{64.2}$$

а энергия связи α-частицы в ядре (или энергия отделения ее)

$$\mathscr{E}_{\alpha} = \mathscr{E}_{\text{св}}(Z, A) - \mathscr{E}_{\text{св}}(Z-2, A-4) - \mathscr{E}_{\text{св}}(\alpha), \tag{64.3}$$

где $\mathscr{E}_{\text{св}}(\alpha)$ — энергия связи α-частицы.

Понятно, что исходное ядро в этих формулах не может быть соответственно протоном, нейтроном или α-частицей. Иначе $\mathscr{E}_{\text{р}}$, например, означала бы энергию отделения протона от ядра, которое само состоит только из одного протона, а такая постановка вопроса лишена смысла.

В силу соотношения между массой и энергией энергия связи ядра может быть вычислена по формуле

$$\mathscr{E}_{\text{св}}(Z, A) = Z M_{\text{р}} + N M_{\text{n}} - M(Z, A), \tag{64.4}$$

если массы выражены в энергетических единицах. Предполагается, что массы всех частиц в формуле (64.4) — *массы покоя* (индекс нуль опущен, как это принято в ядерной физике и физике элементарных

частиц). Массу заряженной частицы можно измерить масс-спектрографическим методом, основанным на измерении отклонений заряженных частиц в статических магнитных и электрических полях. Если же частица не заряжена (например, нейтрон), то измерение ее массы может быть сведено к измерению масс заряженных частиц.

2. Существенно заметить, что в таблицах приводятся не массы ядер, а величины, выражающиеся через массы *нейтральных атомов*. Поэтому для удобства вычислений формулу (64.4) целесообразно преобразовать так, чтобы в нее входили массы атомов, а не ядер. С этой целью в правой части формулы (64.4) прибавим и вычтем массу Z электронов. Пренебрежем далее разницей энергий связи этих электронов в Z атомах водорода, с одной стороны, и в атоме (Z, A) — с другой (такая разница только за последнее время стала доступной измерениям). Тогда формула (64.4) перейдет в

$$\mathscr{E}_{\text{св}}(Z, A) = Z M_{\text{ат}}({}^1_1H) + N M_{\text{n}} - M_{\text{ат}}(Z, A), \tag{64.4а}$$

где $M_{\text{ат}}({}^1_1H)$ — масса атома водорода, а $M_{\text{ат}}(Z, A)$ — масса атома с порядковым номером Z и массовым числом A.

Полезным понятием в ядерной физике является *дефект массы ядра*, связанный с его энергией связи. Дефектом массы ядра называется разность между массой рассматриваемого ядра, выраженной в атомных единицах массы (см. § 63, п. 3), и соответствующим массовым числом A:

$$\Delta(Z, A) = M_{\text{яд}}(Z, A) - A. \tag{64.5}$$

Для установления зависимости между дефектом массы и энергией связи ядра используем формулу (64.4), считая, что вся масса в ней выражена в атомных единицах массы (а. е. м.). Далее, учтем, что из формулы (64.5) следует, что $M_{\text{яд}} = \Delta + A$. В частности, для нейтрона $M_{\text{n}} = \Delta_{\text{n}} + 1$, а для протона $M_{\text{p}} = \Delta_{\text{p}} + 1$. Подставив эти значения в (64.4), получим

$$\mathscr{E}_{\text{св}} = Z(\Delta_{\text{p}} + 1) + N(\Delta_{\text{n}} + 1) - (\Delta + A),$$

или

$$\mathscr{E}_{\text{св}} = Z\Delta_{\text{p}} + N\Delta_{\text{n}} - \Delta(Z, A), \tag{64.4б}$$

так как $Z + N = A$. Отсюда видно, что при надлежащем сдвиге начала отсчета энергии (зависящем только от Z и N) дефект массы отличается от энергии связи ядра только знаком. Применим (64.4б) к расчету энергии связи ядра атома ${}^4_2\text{He}$.

Масса протона $M_{\text{p}} = 938{,}2796$ МэВ $= 1{,}0072764$ а. е. м., масса нейтрона $M_{\text{n}} = 939{,}5731$ МэВ $= 1{,}008665$, масса α-частицы (ядра ${}^4\text{He}$) $M_\alpha = 4{,}001506$ а. е. м. Следовательно, для соответствующих дефектов масс получаем $\Delta_{\text{p}} = 0{,}007276$, $\Delta_{\text{n}} = 0{,}008665$, $\Delta_\alpha = 0{,}001506$, а для энергии связи α-частицы $\mathscr{E}_{\text{св}} = 2(0{,}007276 + 0{,}008665) - 0{,}001506 = 0{,}030$ а. е. м. $= 28{,}38$ МэВ.

Дефект массы, определяемый формулой (64.5), есть величина безразмерная. Но ему искусственно можно приписать размерность массы (энергии), если условиться, что формула (64.5) определяет Δ только

в атомных единицах массы. После этого простым пересчетом определится значение Δ в мегаэлектронвольтах (или в других единицах массы). В результате получится, например, $\Delta_p = 6{,}77761$ МэВ, $\Delta_n = 8{,}07146$ МэВ, $\Delta_\alpha = 1{,}4028414$ МэВ.

Как уже отмечалось выше, в таблицах обычно приводятся не массы ядер, а массы нейтральных атомов. Последние больше масс ядер на массы электронных оболочек. В соответствии с этим вместо дефектов масс ядер приводятся дефекты масс также *нейтральных атомов*, т. е. величины

$$\delta(Z, A) = M_{\text{ат}}(Z, A) - A. \tag{64.5а}$$

Например, дефект массы атома ^4_2He получится, если к дефекту массы α-частицы добавить массу двух электронов: $2 \cdot 0{,}511003 = 1{,}022006$ МэВ. Таким путем для дефекта массы атома ^4_2He получится $1{,}4028414 + 1{,}022006 = 2{,}42485$ МэВ. Очевидно, формула (64.4б) остается справедливой, если дефекты масс ядер заменить на дефекты масс нейтральных атомов, т. е.

$$\mathscr{E}_{\text{св}} = Z\delta_p + N\delta_n - \delta(Z, A). \tag{64.4в}$$

Интересно сравнить энергию связи α-частицы с относительным изменением массы вещества при химических реакциях. Например, в реакции $H_2 + O \to H_2O$ достигается температура порядка 1000 °C $=$ $= 1273$ K, что соответствует средней энергии $3 \cdot (3/2)kT = (9/2) \times$ $\times 1{,}38 \cdot 10^{-16} \cdot 1273 \approx 7{,}9 \cdot 10^{-13}$ эрг $\approx 0{,}5$ эВ на одну молекулу воды (молекула воды состоит из трех атомов: двух атомов водорода и одного атома кислорода). Так как молекула воды содержит 18 нуклонов, а масса нуклона равна 931 МэВ, то собственная энергия молекулы воды равна $18 \cdot 931 = 16{,}8 \cdot 10^3$ МэВ $= 16{,}8 \cdot 10^9$ эВ. Относительное изменение массы вещества при этой химической реакции составляет примерно $3 \cdot 10^{-11} = 3 \cdot 10^{-9}\,\%$, что недоступно измерению даже масс-спектрографическими методами. Этот пример делает понятным, почему экспериментальная проверка формулы Эйнштейна $\Delta m = \Delta\mathscr{E}/c^2$ на химических реакциях оказалась безнадежной, а стала возможной только на ядерных реакциях.

Отношение $\mathscr{E}_{\text{св}}/A$, т. е. средняя энергия связи, приходящаяся на один нуклон, называется *удельной энергией связи*, а дефект массы, отнесенный к одному нуклону, $f = \Delta/A$, — *упаковочным коэффициентом*.

3. Отметим некоторые свойства атомных ядер, установленные экспериментально, и посмотрим, какие выводы следует сделать из них относительно строения ядра. Оказывается, что для стабильных не слишком легких ядер ($A \gtrsim 20$) удельная энергия связи в грубом приближении постоянна и составляет около 8 МэВ на нуклон. Этот факт определенно свидетельствует о том, что ядерные силы являются *короткодействующими* — их радиус действия порядка размеров самих нуклонов и даже меньше. Такая особенность ядерных сил проявляется в их *насыщении*. Насыщение означает, что каждый нуклон в ядре взаимодействует только с несколькими соседними нуклонами. В этом

отношении ядерные силы аналогичны химическим силам, обусловливающим валентность химических элементов.

Насыщением ядерных сил объясняется, почему энергия связи не слишком легких стабильных ядер в грубом приближении пропорциональна массовому числу A. Если бы насыщения не было, а каждый нуклон эффективно взаимодействовал с остальными $A-1$ нуклонами, то энергия связи ядра оказалась бы пропорциональной $A(A-1)/2$, т. е. при больших A возрастала приблизительно пропорционально второй, а не первой степени A. С насыщением ядерных сил связано и то обстоятельство, что плотность ядерного вещества для не слишком легких ядер приблизительно постоянна, т. е. не зависит от A. Благодаря этому радиус ядра R оказывается пропорциональным $A^{1/3}$. Это дает основание рассматривать атомное ядро как *каплю несжимаемой жидкости*, заряженной положительным электричеством; такое предположение вводится в так называемой *капельной модели ядра.* Такая классическая модель представляется наиболее обоснованной для ядер с большими массовыми числами A.

4. С точки зрения капельной модели следует ожидать уменьшения удельной энергии связи ядра по сравнению с той, какой она была бы, если бы нуклоны в ядре подвергались действию только *насыщенных* ядерных сил притяжения. Такое уменьшение действительно наблюдается. Оно связано, во-первых, с влиянием поверхностных нуклонов. Если нуклон находится вблизи поверхности ядра, то уменьшается число нуклонов, удерживающих его в ядре, т. е. не все связи в этом случае будут действовать полностью. Этот эффект особенно существен для легких ядер, так как его влияние тем больше, чем больше отношение поверхности ядра к его объему (пропорциональное $R^{-1} \sim A^{-1/3}$). Удельные энергии связи для легких ядер, вычисленные по формуле (64.4а), приведены в табл. 7. Разумеется, для легких ядер нет оснований ожидать согласия экспериментальных данных с вычисленными по капельной модели.

Таблица 7

Ядро	${}_1^2\mathrm{H}$	${}_2^3\mathrm{He}$	${}_2^4\mathrm{He}$	${}_3^6\mathrm{Li}$	${}_3^7\mathrm{Li}$	${}_7^{14}\mathrm{N}$	${}_8^{16}\mathrm{O}$	${}_9^{19}\mathrm{F}$
$\mathscr{E}_{\text{св}}/A$, МэВ	1,112	2,573	7,074	5,332	5,606	7,476	7,976	7,632

Во-вторых, энергия связи уменьшается из-за кулоновского отталкивания между протонами. Для легких ядер этот эффект не играет существенной роли, поскольку в этом случае ядерные силы превосходят кулоновские примерно на два порядка. Однако кулоновские силы являются дальнодействующими, их энергия пропорциональна $Z(Z-1)$, или при больших Z приблизительно пропорциональна Z^2, тогда как энергия ядерного взаимодействия пропорциональна Z. Поэтому при больших Z роль кулоновской энергии увеличивается. Этим объясняется уменьшение удельной энергии связи тяжелых ядер с возрастанием Z.

5. Помимо объемной, поверхностной и кулоновской энергий, энергия связи ядра содержит еще два члена. Первый учитывает установлен-

ную на опыте для легких ядер повышенную стабильность ядра с равными числами протонов и нейтронов ($Z = N$) по сравнению с другими ядрами того же массового числа $A = Z + N$. Нарушение такой стабильности для тяжелых ядер связано с кулоновским отталкиванием протонов, которое уже было учтено выше. Поэтому мы примем в качестве обобщения опытных фактов, что указанная тенденция к стабильности сохранится и для тяжелых ядер, если «выключено» электрическое взаимодействие. При отклонении чисел Z и N от равенства в энергии связи ядра, помимо поверхностной и кулоновской энергий, появится еще одна отрицательная добавка, называемая *энергией симметрии.* Физическая природа энергии симметрии не ясна. Однако ее наличие, несомненно, свидетельствует о том, что протон отличается от нейтрона не только электрическим зарядом. Между ними имеются и другие, хотя и менее существенные различия. Оставляя в стороне вопрос о физической природе энергии симметрии, поставим более скромную задачу: на основе эмпирических фактов получить приближенное выражение для энергии симметрии.

Допустим, что в ядре «выключено» электромагнитное взаимодействие, а осталось только ядерное взаимодействие. Если теперь в ядре заменить все протоны нейтронами, а все нейтроны — протонами, то от этого энергия связи не изменится. Это утверждение является обобщением опытных фактов и подтверждается, в частности, существованием зеркальных ядер. Математически оно выражается уравнением

$$\mathscr{E}_{\text{св}}(Z, N) = \mathscr{E}_{\text{св}}(N, Z).$$

Введем в это уравнение новую переменную $\varepsilon = N - Z$. С использованием соотношения $A = Z + N$ находим

$$N = \frac{A + \varepsilon}{2}, \quad Z = \frac{A - \varepsilon}{2},$$

так что

$$\mathscr{E}_{\text{св}}\left(\frac{A - \varepsilon}{2}, \frac{A + \varepsilon}{2}\right) = \mathscr{E}_{\text{св}}\left(\frac{A + \varepsilon}{2}, \frac{A - \varepsilon}{2}\right).$$

Отсюда видно, что при постоянном A энергия связи ядра является четной функцией параметра ε. Считая величину ε малой по сравнению с A, разложим функцию $\mathscr{E}_{\text{св}}$ по степеням ε и оборвем это разложение на квадратичном члене:

$$\mathscr{E}_{\text{св}} = F(A) + f(A)\varepsilon^2 = F(A) + f(A)(N - Z)^2.$$

Опытные факты вынуждают признать, что функция $F(A) \infty A$, о чем уже было сказано выше, а $f(A) \infty A^{-1}$, причем функция $f(A)$ должна быть отрицательной, о чем также было сказано выше.

6. Второй из упомянутых в начале п. 5 членов не может быть истолкован классически и учитывает экспериментально установленный факт систематического изменения энергии связи ядра в зависимости от того, четны или нечетны Z и A. Ядра с четными Z и N называются *четно-четными*, с четными Z и нечетными N — *четно-нечетными*,

с нечетными Z и четными N — *нечетно-четными*, с нечетными Z и нечетными N — *нечетно-нечетными*. Энергия связи максимальна для четно-четных ядер, минимальна для нечетно-нечетных и принимает промежуточные значения для остальных ядер. Этот факт с определенностью свидетельствует о *спаривании* одинаковых нуклонов в ядре, т. е. в каком-то смысле объединении в пары как протонов, так и нейтронов. Спаривание увеличивает энергию связи ядра. Соответствующая поправка в энергии связи называется *энергией спаривания*. При четных Z и N все протоны и все нейтроны ядра спарены. При нечетном A остается один неспаренный протон или один неспаренный нейтрон. Наконец, при нечетных Z и N получится один неспаренный протон и одни неспаренный нейтрон.

Экспериментальные факты удовлетворительно описываются, если при нечетном A энергию спаривания включить в объемный член, т. е. принять ее равной пулю. Тогда для четно-четных ядер энергия спаривания будет положительна, а для нечетно-нечетных отрицательна, причем по абсолютной величине обе энергии практически одинаковы.

7. Таким образом, на основании сказанного для энергии связи ядра можно написать

$$\mathscr{E}_{\text{св}} = C_{\text{об}}A - C_{\text{пов}}A^{2/3} - C_{\text{кул}}Z^2A^{-1/3} - C_{\text{сим}}(A-2Z)^2A^{-1} + C_{\text{спар}}A^{-\varepsilon}\delta. \tag{64.6}$$

Эта полуэмпирическая формула называется *формулой Вейцзеккера* (р. 1912). Последний член установлен на основании эмпирических данных, причем для показателя ε разные авторы приводят различные значения от $+1/3$ до 1. В настоящей книге принимается $\varepsilon = 3/4$. Значение δ равно

$$\delta = \begin{cases} +1 & \text{для четно-четных ядер,} \\ 0 & \text{для ядер с нечетным } A, \\ -1 & \text{для нечетно-нечетных ядер.} \end{cases}$$

Коэффициенты в формуле (64.6) подбираются так, чтобы получилось наилучшее согласие с опытом. В настоящее время приняты следующие значения:

$$\begin{gathered} C_{\text{об}} = 15{,}75 \text{ МэВ}, \quad C_{\text{пов}} = 17{,}8 \text{ МэВ}, \quad C_{\text{кул}} = 0{,}710 \text{ МэВ}, \\ C_{\text{сим}} = 23{,}7 \text{ МэВ}, \quad C_{\text{спар}} = 34 \text{ МэВ}, \quad \varepsilon = 3/4. \end{gathered} \tag{64.7}$$

Формула Вейцзеккера для энергии связи в большинстве случаев справедлива с точностью до нескольких мегаэлектронвольт и чрезвычайно полезна при выяснении всех существенных общих свойств ядер (легкие ядра исключаются из рассмотрения). Однако некоторые детали не отражаются этой формулой должным образом. Сюда относятся, например, особая устойчивость «магических» ядер и флуктуации энергии спаривания.

Магическими называются ядра, у которых число протонов или нейтронов равно одному из чисел 2, 8, 20, (28), 50, 82, 126 (в последнем

случае только для нейтронов). Сами эти числа называются также *магическими.* Если у ядра одновременно являются магическими как число протонов, так и число нейтронов, то такое ядро называется *дважды магическим.* Таких ядер всего пять: $^{4}_{2}\mathrm{He}$, $^{16}_{8}\mathrm{O}$, $^{40}_{20}\mathrm{Ca}$, $^{48}_{20}\mathrm{Ca}$, $^{208}_{82}\mathrm{Pb}$. Магические и в особенности дважды магические ядра отличаются *повышенной устойчивостью* (т. е. обладают большими удельными энергиями связи) и большей распространенностью в природе по сравнению с другими ядрами. Существование магических чисел объясняется оболочечной моделью ядра (см. § 78).

8. Применим формулу Вейцзеккера для определения наиболее стабильного изобара при заданном массовом числе A. Изобары отличаются друг от друга значениями Z. Поэтому задача сводится к определению зарядового числа Z, при котором энергия связи ядра максимальна. Продифференцируем (64.6) по Z при постоянном A и приравняем производную нулю. Разумеется при этом достаточно принять во внимание только третий и четвертый члены формулы (64.6), так как остальные члены от Z не зависят. В результате получим

$$Z = \frac{A}{2 + (C_{\text{кул}}/2C_{\text{сим}})\, A^{2/3}} = \frac{A}{2 + 0{,}0150\, A^{2/3}}. \tag{64.8}$$

Формула Вейцзеккера не учитывает различия масс нейтрона и протона: $m_{\text{n}} - m_{\text{p}} = 1{,}29343$ МэВ. Действительно, масса ядра должна содержать член $Zm_{\text{p}} + (A - Z)m_{\text{n}} = Am_{\text{n}} - Z(m_{\text{n}} - m_{\text{p}})$. Поэтому в формулу для энергии связи должно входить слагаемое $Z(m_{\text{n}} - m_{\text{p}})$. С учетом этого слагаемого получится

$$Z = \frac{[1 + (m_{\text{n}} - m_{\text{p}})/4C_{\text{сим}}]\, A}{2 + (C_{\text{кул}}/2C_{\text{сим}})\, A^{2/3}} = \frac{A}{1{,}97 + 0{,}0150\, A^{2/3}}, \tag{64.8а}$$

что отличается от (64.8) примерно на 1%. Такое различие вряд ли реально ощутимо при той точности, на которую может претендовать полуэмпирическая формула Вейцзеккера.

Ядра, не испытывающие β-распада, называются *β-стабильными ядрами.* Числа нейтронов N и протонов Z в них определяются формулами (64.8) или (64.8а). Эти формулы дают только средние или сглаженные значения N и Z для β-стабильных ядер. На плавный ход изменения, соответствующего формулам (64.8) и (64.8а), накладывается ряд *локальных искажений.* Для $A \lesssim 40$ число Z примерно вдвое меньше A, т. е. числа нейтронов и протонов в ядре примерно равны. При больших A из-за кулоновского отталкивания в ядре содержится больше нейтронов, чем протонов.

На рис. 120 на осях координат отложены числа N и Z. Здесь известные β-стабильные ядра изображены прямоугольниками в функции N и Z. Темные квадратики относятся к ядрам, полупериод α-распада которых больше 10^9 лет, а светлые — к ядрам, у которых этот полупериод меньше. Ядра с избыточным числом нейтронов или протонов β-радиоактивны. От избытка протонов ядро освобождается

путем испускания *позитронов*, а от избытка нейтронов — путем испускания *электронов* (см. § 74). Когда избыток протонов становится столь большим, что энергия отделения протона обращается в нуль, то ядро существовать не может и распадается. Аналогичное заключение относится к нейтронам.

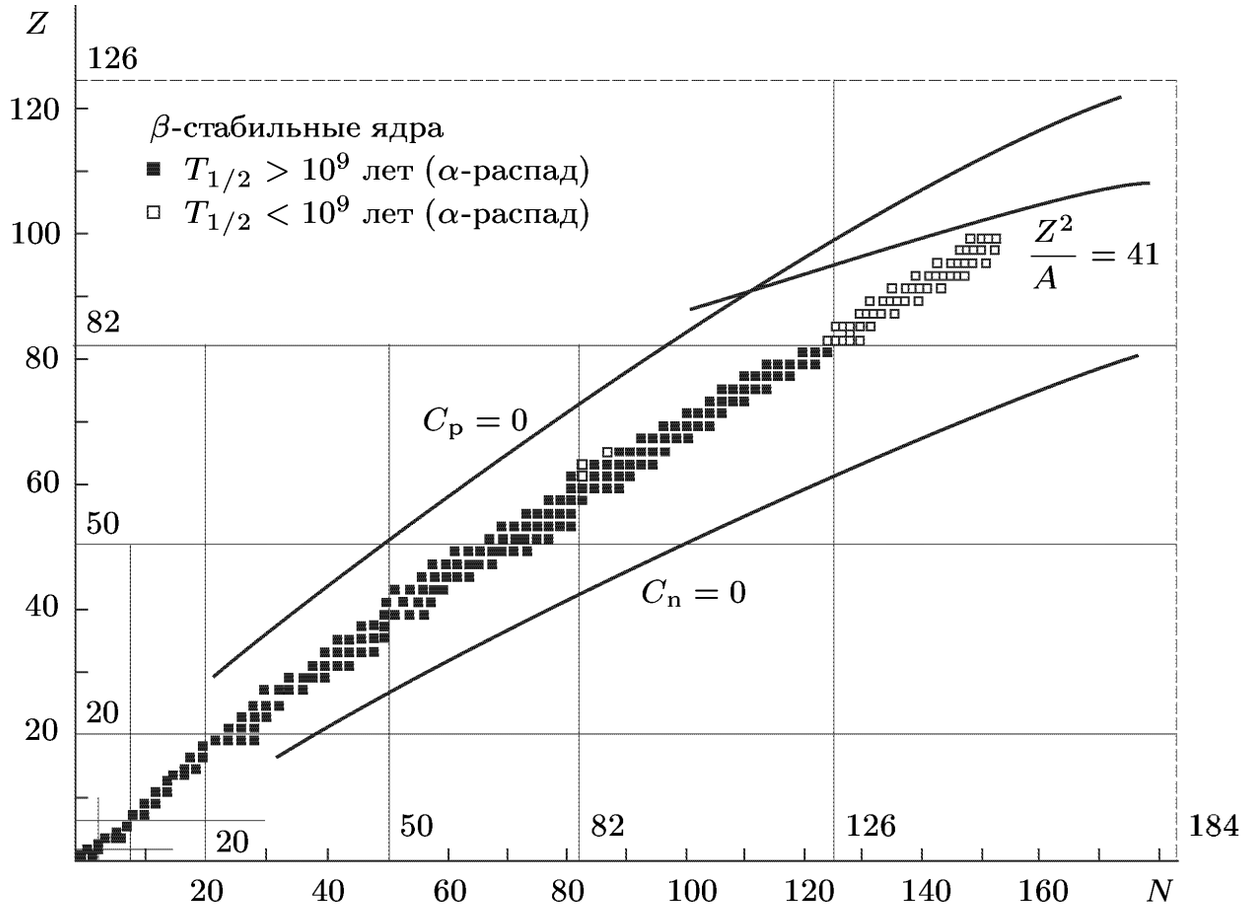


Рис. 120

Соответствующие теоретические границы существования ядер изображены кривыми C_p и C_n на рис. 120. Их можно приближенно получить, приравнивая нулю частные производные по Z (при $N =$ $=$ const) и по N (при $Z =$ const). О кривой $Z^2/A = 41$ будет сказано при рассмотрении вопроса о делении ядра (см. § 93).

9. Зависимость энергии связи ядра от числа нейтронов N и числа протонов Z можно изобразить графически, откладывая на горизонтальных осях значения N и Z, а на вертикальной оси энергии связи $\mathscr{E}_{\text{св}}$. Так как числа N и Z целые, то таким путем получится конечная система точек, каждая из которых соответствует определенному ядру. Если их соединить поверхностью, та она отнюдь не будет гладкой, а получится зубчатой. Положение здесь напоминает то, что получилось бы, если бы подобным образом изобразить зависимость энергии ионизации атома от порядкового номера элемента. Энергия ионизации менялась бы скачкообразно при переходе от одного элемента к соседнему. Особенно велики были бы эти скачки при переходах от соседних атомов к атомам благородных газов, у которых полностью застроены наружные электронные оболочки и именно благодаря этому энергии

ионизации особенно велики. Аналогичное увеличение удельной энергии связи наблюдается и в ядерной физике при переходе к *магическим ядрам.* Этот эффект связан с оболочечной структурой ядра (см. § 78).

Но особенно сильно удельная энергия связи меняется при изменении *четности ядра.* При переходе же от ядер к соседним ядрам той же четности скачки удельной энергии связи относительно меньше. Именно в этом проявляется *энергия спаривания.* Благодаря наличию энергии спаривания поверхность $\mathscr{E}_{\text{св}} = \mathscr{E}_{\text{св}}(N, Z)$ отчетливо расщепляется на три поверхности. Выше всех располагается поверхность для четно-четных ядер, ниже всех — для нечетно-нечетных. Посередине между ними располагается поверхность с нечетными числами A, соответствующая четно-нечетным и нечетно-четным ядрам. Все три поверхности можно аппроксимировать гладкими поверхностями, используя для этого, например, формулу Вейцзеккера. Расстояние между этими поверхностями при $Z \approx 10$–20 и $N \approx 10$–20 составляет примерно 3–2 МэВ, а затем монотонно убывает до 1 МэВ в области самых тяжелых ядер ($Z \approx 100$, $N \approx 150$).

10. На рис. 121 представлена экспериментальная зависимость удельной энергии связи ядра от массового числа A для наиболее стабильных изобаров при всех четных значениях A. Нечетно-нечетных стабильных ядер известно всего пять: это легкие ядра ${}^{2}_{1}\text{D}$, ${}^{6}_{3}\text{Li}$, ${}^{10}_{5}\text{Be}$, ${}^{14}_{7}\text{N}$, а также ${}^{50}_{23}\text{V}$. Они на рис. 121 не представлены. Не представлены и ядра с нечетными значениями A. Тем самым исключены систематические нечетно-четные поправки, связанные с эффектом спаривания нуклонов. Плавная кривая соответствует полуэмпирической формуле Вейцзеккера (64.6). Если исключить из рассмотрения самые легкие ядра, то в грубом приближении, как уже указывалось в п. 3, удельная энергия связи слабо меняется при переходе от ядра к ядру и равна приблизительно 8 МэВ. Удельная энергия связи обращается в максимум при $A \approx 56$ (железо). Этот максимум равен приблизительно 8,8 МэВ. Замедление роста удельной энергии связи с последующим падением ее, как уже указывалось выше (см. п. 3), сначала связано с влиянием поверхностной энергии, а затем с кулоновским отталкиванием. Из рассмотрения кривой видно, что для легких ядер энергетически возможен процесс *слияния* их с выделением ядерной энергии синтеза. Напротив, для тяжелых ядер энергетически выгоден процесс *деления*, сопровождающийся также выделением энергии. На этих процессах основана вся ядерная энергетика.

ЗАДАЧИ

1. Дефект массы атома ${}^{9}_{4}\text{Be}$ равен 11,3484 МэВ. Определить энергию связи ядра ${}^{9}_{4}\text{Be}$ относительно распада его на две α-частицы ${}^{4}_{2}\text{He}$ и нейтрон.

Решение. Воспользовавшись значениями дефектов масс нейтрона и атома ${}^{2}_{4}\text{He}$, приведенными в тексте, получим для искомой энергии связи

$$\mathscr{E}_{\text{св}} = 2\Delta(2,4) + \Delta_{\text{n}} - \Delta(4,9) = 2 \cdot 2{,}42485 + 8{,}07162 - 11{,}3484 = 1{,}5729 \text{ МэВ}.$$

Столь малая энергия связи ядра $^{9}_{4}$Be относительно распада его на две α-частицы и нейтрон позволяет выбивать из этого ядра нейтрон путем облучения его α-частицами. Именно таким путем был открыт нейтрон (см. § 92).

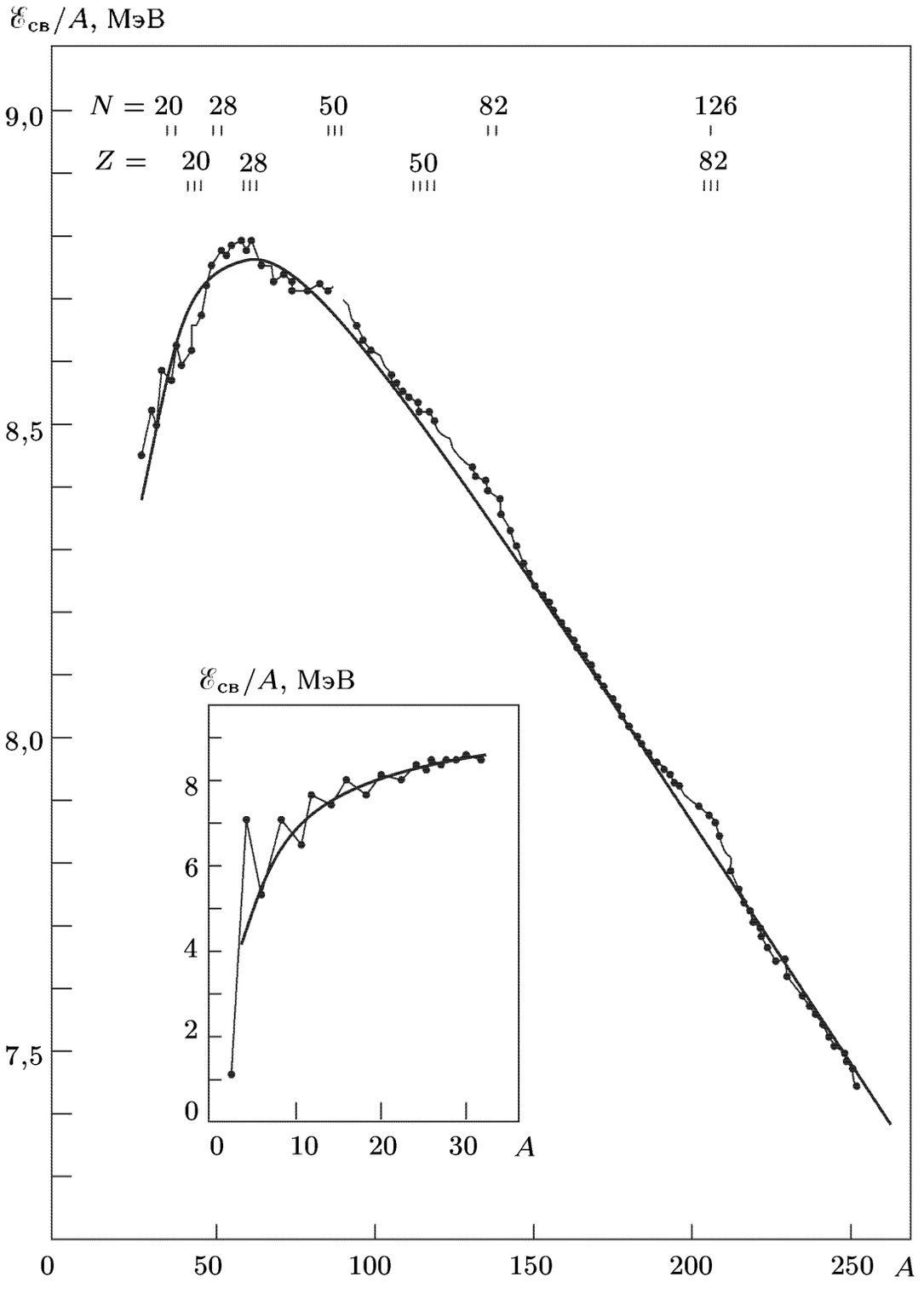


Рис. 121

2. Используя дефекты масс

$$\Delta(1,2) = 13{,}13627 \text{ МэВ}, \quad \Delta(1,3) = 14{,}95038 \text{ МэВ},$$

$$\Delta(2,3) = 14{,}93173 \text{ МэВ},$$

а также дефекты масс, приведенные в тексте, вычислить энергию Q, выделяющуюся в следующих термоядерных реакциях:

$$1)\ {}^2_1\mathrm{D} + {}^2_1\mathrm{D} \to {}^1_1\mathrm{p} + {}^3_1\mathrm{T}, \qquad 3)\ {}^2_1\mathrm{D} + {}^3_1\mathrm{T} \to {}^1_0\mathrm{n} + {}^4_2\mathrm{He},$$

$$2)\ {}^2_1\mathrm{D} + {}^2_1\mathrm{D} \to {}^1_0\mathrm{n} + {}^3_2\mathrm{He}, \qquad 4)\ {}^3_2\mathrm{He} + {}^2_1\mathrm{D} \to {}^4_2\mathrm{He} + {}^1_1\mathrm{p}.$$

Пренебрегая кинетической энергией частиц до реакции, определить, какую энергию уносит каждая частица после реакции.

Ответ.

1) $Q = 2\Delta(1,2) - \Delta(1,1) - \Delta(1,3) = 4{,}033$ МэВ; протон уносит 3,025 МэВ, ядро трития — 1,008 МэВ;

2) $Q = 3{,}27$ МэВ; нейтрон уносит 2,453 МэВ, ядро гелия ${}^3_2\mathrm{He}$ — 0,817 МэВ;

2) $Q = 17{,}59$ МэВ; нейтрон уносит 14,07 МэВ, ядро гелия ${}^4_2\mathrm{He}$ — 3,52 МэВ;

4) $Q = 18{,}35$ МэВ; протон уносит 14,68 МэВ, ядро гелия ${}^4_2\mathrm{He}$ — 3,67 МэВ.

3. Ядро урана ${}^{238}_{92}\mathrm{U}$ делится на два осколка приблизительно одинаковой массы, расположенные в середине периодической системы элементов. Пользуясь кривой рис. 121, определить приближенно освободившуюся при этом кинетическую энергию.

Решение. В результате деления полное число нуклонов 238 остается неизменным. Как видно из рис. 121, средняя энергия нуклона $\mathscr{E}_{\text{об}}/A = -\mathscr{E}_{\text{св}}/A$ до деления равна $-7{,}6$ МэВ, а после деления $-8{,}5$ МэВ. При делении освобождается кинетическая энергия $-238 \cdot 7{,}6 - (-238 \cdot 8{,}5) \approx 200$ МэВ.

4. Ядро урана ${}^{238}_{92}\mathrm{U}$ делится на два одинаковых осколка. Пользуясь формулой Вейцзеккера (64.6), вычислить суммарную кинетическую энергию, которую получили бы оба осколка, если бы между ними действовали только кулоновские силы отталкивания.

Ответ.

$$\mathscr{E}_{\text{кин}} = C_{\text{кул}}\left[Z^2A^{-1/3} - 2\left(\frac{Z}{2}\right)^2\left(\frac{A}{2}\right)^{-1/3}\right] = C_{\text{кул}}Z^2A^{-1/3}(1 - 2^{-2/3}) =$$

$= 0{,}370\, C_{\text{кул}}Z^2A^{-1/3} \approx 360$ МэВ. Кинетическая энергия получилась больше, чем в предыдущей задаче. Это связано с тем, что все прочие силы (за исключением кулоновских) при распаде ядра на осколки удерживают их, т. е. производят отрицательную работу. На полученное значение надо смотреть как на грубо ориентировочное, поскольку формула Вейцзеккера не обоснована теоретически, а ее коэффициенты подобраны на основе экспериментальных данных.

5. Показать, что изотопы ${}^5\mathrm{Li}$ и ${}^8\mathrm{Be}$ нестабильны. С их нестабильностью связано отсутствие в природе стабильных изотопов с массовыми числами 5 и 8. Дефекты масс атомов ${}^5\mathrm{Li}$ и ${}^8\mathrm{Be}$ равны соответственно 11,680 и 4,9418 МэВ.

Указание. Для ядра ${}^5\mathrm{Li}$ рассмотреть процесс ${}^5_3\mathrm{Li} \to {}^4_2\mathrm{He} + \mathrm{p}$, а для ядра ${}^8\mathrm{Be}$ — процесс ${}^8_4\mathrm{Be} \to 2\,{}^4_2\mathrm{He}$.

§ 65. Размеры ядра

1. О размерах ядра нельзя говорить с той же определенностью и однозначностью, как это делается в случае макроскопических тел. Наибольшей определенностью характеризуются размеры тяжелых ядер.

Различные методы определения размеров ядер можно разделить на две группы. В одних методах регистрируется наличие ядерного вещества — в них используются явления, обусловленные ядерными

силами (или так называемыми сильными взаимодействиями). В других используются электромагнитные взаимодействия и исследуется распределение электрического заряда в ядре. Обе группы методов приводят к несколько различным результатам. В точных исследованиях необходимо указывать, в каком смысле употребляется понятие размера ядра и какими методами были определены эти размеры. Однако различия между результатами измерений размеров ядра разными методами не так велики. Когда не требуется особая точность, можно не вдаваться в подробности и говорить о «размерах ядра» вообще, не уточняя, о какой величине идет речь.

Если ядро считать сферическим, то все методы определения его радиуса приводят к формуле

$$R = r_0 A^{1/3}. \tag{65.1}$$

Для постоянной r_0 для тяжелых ядер различными методами получаются несколько отличающиеся результаты, но все они лежат в пределах

$$r_0 = (1{,}2 - 1{,}5) \cdot 10^{-13} \text{ см}. \tag{65.2}$$

Заметим, что за единицу расстояний в ядерной физике и физике элементарных частиц удобно принимать *ферми*, равный 10^{-13} см, а за единицу эффективного сечения *барн* (10^{-24} см^2).

Характерная скорость α-частиц, испускаемых радиоактивными ядрами, порядка 10^9 см/с. Время, в течение которого α-частица пролетает диаметр ядра, порядка $T_{\text{яд}} \approx 10^{-13} : 10^9 \approx 10^{-22}$ с. Время порядка 10^{-23}–10^{-24} с принято называть *ядерным временем.*

Ниже рассматриваются некоторые методы определения R и r_0.

2. Верхний предел радиуса ядра можно грубо определить уже из опытов Резерфорда по рассеянию α-частиц на атомных ядрах (см. § 9). Пусть p — импульс α-частицы, m — ее масса, а $\mathscr{E}_{\text{кин}} = p^2/2m$ — кинетическая энергия. Так как при столкновении импульс сохраняется, а ядро до столкновения можно считать неподвижным, то кинетическая энергия после столкновения, связанная с движением центра масс системы, будет $p^2/2(M + m)$, где M — масса ядра. Для тяжелых ядер этой величиной можно пренебречь, т. е. считать, что при упругом столкновении с ядром кинетическая энергия α-частицы не изменяется. В таком случае расстояние R между центрами ядра и частицы, соответствующее максимальному сближению α-частицы с ядром, определится из формулы $\mathscr{E}_{\text{кин}} = 2Ze^2/R$. При численных расчетах величину $\mathscr{E}_{\text{кин}}$ удобно представить в виде $\mathscr{E}_{\text{кин}} = 2eV$, где $2e$ — заряд α-частицы, а V — «ускоряющий потенциал», соответствующий энергии $\mathscr{E}_{\text{кин}}$. Тогда $R = Ze/V$. Для золота $Z = 79$. «Ускоряющий потенциал» α-частицы $V = 5$ МВ $= (5/3) \cdot 10^4$ СГСЭ. В этом случае

$$R = \frac{79 \cdot 4{,}8 \cdot 10^{-10}}{(5/3) \cdot 10^4} = 2{,}3 \cdot 10^{-12} \text{ см}.$$

Поскольку для энергии $\mathscr{E}_{\text{кин}} = 5$ МэВ (и даже несколько большей) результаты опытов хорошо согласуются с теоретической формулой

Резерфорда, отсюда следует, во-первых, что сумма радиусов ядра и α-частицы во всяком случае меньше $2 \cdot 10^{12}$ см, во-вторых, что на расстояниях $2 \cdot 10^{12}$ см взаимодействие между α-частицей и ядром чисто электрическое и *подчиняется закону Кулона.*

3. Радиус ядра можно оценить с помощью полуэмпирической формулы Вейцзеккера (64.6). Третий член этой формулы $-C_{\text{кул}} Z^2 A^{-1/3}$ связан с кулоновским отталкиванием протонов ядра. Если предположить, что электрический заряд ядра равномерно распределен по его объему, то электрическая энергия ядра будет $(3/5)Z^2e^2/R$. Эта величина должна быть равна $C_{\text{кул}} Z^2 A^{-1/3}$. Постоянную $C_{\text{кул}}$ удобно представить в виде $C_{\text{кул}} = eV_{\text{кул}}$, где согласно (64.7) $V_{\text{кул}} = 0{,}71$ МВ $=$ $= 2370$ СГСЭ. Это дает

$$R = \frac{3}{5}\,\frac{eA^{1/3}}{V_{\text{кул}}} = r_0 A^{1/3},$$

где

$$r_0 = \frac{3}{5}\,\frac{e}{V_{\text{кул}}} = \frac{3}{5}\,\frac{4{,}8 \cdot 10^{-10}}{2370} = 1{,}22 \cdot 10^{-13} \text{ см}.$$

Очевидно, этим методом измеряется «электрический радиус» ядра, т. е. радиус, обусловленный взаимодействием электрических зарядов. Надо заметить, что непрерывность и равномерность распределения электрического заряда в ядре, использованную в приведенной оценке, следует рассматривать не как предположение, а как *точное определение* того, что следует понимать под «электрическим радиусом» ядра.

Особым изяществом рассматриваемый метод отличается в применении к двум зеркальным ядрам, из которых одно, испытав β-превращение, переходит в другое. Допустим, например, что это есть β^+-превращение (позитронный распад). Пусть Z и $A - Z$ — числа протонов и нейтронов исходного ядра. Тогда после β^+-превращения оно переходит в зеркальное ядро с $Z - 1$ протонами и $A - Z + 1 = Z$ нейтронами. Из последнего соотношения для исходного ядра получается $A - 2Z = -1$, а для зеркального ядра $A - 2(Z-1) = +1$. Поэтому для обоих зеркальных ядер четвертый член в формуле Вейцзеккера (64.6) будет одним и тем же. Последнее слагаемое в той же формуле в обоих случаях равно нулю, так как при β-превращении массовое число A не меняется, а оно, как мы видели, нечетное. Таким образом, энергии связи обоих зеркальных ядер отличаются только третьим слагаемым. Поэтому разность энергий связи конечного и исходного ядра будет

$$\Delta\mathscr{E}_{\text{св}} = -C_{\text{кул}}(Z-1)^2 A^{-1/3} + C_{\text{кул}} Z^2 A^{-1/3} = C_{\text{кул}} A^{-1/3}(2Z-1) = \\ = C_{\text{кул}} A^{2/3}.$$

Измерив $\Delta\mathscr{E}_{\text{св}}$ и зная A можно найти $C_{\text{кул}}$, а затем вышеописанным способом и радиус ядра R. Разумеется, приведенное рассуждение применимо и к зеркальным ядрам, одно из которых испытывает β^--распад (электронный распад).

4. Размеры атомных ядер можно исследовать, изучая рассеяние на ядрах нейтронов, электронов и других элементарных частиц. Для достаточно заметного рассеяния необходимо, чтобы длина дебройлевской волны λ рассеиваемой частицы была того же порядка или меньше, что и диаметр ядра. Выразим это условие через энергию частицы. Исходной является формула $\lambda = h/p$. Будем считать нейтрон нерелятивистским и воспользуемся формулой $\mathscr{E}_{\text{кин}} = p^2/2m$. Из нее в комбинации с предыдущей формулой получается

$$\mathscr{E}_{\text{кин}} = \frac{h^2}{2m\lambda^2} = \frac{(hc)^2}{2(mc^2)\lambda^2}.$$

Подставив сюда для нейтрона $mc^2 = 939{,}6$ МэВ, а также $hc = 1{,}2399 \times$ $\times 10^{-10}$МэВ $\cdot$ см, получим

$$\mathscr{E}_{\text{кин}} = \frac{8{,}18}{\lambda^2} \cdot 10^{-24} \text{ МэВ}.$$

Для ультрарелятивистских частиц, к которым относятся быстрые электроны, $\mathscr{E} \approx \mathscr{E}_{\text{кин}} = h\nu$, т. е.

$$\mathscr{E}_{\text{кин}} = \frac{hc}{\lambda} = \frac{1{,}2399}{\lambda_{\text{см}}} \cdot 10^{-10} \text{ МэВ}.$$

Если в качестве λ взять $2 \cdot 10^{-12}$ см, то получится для нейтрона $\mathscr{E}_{\text{кин}} \approx 2$ МэВ, а для ультрарелятивистского электрона $\mathscr{E}_{\text{кин}} \approx 60$ МэВ. Таким образом, кинетическая энергия нейтронов должна превосходить 5 МэВ, а электронов — 100 МэВ.

5. Количественное описание производится наглядно с помощью так называемого *эффективного сечения ядра.* Напомним это понятие. Эффективное сечение вводится, в частности, для характеристики ослабления параллельного пучка частиц в результате того или иного процесса. Говорят, например, об эффективном сечении упругого или неупругого рассеяния электрона на атоме, о полном сечении рассеяния электрона на атоме и т. д. Сейчас нас интересует полное ослабление параллельного пучка нейтронов, электронов и других частиц в результате их рассеяния на атомных ядрах. Действие ядра наглядно можно описать так, как если бы оно представляло собой непроницаемую площадку размером σ, перпендикулярную к падающему пучку, которая выводит из пучка падающие на нее частицы. Площадка σ и называется *эффективным сечением* (или просто сечением) ядра. Рассмотрим плоскопараллельный слой толщиной dx и площадью S, перпендикулярный к падающему пучку частиц, равномерно заполненный рассеивающими ядрами. В таком слое содержится $Sn\,dx$ ядер и связанных с ними площадок σ, где n — число ядер в единице объема. Общая площадь таких площадок равна $Sn\sigma\,dx$, причем из-за малости толщины dx площадки можно считать неперекрывающимися. Относительная доля частиц — dN/N, выводимая из пучка при прохождении рассматриваемого слоя, будет

$Sn\sigma\,dx/S = n\sigma\,dx$. Таким образом,

$$\frac{dN}{N} = -n\sigma\,dx, \tag{65.3}$$

и, следовательно,

$$N = N_0 e^{-n\sigma x}. \tag{65.4}$$

Измеряя ослабление интенсивности потока частиц N при рассеянии на ядрах, можно найти эффективное сечение ядра σ.

6. Как же связано полное эффективное сечение σ с размерами ядра в случае падения на него, например, пучка нейтронов? Это, конечно, зависит от энергии нейтронов и от строения ядра. Простейшей является модель непрозрачного ядра. Для ее применимости необходимо, чтобы энергия нейтронов была не особенно велика. В противном случае (например, при энергиях, больших 100 МэВ) ядро, по крайней мере частично, становится прозрачным, поглощая не все падающие на него нейтроны. Однако необходимо наложить на энергию нейтронов еще и противоположное требование. Она должна быть достаточно велика, чтобы длина дебройлевской волны нейтрона была заметно меньше диаметра ядра $2R$. Обоим условиям удовлетворяют быстрые нейтроны с энергией 20 МэВ. Ядро будет поглощать и рассеивать дебройлевские волны так, как черный экран. Для коротких длин волн вблизи ядра применима геометрическая оптика, а потому сечение поглощения будет равно геометрическому сечению ядра πR^2. Но нейтроны выбывают из пучка не только из-за поглощения, но и из-за *дифракционного рассеяния* в стороны. В случае коротких длин волн, указанных выше, дифракция будет фраунгоферовой, так как условие ее применимости

$$x \gg \frac{(2R)^2}{\lambda} \approx \frac{(2\cdot 10^{-12})^2}{0{,}5\cdot 10^{-12}} \approx 8\cdot 10^{-12}\text{ см},$$

где x — расстояние от ядра до точки наблюдения, для быстрых нейтронов, безусловно, выполняется. Но в случае фраунгоферовой дифракции черный экран рассеивает столько же нейтронов, сколько и поглощает. Это утверждение доказывается в точности так же, как и аналогичное утверждение в оптике (см. задачу к § 41 т. IV). Итак, для полного сечения ядра в рассматриваемой модели можно написать

$$\sigma = 2\pi R^2. \tag{65.5}$$

Измерив σ, можно по этой формуле вычислить R. Опыты с быстрыми нейтронами ($\mathscr{E}_\text{n} \approx$ 15–25 МэВ) привели к результату $r_0 = 1{,}4 \times$ $\times\, 10^{-13}$ см, а с еще более быстрыми ($\mathscr{E}_\text{n} \approx 100$ МэВ и $\mathscr{E}_\text{n} \approx 1000$ МэВ) дали $r_0 = 1{,}37\cdot 10^{-13}$ см и $r_0 = 1{,}28\cdot 10^{-13}$ см. Это указывает на частичную прозрачность ядер для очень быстрых нейтронов.

7. Наиболее точные результаты по измерению размеров ядер получаются при рассеянии быстрых электронов на ядрах. Как показано в п. 4, при энергии электронов порядка 100 МэВ длина дебройлевской волны становится сравнимой с размерами ядер. При длинах волн такого порядка должна отчетливо проявиться дифракция электронов на

ядрах атомов. По угловому распределению быстрых электронов при упругом рассеянии их на ядрах можно судить о размерах ядер. В первых опытах использовались электроны, ускоренные синхротроном до нескольких десятков мегаэлектронвольт. В последующих более точных опытах Хофштадтера (р. 1915) применялись электроны с энергиями до сотен мегаэлектронвольт. В предположении, что электрический заряд равномерно распределен по ядру, обработка результатов измерений дала $r_0 = (1{,}2 - 1{,}3) \cdot 10^{-13}$ см.

Высокая точность опытов по рассеянию быстрых электронов на ядрах ($\mathscr{E} > 500$ МэВ) позволила установить, что электрический заряд неравномерно распределен по объему ядра. Результаты опытов лучше всего согласуются с предположением, что плотность электрического заряда максимальна в центре ядра и для тяжелых ядер монотонно убывает к периферии согласно формуле

$$\rho = \frac{\rho_0}{1 + \exp\left[(r - R_0)/\delta\right]}, \tag{65.6}$$

где R_0 — расстояние от центра ядра, на котором плотность убывает в два раза по сравнению с ρ_0, а величина $\delta \approx 0{,}55 \cdot 10^{-13}$ см, т. е. одинакова для всех ядер. Для всех исследованных ядер $R_0 = 1{,}08 \times \times 10^{-13} A^{1/3}$ см. Отсюда следует, что в центре ядра ρ_0 практически совпадает с ρ.

8. Как уже указывалось во введении (§ 63), в 1937 г. в космических лучах были открыты *мюоны* — нестабильные частицы со временем жизни $2{,}2 \cdot 10^{-6}$ с. Они могут быть положительными и отрицательными. Свойства отрицательного мюона аналогичны свойствам электрона. Эти частицы отличаются одна от другой только массой: $m_\mu \approx 207\, m_e$. Замедляясь в веществе до определенной скорости, отрицательный мюон может захватываться атомом, замещая один из электронов атомной оболочки. Образовавшаяся система называется *мезоатомом*. Так как масса мюона в 207 раз больше массы электрона, то боровский радиус для него в такое же число раз меньше. Он равен $r_{\mu B} = \hbar^2/Zm_\mu e^2$, где m_μ — масса мюона. Таким образом, мюон может очень близко подходить к атомному ядру. Уже при $Z \approx 30$ боровская орбита мюона лежит внутри ядра. Для свинца ($Z = 82$), например, эта формула дает $r_{\mu B} = 3{,}11 \cdot 10^{-13}$ см. При переходе мюона с одного энергетического уровня на другой испускаются жесткие рентгеновские лучи. Их энергию можно измерить и рассчитать теоретически. Результаты вычислений сильно зависят от предположений относительно размеров ядер и поэтому могут служить для определения последних. Особенно точные результаты получаются для тяжелых ядер, поскольку в этих случаях мюон может очень близко подходить к ядру. Например, для свинца получается $r_0 = 1{,}17 \cdot 10^{-13}$ см, а $R(^{207}\mathrm{Pb}) = 1{,}17 \cdot 10^{-13} \times \times 207^{1/3} = 6{,}9 \cdot 10^{-13}$ см.

9. Радиусы α-радиоактивных ядер могут быть найдены по времени их жизни относительно α-распада. Об этом методе сказано в § 73 (п. 11).

10. До открытия нейтрона общепринятой считалась *электронно-протонная модель ядра*, согласно которой ядро состоит из A протонов и C электронов, так что зарядовое число равно $Z = A - C$. Малые размеры ядер являются сильной аргументацией против такой модели. Действительно, возьмем, например, ядро с радиусом $R = 3 \cdot 10^{-13}$ см, т. е. с диаметром $2R = 6 \cdot 10^{-13}$ см. Если бы частица (протон, нейтрон или электрон) находилась внутри ядра, то ее импульс p определялся бы там оценочной формулой

$$p = \frac{h}{2R} \approx 1{,}1 \cdot 10^{-14} \text{ г} \cdot \text{см/с},$$

а энергия электрона по релятивистской формуле $\mathscr{E} = pc$ была бы равна $3{,}3 \cdot 10^{-4}$ эрг ≈ 200 МэВ. Такого же порядка были бы и энергии электронов внутри других ядер. Среди искусственно получаемых ядер встречаются β-активные ядра всех атомных чисел (за исключением протона). Маловероятно, что энергия β-электрона, вылетевшего из ядра, существенно отличалась бы от его энергии внутри ядра. При β-распаде не наблюдаются электроны с большими энергиями (порядка 100 МэВ). Это противоречит протонно-электронной модели ядра. Электроны, получающиеся при β-распаде, *не содержатся* в исходном ядре, а *образуются* в результате этого процесса.

Совсем иначе обстоит дело с протонами и нейтронами. Энергию каждой из таких частиц можно оценить по нерелятивистской формуле $\mathscr{E} = p^2/2m$, где m — масса протона или нейтрона. Подстановка числовых значений дает $\mathscr{E} \approx 10^{-5}$ эрг ≈ 6 МэВ. Это — разумный результат, так как средняя энергия связи в ядре на один нуклон составляет около 8 МэВ. Таким образом, протоны н нейтроны могут содержаться и действительно содержатся в ядре.

Приведенное здесь возражение против нахождения электронов внутри ядра неприменимо к частицам, масса которых составляет несколько десятых масс нуклона, например к π-мезонам или кваркам.

Современные эксперименты по глубоко неупругому рассеянию мюонов на ядрах (т. е. рассеянию с большим изменением импульса мюона и рождением вторичных частиц) свидетельствуют о том, что в ядре могут содержаться *кварковые ассоциации*, более тяжелые, чем нуклоны.

§ 66. Спин ядра и сверхтонкая структура спектральных линий

1. Существование спина, т. е. собственного момента импульса ядра, и связанного с ним магнитного момента было постулировано Паули в 1928 г. для объяснения так называемой *сверхтонкой структуры спектральных линий.* Спектроскопическое изучение этого явления дало первое доказательство справедливости гипотезы Паули.

Как было показано в § 40, так называемая *тонкая структура*, т. е. мультиплетность спектральных линий, объясняется спин-орбитальным

взаимодействием электронов, точнее, взаимодействием магнитных орбитальных моментов электронной оболочки атома с ее спиновыми магнитными моментами. У щелочных металлов, имеющих один валентный электрон, мультиплетная структура наиболее проста: спектральные линии у них двойные — *дублеты.* Для разрешения дублетной структуры в случае лития требуются приборы высокой разрешающей силы (большие дифракционные решетки, интерференционные спектроскопы). У остальных щелочных металлов дублетная структура легко разрешается уже небольшими призменными спектроскопами. У натрия расстояние между компонентами резонансного дублета (D_1, D_2) составляет 0,6 нм, а у цезия оно доходит до 42 нм. В последнем случае по внешнему виду спектр уже не напоминает «тонкую структуру». Только по ряду дополнительных признаков удается установить, что указанные две линии цезия являются дублетом одной и той же спектральной серии, а не одиночными линиями различных спектральных серий.

Оказалось, что сами компоненты тонкой структуры испытывают дальнейшее расщепление, которое не может быть истолковано как результат спин-орбитального взаимодействия электронов электронной оболочки атома. Такое расщепление и получило название сверхтонкой структуры.

Сверхтонкая структура в оптических спектрах была открыта Майкельсоном в 1891 г. с помощью его интерферометра. Фабри и Перо, а затем Луммер (1860–1925) и Герке (1878–1960), используя предложенные и сконструированные ими интерферометры, продолжили первые наблюдения этого явления. Оказалось, что некоторые спектральные линии состоят из нескольких (иногда до 10 и более) тесно расположенных компонент. Расстояния между компонентами обычно не превышают 0,01–0,02 нм. Уже к 1910 г. был накоплен значительный экспериментальный материал. Его осмысливание стало возможным только после создания квантовой теории Бора и введения упомянутой выше гипотезы Паули.

Большое значение для выяснения природы явления имело установление сверхтонкой структуры компонент двойной желтой D-линии натрия. Эта структура была открыта в 1928 г. независимо друг от друга А. Н. Терениным (1896–1967) и Л. Н. Добрецовым, с одной стороны, и Шюлером — с другой. Они нашли, что каждая компонента дублета натрия (D_1, D_2) в свою очередь представляет дублет с расстоянием между составляющими около 0,002 нм, которое примерно в 300 раз меньше расстояния между самими линиями D_1 и D_2 (0,6 нм). Такая структура может быть разрешена спектральными приборами с разрешающей силой не менее 300 000. Дальнейшее изучение сверхтонкой структуры потребовало разрешающей силы порядка нескольких миллионов, а ее могут дать только интерференционные спектральные приборы.

Первые измерения спинов и магнитных моментов ядер были первоначально получены при изучении сверхтонкой структуры спектральных линий. Этот метод отличался небольшой точностью и утратил свое значение. Все последующие точные сведения о спинах и магнитных

моментах ядер были получены методом *ядерного магнитного резонанса* (см. § 42). Но мы подробно остановимся и на методе, основанном на явлении сверхтонкой структуры, поскольку это явление само по себе представляет большой физический интерес.

2. Экспериментальное исследование сверхтонкой структуры — очень трудная задача. Разрешающей способности самого распространенного интерференционного спектрального прибора — интерферометра Фабри–Перо — с избытком достаточно для разрешения компонент сверхтонкой структуры. Главная трудность связана с тем, что расстояние между компонентами сверхтонкой структуры исследуемой спектральной линии перекрывается *шириной самой линии.* Естественная ширина спектральной линии, обусловленная затуханием или обрывом колебаний, порядка десятитысячных долей нанометра, т. е. на порядок меньше самого сверхтонкого расщепления. Однако к естественной ширине добавляется еще *доплеровское уширение*, которое даже при низких температурах увеличивает ширину линии примерно в 10 раз и тем самым делает невозможным разделение компонент сверхтонкой структуры (см. т. IV, § 89). Например, для разрешения сверхтонкой структуры D-линий натрия потребовалось бы охлаждение источника света до температуры порядка 5 K, что практически невозможно. Для преодоления указанной трудности Теренин и Добрецов использовали в качестве источника света пучок возбужденных атомов, движущихся приблизительно в одном определенном направлении. Доплеровское изменение частоты

$$\frac{\Delta\nu}{\nu} = \frac{v}{c}\cos\theta \tag{66.1}$$

зависит от угла между скоростью атома v и направлением наблюдения. Если наблюдать под углом $\theta = 90°$, то в случае параллельного пучка возбужденных атомов доплеровское уширение исчезнет. Конечно, атомный пучок в действительности всегда немного расходится. Но соответствующим расположением щелей удается снизить «эффективную скорость» $v\cos\theta$ в 10–30 раз по сравнению со скоростью v. Таким образом, удается добиться необходимого ослабления доплеровского уширения. Теренин и Добрецов применяли боковое освещение пучка атомов резонансной D-линией натрия. При поглощении света атомы пучка переходили в возбужденное состояние и начинали светиться. Бóльшая интенсивность свечения получается при возбуждении атомов пучка электронными ударами. Применение электронов допустимо из-за их малой массы, так как по этой причине электроны несут малые импульсы, а потому при неупругих столкновениях с атомами пучка практически не меняют направление их движения.

3. Гипотеза Паули, упомянутая выше, предполагает, что сверхтонкая структура спектральных линий возникает из-за взаимодействия магнитного момента ядра с магнитным полем, создаваемым электронной оболочкой (орбитальным и спиновым моментами электронов). Кроме того, у ядра может быть электрический квадрупольный момент,

а также электрические и магнитные мультиполи более высоких порядков, взаимодействующие с электронной оболочкой. Эта гипотеза полностью оправдалась. Основное значение имеет *магнитный дипольный момент ядра* $\mu_{\text{яд}}$. Взаимодействие магнитного момента ядра атома с магнитным полем окружающей его электронной оболочки приводит к расщеплению энергетических уровней атома, с чем главным образом и связана сверхтонкая структура энергетических уровней и спектральных линий. В некоторых случаях для получения полного согласия с опытом необходимо учитывать наличие *электрического квадрупольного момента*, но мы здесь этого делать не будем.

По гипотезе Паули ядро обладает также *собственным моментом* импульса $\mathbf{I}$, который складывается из спиновых и орбитальных моментов импульсов протонов и нейтронов, из которых построено атомное ядро. Величина $\mathbf{I}$ носит название *спинового момента ядра.* Спиновый момент ядра квантуется согласно общим правилам квантования (см. § 31). Если за единицу момента ядра принять постоянную Планка $\hbar$, то его проекция на избранное направление (определяемое внешним магнитным полем) может принимать только дискретные значения: 0, ± 1, $\pm 2, \ldots$ или $\pm 1/2$, $\pm 3/2$, $\pm 5/2, \ldots$ Максимальное значение этой проекции принято обозначать через I и называть *спином ядра.* Его никоим образом нельзя смешивать с длиной $|\mathbf{I}|$ самого вектора $\mathbf{I}$. Полная длина вектора $\mathbf{I}$ определяется правилом квантования квадрата момента импульса

$$\mathbf{I}^2 = I(I+1). \tag{66.2}$$

Например, спин протона равен $1/2$, тогда как длина вектора спина протона (в единицах $\hbar$) равна $\sqrt{(1/2)(1/2+1)} = \sqrt{3}/2$. Таким образом, когда говорят о спине ядра (или любой другой частицы), то всегда имеют в виду *максимальное значение*, которое может принимать проекция вектора $\mathbf{I}$ на избранное направление. Такова установившаяся традиция. Быть может, она не вполне целесообразна, поскольку есть опасение смешивания величин I и $|\mathbf{I}|$. Однако этого можно не опасаться, если правильно понимать правила квантования момента импульса в квантовой механике. (Исторически понятие спина было введено до создания квантовой механики, т. е. в духе теории Бора, где длина вектора момента импульса отождествлялась с максимальным значением его проекции.) Отличие I от $|\mathbf{I}|$ связано с наличием у вектора $\mathbf{I}$ поперечной к избранному направлению компоненты, длина которой имеет определенное значение, но ее направление остается неопределенным (за исключением случая $\mathbf{I} = 0$, когда все три компоненты вектора $\mathbf{I}$ равны нулю).

Магнитный дипольный момент ядра связан с его спиновым моментом соотношением

$$\mu_{\text{яд}} = g_{\text{яд}} I, \tag{66.3}$$

где $g_{\text{яд}}$ — скаляр, называемый *ядерным гиромагнитным отношением.* Малость сверхтонкого расщепления спектральных линий приводит

к заключению, что магнитные моменты ядер по порядку величины примерно в 1000 раз меньше магнитного момента электрона, т. е. магнетона Бора μ_B. Естественной единицей магнитного момента ядра является *ядерный магнетон* μ_N, определяемый соотношением

$$\mu_N = \frac{e\hbar}{2m_{\text{p}}c} = 3{,}1524915(53) \cdot 10^{-18} \text{ МэВ/Гс}. \qquad (66.4)$$

Это такое же соотношение, каким определяется магнетон Бора μ_B. Только масса электрона в нем заменена массой протона, поэтому

$$\frac{\mu_N}{\mu_B} = \frac{m_{\text{e}}}{m_{\text{p}}} \approx \frac{1}{1836}.$$

4. Для наглядного объяснения сверхтонкой структуры применяется векторная модель, вполне аналогичная той, которая использовалась в § 40, 41 для объяснения тонкой структуры и эффекта Зеемана. Из-за малости магнитного момента ядра воздействие его на электронную оболочку атома не нарушает связи между ее орбитальным $\mathbf{L}$ и спиновым $\mathbf{S}$ моментами. Они складываются в полный момент импульса оболочки $\mathbf{J} = \mathbf{L} + \mathbf{S}$. В свою очередь спиновый момент ядра $\mathbf{I}$ складывается с моментом $\mathbf{J}$ в полный момент импульса атома $\mathbf{F} = \mathbf{I} + \mathbf{J}$. При отсутствии внешних воздействий на атом все изменения векторов $\mathbf{F}$, $\mathbf{I}$, $\mathbf{J}$ во времени сводятся к их прецессии, при которой длины всех трех векторов остаются неизменными; вектор $\mathbf{F}$ сохраняется по модулю и направлению, а векторы $\mathbf{I}$ и $\mathbf{J}$ прецессируют вокруг него с одной и той же угловой скоростью (поскольку $\mathbf{F}$, $\mathbf{I}$, $\mathbf{J}$ в любой момент лежат в одной плоскости). При наличии внешнего (по отношению к атому) постоянного магнитного поля, если только последнее достаточно слабое, вектор $\mathbf{F}$ медленно прецессирует вокруг этого поля, а векторы $\mathbf{I}$ и $\mathbf{J}$ более быстро прецессируют вокруг $\mathbf{F}$. Разумеется, это только классическая аналогия, так как в квантовом рассмотрении вектор момента не имеет определенного направления.

Взаимодействие между электронной оболочкой и ядром атома, вызывающее сверхтонкую структуру спектральных линий, обусловлено наличием у ядра не только магнитного дипольного момента, но и наличием электрического квадрупольного момента, а также электрических и магнитных моментов высшей мультипольности. Основное значение имеет магнитный дипольный момент ядра $\boldsymbol{\mu}_{\text{яд}}$, который только и учитывается в последующих рассуждениях. В этом приближении энергия взаимодействия ядра с электронной оболочкой атома — чисто магнитная и равна

$$W = -(\boldsymbol{\mu}_{\text{яд}}\mathbf{B}_{\text{об}}),$$

где $\mathbf{B}_{\text{об}}$ — магнитное поле, создаваемое электронной оболочкой атома в месте нахождения ядра. Векторы $\mathbf{B}_{\text{об}}$ и $\mathbf{J}$ пропорциональны между собой и ввиду отрицательности заряда электрона направлены противоположно. Магнитный момент ядра $\boldsymbol{\mu}_{\text{яд}}$ совпадает по направлению с вектором $\mathbf{I}$ и также пропорционален ему. Поэтому энергия W может

быть представлена в виде

$$W = A(\mathbf{I}\,\mathbf{J}), \tag{66.5}$$

где A — постоянная магнитного взаимодействия ядра с электронной оболочкой атома, пропорциональная магнитному моменту ядра. Из формулы $\mathbf{F} = \mathbf{I} + \mathbf{J}$ возведением в квадрат получаем $\mathbf{F}^2 = \mathbf{J}^2 + \mathbf{I}^2 + 2(\mathbf{I}\,\mathbf{J})$. С использованием правила квантования квадрата момента отсюда находим $(\mathbf{IJ})$, а затем и энергию W:

$$W = \frac{A}{2}\,[F(F+1) - I(I+1) - J(J+1)], \tag{66.6}$$

где F означает квантовое число, определяющее максимальное значение проекции вектора $\mathbf{F}$. Формула (66.6) — основная формула в теории сверхтонкой структуры энергетических уровней, обусловленной магнитным моментом ядра.

Таким образом, наличие магнитного момента у ядра приводит к тому, что каждый энергетический уровень атома, каким он был бы в кулоновском электрическом поле ядра, получает добавочную энергию W, определяемую выражением (66.5). При фиксированных значениях J и I эта энергия зависит от угла между векторами $\mathbf{J}$ и $\mathbf{I}$. Действительно, этот угол определяет длину вектора $\mathbf{F}$, а с ней, согласно формуле (66.6), и добавочную энергию W. В зависимости от указанного угла квантовое число F может принимать следующие значения:

$$F = I + J, |I + J - 1|, \ldots, |I - J|. \tag{66.7}$$

Если $I \leqslant J$, то число значений числа F будет $2I + 1$, а если $I \geqslant J$, то их будет $2J + 1$. В результате в первом случае энергетический уровень расщепится на $2I + 1$, а во втором — на $2J + 1$ подуровней. Это и есть сверхтонкая структура энергетических уровней.

5. В спектральном приборе, конечно, наблюдается сверхтонкая структура не энергетических уровней, а *спектральных линий.* Каждая же спектральная линия сверхтонкой структуры возникает в результате перехода атома с одного подуровня на другой. Допустимые переходы определяются правилами отбора. Одно из них определяет возможные значения ΔF:

$$\Delta F = 0, \pm 1 \quad (\text{кроме } F_1 = 0 \to F_2 = 0). \tag{66.8}$$

Невозможность перехода, указанного в скобке, подробно объяснена в § 39. В том же параграфе установлено, что излучательные переходы в оптической области спектра практически никогда не сопровождаются изменением вектора спина $\mathbf{S}$ электронной оболочки атома. Рассуждения, приведенные там, ввиду малости ядерных магнитных моментов еще в большей мере справедливы для спинов ядер. Поэтому при излучательных переходах атома практически всегда должен сохраняться спин ядра. Однако спин ядра $\mathbf{I}$ связан со спином $\mathbf{J}$ электронной оболочки. Если $I \leqslant J$, то сохранение ядерного спина означает, что при излучательных квантовых переходах сохраняются все проекции вектора $\mathbf{I}$ на

направление вектора **J**. (Эти проекции, измеренные в единицах $\hbar$, обозначаются ниже через I_J.) Ввиду этого при $I \leqslant J$ каждая спектральная линия расщепляется на столько же компонент, что и уровень энергии, т. е. на $2I+1$.

Рисунок 122 поясняет последнее утверждение. На нем слева без учета сверхтонкой структуры представлены два уровня энергии J_1 и J_2, между которыми возможен переход с испусканием света определенной длины волны. Справа изображена сверхтонкая структура обоих уровней. Число подуровней сверхтонкой структуры для обоих уровней одно и то же. Оно определяется только спином ядра и равно $2I+1$.

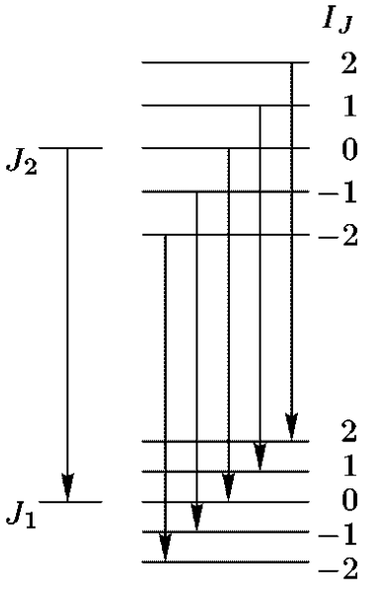


Рис. 122

Допустимы только переходы между подуровнями с одинаковыми значениями I_J. Поэтому число переходов, а с ним и число сверхтонких компонент спектральной линии, будет $N = 2I + 1$. Сосчитав это число, найдем спин ядра I. В нашем примере $N = 5$, а потому $I = 2$.

6. При $I > J$ изложенный способ определения спина ядра неприменим, так как тогда число сверхтонких компонент, на которые расщепляется спектральная линия, уже не равно $2I+1$. Однако изучение сверхтонкой структуры спектральных линий дает еще два способа определения спина ядра, которыми и можно воспользоваться. Оба способа предполагают, что квантовое число J известно. Оно может быть найдено при изучении спектра атома без учета сверхтонкой структуры.

Один из способов основан на так называемом *правиле интервалов* для мультиплетов сверхтонкой структуры. Это правило получается следующим образом. Все подуровни, на которые расщепляется тот или иной одиночный уровень, характеризуются одними и теми же значениями квантовых чисел I и J, но различными значениями квантового числа F, соответствующими всем допустимым значениям угла между векторами **I** и **J**. Расстояния между соседними подуровнями с квантовыми числами $F+1$ и F, согласно формуле (66.6), равны

$$W_{F+1} - W_F = (A/2)[(F+1)(F+2) - F(F+1)] = A(F+1). \quad (66.9)$$

Поэтому частоты переходов между соседними подуровнями с квантовыми числами F, $F+1$, $F+2$ удовлетворяют соотношениям

$$\delta\nu_1 : \delta\nu_2 : \delta\nu_3 : \ldots = (F+1) : (F+2) : (F+3) : \ldots \quad (66.10)$$

Это и есть правило интервалов. Измеряя на опыте длины волн соседних линий, можно найти отношения $(F+1) : (F+2) : (F+3) : \ldots$, а затем и квантовое число F. Совпадение значений F, вычисленных различными способами, может служить критерием правильности исходных положений, из которых были получены формулы (66.6) и (66.10).

Напомним, что при неизменных I и J квантовое число F может принимать значения, перечисленные в (66.7). Все эти значения можно

определить с помощью правила интервалов. Наибольшее из них равно $F_{\text{макс}} = I + J$. По этой формуле и может быть вычислен спин ядра I, поскольку квантовое число J должно предполагаться известным. Этот метод пригоден как при $I \leqslant J$, так и при $I \geqslant J$.

7. Очевидно, для применимости метода интервалов необходимо, чтобы число интервалов между сверхтонкими компонентами спектральной линии, а следовательно, и число самих компонент было не менее двух. Для этого число подуровней, на которые расщепляется энергетический уровень, должно быть не менее трех. Если это условие не выполняется, то можно воспользоваться третьим методом, основанным на сравнении относительных интенсивностей компонент сверхтонкой структуры, на которые расщепляются спектральные линии.

Интенсивности спектральных линий существенно зависят от *кратностей вырождения энергетических уровней*, между которыми происходят квантовые переходы. При неизменных квантовых числах I и J уровень определяется углом между векторами **I** и **J**, т. е. значением квантового числа F. Во внешнем магнитном поле (если только оно не настолько сильное, что связь между **I** и **J** не разрывается) поведение атома определяется полным моментом **F**, а не моментами **I** и **J** в отдельности. Момент же **F** может ориентироваться во внешнем поле $2F + 1$ способами, поскольку при заданном F проекции вектора **F** на направление поля могут принимать только следующие значения: $-F, -(F-1), \ldots, +(F-1), +F$. Число возможных проекций $2F + 1$ и есть кратность *вырождения* уровня с квантовым числом F. Интенсивность спектральной линии, излучаемой с уровня F или на него, при прочих равных условиях пропорциональна кратности вырождения уровня $2F + 1$.

Метод сравнения интенсивностей дает меньшую точность, чем первые два метода. Поэтому мы ограничимся рассмотрением только случая, когда он необходим (т. е. случая, когда неприменимы первые два метода). Идея метода полностью выясняется на примере вычисления спина ядра натрия из сверхтонкой структуры компонент дублета натрия (D_1, D_2), исследованной Терениным и Добрецовым.

На рис. 123 *а* воспроизведено уже рассмотренное в § 40 происхождение тонкой структуры двойной желтой линий натрия (D_1, D_2). Нижний уровень $3^2s_{1/2}$ — синглетный. Следующий уровень $3p$ из-за наличия спина электрона расщеплен на два подуровня $3^2p_{1/2}$ и $3^2p_{3/2}$. Переходы между этими подуровнями и одиночным уровнем $3^2s_{1/2}$ и дают спектральные линии тонкой структуры D_1 и D_2. На соседнем рис. 123 *б* показано расщепление уровня $3^2s_{1/2}$ на два из-за наличия магнитного момента у ядра и соответствующее сверхтонкое расщепление линии D_1 на две компоненты. Расщепление подуровня $3^2p_{1/2}$ из-за его малости в опытах Теренина и Добрецова не было разрешено и не показано на рисунке. (На самом деле каждый из этих подуровней имеет тонкую структуру, определяющую статистический вес неразрешенного подуровня.) Обозначим через F полный момент импульса атома для верхнего подуровня $3^2s_{1/2}$. Тогда из-за противоположной ориентации

спина ядра та же величина для нижнего подуровня $3^2s_{1/2}$ будет $F - 1$. Следовательно, статистические веса указанных подуровней равны соответственно $2F + 1$ и $2(F - 1) + 1 = 2F - 1$, а отношение интенсивностей рассматриваемых сверхтонких компонент спектральной линии $(2F + 1) : (2F - 1)$. По измерениям Теренина и Добрецова это отношение оказалось равным 1,7, т. е.

$$\frac{2F+1}{2F-1} = 1{,}7.$$

Отсюда $F = 1{,}929$. А так как число F может принимать только целые и полуцелые значения, то с учетом неизбежных экспериментальных ошибок следует положить $F = 2$. Значит, спин ядра натрия равен $I = F - J = 2 - 1/2 = 3/2$.

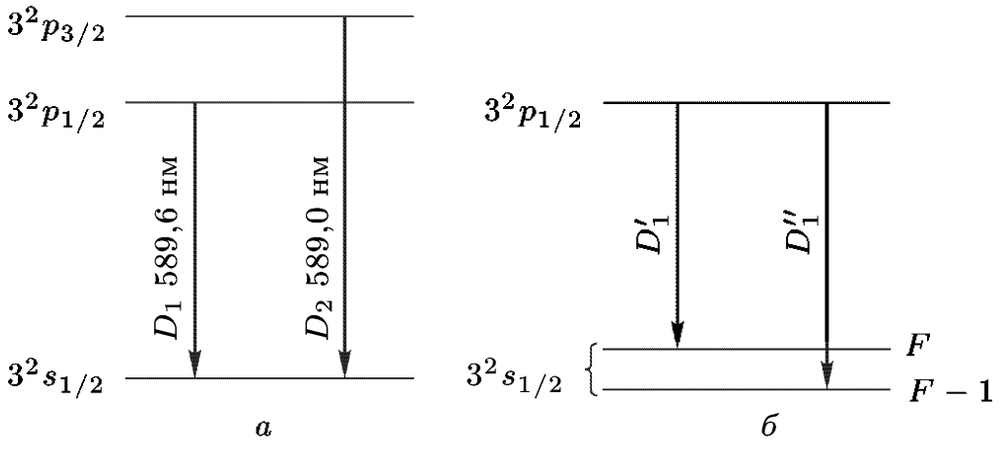


Рис. 123

8. Знание спина ядра натрия позволяет нарисовать детальную схему сверхтонкого расщепления энергетических уровней и спектральных линий для этого элемента. Уровни $^2S_{1/2}$ и $^2P_{1/2}$ должны расщепиться каждый на два подуровня, так как в этом случае $J = 1/2$, а потому квантовое число F может принимать два значения: $F = I - 1/2 = 1$ и $F = I + 1/2 = 2$ (рис. 124 *а*). Уровень $^2P_{3/2}$ должен расщепиться на четыре подуровня, так как при $J = 3/2$ возможны четыре значения числа F: $F = I + J = 3$, $F = I + J - 1 = 2$, $F = I + J - 2 = 1$, $F = I + J - 3 = 0$ (рис. 124 *б*). (Как уже указывалось, в опытах Теренина и Добрецова расщепление p-уровней из-за его малости разрешено не было.) На рис. 124 в соответствии с правилом отбора (66.8) изображены возможные переходы (без соблюдения масштаба). Внизу приводится наблюдаемая картина сверхтонкого расщепления линий. Об относительной интенсивности компонент дают представление длины вертикальных отрезков, изображенных под соответствующими квантовыми переходами. (В этом случае масштаб на правом рисунке выбран примерно вдвое меньше, чем на левом.)

После установления схемы уровней становится понятным, почему для нахождения спина ядра мы воспользовались третьим методом. Дело в том, что единственным уровнем, сверхтонкое расщепление которого может быть установлено на опыте с достаточной точностью, является

уровень $3^2S_{1/2}$. Но он расщепляется всего на два подуровня, а этого недостаточно для применения правила интервалов. Метод подсчета числа

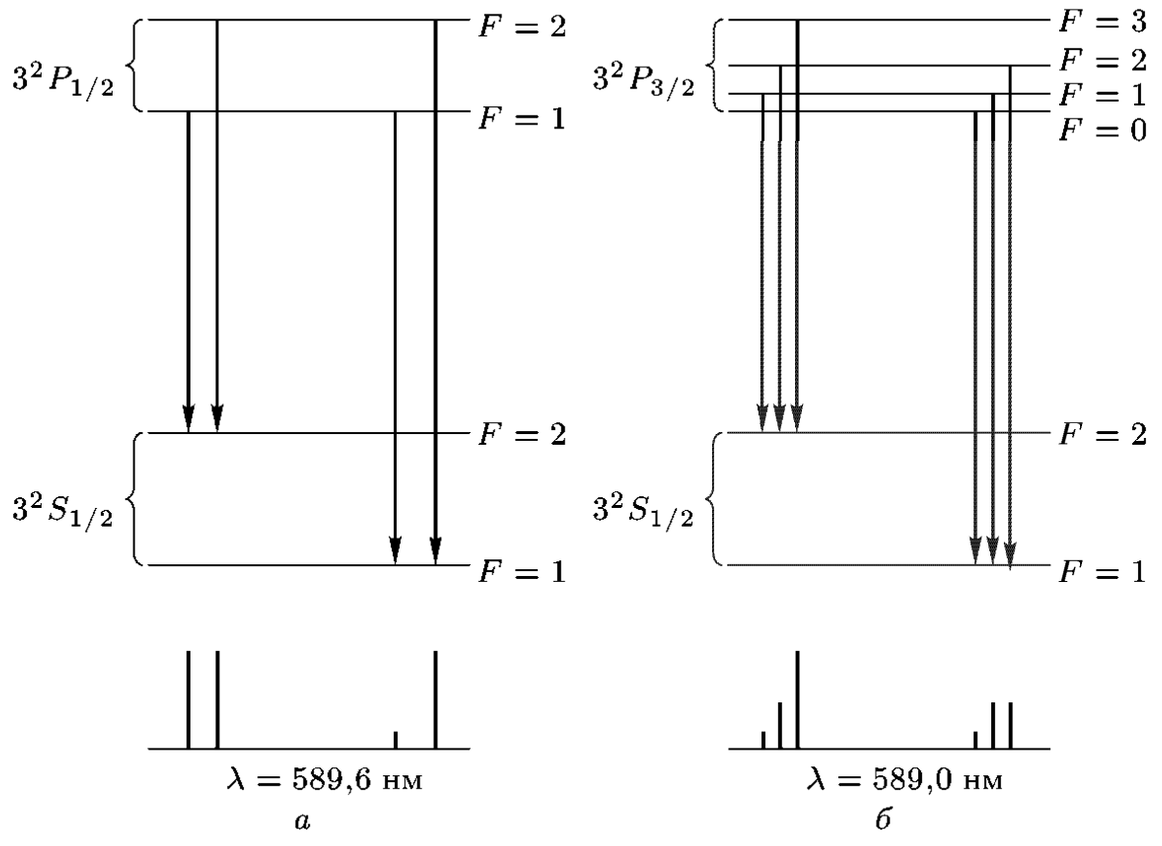


Рис. 124

сверхтонких компонент спектральной линии неприменим потому, что для натрия $I = 3/2$, а $J = 1/2$, т. е. $I > J$. Можно было бы применить правило интервалов, исследовав на опыте расщепление уровня $3^2P_{3/2}$ на четыре подуровня, но расщепление в этом случае мало и трудно поддается экспериментальному измерению. Применим также и первый способ, так как для уровня $3^2P_{3/2}$ спин $I = J = 3/2$.

9. Квантовое число F может принимать только целые или полуцелые значения, и притом небольшие. Поэтому стоящие в правой части формулы (66.10) отношения $(F + 1) : (F + 2) : \ldots$ являются отношениями небольших целых чисел. Экспериментально измеренные отношения между сверхтонкими компонентами спектральных линий отнюдь не являются таковыми. Они должны быть аппроксимированы отношениями небольших целых чисел. Но при выводе правила интервалов учитывалось только магнитное дипольное взаимодействие ядра с электронной оболочкой атома. По этой причине правило интервалов может оказаться неточным. Аппроксимация, указанная выше, если формально руководствоваться только соображениями точности, может быть произведена не в ту сторону, что приведет к ошибке в определении спина ядра. Поэтому определение спина ядра по правилу интервалов должно контролироваться другими методами.

В качестве примера рассмотрим висмут, на котором исторически впервые была подтверждена теоретическая схема сверхтонкой структуры. У висмута тщательно исследована сверхтонкая структура большо-

го числа линий, однозначно сводимая к расщеплению энергетических уровней на подуровни. На рис. 125 представлена схема сверхтонкого расщепления уровней висмута $^2S_{1/2}$ и $^2D_{3/2}$, в результате переходов между которыми возникает линия $\lambda =$ $= 472{,}2$ нм. Уровень $^2S_{1/2}$ расщепляется на два, а уровень $^2D_{3/2}$ — на четыре подуровня. Так как для первого уровня $J = 1/2$, а для второго $J = 3/2$, то в обоих случаях число подуровней равно $2J + 1$. Поэтому $I \geqslant 3/2$ и приходится применять метод интервалов. По экспериментальным измерениям интервалы между подуровнями уровня $^2D_{3/2}$ равны 0,152, 0,198 и 0,255 см$^{-1}$. Если $F_{\text{мин}}$ — минимальное квантовое число, которым характеризуются подуровни, то в силу (66.10) должно быть

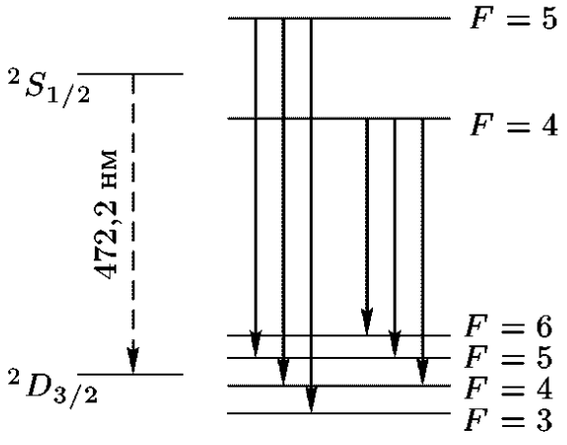


Рис. 125

$$(F_{\text{мин}} + 1) : (F_{\text{мин}} + 2) : (F_{\text{мин}} + 3) \approx 152 : 198 : 255.$$

Но это отношение приближенное и должно быть аппроксимировано отношением небольших целых чисел. Если руководствоваться чисто арифметическими соображениями, то следовало бы взять 3 : 4 : 5. Можно также взять 4 : 5 : 6, хотя арифметически это и менее точно. В первом случае получилось бы $F_{\text{мин}} = 2$, $F_{\text{макс}} = 2 + 3 = 5$. Так как $F_{\text{мин}} = I - 3/2$, а $F_{\text{макс}} = I + 3/2$, то спин ядра определится из пропорции

$$\frac{I - 3/2}{I + 3/2} = \frac{2}{5},$$

откуда $I = 7/2$, что не согласуется с величиной, найденной по расщеплению сверхтонких компонент во внешнем магнитном поле (эффект Зеемана, см. § 67). Поэтому мы возьмем 4 : 5 : 6. Тогда $F_{\text{мин}} = 3$, $F_{\text{макс}} = 6$, а для спина получается правильное значение $I = 6 - 3/2 =$ $= 9/2$.

Таким образом, подуровни уровня $^2D_{3/2}$ можно занумеровать квантовыми числами $F = 3, 4, 5, 6$, что и сделано на рис. 125. После этого квантовые числа F для подуровней уровня $^2S_{1/2}$ однозначно устанавливаются на основании правила отбора (66.8) и равны 5 и 4, как указано на рис. 125.

10. Мы рассмотрели сверхтонкую структуру энергетических уровней и спектральных линий, предполагая, что рассматриваемый химический элемент состоит всего из одного изотопа. Как изменится картина сверхтонкого расщепления, если элемент, как это бывает в большинстве случаев, представляет собой смесь нескольких изотопов? Спины и магнитные моменты ядер различных изотопов даже одного и того же элемента, вообще говоря, различны. Поэтому изотопы будут отличаться и характером сверхтонкого расщепления. Наблюдаемая картина

усложнится и будет состоять из наложения картин, получающихся от каждого изотопа в отдельности.

11. Отметим в заключение, что в 1951 г. в космическом радиоизлучении была открыта линия $\lambda = 21$ см. Она представляет собой линию сверхтонкой структуры атомарного водорода, находящегося на самом глубоком энергетическом уровне. Этот уровень состоит из двух близких подуровней, между которыми совершаются переходы с испусканием указанной линии. Время жизни атома водорода на верхнем подуровне исключительно велико ($3{,}5 \cdot 10^{14}$ с $\approx 1{,}1 \cdot 10^7$ лет). (Для сравнения напомним, что время жизни в возбужденном состоянии атома, испускающего оптические линии, в обычных условиях всего порядка 10^{-8} с.) Между этими подуровнями совершаются и безызлучательные переходы из-за столкновений между атомами. В облаках межзвездного газа они совершаются за время порядка нескольких сотен лет. Хотя такое время и очень мало по сравнению с 10^7 лет, радиационные переходы с испусканием линии $\lambda = 21$ см все же не полностью подавлены столкновениями. Это возможно благодаря тому, что водород является самым распространенным элементом Вселенной. Исследования космического радиоизлучения с $\lambda = 21$ см играют важную роль в радиоастрономии.

§ 67. Влияние спина ядра на эффект Зеемана

1. В предыдущем параграфе предполагалось, что внешнего магнитного поля нет. Допустим теперь, что атом находится в постоянном однородном внешнем магнитном поле $\mathbf{B}_{\text{внеш}}$. Тогда будет наблюдаться эффект Зеемана. Посмотрим, какое влияние на характер этого эффекта оказывают спин и магнитный момент ядра. Для наших целей достаточно ограничиться наиболее важным случаем, когда поле $\mathbf{B}_{\text{внеш}}$ *сильное.* Это значит, что энергия взаимодействия магнитного момента ядра с магнитным полем электронной оболочки мала по сравнению с энергией взаимодействия магнитного момента оболочки с внешним полем. Магнитное поле электронной оболочки $\mathbf{B}_{\text{об}}$ в месте нахождения ядра легко оценить. Оно довольно велико — порядка 10^5–10^6 Гс. Но зато магнитный момент ядра примерно в тысячу раз меньше магнетона Бора μ_B, тогда как для оболочки он порядка μ_B. Если поле $\mathbf{B}_{\text{внеш}}$ значительно превосходит, например, 1000 Гс, то его следует считать сильным. В общем случае критерий сильного поля можно получить из следующих соображений. Энергия взаимодействия магнитного момента ядра с магнитным полем оболочки порядка $\mu_{\text{яд}} B_{\text{об}} \approx A(IJ)$, тогда как энергия взаимодействия электронной оболочки с внешним полем порядка $\mu_{\text{об}} B_{\text{внеш}} \approx g_{\text{об}} J B_{\text{внеш}}$, где $g_{\text{об}}$ выражено в магнетонах Бора. Поэтому искомый критерий можно записать в виде

$$g_{\text{об}} B_{\text{внеш}} \gg AI. \tag{67.1}$$

Необходимо заметить, что этот критерий более слабый, чем аналогичный критерий в случае эффекта Пашена–Бака (см. § 41).

2. Внешнее магнитное поле разрывает связь между вектором $\mathbf{J}$ и вектором $\mathbf{I}$. Электронная оболочка начинает прецессировать вокруг

направления поля $\mathbf{B}_{\text{внеш}}$. Поскольку связь вектора $\mathbf{I}$ с $\mathbf{J}$ разорвана, вектору $\mathbf{I}$ не остается ничего другого, как совершать независимую прецессию вокруг того же направления. Магнитное квантовое число m_I, определяющее проекции вектора ядерного спина I на направление поля $\mathbf{B}_{\text{внеш}}$, может принимать значения $-I, -(I-1), \ldots, (I-1), I$. Таких значений всего $2I+1$. Поэтому каждый энергетический уровень в магнитном поле расщепляется на $2I+1$ подуровня. Пусть 1 и 2 — какие-либо два энергетических уровня атома в магнитном поле, какими они были бы без учета спина и магнитного момента ядра. Если между уровнями 1 и 2 возможен излучательный переход, то в результате такого перехода в спектре появляется соответствующая зеемановская линия. С учетом спина и магнитного момента ядра каждый из уровней 1 и 2 расщепляется на $2I+1$ подуровней, квантовые переходы между которыми подчиняются правилу отбора $\Delta m_I \approx 0$. Эти переходы приводят к сверхтонкому расщеплению каждой зеемановской линии на $2I+1$ компонент.

Таким образом, в магнитном поле, если отвлечься от наличия спина и магнитного момента ядра, должно наблюдаться обычное (простое или сложное) явление Зеемана. Влияние спина и магнитного момента ядра проявляется наиболее просто, когда поле сильное. В этом случае спин и магнитный момент ядра приводят к дальнейшему — сверхтонкому — расщеплению каждой зеемановской линии на $2I+1$ компонент. Расстояние между ними малы по сравнению с расстояниями между обычными зеемановскими компонентами (т. е. компонентами, какими они получились бы, если бы не было спина и магнитного момента ядра). Сосчитав число сверхтонких зеемановских компонент $2I+1$, можно определить спин ядра I. Этот метод не накладывает никаких ограничений на значение спина I.

3. Вопрос о более детальной структуре зеемановской линии сводится к вычислению энергии атома во внешнем магнитном поле. Если внешнее поле $\mathbf{B}_{\text{внеш}}$ сильное, то магнитная энергия атома слагается из магнитной энергии электронной оболочки

$$-(\boldsymbol{\mu}_{\text{об}}\mathbf{B}_{\text{внеш}}) \text{ и ядра } -(\boldsymbol{\mu}_{\text{яд}}\mathbf{B}_{\text{внеш}}) - (\boldsymbol{\mu}_{\text{яд}}\mathbf{B}_{\text{об}}).$$

Первое слагаемое в обсуждаемом нами вопросе не играет существенной роли, так как оно вызывает обычное зеемановское расщепление, уже изученное нами. Его можно отбросить. Что касается второго слагаемого, то им можно пренебречь, так как в обычных условиях $\mathbf{B}_{\text{внеш}} \ll \mathbf{B}_{\text{об}}$. Остается только слагаемое $-(\boldsymbol{\mu}_{\text{яд}}\mathbf{B}_{\text{об}})$, которое и следует учесть. Это слагаемое выражается прежней формулой (66.5). Однако при наличии сильного внешнего магнитного поля вектор $\mathbf{I}$ квантуется иначе, чем в случае свободного атома, поскольку он прецессирует не вокруг $\mathbf{J}$, а вокруг $\mathbf{B}_{\text{внеш}}$. Прецессирующие векторы $\mathbf{I}$ и $\mathbf{J}$ имеют определенные проекции только на направление поля $\mathbf{B}_{\text{внеш}}$. Они определяются магнитными квантовыми числами m_I и m_J. Перпендикулярные проекции остаются неопределенными. А поскольку обе прецессии совершаются независимо, среднее по времени произведение перпендикулярных проекций равно нулю. Следовательно, в среднем $(\mathbf{IJ}) = m_I m_J \hbar^2$. Тогда

формула (66.5) переходит в

$$W = Am_I m_J \hbar^2. \tag{67.2}$$

Если теперь (без учета сверхтонкого расщепления) снова рассмотреть два энергетических уровня 1 и 2, при переходе между которыми излучается какая-либо зеемановская линия, то с учетом сверхтонкого расщепления между уровнями возникнут переходы, при которых будет излучаться энергия

$$\delta W = [(Am_I m_J)_2 - (Am_I m_J)_1]\hbar^2 = [(Am_J)_2 - (Am_J)_1]m_I \hbar^2, \tag{67.3}$$

так как в силу правила отбора $\Delta m_I = 0$. В результате таких переходов зеемановская линия и претерпит сверхтонкое расщепление на $2I + 1$ компонент, соответствующих значениям квантового числа $m_I = I$, $(I-1), \ldots, -(I-1), -I$. Расстояния между сверхтонкими компонентами будут одни и те же и равны $[(Am_J)_2 - (Am_J)_1]\hbar^2$.

Описанная картина, в частности, отчетливо наблюдается у висмута на линии 472,2 нм (см. предыдущий параграф, п. 9). В достаточно сильном магнитном поле (порядка 10 000 Гс) получается обычный простой эффект Зеемана. Но каждая зеемановская составляющая состоит из 10 равноотстоящих компонент. Из соотношения $2I + 1 = 10$ получается $I = 9/2$, как и было указано в § 66, п. 9.

§ 68. Измерения спинов и магнитных моментов ядер методом магнитного резонанса. Опытные данные о спинах и магнитных моментах ядер

1. Зная спин ядра, можно по магнитному взаимодействию ядра с электронной оболочкой атома определить и магнитный момент ядра. Вопрос сводится к нахождению постоянной магнитного взаимодействия A, входящей в формулу (66.5). Но эта постоянная, даже для легких атомов, теоретически может быть вычислена с малой погрешностью (не превышающей примерно 10 %). Более надежно спины и магнитные моменты ядер могут быть найдены при исследовании поведения атомных ядер во внешних магнитных полях. Сюда относятся исследования сверхтонкого расщепления зеемановских спектральных линий в таких полях, о чем говорилось в предыдущем параграфе.

Можно также воспользоваться методом Штерна и Герлаха, исследуя отклонения атомных и молекулярных пучков в сильных и резко неоднородных магнитных полях (см. § 36). По числу компонент, на которые расщепляется пучок, можно определить спин, а по значению расщепления — магнитный момент ядра. Однако определение магнитных моментов ядер методом Штерна и Герлаха много труднее, чем магнитных моментов атомов, так как из-за малости магнитных моментов ядер ожидаемое отклонение примерно в тысячи раз меньше, чем у атомов с отличными от нуля магнитными моментами электронных оболочек. Влияние ядер совершенно исчезает на фоне более сильного

эффекта атомных оболочек. Для преодоления возникшей трудности опыты надо производить на атомах с замкнутыми электронными оболочками или с молекулами (например, H_2, H_2O), у которых магнитные моменты электронных оболочек взаимно скомпенсированы. Кроме того, надо создавать резко неоднородные магнитные поля с градиентами порядка 10^5 Гс/см. И даже в этих случаях наблюдаемое расщепление (около 0,05 мм) сравнимо с разбросом из-за максвелловского распределения скоростей. Фактическое расщепление пучка в подобных опытах не наблюдается, и для определения магнитных моментов приходится тщательно исследовать плотность распределения частиц пучка в месте попадания их на детектор. Хотя методом Штерна и Герлаха и удалось измерить магнитный момент протона, но для определения магнитных моментов ядер этот метод в большинстве случаев непригоден.

2. Прецизионную точность (примерно до семи знаков) дает *метод магнитного резонанса*, наблюдаемый как на нейтральных пучках атомов или молекул (со скомпенсированными магнитными моментами), так и методом поглощения. В случае нейтронов можно пользоваться только пучками, так как нейтроны нельзя содержать в ампулах. Магнитный резонанс подробно изложен в § 42, а потому нет необходимости его еще раз излагать здесь. Заметим только, что по числу резонансов можно определить спин, а по резонансным частотам — расстояние между энергетическими уровнями ($\mu_{\text{яд}}B$) и магнитный момент ядра. Методом магнитного резонанса и получены все точные данные о магнитных моментах ядер.

3. Приведем теперь опытные данные относительно спинов и магнитных моментов ядер.

1. Протон и нейтрон, как и электрон, обладают спином, равным $1/2$ (в единицах $\hbar$). Полный момент импульса каждого нуклона в ядре складывается из его спинового и орбитального моментов по квантовомеханическому правилу сложения моментов. В свою очередь полный момент ядра $\mathbf{I}$ по тому же правилу складывается из моментов импульса составляющих его нуклонов.

2. При четных A спин ядра I всегда целый, а при нечетных — полуцелый. Исторически этот факт был решающим при переходе от протонно-электронной к протонно-нейтронной модели ядра. В самом деле, например, ядро азота $^{14}_{7}\mathrm{N}$, состоящее по протонно-электронной модели из 21 частицы, должно было бы иметь полуцелый спин, поскольку спин каждой частицы равен $1/2$, а их орбитальные моменты целочисленны. Экспериментально же измеренный спин ядра азота оказался равным 1. В свое время этот факт получил название «азотной катастрофы». В протонно-нейтронной модели ядра противоречия с опытом не получается, так как по этой модели ядро азота состоит из 7 протонов и 7 нейтронов, т. е. из четного числа частиц, а потому его спин, в согласии с опытом, должен быть целым.

3. Для четно-четных стабильных ядер (Z и N четные) спин всегда равен нулю. К таким ядрам относится больше половины всех стабильных ядер. Почти все остальные стабильные ядра либо четно-нечетные

(Z четное, N нечетное), либо нечетно-четные (Z нечетное, N четное). Ядер указанных типов имеется примерно поровну. Спины всех этих ядер отличны от нуля, так как все они имеют нечетные A. Минимальное значение спина у этих ядер равно $1/2$. Нечетно-нечетных стабильных ядер (Z и N нечетные), как уже указывалось в § 64, имеется всего пять (^{2}D, ^{6}Li, ^{10}B, ^{14}N, ^{50}V) [1]. Все они имеют целочисленные спины, отличные от нуля (спин для ^{50}V равен 6, для ^{6}Li — 3, для остальных ядер — 1).

4. Спины всех известных стабильных ядер не превышают $9/2$, за исключением ванадия ^{50}V, спин которого равен 6. Таким образом, спины ядер очень малы по сравнению с суммой абсолютных значений спинов и орбитальных моментов всех частиц, входящих в ядро. Наряду с преобладанием четно-четных ядер, отмеченным выше, этот факт указывает на то, что большинство нуклонов в ядре прочно связано в замкнутых оболочках, имеющих нулевой суммарный момент импульса, и не участвует в создании спина ядра.

5. Ядра со спинами $I \geqslant 1/2$ обладают магнитными моментами. Магнитные моменты ядер, о чем уже неоднократно говорилось выше, примерно в тысячи раз меньше магнетона Бора, определяющего магнитный момент электрона. Естественной единицей ядерных магнитных моментов является ядерный магнетон. По определению он в $m_\mathrm{p}/m_\mathrm{e} \approx 1836$ раз меньше магнетона Бора. Магнитные моменты ядер с ненулевыми спинами — порядка ядерного магнетона. Это указывает на то, что магнитные моменты отдельных нуклонов в ядре, как и их угловые моменты, в основном компенсируют друг друга. Малость же магнитных моментов ядер еще раз свидетельствует против наличия в ядре электронов, поскольку магнитный момент электрона в 1836 раз больше ядерного магнетона.

6. Собственные магнитные моменты нуклонов *не аддитивны.* Например, дейтрон состоит из протона и нейтрона, магнитные моменты которых (в ядерных магнетонах) равны $\mu_\mathrm{p} = 2{,}79$, $\mu_\mathrm{n} = -1{,}91$. Если бы эти моменты были аддитивны, то магнитный момент дейтрона был бы равен $\mu_\mathrm{d} = 2{,}79 - 1{,}91 = 0{,}88$, тогда как опыт дает $\mu_\mathrm{d} = 0{,}86$. Это расхождение далеко выходит за пределы погрешностей. Неаддитивность магнитных моментов находит свое истолкование в нецентральности сил, действующих между нуклонами.

§ 69. Четность. Закон сохранения четности

1. Понятие *четности* возникает в связи с операцией *инверсии.* При инверсии относительно начала координат знаки декартовых координат всех частиц системы меняются на противоположные, т. е. x, y, z переходят в $-x, -y, -z$ или $\mathbf{r}$ заменяется на $-\mathbf{r}$. В дальнейшем для

[1] Изотоп $^{50}_{23}V$ β^--радиоактивен, но период полураспада для него равен 6×10^{15} лет, т. е. очень велик. По этой причине он и отнесен нами к стабильным изотопам.

сокращения записи под $\mathbf{r}$ обычно будет пониматься радиус-вектор не одной частицы, а совокупность радиус-векторов частиц всей системы. Если же в рассуждении требуется явно указать, что частиц несколько, то мы (также для сокращения записи) ограничимся случаем двух частиц, нумеруя их индексами 1 и 2. Это не вводит никаких ограничений.

Оператор инверсии обозначается через $\widehat{P}$. Таким образом, по определению $\widehat{P}\psi(\mathbf{r}) = \psi(-\mathbf{r})$. Операцию инверсии $\mathbf{r} \to -\mathbf{r}$ можно представить как зеркальное отражение относительно плоскости, проходящей через начало координат, с последующим поворотом на 180° вокруг оси, перпендикулярной к этой плоскости.

Найдем прежде всего собственные значения P оператора $\widehat{P}$. Они определяются уравнением

$$\widehat{P}\psi(\mathbf{r}) = P\psi(\mathbf{r}).$$

Повторное применение оператора $\widehat{P}$ дает

$$\widehat{P}^2\psi(\mathbf{r}) = P\widehat{P}\psi(\mathbf{r}) = P^2\psi(\mathbf{r}).$$

Но оператор $\widehat{P}^2$ есть тождественное преобразование, при котором ничего не меняется. Значит, $\psi(\mathbf{r}) = P^2\psi(\mathbf{r})$, а потому $P^2 = 1$, $P = \pm 1$. Таким образом, собственные значения оператора $\widehat{P}$ будут $+1$ и -1. В соответствии с этим собственные функции оператора $\widehat{P}$ разделяются на *четные* и *нечетные*. Четная функция определяется соотношением $\psi(\mathbf{r}) = \psi(-\mathbf{r})$, а нечетная — соотношением $\psi(\mathbf{r}) = -\psi(-\mathbf{r})$. Число P принято называть *четностью* функции $\psi(r)$ или состояния системы. Для четных функций $P = +1$, для нечетных $P = -1$.

2. В уравнении Шредингера

$$i\hbar\frac{\partial\Psi}{\partial t} = \widehat{\mathscr{H}}\Psi \tag{69.1}$$

гамильтониан $\widehat{\mathscr{H}}$ определяется выражением

$$\widehat{\mathscr{H}} = -\frac{\hbar^2}{2m_1}\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial y_1^2} + \frac{\partial^2}{\partial z_1^2}\right) + \\ + \frac{\hbar^2}{2m_2}\left(\frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial y_2^2} + \frac{\partial^2}{\partial z_2^2}\right) + U(\mathbf{r}_1, \mathbf{r}_2). \tag{69.2}$$

Первые два слагаемых представляют оператор кинетической энергии и не меняются при инверсии, если начало координат поместить в центре масс системы, что и будет делаться в дальнейшем. В этом случае оператор кинетической энергии не меняется при инверсии относительно начала координат, поскольку дифференциалы координат в него входят во второй степени. До 1956 г. считали, что оператор потенциальной энергии $U(\mathbf{r}_1, \mathbf{r}_2)$ при инверсии также никогда не меняется. Действительно, при инверсии не изменяется относительное расположение любой пары частиц системы. Меняется на прямо противоположное только направление соединяющей их прямой. А от этого, как думали, потенциальная

функция системы $U(\mathbf{r}_1, \mathbf{r}_2)$ не претерпевает никаких изменений. После открытия в 1956 г. *несохранения четности в слабых взаимодействиях* было установлено, что это заключение справедливо для электромагнитных и сильных взаимодействий и нарушается для слабых. Таким образом, при сильных и электромагнитных взаимодействиях гамильтониан $\widehat{\mathscr{H}}$ не меняется при инверсии. В этом случае имеет место *закон сохранения четности волновой функции.* Это приближенный закон, справедливый с точностью до слабых взаимодействий.

Закон сохранения четности является следствием уравнения Шредингера (69.1). Действительно, допустим, что в момент времени $t = 0$ волновая функция $\Psi = \Psi_0(\mathbf{r})$ либо четная, либо нечетная. Для приращения $d\Psi$ за время dt уравнение (69.1) дает

$$d\Psi = \frac{\partial \Psi}{\partial t}\, dt = \frac{dt}{i\hbar}\, \widehat{\mathscr{H}} \Psi(r),$$

или с точностью до членов более высокого порядка малости

$$d\Psi = \frac{dt}{i\hbar}\, \widehat{\mathscr{H}} \Psi_0(\mathbf{r}).$$

Но гамильтониан $\widehat{\mathscr{H}}$ не меняется при инверсии координат. Значит, функция $\widehat{\mathscr{H}}\Psi_0(\mathbf{r})$, а с ней и функция $\Psi_{dt}(r) \equiv \Psi_0(r) + d\Psi$ обладают той же четностью, что и начальная функция $\Psi_0(\mathbf{r})$. Применяя этот процесс дальше, докажем, что это справедливо и для функции $\Psi_t(\mathbf{r})$ при любом конечном значении времени t.

Доказательство предполагало, что волновая функция $\Psi(\mathbf{r})$ либо четная, либо нечетная. Определенной четностью волновая функция обладает только для невырожденного состояния системы (например, для основного состояния ядра), описываемого единственной собственной волновой функцией (разумеется, определенной с точностью до несущественного фазового множителя $e^{i\delta}$). Во всяком вырожденном состоянии волновая функция в общем случае может быть представлена линейной суперпозицией двух функций, из которых одна четная, а другая нечетная. В таком случае закон сохранения четности означает сохранение относительной доли обоих состояний с определенной четностью. Доказательство, приведенное выше, без всяких затруднений обобщается и на этот случай.

Заметим, что это доказательство основано на уравнении (69.2), а оно не учитывает возможности рождения частиц. Более общее рассмотрение показывает, что с точностью до слабых взаимодействий четность волновой функции системы не меняется при любых процессах (включая рождение и поглощение частиц).

3. Из приведенного нами доказательства видно, что закон сохранения четности есть свойство гамильтониана $\widehat{\mathscr{H}}$, т. е. свойство самой системы, а не функции ψ, характеризующей ее состояние. Поэтому-то из закона сохранения четности, как из всякого закона, можно вывести определенные физические следствия, доступные экспериментальной проверке. В качестве примера в конце этого параграфа мы приводим

вывод правила отбора при излучении по орбитальному квантовому числу l.

4. Важное значение имеет задача определения четности волновой функции системы, состоящей из нескольких составных частей. Допустим для простоты, что система состоит из двух частей A и B. Если можно пренебречь взаимодействием между ними, то волновая функция сложной системы может быть представлена в виде

$$\Psi_{A+B} = \Psi_A \Psi_B \Psi_{l_A} \Psi_{l_B}, \tag{69.3}$$

где Ψ_A и Ψ_B — волновые функции, описывающие внутренние движения подсистем относительно их центров масс, а Ψ_{l_A} и Ψ_{l_B} — движения тех же центров масс относительно центра масс всей сложной системы. Испытание на четность полной волновой функции Ψ_{A+B} сводится к последовательному повторению того же испытания для каждой из четырех функций $\Psi_A, \Psi_B, \Psi_{l_A}, \Psi_{l_B}$ в отдельности. Поэтому для четности всей системы можно написать

$$P_{A+B} = P_A P_B P_{l_A} P_{l_B}. \tag{69.4}$$

Чтобы определить четность сложной системы по четностям составляющих ее подсистем, надо знать явный вид волновых функций Ψ_{l_A} и Ψ_{l_B} для относительного движения центров этих подсистем. Эта задача сводится к нахождению волновой функции частицы при ее движении относительно неподвижного центра. Она решается в квантовой механике. Мы не предполагаем ее решать, а лишь заимствуем необходимые результаты из квантовой механики. В сферической системе координат положение частицы относительно неподвижного центра задается расстоянием до него r, полярным углом θ и азимутальным углом φ. Волновая функция частицы в такой системе имеет вид

$$\Psi_l = R(r) P_l^m(\cos\theta) e^{im\varphi},$$

где l и m — квантовые числа орбитального момента и его проекции на полярную ось, а $P_l^m(\cos\theta)$ — так называемые *присоединенные полиномы Лежандра* (1752–1833).

Явный вид функции $R(r)$, а также присоединенных полиномов Лежандра в рассматриваемой нами задаче не требуется. Достаточно указать, что при замене $\cos\theta$ на $-\cos\theta$ полином $P_l^m(\cos\theta)$ приобретает множитель $(-1)^{l-m}$. При инверсии значение радиуса r не меняется, а углы θ и φ заменяются соответственно на $\pi - \theta$ и $\varphi + \pi$, так что

$$\cos\theta \to \cos(\pi - \theta) = -\cos\theta, \quad e^{im\varphi} \to e^{im(\varphi+\pi)} = (-1)^m e^{im\varphi}.$$

Поэтому

$$\Psi_l \to R(r)(-1)^{l-m} P_l^m(\cos\theta) \cdot (-1)^m e^{im\varphi} = (-1)^l \Psi_l.$$

Таким образом, четность волновой функции относительного движения равна $P_l = (-1)^l$, а четность системы $A + B$

$$P_{A+B} = (-1)^{l_A}(-1)^{l_B} P_A P_B, \tag{69.5}$$

если моменты имеют определенные значения. Эта формула тривиально обобщается на случай сложной системы, состоящей из произвольного числа частей.

5. Части A и B, из которых состоит система, могут быть и элементарными частицами с отличной от нуля массой покоя. Как показывает опыт, каждая элементарная частица с точностью до слабых взаимодействий характеризуется определенной четностью, не связанной с ее движением как целого. Такая четность называется *внутренней четностью* частицы. Внутренняя четность — такое же неотъемлемое свойство частицы, как и ее спин. Частицы, у которых внутренняя четность равна $+1$, называются *четными*, а частицы с внутренней четностью -1 — *нечетными*. Правило (69.5), если пренебречь эффектами слабых взаимодействий, распространяется и на системы, состоящие из элементарных частиц (с отличными от нуля массами покоя), но с учетом их внутренних четностей. Внутренние четности протона, нейтрона и электрона могут быть заданы произвольно. Это соглашение и применяется в дальнейшем. Обычно они принимаются равными $+1$. После этого внутренние четности атома, ядра и большинства элементарных частиц однозначно определятся из экспериментальных данных на основе закона сохранения четности.

Атомное ядро является сложной системой, состоящей из движущихся внутри него нуклонов. Если взаимодействием между ними можно пренебречь, то четность ядра будет $(-1)^{\sum l_i}$,где l_i — орбитальное квантовое число, определяющее характер движения i-го нуклона. Состояние нуклона в ядре будет четным, если его орбитальное квантовое число l четное, и нечетным в противоположном случае. Так, протоны и нейтроны в s-состоянии являются четными нуклонами, а в p-состоянии — нечетными. В качестве примера определим четность ядра $^7_3\mathrm{Li}$. В модели ядерных оболочек (см. гл. X) показывается, что это ядро в основном состоянии состоит из четырех s-нуклонов и трех p-нуклонов. Поэтому четность такого ядра равна $(-1)^3 = -1$. Напротив, α-частица состоит из четырех нуклонов в s-состоянии — ее внутренняя четность равна $+1$. Опыт показывает, что при энергиях падающего протона, меньших примерно 0,5 МэВ, ядерная реакция

$$\mathrm{p} + {}^7_3\mathrm{Li} \to \alpha + \alpha + 17{,}1\ \text{МэВ},$$

несмотря на ее высокую экзотермичность, подавлена (т. е. идет с малой вероятностью). Дело в том, что четность двух α-частиц равна $+1$. Такова же четность протона при указанных энергиях. Четность же ядра $^7_3\mathrm{Li}$ равна -1, так что в рассматриваемой реакции закон сохранения четности нарушается.

Основное состояние четно-четных ядер имеет положительную четность. Основные состояния других ядер могут быть как четными,

так и нечетными. Ядра в возбужденных состояниях могут иметь различную четность, не обязательно совпадающую с четностью основного состояния. На схемах ядерных уровней обычно указываются спин и четность каждого уровня. Спин обозначается числом, а четность — знаком «+» или «−». Например, символ 2^+ означает четный уровень со спином 2, а символ $(1/2)^-$ — нечетный уровень со спином $1/2$. Совокупность значений спина и четности называется *характером уровня ядра.*

6. Все изложенное относится к частицам с ненулевой массой покоя. Для фотонов, как и для всяких релятивистских частиц с нулевой массой покоя, понятия состояния с определенным значением орбитального момента l не существует. Вместо этого вводится аналог этого понятия, называемый *мультиполем.* Мультиполь электромагнитного поля — это состояние свободно распространяющегося поля, обладающего определенным полным моментом L и определенной четностью P. Для свободного фотона возможны состояния с полным моментом $L = 1, 2, 3, \ldots$ Частный случай $L = 1$ уже был подробно рассмотрен в § 37. Состояния с нулевым полным моментом L для фотона не существует. Состояние фотона с моментом L и четностью $(-1)^L$ называется *электрическим* 2^L*-полем*, а состояние с таким же моментом и четностью $(-1)^{L+1}$ — *магнитным* 2^L*-полем.* Состояние с $L = 1$ называется *дипольным*, с $L = 2$ — *квадрупольным*, с $L = 3$ — *октупольным* и т. д. В соответствии с этим электрический диполь и магнитный квадруполь нечетны, а магнитный диполь и электрический квадруполь — четны. Для обозначения кванта электрического мультиполя ставится буква E, которой приписывается значение полного момента L. В случае кванта магнитного мультиполя буква E заменяется на M. Например, электрический дипольный квант обозначается через $E1$, магнитный дипольный квант — через $M1$, электрический квадрупольный — через $E2$ и т. д.

Мультипольная терминология основана на классическом понятии мультиполя (см. следующий параграф). Так, при колебаниях электрического дипольного момента возникает электромагнитное излучение, которое с квантовой точки зрения состоит из $E1$ фотонов.

Если приведенная длина фотона $\lambda\!\!\!^{-} \equiv \lambda/2\pi$ много больше размеров R физической системы, с которой он взаимодействует ($R \ll \lambda\!\!\!^{-}$), то в этом взаимодействии участвуют преимущественно мультиполи наинизшего порядка, допускаемые законами сохранения момента и четности. При прочих равных условиях отношение вероятности испускания (или поглощения) электрического квантового мультиполя 2^L к соответствующей вероятности испускания (или поглощения) кванта $E1$ порядка $(R/\lambda\!\!\!^{-})^{2(L-1)}$. В случае испускания (поглощения) магнитного кванта той же мультипольности 2^L то же отношение — порядка $(R/\lambda\!\!\!^{-})^{2L}$, т. е. в $(\lambda\!\!\!^{-}/R)^2$ раз меньше. Поэтому, например, вероятности испускания квантов $E2$ и $M1$ обычно близки между собой. Это связано с тем, что по порядку величины отношение $(R/\lambda\!\!\!^{-})^2$ равно $(v/c)^2$, где v — скорость заряженной частицы в системе (например, протона в яд-

ре), а отношение напряженностей электрического и магнитного полей, генерируемых движущимся зарядом, — порядка v/c.

Изложенное в этом пункте в равной степени применимо к мультиполям молекул, атомов, ядер и элементарных частиц. В качестве примера рассмотрим правило отбора (40.1). Оно относится к испусканию (или поглощению) при наличии в атоме одного внешнего (валентного) электрона. Испускаемый электрический дипольный фотон, как мы видели, — нечетный. Четность атома в результате испускания такого фотона меняется на множитель $(-1)^{\Delta l}$, а всей системы «атом — испущенный фотон» — на множитель $(-1)^{\Delta l \pm 1}$. Закон сохранения четности при дипольном излучении допускает только значения $\Delta l = \pm 1$. Значение $\Delta l = 0$ (хотя и допускаемое законом сохранения момента) запрещено законом сохранения четности. Следует, однако, еще раз подчеркнуть, что этот запрет относится к *электрическому дипольному испусканию.* Испускание электрических квадрупольных и магнитных дипольных квантов возможно и приводит к появлению в спектре так называемых запрещенных линий. Но вероятность электрического квадрупольного и магнитного дипольного испускания примерно в $(\lambda\!\!\!^{-}/R)^2$ меньше, чем вероятность электрического дипольного испускания. Она проявляется существенно только тогда, когда последнее излучение по каким-либо причинам запрещено.

За счет слабых взаимодействий волновая функция системы всегда содержит малую примесь состояния с противоположной четностью. Поэтому если, например, разрешен по четности и моменту $M1$-переход, то он будет сопровождаться слабым $E1$-переходом. Интерференция $M1 + E1$ приводит к циркулярной поляризации квантов или к асимметрии их вылета по спину и против спина.

§ 70. Электрические свойства и форма ядра

1. Величины, характеризующие электрические свойства ядра, могут быть введены совершенно так же, как это делается в электростатике для системы точечных зарядов, занимающих небольшую область пространства. Поэтому нуклоны в ядре будем считать точечными, хотя это вовсе не обязательно. Во внешнем постоянном электрическом поле с потенциалом φ потенциальная энергия ядра определяется выражением

$$U = \sum_{\alpha} e\varphi(x_{\alpha i}), \tag{70.1}$$

где суммирование производится только по протонам ядра, так как нейтроны, поскольку они не имеют электрического заряда, не вносили бы в эту сумму никакого вклада. Функция $\varphi(x_{\alpha i})$ означает потенциал внешнего поля в точке нахождения протона α, а x — совокупность декартовых координат того же протона ($i = 1, 2, 3$; $x_1 \equiv x$, $x_2 \equiv y$, $x_3 \equiv z$). Таким образом, в подробной записи

$$U = \sum_{\alpha} e\varphi(x_\alpha, y_\alpha, z_\alpha) = \sum_{\alpha} e\varphi(\mathbf{r}_\alpha). \tag{70.1а}$$

Поместим начало координат в центре масс всего ядра (т. е. учитывая и нейтроны) и примем во внимание, что на расстояниях порядка линейных размеров ядра внешнее электрическое поле меняется мало. Тогда потенциал $\varphi(x_{\alpha i})$ целесообразно разложить в степенной ряд по координатам:

$$\varphi(x_{\alpha i}) = \varphi(0) + x_{\alpha i}\left(\frac{\partial\varphi}{\partial x_i}\right)_0 + \frac{1}{2}\,x_{\alpha i}x_{\alpha k}\left(\frac{\partial^2\varphi}{\partial x_i\partial x_k}\right)_0 + \ldots,$$

где в соответствии с общепринятой тензорной символикой по дважды встречающимся координатным индексам (но не по индексу α, который означает номер протона) производится суммирование. Подставляя это разложение в формулу (70.1), получим

$$U = \varphi(0)\sum_\alpha e + \left(\frac{\partial\varphi}{\partial x_i}\right)_0\sum_\alpha e x_{\alpha i} + \frac{1}{2}\left(\frac{\partial^2\varphi}{\partial x_i\partial x_k}\right)_0\sum_\alpha e x_{\alpha i}x_{\alpha k} + \ldots \tag{70.2}$$

Первый — главный — член этой суммы давал бы энергию заряженного ядра во внешнем электрическом поле, если бы весь заряд был сконцентрирован в одной точке — начале координат. Этот член может быть записан в виде $Ze\varphi(0)$. Он характеризует электрические свойства ядра суммарно, но не дает никаких указаний относительно распределения электричества по объему ядра.

2. Второй член суммы (70.2) содержит три компоненты вектора $\mathbf{d} = \sum_\alpha e\mathbf{r}_\alpha$, где $\mathbf{r}_\alpha = \mathbf{r}_\alpha(x_\alpha, y_\alpha, z_\alpha)$. Это есть *электрический дипольный момент ядра.* Выражение $\mathbf{d} = \sum_\alpha e\mathbf{r}_\alpha$, конечно, не инвариантно относительно выбора начала координат, поскольку полный заряд ядра $\sum_\alpha e$ отличен от нуля. Поэтому для однозначного определения вектора $\mathbf{d}$ начало координат и было выбрано не произвольно, а помещено в центре масс ядра. Можно было бы думать, что после заряда дипольный момент $\mathbf{d}$ является главной электрической характеристикой ядра в основном состоянии. Однако, по-видимому, центр масс ядра в основном состоянии является и центром симметрии распределения зарядов. Это значит, что каждому заряду в точке $\mathbf{r}$ соответствует равный по модулю и одинаковый по знаку заряд в точке $-\mathbf{r}$. Поэтому электрический дипольный момент ядра в основном состоянии равен нулю. В возбужденном состоянии это, вообще говоря, не так, хотя бы из-за движения нуклонов в ядре, нарушающего симметричное распределение протонов относительно центра масс ядра. (Заметим, что это не относится к магнитному моменту ядра. Классическим аналогом может служить равномерно заряженный шарик, вращающийся вокруг диаметра. В этом случае появляется магнитный дипольный момент, хотя и сохраняется полная симметрия относительно центра шарика.)

3. Из-за отсутствия электрического дипольного момента у ядра в основном состоянии главную роль во взаимодействии его с внешним электрическим полем, после самого заряда, играет третий член в формуле (70.2), определяющий *квадрупольное взаимодействие.* Следующие члены, соответствующие более высоким *мультипольным моментам*, играют малую роль и не учитываются нами. Член же с квадрупольным моментом содержит вторые производные потенциала φ по координатам, а потому квадрупольное взаимодействие, в отличие от дипольного, в однородном электрическом поле не существует.

Преобразуем квадрупольный член в (70.2) к обычно применяемому стандартному виду. Для избежания громоздкости написания формул опустим индекс суммирования α у всех координат частиц. В силу уравнения Лапласа

$$\Delta\varphi \equiv \frac{\partial^2\varphi}{\partial x^2} + \frac{\partial^2\varphi}{\partial y^2} + \frac{\partial^2\varphi}{\partial z^2} \equiv \frac{\partial^2\varphi}{\partial x_k^2} = 0,$$

или

$$\delta_{ik}\frac{\partial^2\varphi}{\partial x_i \partial x_k} = 0,$$

где δ_{ik} — единичный тензор ($\delta_{ik} = 1$ при $i = k$ и $\delta_{ik} = 0$ при $i \neq k$). На основании этого

$$\frac{1}{2}\sum x_i x_k \frac{\partial^2\varphi}{\partial x_i \partial x_k} = \frac{1}{2}\sum (x_i x_k + \lambda\delta_{ik})\frac{\partial^2\varphi}{\partial x_i \partial x_k},$$

где λ — произвольное число. Его удобно выбрать так, чтобы след тензора $(\sum x_i x_k + \lambda\delta_{ik})$, т. е. сумма его диагональных членов $\sum(x_i x_i + \lambda\delta_{ii}) = \sum(r^2 + 3\lambda)$, обратился в нуль. При таком выборе энергия квадрупольного взаимодействия ядра с внешним электрическим полем запишется в виде

$$U_{\text{квад}} = \frac{e}{6}\sum(3x_i x_k - r^2\delta_{ik})\frac{\partial^2\varphi}{\partial x_i \partial x_k},$$

или

$$U_{\text{квад}} = \frac{e}{6}Q_{ik}\frac{\partial^2\varphi}{\partial x_i \partial x_k}, \tag{70.3}$$

где

$$Q_{ik} = \sum(3x_i x_k - r^2\delta_{ik}). \tag{70.4}$$

В компонентах

$$\begin{aligned} Q_{xx} &= \sum(3x^2 - r^2) = \sum(2x^2 - y^2 - z^2),\\ Q_{yy} &= \sum(3y^2 - r^2) = \sum(2y^2 - z^2 - x^2),\\ Q_{zz} &= \sum(3z^2 - r^2) = \sum(2z^2 - x^2 - y^2), \end{aligned} \tag{70.5}$$

$$Q_{xy} = Q_{yx} = \sum 3xy, \quad Q_{yz} = Q_{zy} = \sum 3yz, \quad Q_{zx} = Q_{xz} = \sum 3zx.$$

Тензор Q_{ik} называется *тензором квадрупольного момента ядра.* Он обращается в нуль для сферически симметричного тела. При другом выборе постоянной λ этого бы не получилось, чем и оправдывается сделанный выбор.

Как уже говорилось в § 62, взаимодействие магнитного момента ядра с магнитным полем электронной оболочки атома вызывает сверхтонкую структуру спектральных линий. Однако такое взаимодействие не всегда достаточно для объяснения этого явления. Дополнительной причиной его является квадрупольное взаимодействие атомного ядра с градиентом электрического поля оболочки. Изучение сверхтонкой структуры спектральных линий и дает один из методов определения электрических квадрупольных моментов ядер. Применяются также резонансные радиоспектроскопические методы.

4. Под квадрупольным моментом ядра обычно понимают не самый тензор Q_{ik}, а *значение его наибольшей компоненты* в системе координат, в которой Q_{ik} диагонален. Если за ось Z принять соответствующую главную ось, то

$$Q = \sum(3z^2 - r^2). \tag{70.6}$$

Эта величина имеет размерность площади. Удобной единицей ее является барн, равный 10^{-24} см^2.

Различают *внешний* (или *наблюдаемый*) и *внутренний* (или *собственный*) квадрупольные моменты ядра. Внешним называется квадрупольный момент (обозначаемый через Q), измеренный в лабораторной системе координат. Внутренним называют и обозначают через Q_0 квадрупольный момент, измеренный в системе координат, вращающейся вместе с атомным ядром вокруг его центра масс. Из-за нулевых колебаний оси атомного ядра относительно лабораторной системы координат эти два момента, вообще говоря, не совпадают между собой. Внешний квадрупольный момент есть среднее значение квадрупольного момента ядра в состоянии, которое характеризуется квадратом полного момента импульса ядра $I(I+1)$ и его максимальной проекции I на выделенное направление в пространстве. Поэтому $Q_0 \geqslant Q$. Сверхтонкая структура спектральных линий и радиоспектроскопические методы, упомянутые выше, позволяют экспериментально определить только внешний квадрупольный момент. Зная Q, можно вычислить и внутренний квадрупольный момент Q_0 по формуле

$$Q = Q_0 \frac{I(2I-1)}{(I+1)(2I+1)}, \tag{70.7}$$

которая выводится в квантовой механике. Для этого, конечно, спин ядра I должен быть отличен от 0 и $1/2$. Внешний квадрупольный момент Q ядра со спином 0 или $1/2$ равен нулю. О внутреннем квадрупольном моменте Q_0 в этом случае на основании формулы (70.7) ничего сказать нельзя. Однако существует и прямой метод измерения Q_0. Собственный квадрупольный момент является мерой отклонения распределения электрического заряда в ядре от сферического.

Многие ядра обладают осью симметрии вращения и имеют плоскость симметрии, перпендикулярную к этой оси и проходящую через центр масс ядра. Обычно принимают, что ядро имеет форму эллипсоида вращения. Квадрупольный момент ядра положителен, если оно имеет вытянутую форму, и отрицателен для сплющенного ядра. Несферичность формы ядра проявляется в появлении в энергетическом спектре ядра *вращательных энергетических уровней.* Они возникают из-за вращения вокруг оси, перпендикулярной к аксиальной оси ядра. Более сложные ядра в основном состоянии могут иметь форму трехосного эллипсоида. У таких ядер энергетическая структура уровней усложняется.

Таблица 8

Внешние квадрупольные моменты некоторых атомных ядер

Ядро	Q, 10^{-24} см^2	Ядро	Q, 10^{-24} см^2	Ядро	Q, 10^{-24} см^2
$^{2}_{1}$H	+0,0027	$^{79}_{35}$Br	+0,33	$^{138}_{57}$La	+2,7
$^{10}_{5}$B	+0,074	$^{85}_{37}$Rb	+0,27	$^{175}_{71}$Lu	+5,9
$^{14}_{7}$N	+0,0071	$^{91}_{40}$Zr	−0,46	$^{179}_{72}$Hf	+3
$^{17}_{8}$O	−0,027	$^{93}_{41}$Nb	−0,3	$^{181}_{73}$Ta	+6
$^{35}_{17}$Cl	−0,084	$^{99}_{43}$Tc	+0,34	$^{204}_{81}$Bi	−0,19
$^{37}_{17}$Cl	−0,066	$^{115}_{49}$In	+1,198	$^{235}_{92}$U	+4,0
$^{69}_{31}$Ga	+0,243	$^{119}_{50}$Sn	−0,08	$^{241}_{95}$Am	+4,9

В табл. 8 приведено несколько значений экспериментально найденных внешних квадрупольных моментов ядер. У некоторых из них величины Q аномально велики и намного превосходят квадрат радиуса ядра R^2. Это указывает на значительное отклонение формы таких ядер от сферической симметрии.

Глава IX

РАДИОАКТИВНОСТЬ

§ 71. Введение

1. *Радиоактивность* есть самопроизвольное изменение состава атомного ядра, происходящее за время, существенно большее характерного ядерного времени (10^{-22} с). Несколько произвольно условились считать, что изменение состава ядра должно происходить не раньше чем через 10^{-12} с после его рождения. Распады ядер часто происходят значительно быстрее, но такие распады уже не принято относить к радиоактивным. Так, при бомбардировке различных мишеней на ускорителях может возникать огромное многообразие ядер, которые мгновенно распадаются и по этой причине не могут считаться сложившимися атомными ядрами. Время 10^{-12} с, ничтожное с общежитейской точки зрения, в ядерных масштабах должно считаться очень большим. За такое время совершается множество внутриядерных процессов и ядро успевает полностью сформироваться.

Ядерные силы, действующие между нуклонами, удерживают нуклоны в ядре, поскольку эти силы являются силами притяжения. Противоположное — дестабилизирующее — действие оказывают отталкивающие кулоновские силы между протонами ядра. Под действием сил обоего рода главным образом и формируются атомные ядра.

Ядра, подверженные радиоактивным превращениям, называются *радиоактивными*, а не подверженные — *стабильными*. Такое деление условно, так как, в сущности, все ядра могут самопроизвольно распадаться, но этот процесс в разных ядрах идет с различной скоростью.

Большая часть радиоактивных ядер получена искусственно путем бомбардировки мишеней различными частицами. В этом случае образующееся радиоактивное ядро можно рассматривать как *долгоживущее составное ядро*, подверженное дальнейшему превращению. Никакой принципиальной разницы между радиоактивным распадом и превращением составного ядра нет. *Искусственная радиоактивность* впервые (в 1934 г.) наблюдалась супругами Ирен (1897–1956) и Фредериком (1900–1958) Жолио-Кюри.

Различают следующие виды радиоактивного распада: 1) *α-распад*; 2) *β-распад*; 3) *спонтанное деление* атомного ядра; 4) *протонный распад*; 5) двухпротонный распад и др.

При α-распаде из ядра спонтанно вылетает α-частица ($^4_2\mathrm{He}$). При этом зарядовое число ядра уменьшается на две единицы и образуется новый элемент, сдвинутый относительно исходного влево на две клетки периодической системы.

Бета-распад может быть трех видов: 1) *электронный*, или β^--*распад*; 2) *позитронный*, или β^+-*распад*; 3) *электронный захват*. При β^--распаде из ядра вылетают электрон и электронное антинейтрино $\overline{\nu}_e$ (см. § 74, п. 8). Массовое число ядра не меняется, а зарядовое число возрастает на единицу (сдвиг в периодической системе вправо на одну клетку). При β^+-распаде из ядра вылетают позитрон e^+ и электронное нейтрино $\overline{\nu}_e$ (сдвиг в периодической системе влево на одну клетку без изменения массового числа). Позитронная радиоактивность была открыта в 1934 г. И. и Ф. Жолио-Кюри практически одновременно с открытием искусственной радиоактивности. При электронном захвате ядро захватывает электрон из электронной оболочки атома; зарядовое число, как и при β^+-распаде, уменьшается на единицу, а массовое число не меняется. Если электрон захватывается с K-оболочки атома, то электронный захват называют также *K-захватом.*

В 1940 г. Г.И. Флеровым и К.А. Петржаком было открыто спонтанное деление ядер урана. Примером может служить процесс

$$^{238}_{92}\mathrm{U} \to {}^{139}_{54}\mathrm{Xe} + {}^{96}_{38}\mathrm{Sr} + 3\,{}^{1}_{0}\mathrm{n}.$$

У ядер с большим избытком протонов, в принципе, возможна *протонная* и даже *двухпротонная радиоактивности*, но эти процессы очень трудно обнаружить из-за сильного фона конкурирующих α- и β^+-распадов, приводящих к образованию изотопов таких же химических элементов. Протонная радиоактивность возможна лишь у небольшого числа искусственно получаемых легких ядер с относительно короткими временами жизни, которые обладают большим избытком протонов. Тем не менее, протонная и двухпротонная радиоактивности обнаружены. По мнению В.И. Гольданского (р. 1923), двухпротонная радиоактивность даже более вероятна, чем однопротонная. Дело в том, что между протонами действуют силы спаривания. В результате из ядра может вылететь не только одиночный протон, но и бипротон, т. е. два спаренных протона с противоположно направленными спинами.

Из ядра могут вылетать и нуклоны, объединенные в более крупные частицы, чем ядра ^{4}He (α-частицы), например ядра углерода. Такой процесс затруднен тем, что в исходном ядре нет готовых ядер углерода. Впрочем, и готовых α-частиц в ядре также нет — они образуются перед самым вылетом из ядра. Только формирование α-частиц несравненно более вероятно, чем формирование ядер углерода. Тем не менее, в 1984 г. в Оксфордском университете было зафиксировано самопроизвольное превращение нестабильных изотопов радия в свинец с испусканием ядер изотопа углерода ^{14}C:

$$^{222}_{88}\mathrm{Ra} \to {}^{208}_{82}\mathrm{Pb} + {}^{14}_{6}\mathrm{C},$$

$$^{223}_{88}\mathrm{Ra} \to {}^{209}_{82}\mathrm{Pb} + {}^{14}_{6}\mathrm{C},$$

$$^{224}_{88}\mathrm{Ra} \to {}^{210}_{82}\mathrm{Pb} + {}^{14}_{6}\mathrm{C}.$$

Все ядра изотопов ^{222}Ra, ^{223}Ra и ^{224}Ra радиоактивны с периодами полураспадов соответственно 39 с, 11,7 дня и 3,6 дня. Вероятность вылета ядра ^{14}C примерно в 10^{10} раз меньше вероятности вылета α-частицы. Поэтому экспериментально зафиксировать распад рассматриваемых ядер радия с вылетом ядер ^{4}C необычайно трудно. То обстоятельство, что одним из продуктов распада является изотоп углерода ^{14}C, а не ^{12}C, можно объяснить тем, что энергетически выгодно, чтобы образовалось дважды магическое ядро $^{208}_{82}$Pb или соседние с ним ядра (см. § 78). В 1985 г. в Дубне, а также группой американских физиков были открыты распады с излучением еще более тяжелых ядер — ядер неона:

$$^{233}_{92}\mathrm{U} \to {}^{209}_{82}\mathrm{Pb} + {}^{24}_{10}\mathrm{Ne}, \quad {}^{233}_{92}\mathrm{U} \to {}^{208}_{82}\mathrm{Pb} + {}^{25}_{10}\mathrm{Ne},$$

$$^{232}_{92}\mathrm{U} \to {}^{208}_{82}\mathrm{Pb} + {}^{24}_{10}\mathrm{Ne}, \quad {}^{231}_{91}\mathrm{U} \to {}^{207}_{81}\mathrm{Tl} + {}^{24}_{10}\mathrm{Ne}.$$

Вероятность первых трех процессов примерно в 10^{12}, а последнего — в 10^{11} раз меньше вероятности α-распада.

В результате радиоактивного распада ядер, образующихся при делении материнского ядра, образуются нейтроноизбыточные возбужденные ядра. При их распаде образуются так называемые *запаздывающие нейтроны.* Такой процесс соблазнительно назвать *нейтронной радиоактивностью.* Однако это не делается, так как снятие возбуждения путем испускания нейтронов происходит практически мгновенно.

2. В процессе радиоактивного распада, конечно, должен выполняться закон сохранения энергии. Если начальное ядро неподвижно, то этот закон можно записать в виде

$$M_{\text{нач}}c^2 = M_{\text{кон}}c^2 + \sum_i M_i c^2 + \mathscr{E}, \tag{71.1}$$

где $M_{\text{нач}}$ и $M_{\text{кон}}$ — массы начального и конечного ядер, M_i — массы образовавшихся частиц, $\mathscr{E}$ — кинетическая энергия, выделившаяся при радиоактивном распаде. Самопроизвольно могут идти только реакции с выделением энергии, а потому энергия $\mathscr{E}$ существенно положительна. Таким образом, радиоактивный распад возможен лишь тогда, когда $M_{\text{нач}} > M_{\text{кон}} + \sum M_i$. Это условие необходимо, но не достаточно, так как для возможности процесса, помимо закона сохранения энергии, должны выполняться и другие законы сохранения (импульса, момента импульса, электрического, барионного и лептонного зарядов и пр.).

3. Все α- и β-радиоактивные элементы можно разделить на четыре радиоактивных ряда, или радиоактивных семейства. В каждом ряду массовое число A выражается формулой

$$A = 4n + C, \tag{71.2}$$

где C — постоянная для рассматриваемого ряда величина, а n — переменное целое число. Каждый элемент ряда получается из предыдущего путем α- или β-превращения. Поэтому в каждом ряду два соседних элемента имеют либо одинаковые массовые числа, либо эти

числа отличаются на четыре. Значению $C = 0$ ($n \leqslant 59$) соответствует ряд тория, $C = 1$ ($n \leqslant 60$) — ряд нептуния, $C = 2$ ($n \leqslant 60$) — ряд урана, $C = 3$ ($n \leqslant 59$) — ряд актиноурана. Ряд с $C = 1$ (нептуния) состоит из изотопов, не встречающихся в природе, но получающихся искусственно.

4. Все четыре радиоактивных ряда представлены в табл. 9. Рассмотрим в качестве примера ряд урана. Он начинается с α-радиоактивного изотопа урана ^{238}U [1]. Испытав α-распад с периодом полураспада $4{,}56 \cdot 10^9$ лет, этот изотоп переходит в β^--активный изотоп тория $^{234}_{90}$Th. После последовательных двух β^--распадов последнего с периодами полураспада 24 дня и 6,74 часа получается другой радиоактивный изотоп урана $^{234}_{92}$U с периодом полураспада $2{,}48 \cdot 10^5$ лет. Он последовательно претерпевает пять α-превращений. Среди радиоактивных изотопов, получающихся в результате этих превращений, содержится и α-радиоактивный изотоп радия $^{226}_{88}$Ra с периодом полураспада 1617 лет, за которым следует радиоактивный газ радон $^{222}_{86}$Rn, и т. д. Заметим, что радиоактивные изотопы $^{214}_{83}$Bi и $^{210}_{83}$Bi могут переходить соответственно в $^{210}_{82}$Pb и в $^{206}_{82}$Pb двумя различными путями, указанными в табл. 9. Ряд урана заканчивается стабильным изотопом свинца $^{206}_{82}$Pb. Аналогичное строение имеют и остальные радиоактивные ряды. В естественных рядах тория и актиноурана содержатся другие изотопы радия и радона. Все естественные радиоактивные ряды заканчиваются различными стабильными изотопами свинца $^{208}_{82}$Pb, $^{206}_{82}$Pb, $^{207}_{82}$Pb. Существование трех различных стабильных изотопов свинца указывает на особую устойчивость ядер свинца, содержащих магическое число протонов — 82.

На самом тяжелом элементе, встречающемся на Земле в естественных условиях, уране $_{92}$U периодическая система элементов, не кончается, как об этом уже указывалось в § 47. Однако все трансурановые, т. е. заурановые, элементы радиоактивны. Период полураспада наиболее долгоживущего изотопа плутония $^{239}_{94}$Pu, например, составляет 24 000 лет. Если такие элементы и существовали когда-то на Земле в естественных условиях, то за время существования Земли они полностью исчезли из-за радиоактивных превращений. Все трансурановые элементы получаются искусственным путем. С учетом этих элементов радиоактивные ряды, представленные в табл. 9, могут быть продолжены в сторону больших Z. И все элементы тяжелее свинца и висмута, в конце концов, правда в отдаленном будущем, должны исчезнуть на Земле, если только к этому времени сама Земля еще будет существовать. Наличие в настоящее время таких элементов является одним из подтверждений ограниченного возраста Земли, который по различным оценкам составляет приблизительно 4–4,5 млрд лет.

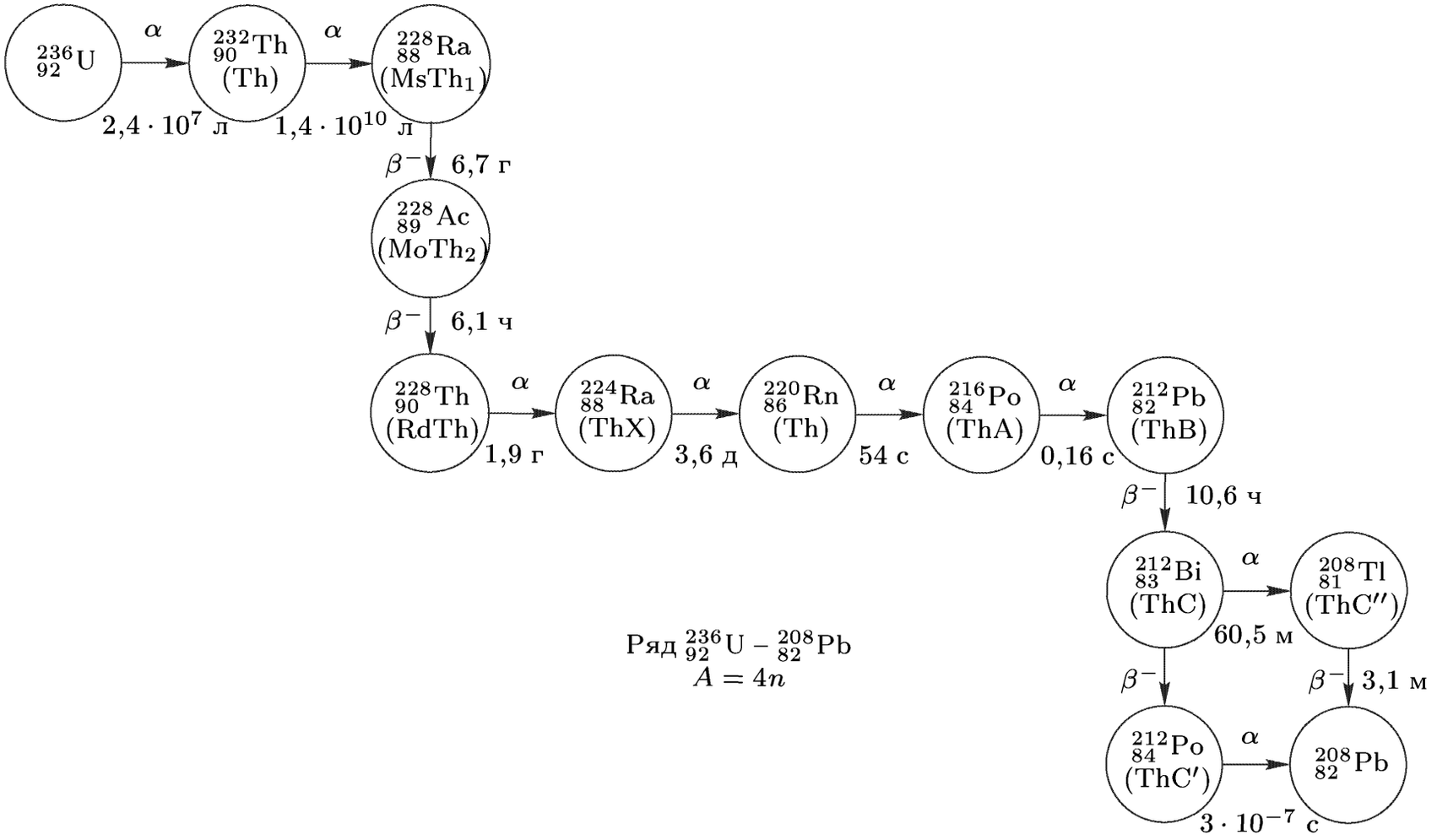
О способах получения трансурановых элементов будет сказано в § 94.

---

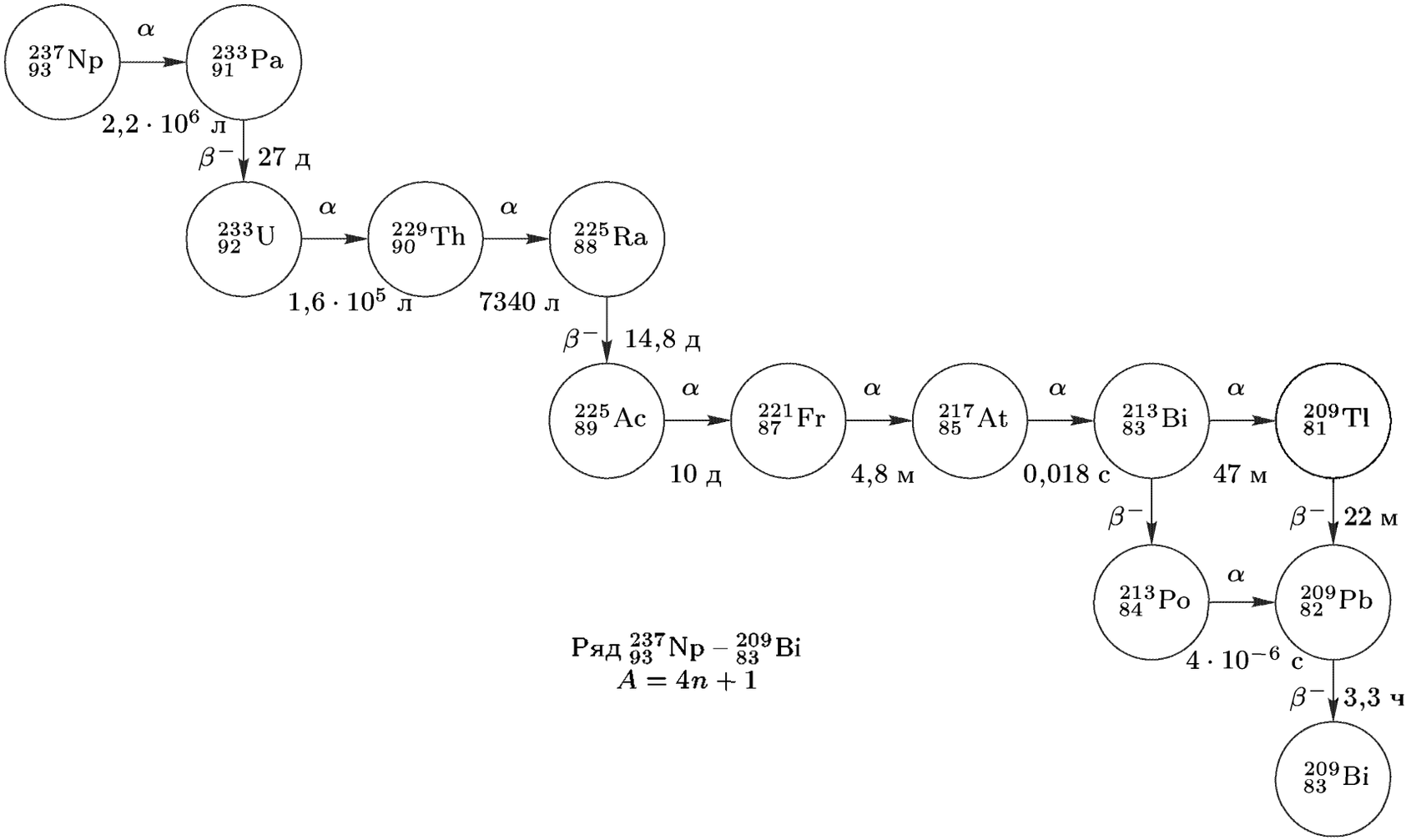
[1]) В табл. 9 в скобках указаны обозначения изотопов, использующиеся в радиохимии.

Таблица 9

Радиоактивные семейства

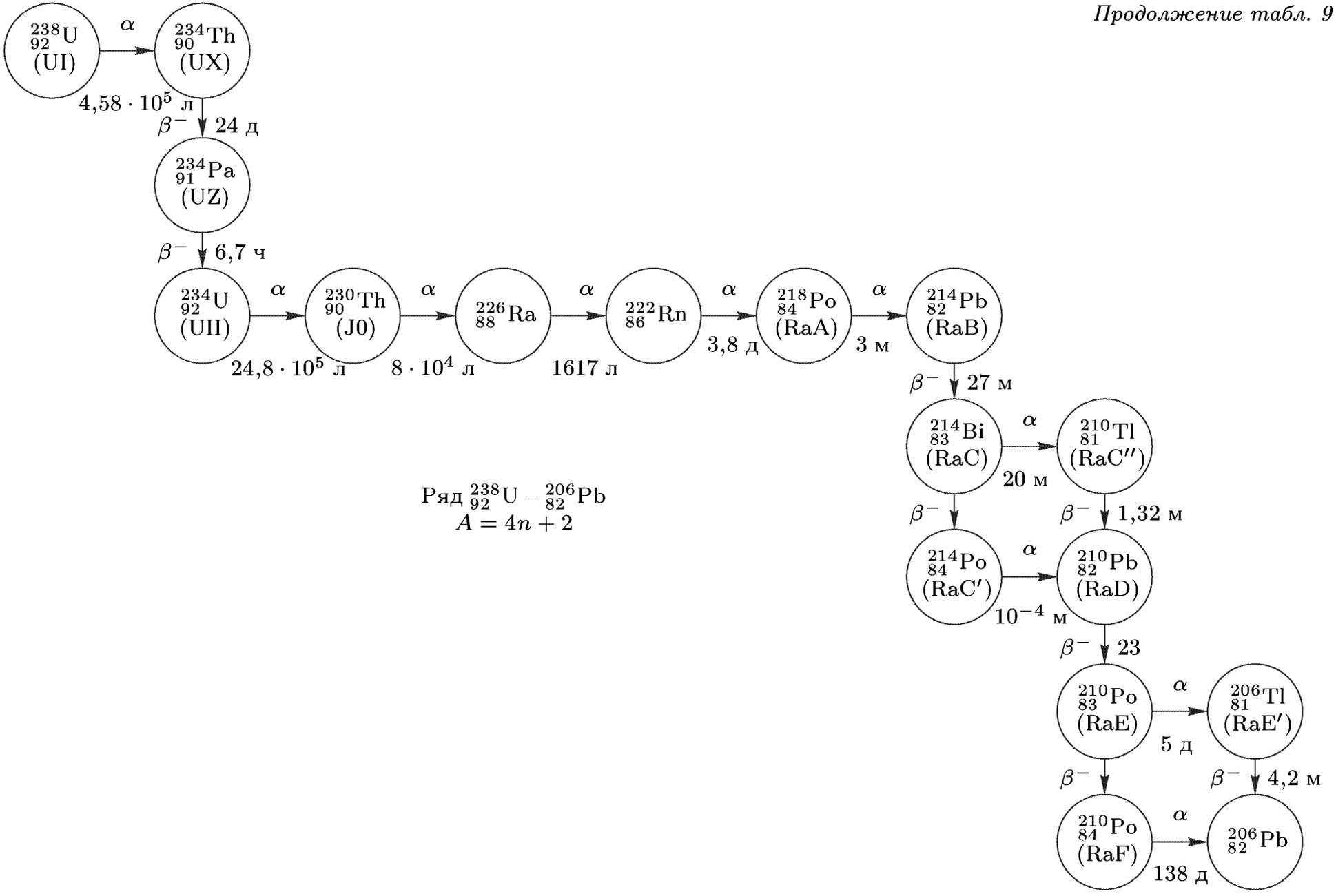


Ряд $^{236}_{92}\mathrm{U} - {}^{208}_{82}\mathrm{Pb}$
$A = 4n$

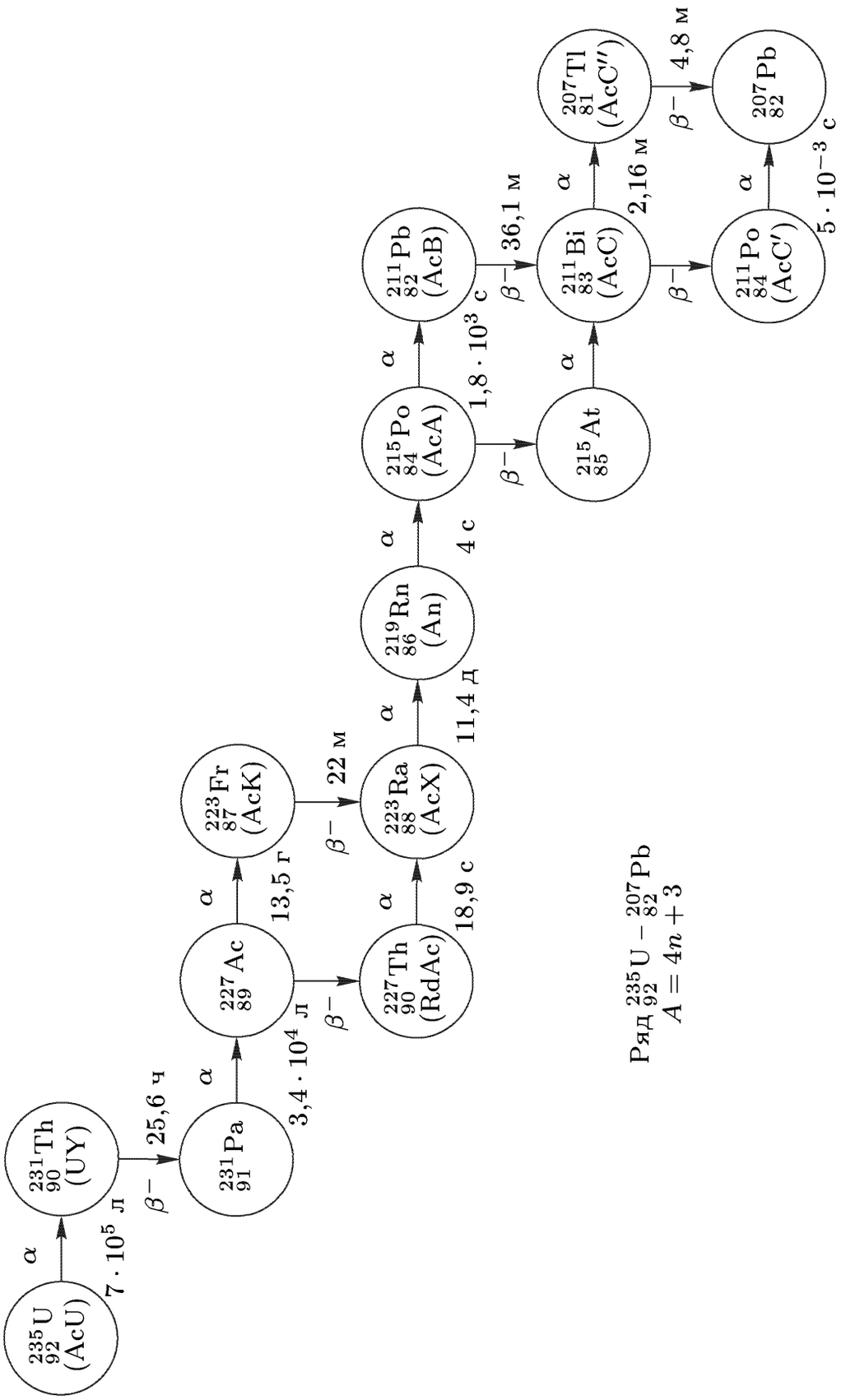


Ряд $^{237}_{93}\mathrm{Np} - {}^{209}_{83}\mathrm{Bi}$
$A = 4n + 1$

Продолжение табл. 9



Ряд $^{238}_{92}\mathrm{U} - {}^{206}_{82}\mathrm{Pb}$
$A = 4n + 2$


$^{235}_{92}$U (AcU)
α
$7 \cdot 10^5$ л
$^{231}_{90}$Th (UY)
β−
25,6 ч
$^{231}_{91}$Pa
α
$3{,}4 \cdot 10^4$ л
$^{227}_{89}$Ac
β−
$^{227}_{90}$Th (RdAc)
α
18,9 с
$^{223}_{87}$Fr (AcK)
α
13,5 г
β−
22 м
$^{223}_{88}$Ra (AcX)
α
11,4 д
$^{219}_{86}$Rn (An)
α
4 с
$^{215}_{84}$Po (AcA)
α
$1{,}8 \cdot 10^3$ с
$^{211}_{82}$Pb (AcB)
β−
36,1 м
$^{215}_{85}$At
α
$^{211}_{83}$Bi (AcC)
α
2,16 м
β−
$^{211}_{84}$Po (AcC′)
α
$5 \cdot 10^{-3}$ с
$^{207}_{81}$Tl (AcC″)
β−
4,8 м
$^{207}_{82}$Pb
Ряд $^{235}_{92}$U – $^{207}_{82}$Pb
$A = 4n + 3$

§ 72. Законы радиоактивного распада

1. Радиоактивный распад — *явление статистическое.* Все предсказания, которые могут быть сделаны на основе законов радиоактивного распада, носят принципиально *вероятностный характер.* Нельзя сказать, какие атомы в радиоактивном образце распадутся за рассматриваемое время. Но можно практически с полной достоверностью предсказать, сколько атомов распадется за это время. Например, в случае радона половина атомов распадется за 3,8 дня. И это вероятностное предсказание будет выполняться тем точнее, чем с большим количеством радона имеют дело.

Вероятность распада ядра за единицу времени называется *постоянной распада* λ радиоактивных ядер данного сорта. Это значит, что из N имеющихся радиоактивных ядер за единицу времени в среднем распадается λN, а за время dt — $\lambda N\,dt$ ядер. Величина λN называется *активностью* радиоактивного источника (радиоактивностью). Старейшей, до сих пор наиболее употребительной единицей радиоактивности является *кюри* (Ки) и ее дольные единицы: *милликюри* (1 мКи $= 10^{-3}$ Ки) и *микрокюри* (1 мкКи $= 10^{-6}$ Ки). По первоначальному определению кюри есть активность одного грамма изотопа радия $^{226}_{88}$Ra. Однако для удобства измерений это определение в дальнейшем было заменено следующим:

$$1\text{ Ки} = 3{,}700 \cdot 10^{10}\text{ расп/с}\ \ \text{(точно)}.$$

Активность же грамма радия лишь приближенно составляет 1 Ки. Естественной единицей активности является 1 распад в секунду. Эта единица получила название *беккерель* (Бк) и принята в Международной системе СИ. В литературе употребляется также единица *резерфорд*: 1 Рд $= 10^6$ Бк.

Поскольку радиоактивные превращения совершаются внутри ядра, внешние условия (температура, давление, химические реакции и пр.) на ход радиоактивных превращений практически не оказывают никакого влияния. Во всяком случае такое влияние не удалось обнаружить самыми точными способами, которыми располагала физика до открытия эффекта Мёссбауэра. В частности, не удавалось обнаружить зависимости от внешних условий постоянной радиоактивного распада λ. Исключением являлся только e-захват. Для него еще до использования мёссбауэровской спектроскопии была обнаружена очень слабая зависимость величины λ от внешних условий. Но в этом случае явление определяется не только тем, что происходит внутри ядра, но и в ближайших к нему участках электронной оболочки. Только методами мёссбауэровской спектроскопии (см. § 76) удалось отчетливо обнаружить влияние электронной оболочки атома на явления, происходящие внутри атомного ядра. Но в громадном большинстве случаев это влияние не играет никакой роли. Постоянная λ не зависит и от времени. Образно говоря, радиоактивные ядра могут только умирать, но они никогда не стареют.

2. После этих замечаний сформулируем *основной закон радиоактивного распада.* Пусть N — число (очень большое) радиоактивных ядер в момент времени t, а $N + dN$ — в более поздний момент $t + dt$. Величина dN отрицательна, поскольку ядра могут только распадаться, т. е. число их убывает. На основании изложенного выше

$$dN = -\lambda N\, dt. \tag{72.1}$$

Поскольку λ не зависит от времени, после интегрирования получаем

$$N = N_0 e^{-\lambda t}, \tag{72.2}$$

т. е. число нераспавшихся ядер убывает во времени экспоненциально. Время t можно отсчитывать от любого момента, принимаемого за начальный. Постоянная N_0 означает число нераспавшихся ядер в начальный момент времени. Формула (72.2) и выражает основной закон радиоактивного распада. Разумеется, она относится к тем атомам радиоактивного вещества, которые могут только распадаться, но не могут появляться или исчезать в результате каких-либо других процессов.

Постоянную распада λ можно выразить через *среднее время жизни* радиоактивного ядра. Так как за промежуток времени между t и $t + dt$ распадается $-dN$ ядер, то можно сказать, что каждое из этих ядер «живет» время t, считая от начала отсчета времени. Суммарное время жизни этих $-dN$ ядер составляет $-t\, dN$, а суммарное время жизни всех N_0 ядер определяется интегралом

$$-\int\limits_{N_0}^{0} t\, dN = \lambda \int\limits_{0}^{\infty} tN\, dt = \lambda N_0 \int\limits_{0}^{\infty} te^{-\lambda t}\, dt = \frac{N_0}{\lambda}.$$

Таким образом, среднее время жизни одного радиоактивного ядра будет $\tau = (N_0/\lambda) : N_0$, т. е.

$$\tau = 1/\lambda. \tag{72.3}$$

Оно не зависит от выбора начала отсчета времени. Это вполне естественно, поскольку все моменты времени в отношении радиоактивного распада полностью равноправны. Различные моменты характеризуются различными значениями полного числа радиоактивных ядер N, но относительное число ежесекундно распадающихся ядер $-\dot{N}/N$ одно и то же для всех моментов времени и равно постоянной распада λ. Заметим еще, что время τ для сокращения обычно называют просто *временем жизни* ядра, опуская прилагательное «среднее».

С введением времени жизни формула (72.1) представится в виде

$$N = N_0 e^{-t/\tau}. \tag{72.4}$$

Время $T_{1/2}$, по истечении которого число наличных радиоактивных атомов убывает в два раза, называется *периодом* или *временем полураспада.* Для его определения на основании (72.4) получим

$$N = N_0/2 = N_0 e^{-T_{1/2}/\tau},$$

откуда

$$T_{1/2} = \tau \ln 2 = 0{,}6931\tau. \tag{72.5}$$

Если одновременно происходят два конкурирующих процесса, так что ядра N могут одновременно испускать частицы одного сорта N_1, согласно уравнению $dN_1 = -\lambda_1 N_1\, dt$, и частицы другого сорта N_2, согласно уравнению $dN_2 = -\lambda_2 N_2\, dt$, то

$$dN = dN_1 + dN_2 = -(\lambda_1 + \lambda_2) N\, dt.$$

Отсюда следует, что обратная величина «результирующего» времени жизни τ равна сумме обратных величин времен жизни τ_1 и τ_2 обоих конкурирующих процессов:

$$1/\tau = 1/\tau_1 + 1/\tau_2.$$

3. При радиоактивном распаде ядер исходного вещества могут возникать новые радиоактивные ядра. В таком случае первые ядра называются *материнскими*, а вторые — *дочерними*. Обозначим числа этих ядер соответственно через N_1 и N_2, а их постоянные распада — через λ_1 и λ_2. Тогда изменения N_1 и N_2 будут описываться уравнениями

$$\frac{dN_1}{dt} = -\lambda_1 N_1, \quad \frac{dN_2}{dt} = \lambda_1 N_1 - \lambda_2 N_2. \tag{72.6}$$

Первое из этих уравнений с точностью до обозначений полностью совпадает с (72.1), поскольку число N_1 может только убывать за счет радиоактивного распада материнских ядер. При этом из каждого материнского ядра возникает дочернее ядро. Это обстоятельство учитывается первым слагаемым в правой части второго уравнения системы (72.6). Другое же слагаемое $(-\lambda_2 N_2)$ учитывает убыль дочерних ядер из-за их радиоактивного распада.

Если дочерние ядра также радиоактивны, то при их распаде возникают новые ядра, число которых обозначим через N_3, а постоянную распада — через λ_3. В этом случае к системе уравнений (72.6) добавляется третье уравнение

$$\frac{dN_3}{dt} = \lambda_2 N_2 - \lambda_3 N_3 \tag{72.7}$$

и т. д.

Важнейшим является случай системы уравнений (72.6), когда рассматриваются только материнские и соответствующие им дочерние также радиоактивные ядра. Этим случаем мы и ограничимся. Решение системы уравнений (72.6) имеет вид

$$N_1 = N_{10} e^{-\lambda_1 t}, \tag{72.8}$$

$$N_2 = N_{10} \frac{\lambda_1}{\lambda_2 - \lambda_1}\, e^{-\lambda_1 t} + \left(N_{20} - N_{10} \frac{\lambda_1}{\lambda_2 - \lambda_1} \right) e^{-\lambda_2 t}, \tag{72.9}$$

где N_{10} и N_{20} — начальные значения чисел атомов N_1 и N_2 материнского и дочернего вещества. В частном случае, когда в начальный момент дочернее вещество еще не образовалось ($N_{20} = 0$), формула (72.9) упрощается и переходит в

$$N_2 = N_{10} \frac{\lambda_1}{\lambda_2 - \lambda_1} (e^{-\lambda_1 t} - e^{-\lambda_2 t}). \tag{72.10}$$

Полное число атомов $N_1 + N_2$, как видно из формул (72.8) и (72.10), не сохраняется, если только дочерние ядра испытывают распад ($\lambda_2 \neq 0$). Но если дочерние ядра не распадаются ($\lambda_2 = 0$), то из тех же формул получается $N_1 + N_2 = N_{10} = \text{const}$, т. е. полное число атомов $N_1 + N_2$ сохраняется. Тот же результат немедленно получается, если почленно сложить уравнения (72.6) и учесть, что $\lambda_2 = 0$. Аналогично, если не распадаются ядра, возникающие из дочерних ($\lambda_3 = 0$), то сложением уравнений (72.6) и (72.7) получим $N_1 + N_2 + N_3 = \text{const}$, и т. д.

Особенно важным является случай, когда материнское вещество — долгоживущее, а дочернее вещество по сравнению с ним распадается быстро ($\lambda_1 \ll \lambda_2$), причем время наблюдения t пренебрежимо мало по сравнению со временем жизни τ_1 материнского вещества ($\lambda_1 t \ll 1$). За это время изменением N_1 можно пренебречь, т. е. N_1 считать величиной постоянной. В таком случае из (72.8) и (72.10) получается

$$N_1 = \text{const}, \quad N_2 = N_1 \frac{\lambda_1}{\lambda_2}(1 - e^{-\lambda_2 t}), \tag{72.11}$$

так как величиной λ_1 в знаменателе формулы (70.10) можно пренебречь. Число атомов N_2 при $t \to \infty$ асимптотически стремится к насыщению $N_2(\infty) = N_1\lambda_1/\lambda_2$. Насыщение наступает практически через промежуток времени $\tau \approx 3\tau_2$ (рис. 126). Таким образом, в состоянии насыщения выполняется условие

$$\lambda_1 N_1 = \lambda_2 N_2. \tag{72.12}$$

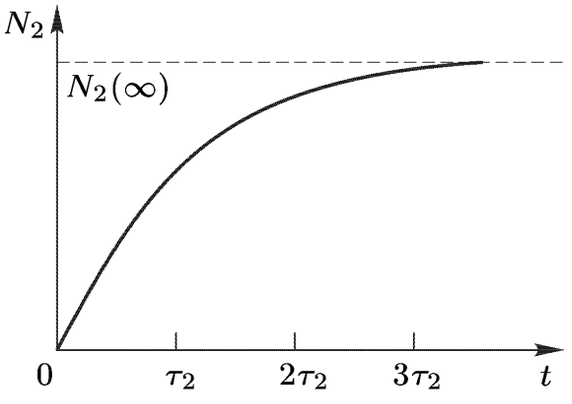


Рис. 126

Это равенство называют также *условием радиоактивного равновесия.* Как видно из (72.6), физический смысл его состоит в том, что распад атомов дочернего вещества в любой момент времени компенсируется увеличением их числа за счет распада атомов материнского вещества.

4. Статистический закон радиоактивного распада при наличии очень большого количества радиоактивных атомов — практически абсолютно точный закон. На его принципе работают «атомные часы», служащие в некоторых случаях, например в геологии или археологии, для измерения промежутков времени. Опишем два применения радиоактивности с этой целью.

Для оценки возраста мертвых организмов (древесины, костей животных и пр.), не превышающего примерно 50 000 лет, используется радиоактивный изотоп углерода $^{14}_{6}C$. Испытывая β-распад, он превращается в азот $^{14}_{7}N$. Период полураспада для $^{14}_{6}C$ составляет 5800 лет. Зная первоначальное количество атомов $^{14}_{6}C$ в образце и измерив оставшееся количество их, можно вычислить время, прошедшее с момента появления образца. Радиоактивный изотоп $^{14}_{6}C$ образуется в верхних слоях атмосферы при столкновениях нейтронов, образовавшихся под

действием космических лучей, с ядрами атомов азота $^{14}_{7}\mathrm{N}$, составляющими основную часть атомов воздуха (см. § 103, п. 12). Образовавшийся углерод $^{14}_{6}\mathrm{C}$ быстро попадает в нижние слои атмосферы, где перемешивается с обычным углеродом $^{12}_{6}\mathrm{C}$. Обычный нерадиоактивный углерод $^{12}_{6}\mathrm{C}$ поглощается животными и растениями, а вместе с ним поглощается и небольшое количество радиоактивного изотопа $^{14}_{6}\mathrm{C}$. Можно считать, что за времена геологического порядка интенсивность космических лучей в земной атмосфере не изменилась. А так как по сравнению с этими временами период полураспада $^{14}_{6}\mathrm{C}$ (5800 лет) относительно мал, то в земной атмосфере установилось равновесное соотношение между радиоактивным $^{14}_{6}\mathrm{C}$ и нерадиоактивным $^{12}_{6}\mathrm{C}$ изотопами углерода, когда вместо каждого распадающегося радиоактивного атома $^{14}_{6}\mathrm{C}$ космические лучи в среднем порождают такой же новый атом. Это соотношение примерно одинаково и в живом организме, поскольку последний частично состоит из атмосферного углерода. После гибели организма он, естественно, не в состоянии больше поглощать ни углерод $^{12}\mathrm{C}$, ни углерод $^{14}\mathrm{C}$. При этом количество углерода $^{12}\mathrm{C}$, накопленного организмом в течение времени жизни, остается неизменным, тогда как половина атомов $^{14}\mathrm{C}$ убывает за каждые 5800 лет. По меняющемуся соотношению между количествами углерода $^{12}\mathrm{C}$ и $^{14}\mathrm{C}$ и можно относительно точно определить возраст мертвого организма.

Второй пример касается определения возраста Земли. Принципиальное (но численно грубое) решение его было дано еще на заре исследований явления радиоактивности. «Атомными часами», пригодными для решения подобных вопросов, могут служить долгоживущие ядра $^{238}\mathrm{U}$ (период полураспада 4,56 млрд лет) и $^{232}\mathrm{Th}$ (период полураспада 14 млрд лет). Конечными продуктами их радиоактивного распада являются соответственно стабильные изотопы свинца $^{206}\mathrm{Pb}$ и $^{208}\mathrm{Pb}$. Они называются *радиогенными* в отличие от так называемого изначального свинца $^{204}\mathrm{Pb}$, не являющегося конечным продуктом радиоактивных превращений. Если ввести предположение, что весь радиогенный свинец получился в результате радиоактивного распада урана и тория, то можно вычислить возраст Земли. Для надежного вычисления надо точно измерить количество различных изотопов радиогенного свинца, содержащихся, например, в радий-урановых рудах. В настоящее время такой метод дает для возраста Земли приблизительно 4,5 млрд лет. Конечно, в основе этого метода лежит *предположение*, что в момент возникновения Земли на ней не существовало радиогенного свинца. Однако определение возраста Земли, основанное на этом предположении, хорошо согласуется с другими методами.

ЗАДАЧИ

1. Через равные малые промежутки времени производится счет α-частиц долгоживущего радиоактивного препарата. Найти вероятность P_n того, что в одном из этих промежутков времени будет зарегистрировано n α-частиц, если среднее число зарегистрированных в одном промежутке времени α-частиц равно $\overline{n}$.

Р е ш е н и е. Пусть за длительный промежуток времени радиоактивный препарат испустил всего N α-частиц, причем за это время его количество практически не изменилось. Обозначим через p вероятность того, что атом радиоактивного вещества испустил α-частицу в рассматриваемый промежуток времени. Тогда вероятность испускания во все остальные промежутки времени будет $1-p$. Искомая вероятность определится соотношением

$$P_n = \frac{N!}{n!(N-n)!}\, p^n (1-p)^{N-n} = \frac{N!}{n!(N-n)!} \left(\frac{\overline{n}}{N}\right)^n \left(1-\frac{\overline{n}}{N}\right)^{N-n}.$$

При $N \to \infty$ это выражение асимптотически переходит в формулу Пуассона

$$P_n = \frac{(\overline{n})^n}{n!}\, e^{-\overline{n}}. \tag{72.13}$$

В самом деле,

$$P_n = \left[\frac{N(N-1)\ldots(N-n+1)}{N^n}\right] \frac{\overline{n}^n}{n!} \left(1-\frac{\overline{n}}{N}\right)^{-(N/\overline{n})\overline{n}(n-N)/N}.$$

Асимптотически при $N \to \infty$

$$\left(1-\frac{\overline{n}}{N}\right)^{-N/\overline{n}} = e,$$

выражение в квадратных скобках стремится к 1, а дробь $(n-N)/N$ стремиться к -1. В результате получается формула (72.13).

2. В настоящее время в природном уране содержится 99,28 % ^{238}U и 0,72 % ^{235}U. Вычислить возраст Земли в предположении, что в момент образования Земли количества ^{238}U и ^{235}U были одинаковыми.

О т в е т. $t = \dfrac{\ln\left(N^{238}/N^{235}\right)}{\lambda^{235}-\lambda^{238}} \approx \dfrac{1}{\lambda^{235}} \ln\dfrac{N^{238}}{N^{235}} \approx \tau^{235} \ln\dfrac{N^{238}}{N^{235}} =$

$= T_{1/2}^{235} \dfrac{\ln\left(N^{238}/N^{235}\right)}{\ln 2} \approx 5{,}5 \cdot 10^9$ лет.

3. Период полураспада ^{234}U равен $T_{1/2}^{234} = 2{,}48 \cdot 10^5$ лет. Какое количество атомов ^{234}U осталось бы на Земле в настоящее время, если бы происходил только процесс радиоактивного распада этого элемента? Как объяснить, что в природном уране содержится примесь ^{234}U в количестве 0,055 %? Возраст Земли $t = 4{,}5 \cdot 10^9$ лет.

О т в е т. $N = N_0 e^{-t/\tau} = N_0 e^{-t \ln 2/T_{1/2}} = N_0 \cdot 10^{-5460}$, где N_0 — количество атомов ^{234}U в момент образования Земли. Если даже предположить, что в этот момент Земля состояла только из ^{234}U, то и тогда на Земле уже давно не осталось бы ни одного атома ^{234}U. Изотоп ^{234}U существует на Земле благодаря α-распаду ^{238}U и β^--распаду ^{234}Th и ^{234}Pa.

§ 73. Альфа-распад

1. Альфа-распад есть *самопроизвольный процесс* испускания ядрами α-частиц, в результате которого массовое число ядра A уменьшается на четыре, а зарядовое число Z уменьшается на два:

$${}^A_Z\mathrm{X} \to {}^{A-4}_{Z-2}\mathrm{X} + {}^4_2\mathrm{He}. \tag{73.1}$$

В настоящее время известно более двухсот α-активных ядер, из которых большинство получается искусственно.

Чтобы α-распад происходил, необходимо (но недостаточно), чтобы энергия связи исходного материнского ядра была меньше суммы энергий связи дочернего ядра и испускаемой α-частицы. При выполнении этого условия кинетическая энергия Q, выделяющаяся при α-распаде, определяется соотношением

$$Q = \mathscr{E}_{\text{св}}(A-4, Z-2) + \mathscr{E}_{\text{св}}(\alpha) - \mathscr{E}_{\text{св}}(Z, A). \qquad (73.2)$$

В основном Q есть кинетическая энергия α-частицы, поскольку масса дочернего ядра всегда много больше массы α-частицы. Предполагается, конечно, что материнское ядро неподвижно.

Альфа-распад возможен только при $Q > 0$ и невозможен в противоположном случае. Возьмем в качестве примера изотопы урана $^{234}_{92}\text{U}$ и тория $^{230}_{90}\text{Th}$. Табличные значения энергий связи этих ядер равны соответственно $\mathscr{E}_{\text{св}}(92{,}234) = 1\,778\,630$ кэВ, $\mathscr{E}_{\text{св}}(90{,}230) = 1\,755\,190$ кэВ. Энергия связи α-частицы $U_{\text{св}}(\alpha) = 28\,296{,}10$ кэВ. По формуле (73.2) находим, что в процессе распада ядра ^{234}U на ядро ^{230}Th и α-частицу выделяется энергия $Q = 4856$ кэВ. Она положительна, а потому указанный процесс энергетически возможен. И действительно он идет: уран ^{234}U превращается в ^{230}Th с испусканием α-частицы. Кинетическая энергия Q распределяется обратно пропорционально массам образовавшихся частиц: α-частица уносит энергию 4773, а ядро ^{230}Th — только 83 кэВ.

Условие $Q > 0$ можно представить в другой форме. Обозначим средние энергии связи, приходящиеся на один нуклон ядра, малыми буквами e. Тогда формула (73.2) преобразуется в

$$Q = (A-4)e_{\text{д}} - Ae_{\text{м}} + 4e_\alpha = A(e_{\text{д}} - e_{\text{м}}) + 4(e_\alpha - e_{\text{д}}),$$

где индексы α, м, д относятся соответственно к α-частице, материнскому и дочернему ядрам. Условие $Q > 0$ приводится к виду

$$e_{\text{д}} - e_{\text{м}} > \frac{4}{A}(e_{\text{д}} - e_\alpha). \qquad (73.3)$$

Для α-частицы $e_\alpha = \mathscr{E}_{\text{св}}(\alpha)/4 = 7074$ кэВ. В приведенном выше примере $e_{\text{м}} = 7601$ кэВ, $e_{\text{д}} = 7631$ кэВ, так что условие (73.3) выполняется.

Для всех ядер, если не считать самые легкие, энергия связи, приходящаяся на один нуклон, заметно превосходит e_α (это видно из только что приведенного примера). Поэтому для возможности α-распада необходимо выполнение условия $e_{\text{д}} > e_{\text{м}}$, т. е. каждый нуклон в дочернем ядре должен быть в среднем связан более прочно, чем в материнском. С возрастанием массового числа средняя энергия связи, приходящаяся на один нуклон, должна убывать, и притом настолько быстро, чтобы выполнялось условие (73.3). Это действительно происходит, и причина этого в том, что с возрастанием Z увеличивается относительная роль кулоновского отталкивания, уменьшающего энергию связи ядра.

2. Характерной особенностью α-распада является очень сильная зависимость периода полураспада $T_{1/2}$ от энергии $\mathscr{E}_\alpha$ вылетающей α-частицы. Уменьшение $\mathscr{E}_\alpha$ всего на 1 % может увеличить период $T_{1/2}$ в 10 раз, а уменьшение $\mathscr{E}_\alpha$ на 10 % может увеличить $T_{1/2}$ на 2–3 порядка.

Периоды полураспада α-активных ядер изменяются в широчайших пределах. Так, для изотопа свинца $^{204}_{82}\mathrm{Pb}$ $T_{1/2} = 14 \cdot 10^{17}$ лет, а для изотопа радона $^{215}_{86}\mathrm{Rn}$ $T_{1/2} = 10^{-6}$ с. Энергии же вылетающих α-частиц заключены в довольно узких пределах, а именно 4–9 МэВ для тяжелых ядер и 2–4,5 МэВ для ядер редкоземельных элементов.

Связь между величинами $T_{1/2}$ и $\mathscr{E}_\alpha$ была эмпирически установлена Гейгером и Неттолом еще в 1911–1912 гг. и получила название *закона Гейгера–Неттола.* Физический смысл этого закона был понят только после того, как к теории α-распада была применена квантовая механика (см. п. 10). В современной форме закон Гейгера–Неттола имеет вид

$$\lg T_{1/2} = C + \frac{D}{\sqrt{\mathscr{E}_\alpha}}, \tag{73.4}$$

где C и D — постоянные, не зависящие от A и слабо меняющиеся с изменением Z. Закон (73.4) хорошо подтверждается для четно-четных ядер. Если $T_{1/2}$ измерять в секундах, а $\mathscr{E}_\alpha$ в мегаэлектронвольтах, то при некоторых Z для таких ядер величины C и D имеют значения, приведенные в табл. 10. Для нечетных ядер наблюдаются отступления от закона (73.4), иногда очень значительные. Отношение $T_{1/2}^{\text{эксп}}/T_{1/2}^{\text{расч}}$ изменяется от единицы до нескольких тысяч (см. п. 14).

Таблица 10

Z	C	G
84	−50,15	128,8
90	−51,94	139,4
98	−55,40	154,7

3. Чтобы выяснить, для каких элементов периодической системы возможна α-радиоактивность с энергетической точки зрения, представим графически зависимость экспериментально найденной энергии Q α-распада от массового числа A. Рисунок 127 дает такую зависимость для наиболее устойчивых (β-стабильных) изобаров, соответствующих рассматриваемому массовому числу A. Тонкие наклонные прямые линии позволяют приближенно определить энергию α-распада, какой она должна была бы быть, если бы период полураспада составлял 1 час и 10^8 лет соответственно. Видно, что α-распад становится энергетически возможным только при $A \gtrsim 140$. Далее, видно, что кривая имеет два резко выраженных максимума. В этих максимумах и их окрестностях, в соответствии с законом Гейгера–Неттола, период полураспада минимален. В остальных случаях период полураспада получается слишком большим и α-распад практически невозможно наблюдать. Один максимум лежит в области тяжелых элементов, другой — в области редкоземельных элементов. Для этих элементов (в основном тяжелых) и наблюдается α-радиоактивность. Максимум в окрестности $A = 145$ связан с заполнением нейтронной оболочки до магического числа $N = A - Z = 82$, а максимум при $A = 215$ — с заполнением протонной оболочки до того же магического числа $Z = 82$. Заполненным оболочкам, как известно, соответствуют максимальные энергии связи, чем и объясняется происхождение максимумов на кривой рис. 127.

Более полные выводы можно получить, если рассмотреть дополнительные данные, относящиеся к энергии α-распада для различных изотопов одного и того же элемента. Это сделано на рис. 128 для тяжелых элементов. Различные изотопы одного и того же элемента

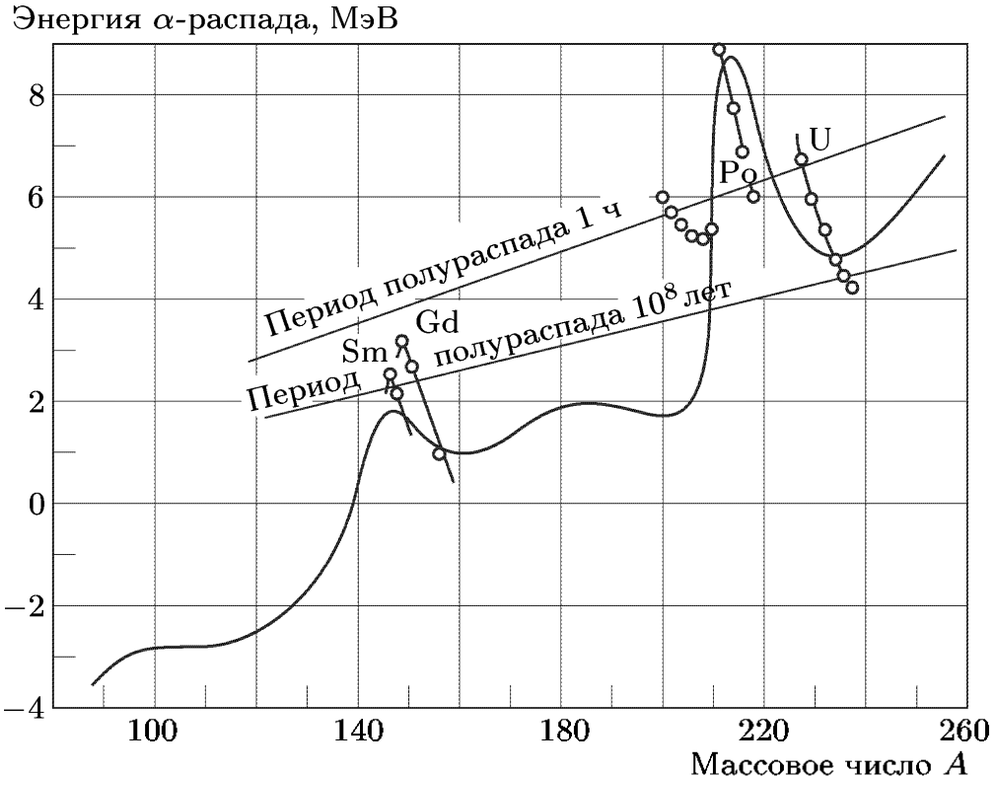


Рис. 127

соединены поочередно сплошными и штриховыми линиями. Светлые точки с направленными вверх стрелками соответствуют случаям, когда энергия α-распада установлена недостаточно точно.

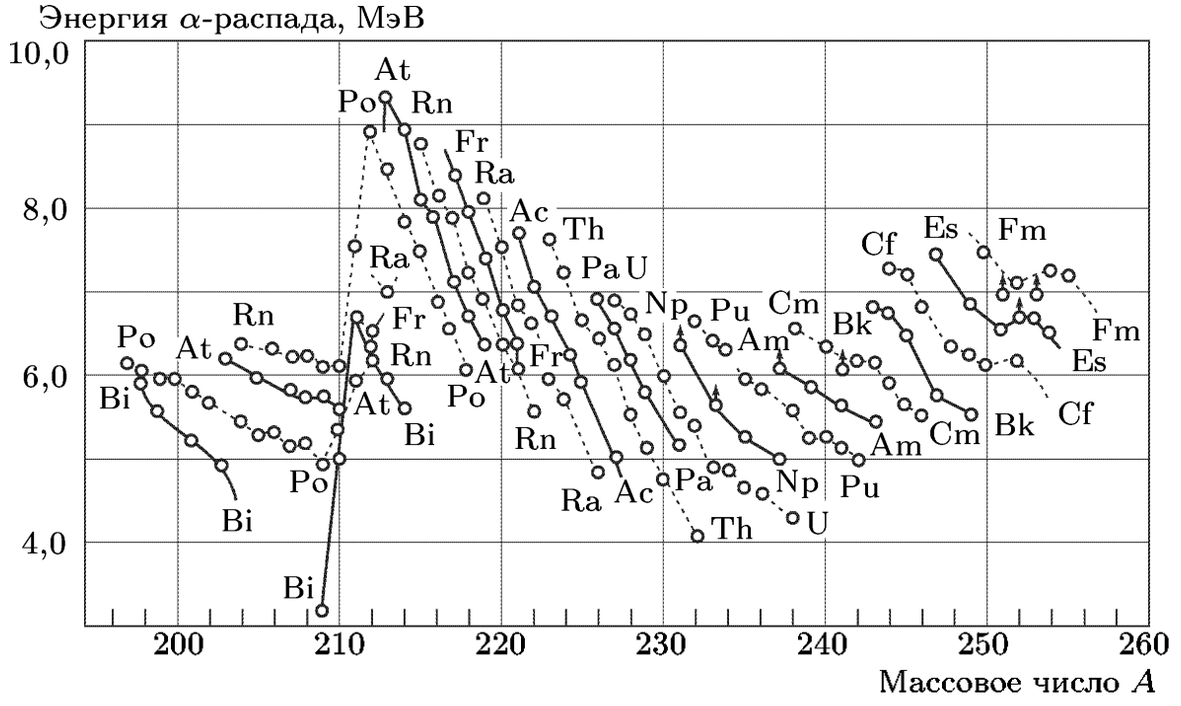


Рис. 128

4. Уменьшенное содержание нейтронов в ядре по сравнению с нормальным или равновесным (дефицит нейтронов) способствует α-распаду, поскольку оно увеличивает относительную роль кулоновского отталкивания между протонами ядра. Противоположно действует относительный избыток нейтронов. Так, редкоземельный элемент гольмий имеет единственный стабильный изотоп $^{165}_{67}$Ho, содержащий 98 нейтронов, тогда как его α-активные изотопы ^{152}Ho, ^{153}Ho, ^{154}Ho, ^{155}Ho нейтронодефицитны: в них числа нейтронов лежат между 85 и 88. Дефицит нейтронов в ядре может уменьшаться за счет процессов α-распада, позитронного β^+-распада и e-захвата электронов электронной оболочки, которые часто конкурируют между собой. Примером могут служить упомянутые выше изотопы гольмия, в которых периоды полураспада по отношению к этим трем процессам сравнимы между собой.

Для ядер, перегруженных нейтронами, в конкуренцию вступает электронный β^--распад. Таким образом, приближение чисел протонов и нейтронов в ядре к их нормальным (равновесным) значениям может идти за счет всех упомянутых процессов: α-распада, $\beta^\pm$-распада и e-захвата. Если период полураспада одного из этих конкурирующих процессов отличается от других на много порядков, то часто практически удается наблюдать только наиболее быстро идущий процесс. Например, долгоживущий изотоп нептуния $^{237}_{93}$Np α-активен с периодом полураспада $2\cdot10^6$ лет, у изотопов же ^{232}Np и ^{233}Np α-распад наблюдать не удается. В первом случае он подавлен позитронным распадом с периодом полураспада 13 мин, а во втором — электронным с периодом 2,3 дня.

5. У ядер, начиная с массового числа $A = 232$, к перечисленным типам радиоактивного распада присоединяется *спонтанное деление* (не смешивать с вынужденным делением). Этот процесс вначале идет очень медленно. Так, период полураспада по отношению к делению ядер для изотопа урана $^{238}_{92}$U составляет $8 \cdot 10^{15}$ лет. Но с увеличением массового числа он быстро уменьшается. Для изотопа курчатовия $^{260}_{104}$Ku период полураспада по отношению к спонтанному делению составляет десятые доли секунды. Спонтанное деление, по-видимому, и наложит предел на возможность получения новых трансурановых элементов.

6. Познакомимся теперь с главными чертами теории α-распада, основы которой были заложены в 1928 г. независимо друг от друга Г. А. Гамовым (1904–1968), с одной стороны, и Гёрни (1899–1953) и Кондоном (1902–1974) — с другой. Введем упрощающее предположение, что α-частицы уже существуют внутри атомных ядер. При такой идеализации материнское ядро состоит из дочернего ядра и α-частицы. Эта идеализация, вероятно, не соответствует действительности. Скорее всего, α-частица образуется из протонов и нейтронов перед вылетом из ядра. Однако указанная идеализация приводит в основном к правильным результатам. Мы вернемся к этому вопросу в п. 14.

На расстояниях от ядра, где практически перестают действовать ядерные силы, остается только кулоновское отталкивание и потенциальная функция U представляется формулой $U = zZe^2/r$, где Ze —

заряд дочернего ядра, а ze — заряд α-частицы. Для α-частицы $z = 2$. Однако мы будем писать z вместо 2, чтобы учесть гипотетическую возможность вылета из ядра не только настоящих α-частиц, но и других образований из нуклонов, для которых $z \neq 2$. Но кулоновское отталкивание на малых расстояниях от ядра должно перейти в притяжение, обусловленное *ядерными силами*, иначе α-частицы в ядре не могли бы удерживаться. Точный закон действия ядерных сил не известен. Известно только, что ядерные силы очень резко убывают с расстоянием. Поэтому для упрощения и самой возможности расчета потенциальная функция U аппроксимируется модельной. Считая ядро сферическим, можно предполагать, что U зависит только от расстояния r до центра дочернего ядра. Модельная функция $U(r)$ принимается равной

$$U(r) = \begin{cases} U_0 = \text{const} & \text{при } r < R, \\ zZe^2/r & \text{при } r > R, \end{cases} \tag{73.5}$$

где R — радиус дочернего ядра. Она представлена на рис. 129 жирной кривой, верхняя часть которой изображена штриховой линией, чтобы отметить, что в переходной области левая часть кривой $U(r)$ в действительности плавно переходит в кулоновскую часть, расположенную правее. Вертикально поднимающийся участок кривой при $r = R$ есть, конечно, идеализация. В действительности этот участок поднимается вверх очень круто, но не вертикально.

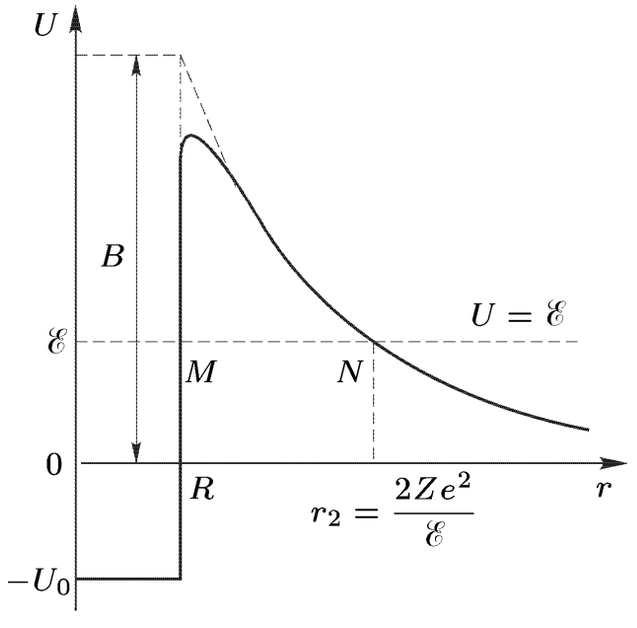


Рис. 129

Кривая $U(r)$ представляет собой *потенциальный барьер*, который должна преодолеть α-частица, чтобы вылететь из ядра. Какова высота этого барьера, до каких расстояний можно применять закон Кулона — на эти вопросы, конечно, нельзя дать вполне точного ответа. Однако Резерфорд в 1927 г. установил, что длиннопробежные α-частицы $^{212}_{84}$Po (старое обозначение — ThC′), обладавшие энергией 8,8 МэВ, рассеиваются тяжелыми радиоактивными ядрами (испускающими α-частицы меньшей энергии) в соответствии с формулой (9.3), выведенной в предположении применимости закона Кулона. Никаких аномалий рассеяния, которые свидетельствовали бы о ядерных взаимодействиях α-частиц $^{212}_{84}$Po с рассматриваемыми ядрами, не наблюдалось. Отсюда следует, что на всех расстояниях, до которых может сближаться α-частица $^{212}_{84}$Po с рассеивающим ядром, действуют *только кулоновские* силы отталкивания, а высота кулоновского потенциального барьера во всяком случае не меньше 8,8 МэВ.

Высоту барьера B можно оценить по формуле

$$B = \frac{zZe^2}{R}, \tag{73.6}$$

в которую радиус ядра входит в первой степени. Поэтому формула мало чувствительна к изменениям R. Полагая $Z = 90$ и принимая для R (возможно, несколько завышенное) значение 10^{-12} см, получим $B \approx 26$ МэВ.

7. Отметим теперь парадокс, разрешить который классическая физика оказалась бессильной. Уран ^{238}U, например, испускает α-частицы с энергией 4,2 МэВ, а радий ^{226}Ra — с энергией 4,8 МэВ. Эти значения много меньше максимальной потенциальной энергии $U_{\text{макс}}$ и во всяком случае меньше энергии α-частиц 8,8 МэВ, которые использовались в опытах Резерфорда. Так же обстоит дело с подавляющим числом α-активных ядер. Но для преодоления потенциального барьера полная энергия α-частицы по классическим представлениям должна быть не меньше $U_{\text{макс}}$. Энергия сохраняется. Поэтому после вылета из ядра кинетическая энергия α-частицы (а только таковой она и обладает на достаточно большом расстоянии от ядра) должна быть не меньше $U_{\text{макс}}$. В действительности же эта энергия много меньше.

Парадокс возникает потому, что к движению α-частицы внутри ядра и вблизи его границы были применены законы и понятия *классической механики.* А в этом случае так поступать нельзя. В самом деле, будем рассуждать классически и посмотрим, будет ли при этом выполняться принцип неопределенностей Гейзенберга. Характерная энергия α-частицы при α-распаде составляет 5 МэВ. В ядре ей соответствует классическая скорость $15 \cdot 10^8$ см/с и импульс 10^{-14} г $\cdot$ см/с. Глубина проникновения x частицы внутрь барьера — порядка радиуса ядра, т. е. 10^{-12} см. Таким образом, $px \sim 10^{-26}$ г $\cdot$ см^2/с. Произведение неопределенностей координаты и импульса $\Delta x \cdot \Delta p$ еще меньше. Оно порядка и, может быть, даже меньше постоянной Планка. Это указывает на неприменимость понятий и законов классической механики к движению α-частицы внутри ядра и вблизи его границ.

8. Теория α-распада должна строиться на основе *квантовой механики*, что и было сделано Г. А. Гамовым после того, как он на семинаре, руководимом Л. И. Мандельштамом, ознакомился с тогда еще не опубликованной работой Мандельштама и Леонтовича, в которой были заложены основы прохождения ψ-волн и связанных с ними частиц через потенциальный барьер (см. § 28). Как уже указывалось, независимо теория α-распада была разработана Гёрни и Кондоном.

Уточняя терминологию, мы в этом параграфе будем понимать под потенциальным барьером часть потенциальной кривой U, заключенную между классическими точками поворота M и N (см. рис. 129), в которых потенциальная функция U равна полной энергии частицы $\mathscr{E}$. Таким образом, внутри потенциального барьера всюду $U > \mathscr{E}$. Часто говорят, что внутри барьера кинетическая энергия частицы отрицательна, а ее скорость чисто мнимая. Мы будем избегать подобных выражений, поскольку они физически бессмысленны и основаны

на распространении классических соотношений и понятий на область пространства, где они неприменимы.

Суть дела была разъяснена в § 28 и заключается в следующем. Состояние частицы описывается волновой функцией ψ. Прохождение волны ψ через барьер есть *детерминистический процесс*, описываемый уравнением Шредингера. Для падающей волны область $U > \mathscr{E}$ представляет какое-то препятствие, но через это препятствие волна может проходить, хотя и с некоторым ослаблением. Однако волновая функция ψ есть величина вспомогательная: все реально наблюдаемые величины связаны с ней *вероятностными соотношениями.* Поскольку функция ψ всюду отлична от нуля, существует *конечная вероятность* обнаружить частицу как внутри барьера, так и за его пределами. В этом смысле и говорят о заходе частицы в классически недостижимую область $U > \mathscr{E}$ и о прохождении ее через потенциальный барьер. При наличии такой области говорят о *подбарьерном прохождении* частицы или *туннельном эффекте*, хотя эти термины и не совсем удачны, поскольку они могут породить неверное представление о прохождении частицы как детерминистическом процессе. В случае же, когда всюду $\mathscr{E} > U$, говорят о *надбарьерном прохождении.* Применяя такую терминологию, можно сказать, что *α-распад есть подбарьерное прохождение частицы.* Внутри барьера деление полной энергии $\mathscr{E}$ на кинетическую и потенциальную лишено смысла. Но далеко за пределами атомного ядра движение α-частицы — классическое, а вся энергия ее — кинетическая.

Приближенная формула (28.17) была получена для плоского потенциального барьера из волнового уравнения Шредингера для стационарных состояний. Но если происходит α-распад, то состояние системы из дочернего ядра и α-частицы, строго говоря, *не стационарно*: имеется поток вероятности из центра ядра, не исчезающий на бесконечности. Поэтому формула (28.17) может быть справедлива только для *достаточно медленных* процессов, которые могут рассматриваться как *приближенно стационарные.* К таким процессам и относится α-распад. Для определения проницаемости барьера D сферическую поверхность ядра можно приближенно считать плоской, заменив, однако, в формуле (28.17) пределы интегрирования x_1 и x_2, соответствующие классическим точкам поворота, на $r_1 = R$ и $r_2 = zZe^2/\mathscr{E} = BR/\mathscr{E}$ (см. рис. 129). Дочернее ядро можно считать неподвижным, поскольку его масса значительно превосходит массу α-частицы. Чтобы получить вероятность распада в одну секунду λ (постоянную распада), проницаемость барьера D (28.17) надо умножить еще на предэкспоненциальный множитель ν, учитывающий вероятность образования α-частицы и ее появления на границе ядра. Таким путем получается формула

$$\lambda = \nu \exp\left[-\int_{r_1}^{r_2} \frac{2}{\hbar}\sqrt{2m(U-\mathscr{E})}\,dr\right]. \tag{73.7}$$

Наибольшие трудности вызывает вычисление величины ν. Однако для наиболее существенного понимания эту величину достаточно оценить грубо, так как постоянная распада λ зависит от нее несравненно слабее, чем от показателя экспоненты. Оценим ν из классических соображений, как это делалось в первоначальных работах Гамова. Положим $\nu = v/R$, где v — средняя скорость α-частицы в ядре. При такой интерпретации ν представляет собой приближенно среднее число соударений, которые испытывает α-частица в одну секунду с поверхностью ядра. Скорость v приближенно оценим из соотношения неопределенностей $mv \cdot R \sim \hbar$. В результате получаем

$$\lambda = \frac{\hbar}{mR^2} \exp\left[-\frac{2}{\hbar}\int_{r_1}^{r_2} \sqrt{2m(U-\mathscr{E})}\, dr\right]. \tag{73.8}$$

Как и следовало ожидать, в классическом пределе ($\hbar \to 0$) формула (73.8) дает $\lambda = 0$, т. е. α-распад становится невозможным. К этому пределу вплотную приближается случай долгоживущих α-радиоактивных ядер.

В случае кулоновского барьера $U = zZe^2/r$. Здесь интеграл в (73.8) вычисляется подстановкой $2m(U-\mathscr{E}) = x^2$. В итоге находим

$$\lambda = \frac{\hbar}{mR^2}\, e^{-\gamma}, \tag{73.9}$$

где

$$\gamma = \frac{2R\sqrt{2mB}}{\hbar}\left(\sqrt{\frac{B}{\mathscr{E}}}\arccos\sqrt{\frac{\mathscr{E}}{B}} - \sqrt{1-\frac{\mathscr{E}}{B}}\right), \tag{73.10}$$

$B = zZe^2/R$ — высота кулоновского барьера, $\mathscr{E}$ — энергия α-частицы, вылетевшей из ядра (или значение потенциальной функции U в классических точках поворота, см. рис. 129).

9. В качестве примера произведем оценку постоянной распада λ для ^{238}U, полагая $Z = 90$, $R = 10^{-12}$ см, $\mathscr{E} = 4{,}2$ МэВ, $B = 26$ МэВ. При таких параметрах выражение в круглых скобках (73.10) равно 1,963. Далее,

$$\frac{2R\sqrt{2mB}}{\hbar} = \frac{2R\sqrt{2mc^2B}}{\hbar c} = \frac{2\cdot 10^{-12}\sqrt{2\cdot 4\cdot 938\cdot 26}}{1{,}973\cdot 10^{-11}} = 44{,}78,$$

$\gamma = 87{,}92$; $e^{-\gamma} = 6{,}59\cdot 10^{-39}$; $\hbar/mR^2 = 1{,}157\cdot 10^{20}$; $\lambda = 7{,}62 \times$ $\times\, 10^{-19}$ с$^{-1}$; $\tau = 1/\lambda = 131\cdot 10^{16}$ с $= 152\cdot 10^{11}$ сут $= 4{,}16\cdot 10^{10}$ лет; $T_{1/2} = \tau/\ln 2 = 2{,}88\cdot 10^{10}$ лет. Для урана ^{238}U экспериментальное значение периода полураспада $T_{1/2} = 4{,}56\cdot 10^{9}$ лет. Это не так уж сильно отличается от вычисленного, если учесть известный произвол при выборе исходных значений параметров для расчета, а также грубо приближенный характер самой теории α-распада.

10. Из формулы (73.9) с учетом соотношения $\tau = 1/\lambda$ получается

$$\ln \tau = \gamma - \ln \frac{\hbar}{mR^2}.$$

Если предположить, что $\mathscr{E}/B \ll 1$, то можно получить и приближенный закон Гейгера–Неттола. В этом приближении

$$\sqrt{1 - \frac{\mathscr{E}}{B}} \approx 1, \quad \arccos \sqrt{\frac{\mathscr{E}}{B}} = \frac{\pi}{2} - \arcsin \sqrt{\frac{\mathscr{E}}{B}} \approx \frac{\pi}{2} - \sqrt{\frac{\mathscr{E}}{B}}.$$

Значит,

$$\ln \tau = \frac{2R\sqrt{mB}}{\hbar} \left(\frac{\pi}{2} \sqrt{\frac{B}{\mathscr{E}}} - 2 \right) - \ln \frac{\hbar}{mR^2},$$

или

$$\ln \tau = C' + \frac{D'}{\sqrt{\mathscr{E}}}, \tag{73.11}$$

где величины C' и D' слабо зависят от Z, а потому могут рассматриваться как постоянные. Но (73.11) есть лишь другая форма закона Гейгера–Неттола. Этот закон объясняет, почему при изменении энергий α-частиц в узких пределах (например, вдвое) периоды полураспада меняются на много порядков (например, на 10–20). Но этот факт, конечно, следует и из общих формул (73.7) и (73.8), которые справедливы и без ограничения $\mathscr{E} \ll B$. Суть дела заключается в том, что в выражении (73.7) для показателя экспоненты γ малая величина $\hbar$ стоит в знаменателе. С этим и связана узость диапазона, в котором могут меняться энергии α-частиц радиоактивных ядер. Из тяжелых ядер α-частицы с энергиями выше 9 МэВ вылетают практически мгновенно, тогда как при энергиях ниже 4 МэВ они живут в ядре настолько долго, что α-распад не удается зарегистрировать. Для редкоземельных α-активных ядер обе эти цифры снижаются из-за уменьшения радиуса ядра и высоты потенциального барьера.

11. Из формулы (73.9) видно, что постоянная распада λ сильно зависит от радиуса ядра, поскольку последний входит не только в предэкспоненциальный множитель, но и в показатель экспоненты $-\gamma$ (через высоту барьера B). Поэтому используя формулы (73.9) и (73.10), из измерений λ и $\mathscr{E}$ для α-распада можно довольно точно определить радиус ядра. Такая возможность сделается еще более ясной, если заметить, что спускающаяся вправо ветвь кривой для U на рис. 129 при заданном Z полностью определяется законом Кулона. Форма потенциального барьера однозначно задается положением левой вертикальной стенки, а она в свою очередь определяет радиус ядра. Полученные таким путем значения радиусов ядер превышают на 20–30 % значения, находимые по рассеянию быстрых электронов. Объясняется это тем, что электроны подвергаются действию только электрически заряженных нуклонов, т. е. протонов, а нейтроны на них не действуют. Метод рассеяния позволяет поэтому определить размеры той области ядра, которая заполнена протонами. Методом же α-распада измеряется

расстояние между центрами ядра и α-частицы, на котором перестают действовать ядерные силы. Поэтому этот метод дает радиус ядра, увеличенный на сумму радиуса α-частицы и радиуса действия ядерных сил.

12. Во всем изложенном выше предполагалось, что α-частица вылетает из ядра с нулевым орбитальным моментом импульса, т. е. в s-состоянии ($l = 0$). Допустим теперь, что $l \neq 0$. В классической физике орбитальный момент можно учесть, перейдя в систему отсчета, вращающуюся вместе с частицей, если к потенциальной функции добавить центробежную потенциальную энергию $U_{\text{цб}} = \mathbf{L}^2/(2mr^2)$, где $\mathbf{L}$ — момент импульса. В квантовой механике можно поступить так же, но учесть квантование момента по формуле $\mathbf{L}^2 = \hbar^2 l(l+1)$. Таким образом, следует положить

$$U = U_{\text{кул}} + U_{\text{цб}} = \frac{zZe^2}{r} + \frac{\hbar^2 l(l+1)}{2mr^2}.$$

Центробежный барьер создается центробежной силой, а эта сила стремится удалить α-частицу от ядра. Казалось бы, что она способствует α-распаду. Но такое заключение было бы правильным, если бы происходил *надбарьерный* процесс. Для *подбарьерного* процесса, каковым является α-распад, все происходит наоборот. *Центробежная сила повышает потенциальный барьер и увеличивает его ширину, т. е. она уменьшает постоянную распада λ и увеличивает период полураспада $T_{1/2}$.*

Однако влияние центробежного барьера не может быть очень значительным. Действительно, момент $\mathbf{L}$ α-частицы в ядре по порядку величины можно оценить из соотношения неопределенностей $mvR \sim \hbar$ или $|\mathbf{L}| \sim \hbar$, так что l может быть не больше нескольких единиц, например $l < 5$. В максимуме

$$U_{\text{цб}} = \frac{\hbar^2 l(l+1)}{2mR^2} = \frac{\hbar^2 c^2 l(l+1)}{2mc^2 R^2} < \frac{(1{,}973)^2 \cdot 5 \cdot 6 \cdot 10^{-22}}{2 \cdot 4 \cdot 938 \cdot 10^{-24}} \approx 1{,}5 \text{ МэВ},$$

что почти в 20 раз меньше высоты кулоновского барьера для ядер с $Z \approx 90$ (см. п. 6). Кроме того, центробежная энергия быстрее убывает с расстоянием, чем кулоновская (соответственно как $1/r^2$ и $1/r$). Ее наличие мало меняет форму потенциальной кривой $U_{\text{кул}}(r)$. Поэтому центробежный барьер, как правило, не меняет порядок величины периода полураспада $T_{1/2}$. В крайнем случае он увеличивает его на порядок.

13. Атомное ядро представляет собой связанную систему, а потому энергетические уровни его дискретны. При испускании α-частицы из материнского ядра, находившегося в определенном энергетическом состоянии, возникает дочернее ядро также в определенном энергетическом состоянии. Разность энергий этих ядер уносится α-частицей и дочерним ядром (ядром отдачи). Если бы переход совершался из основного состояния материнского ядра в основное состояние дочернего ядра, то получилась бы α-частица только одной строго определенной энергии. Поскольку пробег α-частицы определяется ее энергией, то

в камере Вильсона, казалось бы, должны получаться треки α-частиц одной и той же длины. Однако, как правило, длина треков α-частиц при распаде ядер одного и того же сорта, а следовательно, и их энергия оказываются различными. Это явление получило название *тонкой структуры* α-распада. Оно заключается в том, что наряду с основными α-частицами наблюдаются частицы меньших, но очень близких энергий. Такие частицы возникают при переходе материнского ядра из основного состояния в возбужденные состояния дочернего ядра. Однако переходы на высокие уровни возбужденного дочернего ядра порождают α-частицы низких энергий, а потому в соответствии с законом Гейгера–Неттола они маловероятны. По этой причине тонкая структура α-спектра, как правило, связана с переходами на возбужденные уровни несферических дочерних ядер. У таких ядер имеются уровни с небольшими энергиями возбуждения, возникающие из-за вращения ядер. Переходы на такие уровни и порождают α-частицы с близкими энергиями, которые в соответствии с законом Гейгера–Неттола должны происходить с вероятностями, сравнимыми с вероятностями переходов в основное состояние. Поэтому-то тонкая структура α-спектров встречается довольно часто.

У некоторых α-активных ядер при α-распаде могут возникать *длиннопробежные* α-частицы, энергии которых больше энергий основных α-частиц. Примером могут служить длиннопробежные α-частицы, испускаемые ядрами изотопов полония ${}^{212}_{84}\mathrm{Po}(\mathrm{ThC}')$ и ${}^{214}_{84}\mathrm{Po}(\mathrm{RaC}'')$. Длиннопробежные α-частицы возникают при переходах из возбужденных состояний материнского ядра в основные (или близкие к ним вращательные) состояния дочернего ядра. Но возбуждение материнского ядра может быть снято не только в результате испускания длиннопробежных α-частиц, но и в результате испускания γ-квантов. Последний процесс идет со *значительно большей* вероятностью, чем первый. Поэтому испускание длиннопробежных α-частиц наблюдается довольно редко.

14. Как уже указывалось (см. п. 6), изложенная теория α-распада, приводящая к формуле (73.7), предполагает, что α-частица *уже существует* в ядре. Теория приближенно рассчитывает только вероятность выхода α-частицы из ядра. На самом деле этому процессу предшествует *образование* α-частицы в ядре из составляющих ее нуклонов — двух протонов и двух нейтронов.

Учет этого обстоятельства, а также рассмотрение поведения образовавшейся α-частицы внутри ядра потребовали бы подробного рассмотрения физических процессов, происходящих внутри ядра, чего изложенная теория совсем не делает. Однако несомненно, что вероятностный процесс выхода α-частицы из ядра теория описывает правильно. Поэтому и в более полной теории формула (73.7) должна сохраниться, но в ней должно быть выяснено происхождение предэкспоненциального множителя ν и дано его количественное выражение. Этого еще не сделано, хотя множитель ν оценивался и с другой точки зрения, отличной от той, которая была принята Гамовым.

В свете сделанного замечания не приходится удивляться, что α-распады разделяются на *облегченные* и *необлегченные.* Облегченным называется α-распад, для которого достаточно хорошо выполняется формула (73.8). Если же реальный период полураспада превышает рассчитанный по этой формуле более чем на порядок (наблюдаются отклонения примерно на два–три порядка), то процесс называется необлегченным (ср. с изложенным в п. 2). Причин возникновения необлегченных распадов мы касаться не будем, поскольку этот вопрос выяснен еще недостаточно, хотя и имеются качественные соображения относительно указанных причин. Отметим только, что облегченные распады обычно происходят в четно-четных ядрах, а необлегченные — в нечетно-четных и нечетно-нечетных.

§ 74. Бета-распад

1. *Бета-распад* есть самопроизвольный процесс, в котором нестабильное ядро ${}^{A}_{Z}\mathrm{X}$ превращается в ядро-изобар ${}^{A}_{Z+1}\mathrm{X}$ или ${}^{A}_{Z-1}\mathrm{X}$. Конечным результатом этого процесса является превращение в ядре нейтрона в протон или протона в нейтрон. Можно сказать, что β-распад есть не внутриядерный, а *внутринуклонный* процесс. При нем, следовательно, происходят более глубокие изменения вещества, чем при α-превращении. Поэтому и его теория более сложна и разработана с меньшей полнотой, чем теория α-распада.

Различают три вида β-распада:

1) *электронный* β^--распад, в котором ядро испускает электрон, а потому зарядовое число Z увеличивается на единицу;

2) *позитронный* β^+-распад, в котором ядро испускает позитрон и по этой причине его зарядовое число уменьшается на единицу;

3) *электронный захват* (e-захват), в котором ядро поглощает один из электронов электронной оболочки, а потому зарядовое число уменьшается на единицу. Обычно электрон поглощается из K-слоя атома, поскольку этот слой ближе всего находится от ядра. В этом случае e-захват называют также K-захватом. Электрон может поглощаться и из L- или M-слоя и т. д., но эти процессы значительно менее вероятны. Электронный захват имеет особенно существенное значение в случае тяжелых ядер, у которых K-слой расположен близко к ядру. Можно сказать, что в этом случае у электрона K-слоя имеется конечная вероятность находиться внутри ядра. Электронный захват всегда сопровождается рентгеновским излучением. Допустим, например, что был захвачен электрон из K-слоя. На освободившееся место в K-слое переходит электрон из L-слоя. На освободившееся место в L-слое в свою очередь переходит электрон из M-слоя и т. д. В результате возбуждается весь спектр характеристического рентгеновского излучения атома. По наблюдению такого спонтанного характеристического излучения и было открыто явление e-захвата.

2. Сформулируем условие энергетической возможности всех перечисленных процессов. В случае β^--распада в конечном состоянии получается ядро, электрон и, возможно, какая-то одна или несколько

неизвестных частиц, которые при обычных условиях опыта ускользают от наблюдения. Но эти неизвестные частицы, если они действительно существуют, уносят определенную энергию. Поэтому для возможности β^--распада энергия (масса) начального ядра во всяком случае должна превосходить сумму энергий (масс) конечного ядра и электрона:

$$M(Z, A) > M(Z + 1, A) + m_e, \tag{74.1}$$

где M означает массу ядра. Однако в таблицах всегда приводятся не массы ядер M, а массы атомов $M_{ат}$. Поэтому целесообразно выразить условие (74.1) через $M_{ат}$. Для этого к обеим частям неравенства (74.1) прибавим массу Z электронов Zm_e. Тогда получим

$$M_{ат}(Z, A) > M_{ат}(Z + 1, A) \quad (\beta^-\text{-распад}). \tag{74.1а}$$

Здесь мы пренебрегли различием энергий связи электронов в атомах. Это различие находится за пределами точности самых прецизионных измерений. Смысл условия (74.1а) легко также уяснить из следующих соображений. Электронный β^--распад есть процесс, происходящий внутри атома. Один из нейтронов ядра превращается в протон с рождением электрона. Электрон захватывается электронной оболочкой, а заряд ядра возрастает на единицу. В конечном состоянии возникает атом с зарядовым числом $Z + 1$. Этот внутренний процесс может идти самопроизвольно только с уменьшением массы атома, что и выражает условие (74.1а).

Аналогично, для β^+-распада

$$M(Z, A) > M(Z - 1, A) + m_e, \tag{74.2}$$

так как масса позитрона равна массе электрона. После прибавления Zm_e к обеим частям неравенства получается

$$M_{ат}(Z, A) > M_{ат}(Z - 1, A) + 2m_e \quad (\beta^+\text{-распад}). \tag{74.2а}$$

Позитронный β^+-распад есть также процесс, происходящий внутри ядра атома. В результате этого процесса один из протонов ядра превращается в нейтрон с рождением позитрона. Освобождается, кроме того, один электрон из электронной оболочки атома. В результате возникает система, состоящая из атома с числом электронов $Z - 1$, освободившихся электрона и позитрона (общей массой $2m_e$). Условие (74.2а) и выражает, что для возможности β^+-распада необходимо, чтобы масса атома в начальном состоянии превосходила суммарную массу образовавшейся системы.

Наконец, в случае e-захвата в начальном состоянии имеются ядро (Z, A) и электрон, а в конечном — ядро $(Z - 1, A)$ без электрона. Поэтому энергетическое условие e-захвата записывается в виде

$$M(Z, A) + m_e > M(Z - 1, A), \tag{74.3}$$

или после прибавления к обеим частям неравенства по Zm_e

$$M_{ат}(Z, A) > M_{ат}(Z - 1, A) \quad (e\text{-захват}). \tag{74.3а}$$

Это условие также непосредственно очевидно, так как в начальном состоянии имеется атом (Z, A), а в результате e-захвата в конечном состоянии возникает только атом $(Z-1, A)$ и больше ничего, что обычно доступно наблюдению.

3. Для некоторых ядер могут одновременно выполняться условия (74.1) и (74.2). Из второго из них следует, что будет выполняться и условие (74.3). В таком случае ядро (Z, A) может испытывать все три вида β-превращений. Например, ядра $^{64}_{29}\mathrm{Cu}$ в 40 % случаев испускают электроны, в 20 % — позитроны, а в 40 % случаев они испытывают e-захват.

Позитронный β-распад и электронный захват в случае исходных ядер одного и того же сорта приводят к одинаковым окончательным состояниям. Поэтому эти два процесса часто идут одновременно, конкурируя друг с другом. Как уже выяснено, условие (74.2а) позитронного распада вытекает из условия (74.3а) электронного захвата. Поэтому принято говорить, что электронный захват энергетически более выгоден, чем позитронный β-распад. В частности, если

$$M_{\text{ат}}(Z-1, A) < M_{\text{ат}}(Z, A) < M_{\text{ат}}(Z-1, A) + 2m_{\text{e}},$$

или

$$0 < [M_{\text{ат}}(Z, A) - M_{\text{ат}}(Z-1, A)] < 2m_{\text{e}},$$

то электронный захват разрешен, а позитронный β-распад запрещен. Примером может служить ядро бериллия $^{7}_{4}\mathrm{Be}$. В этом случае $Z = 4$, $A = 7$. По таблицам находим $M_{\text{ат}}(4, 7) - M_{\text{ат}}(3, 7) = 0{,}8616$ МэВ, тогда как $2m_{\text{e}} = 1{,}022$ МэВ. Поэтому электронный захват наблюдается, а позитронный β-распад невозможен.

Если энергетически возможны оба конкурирующих процесса — β^+-распад и e-захват, — то для легких и средних ядер обычно преобладает позитронный распад, который часто практически полностью подавляет e-захват. Это связано с тем, что при электронном захвате вероятность обнаружения электрона даже из K-слоя в объеме, занятом ядром, исчезающе мала по сравнению с вероятностью обнаружения его вне ядра. Отношение этих вероятностей порядка $(R/r_0)^3$, где R — радиус ядра (10^{-12}–10^{-13} см), а r_0 — радиус K-слоя (10^{-8}–10^{-10} см), так что указанное отношение меняется в пределах 10^{-15}–10^{-6}. При переходе к тяжелым ядрам R увеличивается, а r_0 уменьшается. Оба эти обстоятельства ведут к увеличению $(R/r_0)^3$. Поэтому для тяжелых ядер, перегруженных протонами, обычно основным β-процессом является электронный захват. Отметим попутно, что в мезоатомах, где электрон замещен отрицательным мюоном, масса которого примерно в 207 раз больше массы электрона (а следовательно, r_0 во столько же раз меньше), отношение $(R/r_0)^3$ возрастает на семь порядков. Поэтому в этом случае вероятность K-захвата (мюона) увеличивается примерно на семь порядков, и в тяжелых ядрах этот процесс становится более вероятным, чем распад мюона, среднее время жизни которого $2{,}2 \cdot 10^{-6}$ с.

4. Масса нейтрона $m_{\text{n}} = 939{,}5731$ МэВ несколько больше массы протона $m_{\text{p}} = 938{,}2796$ МэВ. Разница масс составляет 1,2935 МэВ.

Поэтому энергетически возможно самопроизвольное превращение свободного нейтрона в протон. Такое превращение действительно наблюдается — среднее время жизни свободного нейтрона составляет около 15,3 мин, а период полураспада — 10,6 мин. Самопроизвольное превращение свободного протона в нейтрон энергетически невозможно. Но превращение протонов в нейтроны возможно и действительно идет внутри атомных ядер — именно в этом состоит само явление позитронного β-распада. Недостаток энергии протона пополняется за счет энергии атомного ядра.

Альфа-распад наблюдается только у самых тяжелых ядер и некоторых ядер редкоземельных элементов. Напротив, β-активные ядра гораздо более многочисленны. Для каждого Z (по крайней мере для надежно исследованных $Z < 100$) без исключения существуют нестабильные изотопы, обладающие $\beta^{\pm}$-активностью. Энергия, выделяющаяся в процессе β-распада, лежит в пределах от 0,0186 МэВ ($^3_1\mathrm{H} \to {}^3_2\mathrm{He}$) до 16,6 МэВ ($^{12}_7\mathrm{N} \to {}^{12}_6\mathrm{C}$). Период полураспада β-активных ядер меняется от 10^{-2} с (для $^{12}_4\mathrm{Be}$) до $4 \cdot 10^{12}$ лет (для $^{187}_{75}\mathrm{Re}$).

Простейшим примером электронного β-распада (после β-распада нейтрона) является распад трития:

$$^3_1\mathrm{H} \xrightarrow[12\text{ лет}]{\beta^-} {}^3_2\mathrm{He}.$$

Примером позитронного β-распада может служить распад ядра $^{11}_6\mathrm{C}$:

$$^{11}_6\mathrm{C} \xrightarrow[20{,}4\text{ мин}]{\beta^+} {}^{11}_5\mathrm{B}.$$

Наконец, в качестве примера радиоактивных ядер, испытывающих K-захват, приведем легкий изотоп берилия $^7_4\mathrm{Be}$, который после K-захвата превращается в изотоп лития $^7_3\mathrm{Li}$:

$$^7_4\mathrm{Be} \to \xrightarrow[53{,}6\text{ дня}]{\mathrm{K}} {}^7_3\mathrm{Li}.$$

(Во всех приведенных примерах время означает период полураспада.)

5. При β-распаде ядро превращается в соседний с ним изобар, т. е. меняется зарядовое число Z, но массовое число A остается неизменным. Применим к этому процессу полуэмпирическую формулу Вейцзеккера (64.6). Для исследования зависимости энергии ядра ($-\mathscr{E}_{\text{св}}$) от Z в этой формуле следует положить $A = \text{const}$. Тогда найдем формулу вида

$$-\mathscr{E}_{\text{св}} = aZ^2 + bZ + c,$$

где a, b, c — постоянные. Графиком функции $-\mathscr{E}_{\text{св}} = -\mathscr{E}_{\text{св}}(Z)$ будет парабола, причем, разумеется, аргумент Z может принимать только целочисленные значения. Даже получатся целых три параболы, сдвинутые одна относительно другой вверх и вниз. Ниже всех лежит парабола для четно-четных ядер (Z четное, N четное), выше всех — для нечетно-нечетных ядер (Z нечетное, N нечетное), а посередине между ними — парабола для нечетных A.

Рассмотрим сначала случай нечетных A. Тогда функция $-\mathscr{E}_{\text{св}}(Z)$ однозначна. Возможные значения Z при заданном A изображены кружками на рис. 130. Однако не все ядра, соответствующие этим значениям, стабильны. Стабильным ядрам соответствует минимум энергии $-\mathscr{E}_{\text{св}}$. Поэтому ядра *1*, *2*, *3*, ... будут испытывать последовательные β-превращения с испусканием электронов и с повышением зарядового числа Z на единицу. Ядра же *8*, *7*, *6*, ..., испустив позитрон или испытав e-захват, будут в свою очередь уменьшать зарядовое число на единицу. Одно из оставшихся двух ядер, у которого энергия больше, так же испытает β-превращение. В результате останется, вообще говоря, только одно стабильное ядро. Поэтому при нечетном A, как правило, может существовать только один изобар. Но возможны и исключения из этого правила. Это будет тогда, когда в результате β-превращений возникнут два ядра *4* и *5* с одинаковыми значениями энергии — два стабильных изобара. Зарядовые числа этих изобаров должны отличаться на единицу. Известны всего три таких изобарных пары: $^{113}_{48}\text{Cd}$ и $^{113}_{49}\text{In}$, $^{115}_{49}\text{In}$ и $^{115}_{50}\text{Sn}$, а также $^{187}_{75}\text{Re}$ и $^{187}_{76}\text{Os}$.

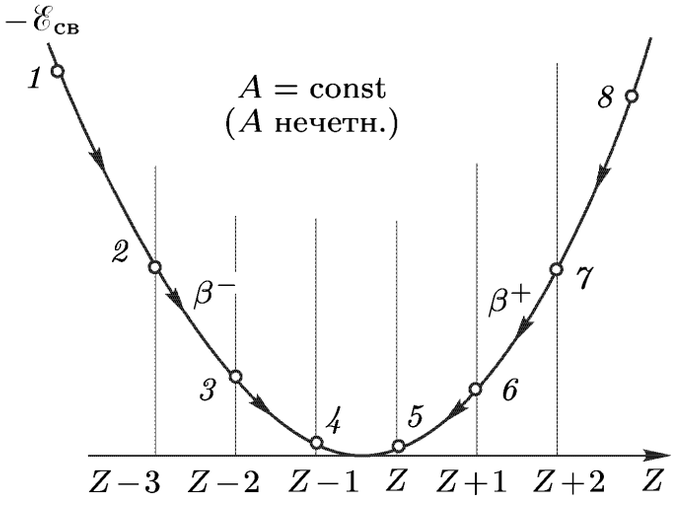


Рис. 130

В случае четных A функция $-\mathscr{E}_{\text{св}}(Z)$ двузначна и представляется двумя параболами. Нижней параболе соответствуют четные, верхней — нечетные значения Z. Здесь также путем соответствующих β-превращений осуществляется переход ядра в энергетически более низкие состояния. На рис. 131 он завершается двумя изобарами *3* и *4*,

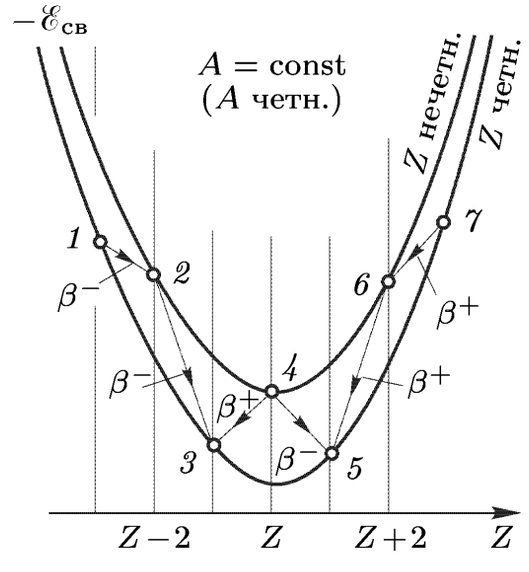


Рис. 131

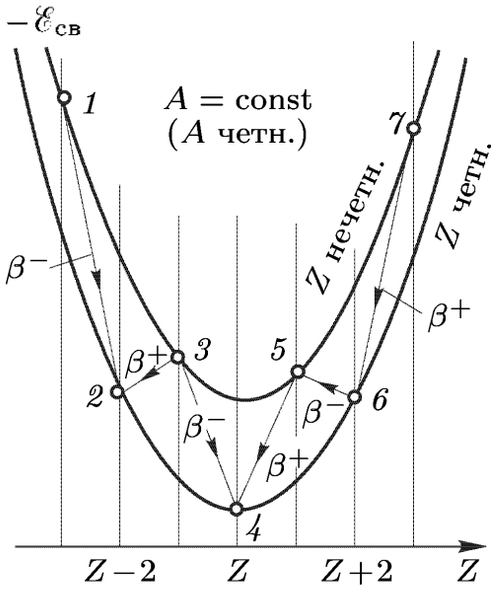


Рис. 132

лежащими на нижней параболе, т. е. имеющими четные значения Z. В подобных случаях массовому числу A соответствуют два изобара, зарядовые числа которых отличаются на два. Переход одного из этих изобар в другой, более устойчивый, может быть осуществлен путем одновременного испускания либо двух электронов, либо двух позитронов. Но такой двойной β-распад хотя в принципе и возможен, но крайне мало вероятен. Рисунок 132 соответствует случаю, когда процессы β-распада заканчиваются тремя стабильными изобарами *2*, *4* и *6*. Все они также лежат на нижней параболе, причем зарядовые числа Z соседних изобаров отличаются на 2. Превращения каждого из изобаров в соседний, энергетически более устойчивый, могли бы осуществляться путем двойного β-распада, а поэтому они не наблюдаются.

Таким образом, в случае четного A могут существовать два и даже три стабильных изобара. Разумеется, может существовать и только один стабильный изобар. Такое положение имеет место, в особенности для легких ядер. Наиболее легкой при четных A является стабильная изобарная пара $^{40}_{18}\mathrm{Ar}$ и $^{40}_{20}\mathrm{Ca}$.

Особо следует отметить случай, изображенный на рис. 133. Здесь четно-четные ядра *1* и *3*, лежащие на нижней параболе, энергетически менее устойчивы, чем нечетно-нечетное ядро *2*, лежащее на верхней параболе. Ядро *1* испытывает электронный, а ядро *3* — позитронный β-распад, и в обоих случаях возникает ядро *2*. Этим можно объяснить существование всех нечетно-нечетных (с четными A) стабильных ядер: $^{2}_{1}\mathrm{H}$, $^{6}_{3}\mathrm{Li}$, $^{10}_{5}\mathrm{B}$ и $^{14}_{7}\mathrm{N}$, о которых уже говорилось выше (см. § 64).

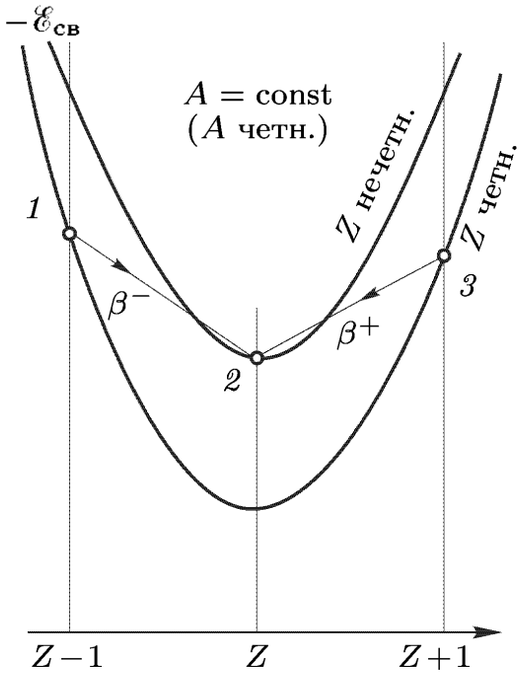


Рис. 133

6. Энергии α- и β-частиц, испускаемых радиоактивными веществами, поскольку эти частицы заряжены, можно измерить методом отклонения их в электромагнитных полях. Такие измерения показали, что каждое радиоактивное вещество испускает α-частицы вполне определенной энергии, присущей только этому веществу. Этот результат вполне естествен. Ядро перед испусканием α-частицы находится в невозбужденном квантовом состоянии с определенной энергией. После испускания получаются α-частица и дочернее ядро также в невозбужденном состоянии с определенной энергией. Разность этих энергий поэтому также имеет вполне определенное значение и проявляется в виде кинетической энергии дочернего ядра и α-частицы. А так как масса последней очень мала по сравнению с массой ядра, то практически вся кинетическая энергия уносится α-частицей. Наличие тонкой структуры в α-спектрах многих α-активных ядер ничего не меняет принципиально, так как оно связано с тем, что при распаде или материнское ядро, или дочернее ядро оказываются

в определенном возбужденном состоянии. Все это можно рассматривать как экспериментальное доказательство, что состояния материнского и дочернего ядер квантованны, а потому вполне определенны.

Совсем иной результат получается при испускании β-частиц. Радиоактивные атомы одного и того же сорта испускают электроны самых различных энергий, начиная от нуля и кончая некоторым предельным значением, характерным для рассматриваемого β-излучателя. Это предельное значение называется *верхней границей β-спектра.* Для примера на рис. 134 приведен β-спектр, возникающий при распаде $^{210}_{83}$Bi. Казалось бы, что и к β-распаду применимо рассуждение, приведенное нами для α-распада. В самом деле, пусть ядро A получилось в результате α-распада и, следовательно, на основании сказанного выше находится во вполне определенном энергетическом состоянии. Пусть затем оно испустило β-частицу и превратилось в ядро B. Допустим, далее, что получившееся ядро B опять α-радиоактивно. Тогда оно также будет находиться в определенном энергетическом состоянии. Но ядро B получается из A путем β-распада. Поэтому естественным представляется заключение, что β-частицы должны получаться вполне определенной энергии, а их энергетический спектр в действительности оказывается сплошным, как это было установлено Чедвиком уже в 1914 г. Получилось противоречие, которое требовало разъяснения.

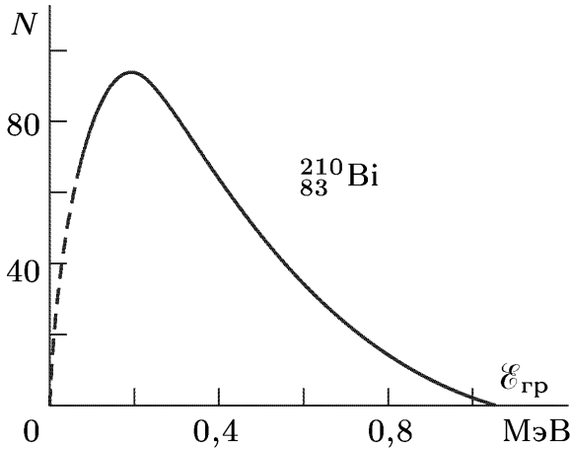


Рис. 134

7. Существование верхней границы β-спектра дало повод некоторым ученым высказать предположение, что в момент испускания все β-частицы получают одинаковые энергии, соответствующие границе β-спектра. Однако на пути к детектору не все они сохраняют эту энергию, а по-разному растрачивают ее на излучение. Другие предполагали, что β-излучение всегда сопровождается γ-излучением, так что энергия определенным образом распределяется между β- и γ-лучами. Правдоподобность последнего предположения, казалось, подтверждалась тем, что β-излучение часто сопровождается γ-излучением, тогда как при α-распаде γ-лучи наблюдаются реже. При справедливости любого из этих предположений полная энергия, освобождающаяся при β-распаде ядра, должна быть строго определенной, поскольку однозначно определены энергетические состояния материнского и дочернего ядер. Этот вывод допускает экспериментальную проверку.

Эллис (1895–1980) и Вустер (р. 1903) поставили соответствующий опыт в 1927 г. Они окружали β-излучатель ($^{210}_{83}$Bi) свинцовой оболочкой такой толщины, что она полностью поглощала электроны, а также рентгеновское и γ-излучение, исходившие от β-излучателя. По нагреванию свинцовой оболочки из-за β-распада радиоактивного препарата и по числу распадов за определенное время можно было определить

среднюю полную энергию, приходящуюся на один акт распада. Правда, калориметрические опыты очень трудны ввиду незначительности нагревания свинцовой оболочки. Эту трудность экспериментаторы преодолели, изготовив два совершенно одинаковых свинцовых калориметра, в один из которых помещался радиоактивный излучатель, а другой оставался пустым. По разности температур калориметров можно было с достаточной точностью измерить искомую величину. Оказалось, что полная энергия, приходящаяся в среднем на один акт β-распада, составляет 0,35 МэВ. С другой стороны, максимальная энергия β-частиц (верхняя граница β-спектра) для $^{210}_{83}$Bi составляет 1,05 МэВ, т. е. в три раза больше среднего значения, найденного калориметрически. Зная же полный β-спектр рассматриваемого излучателя, можно было вычислить среднюю энергию, уносимую одним электроном при β-распаде. Она оказалась равной 0,39 МэВ. Это значение в пределах погрешностей измерений хорошо согласуется с величиной, найденной калориметрически. Отсюда следует, что никакой дополнительной энергии, о которой говорилось выше в обоих предположениях, при β-распаде не выделяется. Оба предположения должны быть отвергнуты.

8. Не останавливаясь более на истории вопроса, приведем сразу решение его, предложенное Паули в 1930 г. Паули выдвинул гипотезу, позднее доказанную экспериментально, согласно которой при β-распаде наряду с электроном вылетает еще какая-то неизвестная нейтральная частица, не регистрируемая в калориметре. Благодаря этому закон сохранения энергии соблюдается и при β-распаде. В редких случаях вся освобождающаяся энергия уносится электроном. Такая энергия соответствует верхней границе электронного β-спектра. Подтверждением этого может служить опытный факт, что указанная граница совпадает с разностью энергий материнского и дочернего ядер. В других редких случаях вся энергия уносится гипотетической частицей. Но чаще всего энергия уносится и этой частицей, и электроном. Распределение энергии между ними носит случайный характер, чем и объясняется форма β-спектра.

Чтобы согласовать свою гипотезу с результатами опытов Эллиса и Вустера, Паули предположил, что гипотетическая частица очень слабо взаимодействует с веществом. Она не ионизует атомы среды, с которыми сталкивается, и поэтому свободно проходит через свинцовую оболочку и не улавливается последней, как это и было в опытах Эллиса и Вустера. Это заставило предположить, что частица Паули не обладает электрическим зарядом, так как заряженная частица с энергией порядка миллиона электронвольт не могла бы пройти через толстый слой свинца. Гипотеза Паули появилась еще до открытия нейтрона. Ферми назвал эту незаряженную гипотетическую частицу *нейтрино* (что в переводе с итальянского означает уменьшительное «нейтрончик»). Масса нейтрино, если таковая имеется у этой частицы, должна быть во всяком случае меньше массы электрона. Это видно уже из того, что при β-распаде, когда должен соблюдаться и закон сохранения импульса, электроны уносят меньше половины освобождающейся энергии. Нейтрино не может быть нейтроном, так как в противном

случае при испускании нейтрино заметно изменялась бы масса ядра, а это при β-распаде не наблюдается. Кроме того, если бы нейтрино и нейтрон были одной и той же частицей, то она обнаружилась бы при столкновении с другими атомными ядрами.

Поскольку нейтрино не ионизует атомы, магнитный момент этой частицы также должен равняться нулю (или, во всяком случае, быть весьма малым). Спин нейтрино должен быть полуцелым. Действительно, характер спина ядра (целый или полуцелый) целиком определяется значением массового числа A (см. § 68). При β-распаде A не меняется, а потому не должен меняться и характер спина ядра. Но вылетающий электрон (позитрон) уносит спин, равный $1/2$. Поэтому спин другой частицы — нейтрино, — во всяком случае, должен быть полуцелым. Теоретические расчеты энергетического спектра электронов (позитронов), рождающихся при β-распаде, выполненные при разных предположениях относительно спина нейтрино, показали, что спин этой частицы равен $1/2$.

9. На основе гипотезы Паули в 1933 г. Ферми построил количественную теорию β-распада. Подробное изложение ее далеко выходит за рамки этой книги; ограничимся общими замечаниями. Основная идея Ферми состоит в том, что β-частицы и нейтрино не существуют в атомных ядрах, а рождаются в них в процессе β-распада, подобно тому как фотон не является составной частью атома, а появляется в процессе лучеиспускания. Ферми потребовалось ввести гипотезу о существовании особого типа короткодействующих сил, которые и вызывают в ядре процессы превращения нейтрона в протон или протона в нейтрон с испусканием β-частиц и нейтрино в обоих случаях. Соответствующие силы называются *слабыми силами*, а взаимодействия, осуществляемые под их действием, — *слабыми взаимодействиями.* Чтобы составить представление о величине слабых взаимодействий, укажем, что теоретические соображения и последующие опытные измерения показали, что средняя длина свободного пробега нейтрино с энергией 1 МэВ в воде равна примерно 10^{21} см (100 световых лет). Это намного превышает линейные размеры звезд, т. е. длины порядка 10^{13} см. Такие нейтрино свободно пронизывают Солнце, а тем более земной шар. С уменьшением энергии нейтрино их проникающая способность становится еще меньше. Радиус действия слабых взаимодействий не превышает 10^{-15} см.

10. Для полноты заметим, что постулированная Паули частица, появляющаяся при β^--распаде вместе с электроном, позднее стала называться *электронным антинейтрино* ($\overline{\nu}_e$), а частица, появляющаяся при β^+-распаде вместе с позитроном, — *электронным нейтрино* (ν_e). Конечно, деление на частицы и античастицы условно. Что называть частицей и что античастицей — это вопрос соглашения, но условиться об этом необходимо. В соответствии с принятым соглашением процессы β^-- и β^+-распадов пишутся в виде

$$n \to p + e^- + \overline{\nu}_e, \tag{74.4}$$

$$p \to n + e^+ + \nu_e. \tag{74.5}$$

Дело в том, что в дальнейшем были открыты другие типы нейтрино и антинейтрино: *мюонные нейтрино* (ν_μ) и *антинейтрино* ($\overline{\nu}_\mu$), *таонные нейтрино* (ν_τ) и *антинейтрино* ($\overline{\nu}_\tau$). Эти частицы появляются в процессах взаимодействия совместно с другими частицами: мюонами и $\tau^\pm$-лептонами (масса τ-лептона $m_{\tau^\pm} = 1{,}78$ ГэВ, среднее время жизни порядка $3 \cdot 10^{-13}$ с).

Лептонами называют класс элементарных частиц, не участвующих в сильных взаимодействиях, т. е. участвующих в электромагнитном, слабом и гравитационном взаимодействиях. Нейтрино и антинейтрино сверх того не участвуют и в электромагнитном взаимодействии (если масса нейтрино отлична от нуля, то не исключено наличие у нейтрино магнитного момента, не превышающего $10^{-10}\ \mu_B$). Все лептоны имеют спин $1/2$, т. е. являются фермионами. К лептонам относятся электрон, мюон, тяжелый лептон (τ), электронное, мюонное и τ-нейтрино и соответствующие им античастицы.

Мюонные нейтрино и антинейтрино появляются при распаде заряженных пионов одновременно с мюонами. *Пионами* (старое название *π-мезоны*) называют семейство из трех нестабильных элементарных частиц — двух заряженных (π^+ и π^-) и одной незаряженной (π^0). Времена жизни обеих заряженных частиц равны $2{,}6 \cdot 10^{-8}$ с, масса $m_{\pi^\pm} = 273 m_e$, а для нейтрального пиона соответственно $0{,}83 \cdot 10^{-16}$ с, $m_{\pi^0} = 264 m_e$. Пионы относятся к классу *мезонов*, который в свою очередь относится к обширной группе *адронов*, т. е. частиц, участвующих в сильных взаимодействиях. Пионы — наиболее легкие адроны. Интересующие нас процессы распада заряженных пионов записываются следующим образом:

$$\pi^- \to \mu^- + \overline{\nu}_\mu, \tag{74.6}$$

$$\pi^+ \to \mu^+ + \nu_\mu. \tag{74.7}$$

Что электронные и мюонные нейтрино (а также их античастицы) не тождественны — это доказано экспериментально. Поэтому заменять ν_e и $\overline{\nu}_e$ соответственно на ν_μ и $\overline{\nu}_\mu$, и обратно нельзя. Так, например, процессы

$$n \not\to p + e^- + \overline{\nu}_\mu, \quad p \not\to n + e^+ + \nu_\mu,$$

$$\pi^- \not\to \mu^- + \overline{\nu}_e, \quad \pi^+ \not\to \mu^+ + \nu_e,$$

получающиеся из (74.4)–(74.7) путем такой замены, не идут, что доказано и экспериментально. Поэтому они и снабжены перечеркнутой стрелкой. Доказано также, что ν_e и ν_μ не тождественны со своими античастицами $\overline{\nu}_e$ и $\overline{\nu}_\mu$. Существование ν_τ и $\overline{\nu}_\tau$ прямыми экспериментами пока не доказано, но подтверждается теоретическими соображениями и косвенными экспериментальными данными.

Каждый лептон характеризуется определенным квантовым числом, называемым *лептонным зарядом* или *лептонным числом*. Следует различать *электронный* (для e^-, e^+, ν_e, $\overline{\nu}_e$), *мюонный* (для μ^-, μ^+, ν_μ, $\overline{\nu}_\mu$) и *таонный* (для τ^-, τ^+, ν_τ, $\overline{\nu}_\tau$) *лептонные заряды*, обозначаемые соответственно через L_e, L_μ, L_τ. Это существенно различные

величины, хотя им условно и приписываются одинаковые числовые значения. Условились для всех отрицательно заряженных частиц (e^-, μ^-, τ^-) считать лептонные заряды равными $+1$. Лептонные заряды всех остальных частиц находятся из экспериментально установленного факта, согласно которому в замкнутой системе разность между числом лептонов и антилептонов каждого типа остается постоянной. Для этого надо этому факту придать форму *закона сохранения лептонного заряда.* При этом лептонные заряды всех других частиц, за исключением лептонов, принимаются равными нулю, так как у этих частиц свойства, связанные с существованием лептонного заряда, не обнаруживаются. Закон сохранения лептонного заряда требует, чтобы все положительно заряженные лептоны (e^+, μ^+, τ^+) имели лептонный заряд -1. Это видно из того, что возможны реакции

$$e^- + e^+ \to 2\gamma, \quad \mu^- + \mu^+ \to 2\gamma, \quad \tau^- + \tau^+ \to 2\gamma.$$

Только тогда суммарный лептонный заряд в левой части будет равен нулю, а это необходимо, так как γ-квант лептонного заряда не имеет. Из возможности процессов (74.5) и (74.7) следует, что лептонный заряд ν_e и ν_μ равен $+1$, а из возможности реакций (74.4) и (74.6) — что лептонный заряд $\overline{\nu}_e$ и $\overline{\nu}_\mu$ равен -1. Аналогично, надо приписать ν_τ лептонный заряд $+1$, а $\overline{\nu}_\tau$ — лептонный заряд -1. Таким образом, лептонные заряды всех отрицательно заряженных лептонов и всех нейтрино равны $+1$, а положительно заряженных лептонов и всех антинейтрино -1.

Примером процессов, в которых соблюдается закон сохранения лептонных зарядов L_e и L_μ, могут служить процессы

$$\mu^- \to e^- + \overline{\nu}_e + \nu_\mu, \quad \mu^+ \to e^+ + \nu_e + \overline{\nu}_\mu,$$

которые действительно происходят.

Процесс (74.4) происходит в ядерных реакторах. При делении ядер освобождаются нейтроны, после распада которых возникают антинейтрино. Антинейтрино $\overline{\nu}_e$ образуется в ядерных реакторах также и в процессах распада β^--активных продуктов реакции деления. Таким образом, ядерный реактор является источником антинейтрино. В недрах Солнца происходят термоядерные реакции, конечным результатом которых является превращение водорода в гелий:

$$4p \to {}^4_2He + 2e^+ + 2\nu_e. \tag{74.8}$$

Закон сохранения лептонного заряда требует, чтобы из недр Солнца исходил поток уже *нейтрино*, а не антинейтрино.

11. Для упрощения теории β-распада Ферми построил ее в предположении, что масса нейтрино равна нулю. Но это предположение не является обязательным требованием теории и не может считаться доказанным. Нет принципа, заставляющего полагать массу нейтрино равной нулю, тогда как в случае фотона такой принцип существует.

Вопрос о массе нейтрино постоянно обсуждался с момента появления гипотезы Паули. Верхний предел для массы электронного нейтрино (и антинейтрино) был получен из анализа формы электронного спектра β-распада трития. Оказалось, что верхний предел для массы электронного нейтрино $m_{\nu_e} < 35$ эВ, что примерно в 15 000 раз меньше массы электрона. В 1980 г. Е.Ф. Третьяков, В.А. Любимов и другие сотрудники Института экспериментальной и теоретической физики опубликовали результаты своих многолетних измерений массы электронного нейтрино. Согласно этим измерениям масса электронного нейтрино хотя и очень мала по сравнению с массой электрона, но все же *отлична от нуля.* (Было найдено, что $14 < m_{\nu_e} < 46$ эВ.) Этот результат не всеми признается как окончательный. Вопрос о массе электронного нейтрино остается открытым.

Оценка верхнего предела массы ν_μ была получена в результате исследования распадов π-мезонов. Оказалось, $m_{\nu_\mu} < 0{,}57$ МэВ. Оценка верхнего предела массы ν_τ производилась из рассмотрения распада $\tau^- \to e^- + \overline{\nu}_e + \nu_\tau$. Оказалось, $m_{\nu_\tau} < 250$ МэВ.

Нейтрино играло первостепенную фундаментальную роль в происхождении Вселенной и продолжает играть эту роль и сейчас в ее развитии, чего, разумеется, не мог предполагать Паули и даже Ферми при создании теории β-распада. Если масса нейтрино отлична от нуля, то многие теоретические представления о процессах с участием нейтрино должны быть пересмотрены. Роль нейтрино во Вселенной окажется более значительной, чем в случае безмассовых нейтрино.

Если бы масса нейтрино была равна нулю, то эта частица не могла бы находиться в состоянии покоя. Она всегда двигалась бы со скоростью света c относительно любой системы отсчета. Проекция вектора спина этой частицы на направление движения может быть равна либо $+1/2$, либо $-1/2$. В первом случае условились говорить, что частица имеет *правую*, или *положительную спиральность* $\lambda = +1$, во втором — *левую*, или *отрицательную*, *спиральность* $\lambda = -1$. Понятие спиральности, конечно, имеет определенный смысл только для безмассовых частиц. Только в этом случае частица имеет определенную спиральность независимо от системы отсчета. Если же масса частицы отлична от нуля, то понятие определенной спиральности лишено смысла: одна и та же частица имела бы то правую, то левую спиральность, в зависимости от того, в какой системе отсчета рассматривается ее движение. Прямым опытом установлено, что $\overline{\nu}_e$ имеет спин, направленный по импульсу. Поэтому если $m_\nu = 0$, то нейтрино будет обладать левой, а антинейтрино — правой спиральностью.

12. Введенное Ферми слабое взаимодействие оказалось более универсальным, чем предполагалось в первоначальной теория β-распада. Оказалось, что слабые взаимодействия с участием нейтрино всех видов проявляются не только при β-распаде, но и во многих других процессах, связанных с превращением элементарных частиц. Не входя в существо вопроса, ограничимся замечанием, что представилось возможным установить *единую природу слабых и электромагнитных сил* и создать теорию *единого электрослабого взаимодействия.* Согласно этой теории

слабые взаимодействия возникают в результате обмена виртуальными тяжелыми частицами со спином, равным 1: промежуточными заряженными $W^{\pm}$-бозонами и нейтральным промежуточным бозоном Z^0, а электромагнитные взаимодействия — в результате обмена виртуальными фотонами — безмассовыми частицами также со спином, равным 1. Например, распад $\mathrm{n} \to \mathrm{p} + \mathrm{e}^- + \overline{\nu}_\mathrm{e}$ происходит следующим образом: нейтрон превращается в протон и на короткое время испускает промежуточный W^--бозон, который распадается затем на электрон и антинейтрино $\overline{\nu}_\mathrm{e}$. Теория предсказала и массы $\mathrm{W}^{\pm}$- и Z^0-бозонов. Бозоны $W^{\pm}$ и Z^0 были открыты в 1983 г. Их массы оказались близкими к предсказанным теорией: $M_{W^{\pm}} \approx 80$ ГэВ, $M_{Z^0} \approx 90$ ГэВ.

Основные представления и главные результаты теории β-распада Ферми сохранили свое значение и в современных теориях (при энергиях, значительно меньших M_W).

13. В заключение этого параграфа остановимся на вопросе об экспериментальном доказательстве существования нейтрино. В дальнейшем имеются в виду только электронные нейтрино и антинейтрино. Соответствующие опыты очень трудно осуществить из-за слабости взаимодействия нейтрино с веществом (см. п. 9). Поэтому вначале опыты по обнаружению нейтрино носили косвенный характер. Первый опыт такого рода был поставлен в 1936 г. А. И. Лейпунским (1903–1972), который исследовал реакцию β-распада $^{11}_{6}\mathrm{C} \to {}^{11}_{5}\mathrm{B} + \mathrm{e}^+$. Не останавливаясь на деталях опыта, ограничимся изложением лишь его идеи. Если бы не было нейтрино, то импульсы позитрона e^+ и ядра отдачи $^{11}_{5}\mathrm{B}$ были бы равны по модулю, но противоположны по знаку. Поскольку обе частицы заряжены, их импульсы можно измерить и таким образом сравнить энергетические спектры ядра $^{11}_{5}\mathrm{B}$ и позитрона. С другой стороны, можно было бы и непосредственно измерить распределение по энергии ядер отдачи $^{11}_{5}\mathrm{B}$. Это распределение оказалось не соответствующим безнейтринной схеме расчета, хотя количественные результаты получить и не удалось.

Более совершенная, но все же косвенная идея опыта принадлежала А. И. Алиханову (1904–1970) и А. И. Алиханьяну (1908–1978), предложившим использовать процесс K-захвата электрона ядром $^{7}_{4}\mathrm{Be}$, при котором возникает ядро $^{7}_{3}\mathrm{Li}$. Спин ядра с нечетным A полуцелый. При K-захвате A не меняется, а потому не должна меняться и четность спина ядра. С другой стороны, электрон имеет спин $1/2$. Поэтому спин ядра после поглощения электрона должен меняться на $1/2$, т. е. стать четным. Противоречие устраняется, если предположить, что и при K-захвате образуется нейтрино, которое и уносит нечетный спин. Преимущество схемы опыта, предложенного Алихановым и Алиханьяном, состоит в том, что при K-захвате образуются только две частицы: ядро отдачи $^{7}_{3}\mathrm{Li}$ и нейтрино, а потому распределение кинетической энергии между ними определяется однозначно. Полная кинетическая энергия, выделяющаяся при K-захвате ядер $^{7}_{4}\mathrm{Be}$, равна

$$\mathscr{E} = [M_{\text{ат}}(^{7}_{4}\mathrm{Be}) - M_{\text{ат}}(^{7}_{3}\mathrm{L})]c^2 = 0{,}864 \text{ МэВ}. \qquad (74.9)$$

В силу закона сохранения импульса $\mathbf{P}_{\text{яд}} + \mathbf{p}_\nu = 0$, где $\mathbf{P}_{\text{яд}}$ — импульс ядра отдачи, а $\mathbf{p}_\nu$ — импульс нейтрино. Ядро отдачи получает кинетическую энергию

$$\mathscr{E}_{\text{яд}} = \frac{\mathbf{P}_{\text{яд}}^2}{2M_{\text{яд}}} = \frac{p_\nu^2}{2M_{\text{яд}}} = \frac{c^2 p_\nu^2}{2M_{\text{яд}} c^2}. \tag{74.10}$$

Но нейтрино — релятивистская частица, так что ее кинетическая энергия $\mathscr{E}_\nu = cp_\nu$. Она уносит подавляющую часть энергии, а потому в предыдущем соотношении с большой точностью можно принять $cp_\nu = \mathscr{E}_\nu = \mathscr{E}$. Таким образом, на долю ядра отдачи ^7_3Li приходится кинетическая энергия

$$\mathscr{E}_{\text{яд}} = \frac{\mathscr{E}^2}{2M_{\text{яд}} c^2}. \tag{74.11}$$

Кинетическую энергию ядра отдачи ^7_3Li можно измерить непосредственно и сравнить ее с выражением (74.11). К этому и сводится идея опыта. При этом, как показывает формула (74.11), выгодно использовать легкие ядра. Вот почему в качестве исходного ядра Алиханов и Алиханьян предложили взять наиболее легкое ядро ^7_4Be, в котором наблюдается K-захват.

Из-за начавшейся войны самими Алихановым и Алиханьяном опыт поставлен не был. Он был осуществлен только в 1942 г. Алленом (р. 1908) в США и привел к положительному результату, т. е. указывал на участие нейтрино в процессе K-захвата. Однако и качественный опыт Лейпунского и количественный опыт Аллена еще не могут рассматриваться как экспериментальное доказательство существования нейтрино, поскольку в этих опытах еще не наблюдалось взаимодействие свободного нейтрино с веществом. Опыт Аллена доказывает только, что в реакции K-захвата гипотеза нейтрино согласуется с законами сохранения энергии и импульса.

14. Зарегистрировать действие свободных антинейтрино стало возможным только после появления мощных ядерных реакторов. Осколки деления тяжелых ядер перегружены нейтронами, а потому они испытывают β^--распад. Радиоактивные продукты β^--распада в свою очередь подвергаются β^--распаду, пока этот процесс не закончится образованием стабильных ядер. В процессе же β^--распада испускаются антинейтрино. При каждом акте деления испускаются в среднем 5–6 антинейтрино. Поэтому ядерные реакторы являются мощными источниками антинейтрино. При мощности реактора около 100 МВт плотность потока антинейтрино за толстой стеной, защищающей от нейтронов и γ-квантов, но легко проницаемой для нейтрино, составляет примерно 10^{19} $\text{с}^{-1} \cdot \text{см}^{-2}$.

Впервые опыт был осуществлен в США Рейнесом (р. 1918) и Коуэном (р. 1919) в 1953 г. и повторен ими в более совершенном виде в 1954 г. (краткое описание этого второго опыта и приводится ниже). Им удалось зарегистрировать реакцию обратного β^--распада

$$\overline{\nu} + \text{p} \to \text{n} + \text{e}^+. \tag{74.12}$$

Эта реакция возможна, если энергия антинейтрино превышает 1,8 МэВ, так как именно на такую величину масса $(n + e^+)$ больше массы протона p.

Схема установки Рейнеса и Коуэна представлена на рис. 135. Она состояла из трех больших ($1{,}9 \times 1{,}3 \times 0{,}6$ м) баков-детекторов D_1, D_2 и D_3, наполненных сцинтиллирующим водным раствором, объем которого «просматривался» 110 фотоумножителями. Между детекторами помещались мишени M_1 и M_2 — два слоя воды (толщиной 7 см каждый), содержащие растворимую соль кадмия $CdCl_2$, для поглощения нейтронов. Каждый слой — мишень с двумя соседними баками-детекторами — образовывал независимую триаду. Для защиты от нейтронного и γ-излучения установка была заключена в свинцово-парафиновую оболочку и помещена глубоко под землей. Там она и облучалась потоком антинейтрино от ядерного реактора.

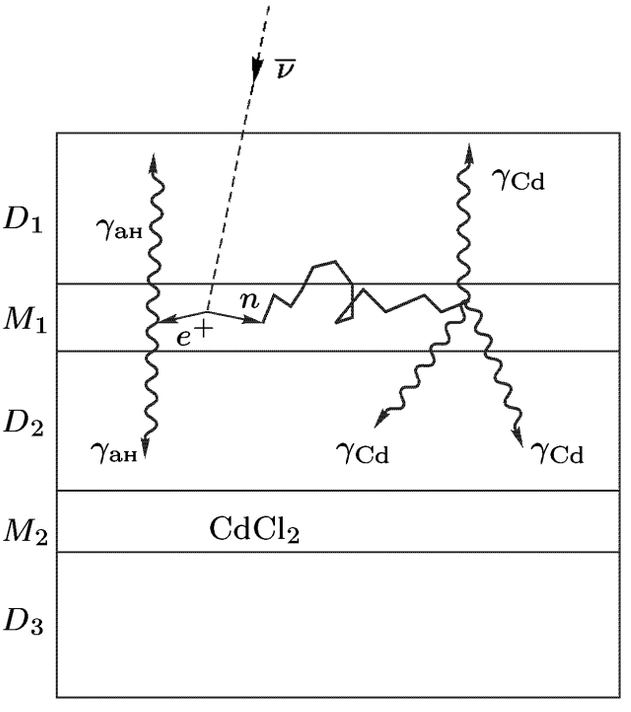


Рис. 135

Позитрон, возникающий в реакции (74.12) через короткое время порядка 10^{-8} с, аннигилировал с электроном с образованием двух γ-квантов (энергия которых равна энергии аннигиляции), которые регистрировались детекторами D_1 и D_2, включенными в схему совпадений. Нейтрон, возникающий в той же реакции (74.12), замедляется в результате последовательных столкновении с протонами, диффундирует и в конце концов за время порядка 10^{-6} с поглощается кадмием. При этом испускаются несколько γ-квантов (с суммарной энергией до 10 МэВ), которые также регистрируются детекторами. Сигналы с детекторов передавались по коаксиальным кабелям на вход специального электронного устройства, где они анализировались по высоте и времени сдвига совпадений. Проводилось фотографирование сигналов с помощью трехлучевого осциллографа, каждая из пластин которого была соединена с одним из детекторов. Акт взаимодействия антинейтрино с веществом (74.12) обнаруживался по появлению в установке двух скоррелированных импульсов, сдвинутых по времени на время замедления и диффузии нейтрона (от 1 до 25 мкс). В процессе длительной (около 1400 ч) работы установки детектор регистрировал в среднем $2{,}88 \pm 0{,}22$ импульса в час. Этому соответствует эффективное сечение реакции (74.12) $\sigma_{\overline{\nu}} \sim 10^{-43}$ см^2.

Для повышения точности полученных результатов авторами ставились многочисленные и разнообразные контрольные опыты, на описании которых мы останавливаться не будем.

Следует заметить, что в противоположность реакции (74.12) реакция

$$\nu + \mathrm{p} \nrightarrow \mathrm{n} + \mathrm{e}^{+} \tag{74.13}$$

запрещена, как это показал Дэвис. Это указывает на то, что нейтрино и антинейтрино — действительно разные частицы, отличающиеся одна от другой характером взаимодействия.

15. В порядке краткой информации скажем еще об обнаружении солнечных нейтрино, т. е. нейтрино, исходящих из недр Солнца при термоядерных реакциях. Здесь речь идет именно о *нейтрино*, а не об антинейтрино, как об этом уже говорилось в п. 10. Вопрос этот систематически исследуется Дэвисом с группой сотрудников из Брукхейвенской национальной лаборатории (США), а сами опыты производятся в шахте на глубине 1490 м для уменьшения фона космических лучей (сами измерения, не считая подготовительной работы, длятся уже более 15 лет). Для регистрации нейтринного излучения от Солнца используется хлор-аргоновый метод, предложенный Б. М. Понтекорво (р. 1913) еще в 1946 г. Детектором служит резервуар, заполненный 615 т жидкого перхлорэтилена (C_2Cl_4). В него предварительно вводится около 1 см^3 неактивного изотопа аргона $^{37}_{18}\mathrm{Ar}$. Под действием солнечных нейтрино происходит реакция

$$^{37}_{17}\mathrm{Cl} + \nu_{\mathrm{e}} \to {}^{37}_{18}\mathrm{Ar} + \mathrm{e}^{-}, \tag{74.14}$$

в которой образуется изотоп $^{37}_{18}\mathrm{Ar}$. Последний радиоактивен и испытывает K-захват с периодом полураспада 35 дней. Накопившийся радиоактивный $^{37}_{18}\mathrm{Ar}$ извлекается методом носителя, в качестве которого применяется неактивный изотоп аргона $^{36}_{18}\mathrm{Ar}$. В этом методе через объем бака с C_2Cl_4 в течение нескольких часов пропускается гелий. Последний в виде мелких пузырьков проходит через жидкий C_2Cl_4 и «вылавливает» из него атомы аргона вместе с радиоактивными атомами $^{37}_{18}\mathrm{Ar}$. Затем гелий поступает в ловушки, заполненные активированным углем и охлаждаемые жидким азотом. Аргон конденсируется и адсорбируется углем, а гелий свободно проходит через ловушки. Повторив такую операцию несколько раз, ловушки отделяют, соединяют с детектором и нагревают. Уголь выделяет адсорбированный аргон и последний поступает в пропорциональный счетчик. Зная количество аргона, добавленного к C_2Cl_4 и поступившего в счетчик, а также число актов, зарегистрированных счетчиком, можно получить и полное количество радиоактивных атомов $^{37}_{18}\mathrm{Ar}$, образовавшихся в C_2Cl_4. Было найдено, что под действием солнечных нейтрино в одной тонне C_2Cl_4 образуется в среднем 0,330 атомов $^{37}_{18}\mathrm{Ar}$ в год. Это примерно в 3–4 раза меньше теоретических предсказаний, основанных на традиционных представлениях о свойствах нейтрино и структуре центральных областей Солнца (см. § 102, п. 4). Такое расхождение получило название «загадки солнечных нейтрино». Мы не будем более останавливаться на этом вопросе, так как его еще нельзя считать решенным ни теоретически, ни экспериментально.

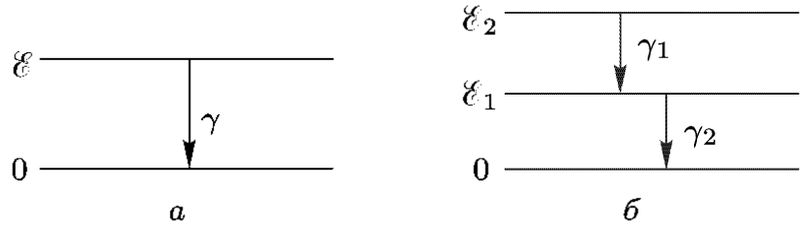
Заметим только, что в Приэльбрусье близится к завершению строительство Баксанской нейтринной обсерватории Института ядерных исследований АН СССР. Обсерватория располагается на глубине около 2 км под горой Андырчи. Хлор-аргоновая установка по объему будет в пять раз превышать установку Дэвиса. С помощью реакции превращения хлора в аргон можно зарегистрировать лишь относительно малоинтенсивную часть спектра солнечных нейтрино, обладающих высокими энергиями. Поэтому будет использован еще и новый метод, предложенный В.А. Кузьминым, основанный на превращении $^{71}_{31}\mathrm{Ga}$ в $^{71}_{32}\mathrm{Ge}$. Метод позволит регистрировать и основную часть нейтринного излучения Солнца — нейтрино более низких энергий от реакции $\mathrm{p} + \mathrm{p} \to d + \mathrm{e}^+ + \nu_\mathrm{e}$, задающей скорость энерговыделения в Солнце.

Можно надеяться, что эти исследования прояснят «загадку солнечных нейтрино», а также приведут к выяснению вопросов о внутреннем строении Солнца, что можно сделать только методами «нейтринной астрономии».

§ 75. Гамма-излучение ядер и внутренняя конверсия электронов

1. *Гамма-излучением* называется электромагнитное излучение, возникающее при переходе атомных ядер из возбужденных в более низкие энергетические состояния. В таких процессах числа протонов и нейтронов в ядре не изменяются, но испускаются γ-кванты. Спектр γ-излучения всегда *дискретный*, так как дискретны энергетические уровни самого ядра. Обычно энергия γ-квантов, испускаемых атомными ядрами, лежит в пределах примерно от 10 кэВ до 5 МэВ ($10^{-8} \gtrsim \lambda \gtrsim 2 \cdot 10^{-11}$ см).

Переход ядра из возбужденного состояния в нормальное при γ-излучении может быть *однократным*, когда ядро после испускания одного кванта сразу переходит в нормальное состояние (рис. 136 *а*). Но снятие возбуждения может быть и *каскадным*, когда переход осуществляется в результате последовательного испускания нескольких γ-квантов (рис. 136 *б*).



а *б*

Рис. 136

Изолированный свободный нуклон испускать γ-квант не может, так как в противном случае было бы нарушено одновременное выполнение законов сохранения энергии и импульса (см. § 1). Однако этот процесс

может происходить и действительно происходит внутри ядра, поскольку испущенный (или поглощенный) γ-квант может обмениваться импульсом не только с рассматриваемым нуклоном, но и с остальными нуклонами ядра. Таким образом, в противоположность β-распаду, испускание γ-лучей есть *внутриядерный*, а не внутринуклонный процесс.

2. Возбужденные ядра образуются при β-распаде в тех случаях, когда распад материнского ядра в основное состояние дочернего ядра оказывается запрещенным. Дочернее ядро в этом случае может получиться как в нижнем, так и в одном из верхних возбужденных состояний. В последнем случае совершается каскадный процесс переходов через несколько возбужденных состояний дочернего ядра, если только такие переходы разрешены. При этих переходах и происходит испускание γ-квантов. Типичным примером может служить изотоп $^{24}_{11}$Na. Основное состояние его имеет характеристику 4^+. В результате β-распада возникает изотоп $^{24}_{12}$Mg. Однако этот изотоп получается не сразу в основном состоянии, так как основной уровень $^{24}_{12}$Mg имеет характеристику 0^+. Поэтому переход на этот уровень практически не происходит, так как при таком переходе спин должен был бы сразу измениться на 4, что крайне маловероятно. Но изотоп $^{24}_{12}$Mg имеет два возбужденных уровня 4,12 и 1,37 МэВ с характеристиками 4^+ и 2^+. Бета-переход ядра $^{24}_{11}$Na на второй из них также запрещен по спину, хотя и не столь сильно, как переход на основной уровень 0^+. Бета-распад (с полупериодом 15 ч) идет почти исключительно на разрешенный уровень 4,12 МэВ (с характеристикой 4^+). С этого уровня происходит испускание γ-кванта при переходе на уровень 1,37 МэВ, а с него — на основной уровень 0. В результате испускаются γ-кванты с энергиями 2,75 и 1,37 МэВ (рис. 137). Если изотоп $^{24}_{11}$Na поместить в стеклянную ампулу, то β^--частицы задержатся стеклом, а γ-излучение выйдет наружу. Поэтому ампула с изотопом $^{24}_{11}$Na может служить удобным портативным источником γ-излучения. Напомним, что 0–0-переходы при излучении γ-квантов абсолютно запрещены (см. § 39, п. 1).

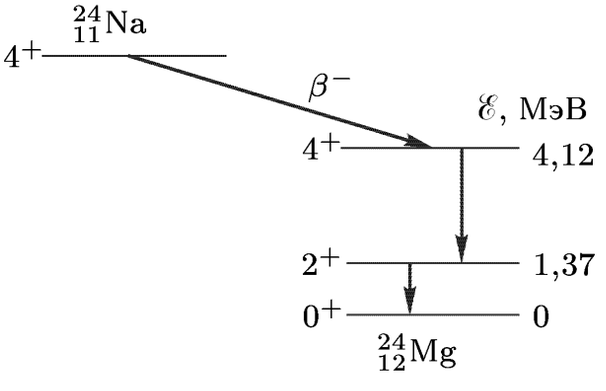


Рис. 137

Возбужденные ядра, способные к γ-излучению, могут возникать также в результате предшествующего α-распада. Однако γ-кванты, испускаемые таким путем, обычно обладают невысокими энергиями ($\mathscr{E}_\gamma \lesssim 0{,}5$ МэВ). Это связано с тем, что для испускания γ-квантов высоких энергий материнские ядра должны испускать α-частицы также очень высоких энергий. Энергия α-частицы должна быть достаточной не только для преодоления потенциального барьера, но и для сильного возбуждения возникающего дочернего ядра. Обычно α-частицы, испускаемые атомными ядрами, этому условию не удовлетворяют. Энергия γ-квантов, испускаемых дочерними ядрами после

β-распада, может быть больше и может достигать 2–2,5 МэВ. Объясняется это тем, что вероятность β-распада определяется более слабо меняющейся функцией энергии β-частицы, чем вероятность α-распада. Возбужденные ядра, способные к испусканию γ-квантов, могут также возникать в результате захвата нейтронов, в результате кулоновского возбуждения ядер при столкновениях с заряженными частицами в различных ядерных реакциях.

3. Возбужденное ядро может перейти в основное состояние не только путем испускания γ-кванта, но и путем непосредственной передачи энергии возбуждения одному из электронов атомных оболочек K-, L-, M-электрону и т. д.). Этот процесс, конкурирующий с γ-излучением, называется *внутренней конверсией* электронов, а сами электроны — *электронами внутренней конверсии.* Внутренняя конверсия может конкурировать с γ-излучепием. Но она может происходить и без него (например, в случае 0–0-переходов, когда испускание γ-квантов вообще невозможно). Отношение среднего числа электронов внутренней конверсии к среднему числу испускаемых γ-квантов для конкретного перехода называется *коэффициентом внутренней конверсии* перехода.

Коэффициент конверсии зависит от энергии и мультипольности перехода. Поэтому измеряя на опыте коэффициент конверсии, можно установить мультипольность и, таким образом, спин возбужденного ядра.

Энергия электрона внутренней конверсии $\mathscr{E}_e$ определяется выражением

$$\mathscr{E}_e = \mathscr{E} - \varepsilon, \tag{75.1}$$

где $\mathscr{E}$ — энергия, освобождаемая при ядерном переходе, а ε — энергия связи электрона в электронной оболочке атома. Очевидно, что электроны внутренней конверсии моноэнергетичны. Это и позволяет отличить их от электронов, испускаемых при β-распаде ядер, спектр которых непрерывен. Если энергия возбуждения ядра $\mathscr{E}$ меньше энергии связи электрона ε_K в K-слое, то, очевидно, внутренняя конверсия на электронах K-слоя энергетически невозможна. Такой случай может иметь место для тяжелых ядер. Однако в этом случае может происходить внутренняя конверсия на электронах других слоев.

Внутренняя конверсия сопровождается *рентгеновским излучением*, которое возникает в результате переходов электрона с вышележащих электронных слоев и оболочек на место, освобожденное электроном внутренней конверсии. Этот процесс вполне аналогичен обычному возбуждению рентгеновского характеристического спектра атомов (см. § 48). В результате внутренней конверсии могут появиться и электроны Оже (см. § 48).

Если энергия $\mathscr{E}$ возбуждения ядра превышает удвоенную собственную энергию электрона, т. е.

$$\mathscr{E} > 2mc^2 = 1{,}02 \text{ МэВ},$$

то может происходить процесс *парной конверсии*, при котором ядро теряет энергию возбуждения путем одновременного испускания электрона и позитрона. Электронная оболочка атома на такой процесс не оказывает никакого влияния, а потому он может происходить на ядре, лишенном атомных электронов. Однако вероятность парной конверсии не превышает примерно тысячной доли вероятности испускания γ-кванта.

4. Среднее время жизни γ-активных ядер обычно невелико (порядка 10^{-7}–10^{-11} с). Однако при сочетании высокой степени запрета с малыми расстояниями между энергетическими уровнями могут возникать долгоживущие или *метастабильные* γ-активные ядра со временами жизни макроскопического масштаба (до нескольких часов и даже больше). Такие возбужденные метастабильные ядра называются *изомерами*. Явление изомерии было открыто в 1921 г. Отто Ганом (1879–1968). Он обнаружил, что при β-превращениях

$$
{}^{234}_{90}\mathrm{Th} \begin{array}{l} \xrightarrow{\beta^-} \mathrm{UZ_1} \\ \xrightarrow[\beta^-]{} \mathrm{UX_2} \end{array}
$$

получаются два радиоактивных вещества, названных им $\mathrm{UZ_1}$ и $\mathrm{UX_2}$, которые состоят из одинаковых ядер ${}^{234}_{91}\mathrm{Pa}$, но имеют различные периоды полураспада (6,7 ч и 1,22 мин соответственно). В 1935 г. аналогичное явление было открыто И. В. Курчатовым с сотрудниками на искусственных радиоактивных ядрах ${}^{80}_{35}\mathrm{Br}$ и ${}^{82}_{35}\mathrm{Br}$, получаемых путем облучения нейтронами естественной смеси изотопов брома ${}^{79}_{35}\mathrm{Br}$ и ${}^{81}_{35}\mathrm{Br}$. Объяснение природы изомерии существованием у ядер метастабильных состояний было дано в 1936 г. Вейцзеккером.

Обычно изомерный уровень имеет спин, сильно отличающийся от спинов нижележащих уровней, и характеризуется низкой энергией возбуждения. Этому условию удовлетворяют значения Z и N, лежащие непосредственно перед магическими числами 50, 82 и 126 со стороны меньших Z и N. Как правило, изомерные состояния совпадают с первым возбужденным уровнем ядра. Такие закономерности следует ожидать и с точки зрения оболочечной модели ядра. Но существуют и изомеры, не подчиняющиеся этим правилам (таков, например, изомер кюрия ${}^{244}_{96}\mathrm{Cm}$ с энергией возбуждения 1,04 МэВ).

В некоторых случаях ядра могут иметь по два метастабильных уровня и, следовательно, обнаруживать три периода полураспада. Примером может служить ядро ${}^{124}_{51}\mathrm{Sb}$, испускающее электроны с периодами полураспада 60 дней, 21 и 1,3 мин.

Ядерная изомерия — не столь редкое явление, как может показаться на первый взгляд. Известно около сотни достаточно долгоживущих ядер. Наибольшее число изомерных состояний встречается у ядер с нечетным массовым числом A. Изомеры достаточно часто встречаются у нечетно-нечетных ядер и очень редко у четно-четных. Время жизни возбужденного ядра изменяется в широких пределах (от очень малых долей секунды до многих тысяч лет). Так, изомер ${}^{236}_{93}\mathrm{Np}$ имеет период

полураспада 5000 лет, а изомер ^{135}Cs — $2{,}8 \cdot 10^{-10}$ с. Можно ожидать, что с развитием методики эксперимента будут обнаружены изомеры с еще более длинными и короткими периодами полураспада.

Метастабильные состояния наблюдаются и у β-стабилыших ядер. В этих случаях метастабильное ядро переходит в основное состояние путем испускания γ-квантов и конверсионных электронов. Примером может служить β-стабильное ядро $^{113}_{49}$In $(9/2^+)$, которое имеет метастабильный изомер с энергией возбуждения 0,393 МэВ и временем жизни 104 мин $(1/2^+)$. Изомерия может также проявляться в форме существования у ядра нескольких периодов полураспада относительно спонтанного деления.

§ 76. Эффект Мёссбауэра

1. В 1958 г. было обнаружено *резонансное поглощение* γ-лучей, получившее название *эффекта Мёссбауэра* (р. 1929) — по имени ученого, который сделал это открытие. Явление это аналогично *оптической резонансной флуоресценции.* Оно состоит в том, что если возбужденный атом (или ядро) испустил фотон, то другой такой же, но невозбужденный атом (или ядро) способен с большой вероятностью его поглощать.

Для выяснения условий, при которых возможно резонансное поглощение γ-квантов (фотонов), надо принять во внимание, что в процессе испускания энергия возбужденного ядра передается не только γ-кванту, но и самому ядру — в виде кинетической энергии поступательного движения последнего, или энергии отдачи. Аналогично, при поглощении энергия γ-кванта идет не только на внутреннее возбуждение ядра, но и на сообщение ему поступательного движения. Допустим, что первое ядро до испускания, а второе до поглощения γ-кванта неподвижны. Тогда энергия испущенного γ-кванта окажется недостаточной, чтобы возбудить второе ядро. Для внутреннего возбуждения поглощающего ядра до того же энергетического уровня, на котором находилось испускающее ядро, требуется γ-квант большей энергии. Рассмотрим этот вопрос более подробно.

Пусть неподвижное ядро испустило γ-квант. Если $\mathscr{E}$ — разность энергий ядра до и после испускания, то на основании законов сохранения энергии и импульса можно написать

$$\mathscr{E} = \mathscr{E}_\gamma + K_{\text{яд}}, \quad \mathbf{P}_\gamma + \mathbf{P}_{\text{яд}} = 0,$$

где $K_{\text{яд}}$ и $\mathbf{P}_{\text{яд}}$ — кинетическая энергия и импульс ядра после испускания γ-кванта, а $\mathbf{P}_\gamma$ — импульс испущенного γ-кванта. (Предполагается, что до испускания γ-кванта ядро покоилось.) Таким образом, кинетическая энергия ядра отдачи

$$K_{\text{яд}} = \frac{P_{\text{яд}}^2}{2M_{\text{яд}}} = \frac{P_\gamma^2}{2M_{\text{яд}}},$$

где $M_{\text{яд}}$ — масса ядра. Импульс и энергия γ-кванта связаны соотношением $\mathscr{E}_\gamma = cP_\gamma$. Поэтому

$$K_{\text{яд}} = \frac{\mathscr{E}_\gamma^2}{2M_{\text{яд}}c^2}.$$

Но подавляющую долю энергии при испускании уносит γ-квант. На долю кинетической энергии ядра (из-за большой массы $M_{\text{яд}}$) приходится ничтожная часть. Следовательно, с достаточной точностью

$$K_{\text{яд}} = \frac{\mathscr{E}^2}{2M_{\text{яд}}c^2}. \tag{76.1}$$

Рассмотрим теперь поглощение γ-кванта ядром. В этом случае все величины будем обозначать теми же, но штрихованными буквами. Исключение оставим для величины $\mathscr{E}$, так как она имеет в точности тот же смысл, что и раньше, а именно равна разности между теми же энергетическими уровнями ядра. Это есть внутреннее свойство ядра и не зависит от того, рассматривается ли оно в процессе испускания или поглощения γ-кванта. Таким образом, при поглощении

$$\mathscr{E}'_\gamma = \mathscr{E} + K'_{\text{яд}}, \quad \mathbf{P}'_\gamma = \mathbf{P}'_{\text{яд}},$$

откуда

$$K'_{\text{яд}} = \frac{P'^2_{\text{яд}}}{2M_{\text{яд}}} = \frac{P'^2_\gamma}{2M_{\text{яд}}} = \frac{\mathscr{E}'^2_\gamma}{2M_{\text{яд}}c^2},$$

или с прежней точностью

$$K'_{\text{яд}} = \frac{\mathscr{E}^2}{2M_{\text{яд}}c^2} = K_{\text{яд}}. \tag{76.2}$$

Линии испускания и поглощения γ-квантов сдвинуты относительно друг друга на величину

$$\Delta\mathscr{E} = \mathscr{E}'_\gamma - \mathscr{E}_\gamma = K'_{\text{яд}} + K_{\text{яд}} = 2K_{\text{яд}}. \tag{76.3}$$

2. Полученные результаты полностью применимы к испусканию и поглощению γ-квантов в оптической области спектра. Только в этом случае испускание и поглощение производится не ядрами, а электронными оболочками атомов. В соответствии с этим во всех формулах, приведенных выше, массу ядра следует заменить на массу атома (что практически не имеет абсолютно никакого значения). Кроме того, вместо термина «γ-квант» в оптической области спектра используется термин «фотон». В идеальном случае для получения резонансного поглощения требуется совпадение линий испускания и поглощения, т. е. $\Delta\mathscr{E} = 0$. Величина $\Delta\mathscr{E}$ в оптической области ничтожна, так как энергия испускаемого фотона (равная энергии возбуждения атома $\mathscr{E}$) порядка одного или нескольких электронвольт. Даже для самого легкого атома — атома водорода ($Mc^2 \sim 10^9$ эВ) — получается

$$\Delta\mathscr{E} \approx 1^2/10^9 \approx 10^{-9} \text{ эВ}.$$

По этой причине в оптической области спектра резонансное поглощение света атомами легко наблюдается. Не так обстоит дело для γ-лучей. Энергия γ-квантов, испускаемых ядрами, примерно в 10^6 раз больше, а следовательно, сдвиг $\Delta\mathscr{E}$ в 10^{12} раз больше, чем в оптической области. Поэтому долгое время считалось, что осуществить резонансное поглощение γ-квантов невозможно.

Казалось бы, что сдвиг $\Delta\mathscr{E}$ можно устранить, приведя в движение излучающее ядро в направлении к поглощающему. Тогда из-за эффекта Доплера энергия излучаемого γ-кванта увеличится, а скорость движения можно подобрать так, чтобы величина $\Delta\mathscr{E}$ обратилась в нуль. Того же самого можно достигнуть приближением поглощающего ядра к испускающему. Однако здесь не принято во внимание, что источник испускает, а поглотитель поглощает не бесконечно тонкую линию, а линию конечной ширины. Для возможности резонансного поглощения необходимо, конечно, чтобы линии испускания и поглощения перекрывались, т. е. должно быть

$$\Gamma \gtrsim K_{\text{яд}}, \tag{76.4}$$

где Γ — полуширина линии.

3. Уширение линии обусловлено различными причинами. Прежде всего существует *доплеровское уширение* спектральных линий, обусловленное тепловым движением атомов. *Доплеровская полуширина линии* $\Delta\nu$ может быть оценена по формуле

$$\frac{\Delta\nu_{\text{доп}}}{\nu} \approx \frac{1}{c}\sqrt{\frac{2kT}{M_{\text{яд}}}}, \tag{76.5}$$

где k — постоянная Больцмана, а T — термодинамическая температура источника (см. т. IV, § 89). Если полуширину выражать в энергетических единицах ($\Gamma = h\Delta\nu$), то эта формула преобразуется в

$$\Gamma_{\text{доп}} \approx \frac{\mathscr{E}}{c}\sqrt{\frac{2kT}{M_{\text{яд}}}}, \tag{76.6}$$

так как в рассматриваемом случае энергию γ-кванта можно с большой точностью положить равной энергии возбуждения ядра. Доплеровское уширение играет основную роль в случае источника, содержащего много атомов или атомных ядер. Оно, очевидно, пропадает, когда излучателем является изолированный атом или изолированное ядро, так как в этом случае говорить о тепловом движении не имеет смысла. Движение изолированного атома или ядра сказывается на смещении спектральных линий, но не на их уширении.

В случае изолированного ядра ширина линии называется *естественной шириной.* Она может быть оценена по времени жизни τ возбужденного ядра с помощью соотношения неопределенностей

$$\Gamma_{\text{ест}} \approx \frac{\hbar}{\tau}. \tag{76.7}$$

Посмотрим теперь на примере, выполняется ли условие (76.4) в оптической области и в области γ-лучей. В качестве примера возьмем

ядро изотопа железа $^{57}_{26}\mathrm{Fe}$. Энергия возбуждения первого уровня этого ядра равна 14 кэВ, т. е. для γ-лучей это совсем малая величина. Время жизни его $\tau \approx 10^{-8}$ с, а естественная ширина линии

$$\Gamma_{\text{ест}} \approx 10^{-8} \text{ эВ}.$$

Кинетическая энергия ядра железа, приобретаемая им согласно формуле (76.1) при испускании γ-кванта, будет

$$K_{\text{яд}} = \frac{(14 \cdot 10^3)^2}{2 \cdot 57 \cdot 938 \cdot 10^6} = 0{,}00183 \text{ эВ},$$

т. е. примерно в 10^5 раз превышает естественную ширину спектральной линии. О выполнении условия (76.4) не может быть и речи. Поэтому резонансное поглощение γ-квантов на изолированных неподвижных ядрах невозможно.

Иначе обстоит дело в случае оптических фотонов. В этом случае, согласно той же формуле (76.1), кинетическая энергия ядра отдачи порядка

$$K_{\text{яд}} \approx \frac{1^2}{2 \cdot 57 \cdot 938 \cdot 10^6} \approx 10^{-11} \text{ эВ}.$$

Взяв для естественной ширины линии прежнее значение 10^{-8} эВ (это — очень узкая линия), видим, что условие (76.4) хорошо выполняется. Поэтому-то резонансное поглощение оптических фотонов происходит и на изолированных атомах.

4. Обратимся теперь к испусканию и поглощению γ-квантов макроскопическими телами — кристаллами. Казалось бы, что в этом случае достаточно естественную ширину спектральной линии заменить на доплеровскую. Для температуры $T = 300$ К в случае изотопа железа $^{57}_{26}\mathrm{Fe}$

$$\Gamma_{\text{доп}} \approx 0{,}018 \text{ эВ},$$

что на порядок больше кинетической энергии ядра $K_{\text{яд}}$. Условие (76.4) выполняется, хотя и на пределе. Поэтому следует ожидать, что в рассматриваемом случае резонансное поглощение γ-квантов на отдельных ядрах должно наблюдаться и в кристаллах. Однако при переходе к достаточно жестким γ-квантам и при понижении температуры условие (76.4) перестает выполняться, а резонансное поглощение в кристалле, казалось бы, должно сделаться невозможным. Например, для ядра иридия $^{194}\mathrm{Ir}$ энергия возбуждения $\mathscr{E} = 129$ кэВ, так что в этом случае

$$K_{\text{яд}} = \frac{(129 \cdot 10^3)^2}{2 \cdot 191 \cdot 938 \cdot 10^6} = 0{,}046 \text{ эВ}.$$

Поэтому даже при $T = 300$ К условие $\Gamma_{\text{доп}} > K_{\text{яд}}$ не выполняется, так как для более тяжелого иридия при одинаковых температурах $\Gamma_{\text{доп}}$ меньше, чем для железа. Но и в тех случаях, когда условие $\Gamma_{\text{доп}} > K_{\text{яд}}$ выполняется, следовало бы ожидать очень широкие и пологие максимумы резонансного поглощения.

При понижении температуры источника и поглотителя область перекрытия доплеровских линий испускания и поглощения уменьшается.

Казалось бы, что при этом должна уменьшаться и доля поглощаемых γ-квантов. На самом деле, как показали опыты Мёссбауэра в 1958 г., она увеличивается. Этот неожиданный результат, как понял сам Мёссбауэр, указывает на *статистический характер* испускания и поглощения γ-квантов в кристалле. Большая часть γ-квантов испускается и поглощается так, как описано выше, т. е. отдельными ядрами. Однако поскольку ядра в кристаллической решетке связаны между собой, наряду с такими индивидуальными процессами происходят и *коллективные процессы*, напоминающие возбуждение квазичастиц в теории теплоемкостей твердых тел, допускаемые квантовой механикой. Какой процесс произойдет — индивидуальный или коллективный, — зависит от случая. Соотношение между числом тех и других процессов управляется *статистическими законами.*

В коллективных процессах возбужденное ядро возвращается в нормальное состояние, энергия возбуждения уносится γ-квантом, но импульс воспринимается *кристаллом в целом* или, во всяком случае, большой группой атомов. Аналогично, энергия испущенного фотона поглощается отдельным ядром, а его импульс передается *кристаллу в целом.* На кинетическую энергию всего кристалла (ввиду большой массы последнего), возникающую в этих процессах, приходится ничтожная доля, малая по сравнению с естественной шириной линии (измеренной в энергетических единицах). Явление происходит так, как если бы какая-то часть ядер испускала и поглощала энергию, но не испытывала отдачи импульса. *Испускание и поглощение γ-квантов без отдачи импульса и составляет сущность эффекта Мёссбауэра.* Поскольку явления испускания и поглощения γ-квантов происходят так, как если бы масса ядра была бесконечно велика, они не сопровождаются доплеровским уширением спектральных линий. Остается только естественная ширина линии. В таких процессах проявляются, таким образом, очень узкие спектральные линии испускания и поглощения γ-квантов.

5. Это объяснение Мёссбауэра убедительно подтверждается опытами по резонансному поглощению γ-квантов. Принципиальная схема опыта для наблюдения этого явления приведена на рис. 138. Источник

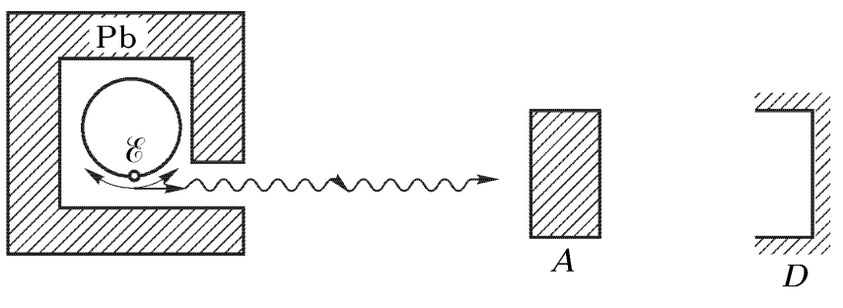


Рис. 138

резонансного γ-излучения $\mathscr{E}$ медленно движется по окружности с помощью часового механизма относительно поглотителя A. За поглотителем расположен счетчик γ-квантов D. Измеряется зависимость скорости счета от скорости движения источника в моменты приближения

и удаления его от поглотителя. Если источник движется достаточно быстро, то линия испускания сдвигается относительно линии поглощения и резонансное поглощение не наблюдается. При уменьшении скорости источника обе эти линии сближаются, а при их совпадении появляется острый максимум поглощения. Это проявляется в резком уменьшении скорости счета счетчика. На рис. 139 изображена экспериментальная кривая, полученная таким путем. Источником излучения является ядро $^{57}_{27}\mathrm{Co}$, которое в результате K-захвата превращается

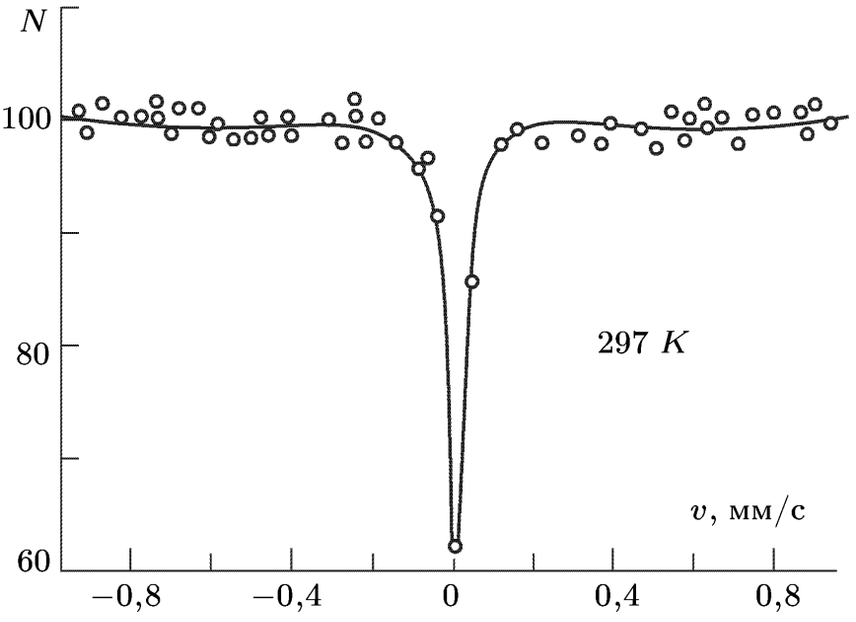


Рис. 139

в ядро железа $^{57}_{26}\mathrm{Fe}$, испускающее γ-кванты с энергией $\mathscr{E} = 14$ кэВ. Поглотителем служит соль $\mathrm{K_3}\ ^{57}\mathrm{Fe}\ (\mathrm{CN})_6$. Кривая получена при $T = 297$ К. По вертикальной оси отложена относительная интенсивность γ-излучения, прошедшего через поглотитель (максимальная интенсивность принята за 100).

Из рисунка видно, что резонанс нарушается уже при ничтожных скоростях источника v — порядка 0,1 мм/с. Отсюда следует, что относительная ширина самих линий испускания и поглощения $v/c \approx 10^{-11}$–10^{-12}, а абсолютная $\Gamma = \mathscr{E}v/c \approx 14\,000v/c \approx 10^{-7}$–$10^{-8}$ эВ, т. е. того же порядка, что и естественная ширина линии. Значит, в опыте действительно наблюдалось резонансное испускание и поглощение *без отдачи импульса.* Впервые такое экспериментальное доказательство эффекта было дано Мёссбауэром в 1958 г. Излучателем и поглотителем γ-квантов у него были изотопы $^{191}\mathrm{Ir}$, охлажденные до 88 К. Постановка этого фундаментального опыта и может считаться временем открытия эффекта Мёссбауэра.

Эффект Мёссбауэра наблюдается на многих веществах, причем для многих из них были зафиксированы еще более узкие линии испускания и поглощения, чем у рассмотренных выше изотопов железа и иридия. Рабочие температуры для разных веществ колеблются в пределах от комнатных до гелиевых (около 4 К и ниже). С ростом температуры

эффект постепенно ослабевает и наконец совсем пропадает. Для наблюдения эффекта Мёссбауэра благоприятным является высокое значение f-коэффициента, определяющего относительную долю процессов испускания γ-квантов, происходящих без отдачи импульса. В свою очередь этот коэффициент тем выше, чем ниже энергия возбуждения ядра $\mathscr{E}$, а также чем выше дебаевская температура T_D, поскольку она характеризует прочность связи ядра в кристаллической решетке.

Разрешающая способность метода мёссбауэровской спектроскопии характеризуется относительной шириной линии $\Gamma/\mathscr{E}$. Так, для изотопа железа $^{57}_{26}\mathrm{Fe}$ $\Gamma \approx 10^{-8}$ эВ, его период полураспада $T_{1/2} = 10^{-7}$ с, $T_D \approx 500$ К, $f > 0{,}6$ вплоть до комнатной температуры 300 К. В связи с такими хорошими характеристиками этот изотоп железа широко используется в работах по эффекту Мёссбауэра. Другим веществом, применяющимся при комнатной температуре, является изотоп олова $^{119}\mathrm{Sn}$ ($\mathscr{E} = 23{,}8$ кэВ, $T_{1/2} = 1{,}8 \cdot 10^{-8}$ с, $\Gamma = 2{,}5 \cdot 10^{-8}$ эВ, $\Gamma/\mathscr{E} \approx 10^{-12}$), а также изотоп $^{181}\mathrm{Ta}$ ($\mathscr{E} = 6{,}2$ кэВ, $T_{1/2} = 6{,}8 \cdot 10^{-8}$ с, $\Gamma = 6{,}7 \cdot 10^{-11}$ эВ, $\Gamma/\mathscr{E} \approx 10^{-14}$, $f \approx 5\,\%$ при комнатной температуре). Уникальной разрешающей способностью обладает $^{67}\mathrm{Zn}$ ($\mathscr{E} = 93$ кэВ, $T_{1/2} = 9{,}4 \cdot 10^{-6}$ с, $\Gamma \approx 5 \cdot 10^{-11}$ эВ, $\Gamma/\mathscr{E} \approx 5 \cdot 10^{-16}$). Однако дебаевская температура у $^{67}\mathrm{Zn}$ настолько низка, что даже при гелиевых температурах (около 4 К) f составляет только около 0,3 %.

6. Основное применение эффекта Мёссбауэра связано с тем, что он дает уникальный метод измерения ничтожных изменений энергии, которые не могут быть измерены никаким другим методом. Ограничимся двумя примерами.

С помощью эффекта Мёссбауэра удалось обнаружить в лабораторных условиях *гравитационное смещение спектральных линий*, предсказанное теорией относительности Эйнштейна (см. § 7, а также т. I, § 72). По этой теории фотон, распространяющийся вертикально в поле тяжести Земли, при прохождении расстояния h меняет свою энергию $\mathscr{E}$ на величину

$$\Delta\mathscr{E} = \frac{\mathscr{E}}{c^2}\, gh, \tag{76.8}$$

что проявляется в изменении его частоты. При падении вниз частота фотона увеличивается (фиолетовое смещение), при поднятии вверх уменьшается (красное смещение). Паунд (р. 1919) и Ребке в 1959 г. поставили соответствующий опыт в башне Гарвардского университета ($h = 22{,}6$ м), использовав в качестве излучателя и поглотителя образцы из изотопа железа $^{59}_{26}\mathrm{Fe}$, охлажденные до гелиевых температур. Относительное изменение энергии фотона в этом случае составляло

$$\frac{\Delta\mathscr{E}}{\mathscr{E}} = \frac{gh}{c^2} \approx 2{,}46 \cdot 10^{-15},$$

а абсолютное $\Delta\mathscr{E} \approx 14\,000 \cdot 2{,}46 \cdot 10^{-15} \approx 3{,}4 \cdot 10^{-11}$ эВ, что примерно в 300 раз меньше естественной ширины линии. Для компенсации этого изменения энергии доплеровским смещением требуется скорость

источника $v \approx c\Delta\mathscr{E}/\mathscr{E} \approx 0{,}75$ мкм/с. Для надежного обнаружения гравитационного смещения необходимо было измерять изменения энергии с погрешностью $10^{-3}\Gamma_{\text{ест}} \approx 10^{-11}$ эВ. Тем не менее эффект был с уверенностью обнаружен, в согласии с предсказанием Эйнштейна. Обнаруженный в лабораторных условиях эффект был примерно в 10^9 раз меньше гравитационного смещения, вызываемого полем тяготения Солнца, который измеряется уже астрономическими методами. В 1965 г. опыт был повторен в усовершенствованной форме Паундом и Снайдером.

Методами мёссбауэровской спектроскопии удалось обнаружить влияние электронных оболочек атомов на процессы, происходящие внутри атомных ядер. Линии мёссбауэровских спектров одного и того же ядра заметно сдвигаются и меняются по ширине при переходе от одного химического соединения к другому, при изменении структуры кристаллической решетки, при изменении температуры, при наложении и снятии механических напряжений и т. п. В качестве примера на рис. 140 приведены мёссбауэровские спектры изотопа железа $^{57}_{26}\text{Fe}$ для нержавеющей стали (слева) и железосодержащего

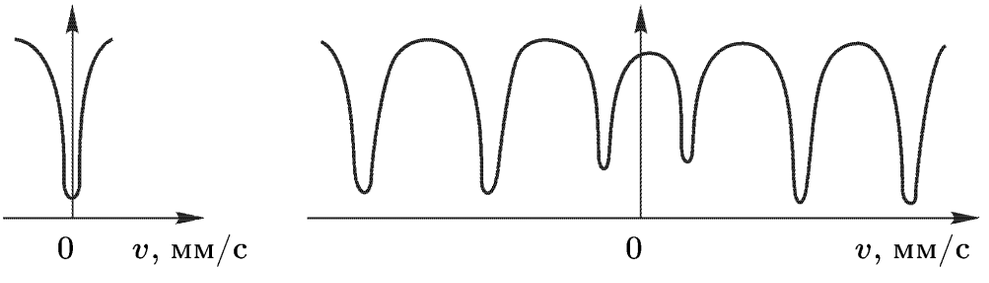


Рис. 140

соединения Fe_2O_3 — гематита (справа). По горизонтальной оси отложена скорость источника относительно поглотителя, по вертикальной — интенсивность γ-лучей, прошедших через поглотитель. Для нержавеющей стали получается одиночная линия. Для окиси железа Fe_2O_3 под действием магнитного поля электронных оболочек линия расщепляется уже на шесть линий. Впрочем, и спектральная линия нержавеющей стали испытывает также расщепление при наклепе вследствие изменения внутренней структуры кристалла из-за пластической деформации. Методы мёссбауэровской спектроскопии нашли широкое применение в исследованиях по физике твердого тела.

Глава X

КРАТКИЕ СВЕДЕНИЯ О ЯДЕРНЫХ МОДЕЛЯХ

§ 77. Общие сведения

1. Необходимость ядерных моделей была обоснована в § 63. Она обусловлена тем, что до настоящего времени мы не располагаем последовательной физической теорией ядра. Трудность построения такой теории связана в первую очередь с недостаточной полнотой и определенностью данных о ядерных силах. Сюда добавляются математические трудности, поскольку ядро является многочастичной квантовой системой с сильным взаимодействием. Поэтому на основе первичных физических принципов до сих пор рассматривалась структура только простейших ядер — дейтрона ^{2}H, тритона ^{3}H и ^{4}He. В случае более сложных ядер для систематизации опытных фактов и предсказания новых (хотя бы качественных) приходится прибегать к *модельным представлениям.* Каждая модель имеет ограниченную область применимости, т. е. может претендовать на приближенное описание только более или менее узкой области явлений. Из всего многообразия параметров, характеризующих атомное ядро, она учитывает, и притом в схематизированной форме, только некоторые, являющиеся для этой области явлений главными и определяющими, и пренебрегает влиянием всех остальных параметров. Конечно, возможность и правильность самого выбора определяющих параметров в конце концов доказывается сопоставлением результатов, к которым приводит принятая модель, с опытными фактами, на объяснение которых она претендует. Модель тем лучше, чем большее количество фактов она объясняет.

2. Модели ядра можно разделить на *одночастичные, коллективные и обобщенные.* В одночастичных моделях учитываются степени свободы, описывающие движение индивидуальных нуклонов; в коллективных — степени свободы, определяющие коррелированное движение всех нуклонов ядра. Одночастичные модели называют также *моделями независимых частиц,* коллективные — *моделями с сильным взаимодействием.* Если пользоваться наглядной терминологией физики сплошных сред, то можно сказать, что коллективные модели следует применять в тех случаях, когда средний свободный пробег каждого нуклона в ядре мал по сравнению с размерами самого ядра. В этом случае определяющее значение имеют частые и интенсивные взаимодействия каждого нуклона с его ближайшими соседями. В противоположном случае, когда средний свободный пробег нуклона значительно больше размеров ядра, применяются одночастичные модели, в которых

принимается, что нуклоны движутся независимо в некотором усредненном *самосогласованном поле.* Более трудным является случай, когда свободный пробег нуклона и размер ядра — величины одного порядка.

Таким образом, если пользоваться классическими представлениями, то в одночастичных и коллективных моделях о движении нуклонов в ядре вводятся противоположные и взаимно исключающие предположения. Допустимость обоих предположений в ядерной физике обосновывается тем, что на движение нуклонов в ядре нельзя буквально переносить представления, заимствованные из классической физики сплошных сред. Во-первых, потому, что в ядре, даже самом тяжелом, не так много частиц, чтобы без всяких оговорок оперировать с ним как со сплошной средой и использовать статистические методы. Во-вторых, и это главное, движение нуклона в ядре существенно квантовое. Действительно, средняя кинетическая энергия нуклона в ядре $\mathscr{E} \approx 25$ МэВ, а масса нуклона $m = 1{,}68 \cdot 10^{-24}$ г. Этому соответствует средняя длина волны де Бройля нуклона в ядре $\lambda = h/\sqrt{2m\mathscr{E}} \approx 0{,}5 \cdot 10^{-12}$ см — величина порядка радиуса ядра. При таких условиях о классическом движении нуклона в ядре говорить не приходится.

Отметим, в частности, что в невозбужденном ядре согласно принципу Паули все нижние состояния заняты. При движении в ядре нуклон должен вытеснить другой одноименный нуклон из занимаемого им состояния и встать на его место. Но согласно квантовой механике такой обмен местами одноименных нуклонов не приводит к новому состоянию, а это не согласуется с представлениями классической физики. Кроме того, в некоторых одночастичных моделях ядра состояния ядра выражаются через одночастичные состояния не самих нуклонов, а квазичастиц. Введение же квазичастиц соответствует уже коллективному способу описания ядра. При таком понимании резкое разграничение моделей ядра на одночастичные и коллективные фактически утрачивается.

Комбинированием одночастичных и коллективных моделей получаются *обобщенные модели ядра*, в которых одновременно принимаются во внимание как одночастичные, так и коллективные степени свободы, существенные для рассматриваемой группы явлений.

Подробное описание ядерных моделей с указанием областей и границ их применимости должно даваться в специальных руководствах по ядерной физике. Здесь, в общем курсе физики, можно затронуть эти вопросы весьма бегло и схематично, ограничиваясь общими представлениями.

3. Простейшей и исторически первой из коллективных моделей является *капельная модель ядра*, о которой уже говорилось в § 64. Она была предложена в 1936 г. Нильсом Бором и независимо от него Я. И. Френкелем (1894–1952). Капельной моделью мы и ограничимся при рассмотрении коллективных моделей. Здесь будет сделано несколько кратких дополнительных замечаний об этой модели.

В капельной модели принимается, что ядро ведет себя подобно капле несжимаемой заряженной жидкости. Воспользуемся формула-

ми (65.1) и (65.2) для определения некоторых параметров «ядерной жидкости». Для концентрации нуклонов в ядре получаем

$$n = \frac{A}{4\pi R^3/3} = \frac{3}{4\pi r_0^3} = 0{,}87 \cdot 10^{38} \approx 10^{38}\ \text{см}^{-3},$$

для плотности вещества в ядре

$$\rho = nm \approx 1{,}45 \cdot 10^{14}\ \text{г/см}^3,$$

где m — масса нуклона, а в качестве r_0 взято $r_0 = 1{,}4 \cdot 10^{-13}$ см. Среднее расстояние между нуклонами равно

$$\delta = \sqrt[3]{\frac{4\pi R^3}{3A}} = \sqrt[3]{\frac{4\pi}{3}} r_0 \approx 2{,}3 \cdot 10^{-13}\ \text{см}.$$

Таким образом, если исключить из рассмотрения самые легкие ядра, то концентрация нуклонов, плотность вещества в ядре, а также среднее расстояние между нуклонами практически одинаковы во всех ядрах. Это и позволяет в капельной модели считать ядро несжимаемым.

Несжимаемость «ядерной жидкости» и отражает тот факт, что между нуклонами ядра существует очень сильное взаимодействие, а потому капельная модель и относится к числу коллективных. Она простейшим образом выясняет физический смысл первых трех членов в полуэмпирической формуле Вейцзеккера (64.6) для энергии связи ядра. Остальные два члена в этой формуле — энергия симметрии и энергия спаривания — уже не могут быть интерпретированы на основе чисто капельной модели. Энергия симметрии обычно связывается с принципом Паули. Наличие энергии спаривания свидетельствует о зависимости сил взаимодействия между нуклонами от относительной ориентации их спинов. С учетом этих двух дополнительных членов капельная модель неплохо описывает зависимость усредненной энергии связи ядра от зарядового и массового чисел Z и A. Более тонкие индивидуальные (а не усредненные) изменения энергии связи, как и другие индивидуальные свойства атомных ядер, в капельной модели остаются необъясненными. В этом один из недостатков этой модели.

4. С точки зрения капельной модели ядро в невозбужденном состоянии должно было бы иметь сферическую форму. И действительно, до 1950 г. в ядерной физике господствовало представление о сферической симметрии равновесной формы ядра. Однако позднее от этого представления пришлось отказаться. По оболочечным соображениям многие ядра уже в невозбужденном состоянии должны иметь не сферическую форму, а форму эллипсоида вращения и даже трехосного эллипсоида (это не общепринято). Такие ядра стали называть *деформированными*. Эта неудачная терминология, к сожалению, применяется иногда и теперь. Неудачность ее заключается в том, что несферическая форма основного состояния ядра есть его внутреннее свойство, а не проявление результата какого-то внешнего деформирующего воздействия. О некоторых фактах, которые заставили признать многие ядра несферическими, будет сказано в следующем параграфе (п. 12).

Согласно квантовой механике, ввиду тождественности одинаковых частиц, не имеет смысла говорить о вращении сферически-симметричного ядра — в сферически-симметричных ядрах не наблюдаются вращательные энергетические уровни. Напротив, в сферически-несимметричных ядрах такие вращательные уровни наблюдаются. Они принадлежат к низким уровням возбуждения.

Другими возбуждениями ядра, допускаемыми капельной моделью, являются *деформации* и *колебания поверхности ядра.* Объемные колебания ядра, ввиду предположенной несжимаемости «ядерной жидкости», практически невозможны. Простейшими являются *квадрупольные* малые колебания ядра, когда его поверхность попеременно принимает форму вытянутого и сплюснутого эллипсоида вращения. При таких колебаниях у ядра возникают *квадрупольные моменты.* Более сложными являются малые *октупольные* колебания, при которых ядро принимает грушевидную форму и приобретает *октупольный момент*, а также колебания с возбуждением высших *мультипольных* моментов. Колебания должны быть квантованы, а энергия возбуждения колебаний, как следовало бы ожидать, должна определяться формулой $\mathscr{E} = n\hbar\omega$, где n — целое квантовое число. Однако даже при малых квантовых числах n (0, 1, 2) полученные таким образом значения в несколько раз расходятся с тем, что дает опыт. При сильных колебаниях нарушается их гармоничность. Такие колебания могут возбуждаться, например, при попадании нейтрона в ядро. В результате ядро может разделиться на две части. Таким образом, капельная модель ядра объясняет деление некоторых тяжелых ядер (см. гл. XIV).

Наконец, в рамках капельной модели возможно возбуждение еще одного типа колебаний. Ядро состоит из смеси двух «несжимаемых жидкостей» — *нейтронной* и *протонной*, проникающих друг в друга. Такие колебания напоминают *оптическую ветвь частот* при колебаниях кристаллических решеток, построенных из различных атомов (см. § 56). Они приводят к возбуждению у ядра электрических дипольных моментов, т. е. к *поляризации ядра.* Поляризационным возбуждениям соответствуют довольно высокие энергии — примерно 15–20 МэВ для тяжелых и 20–25 МэВ для легких ядер.

§ 78. Оболочечная модель ядра

1. Из одночастичных моделей наибольшее значение имеет *оболочечная модель ядра.* Оказывается, что ядра обладают известной периодичностью, аналогичной периодичности атомов, нашедшей свое отражение в периодической системе Менделеева. Ядра, содержащие магическое число нейтронов или протонов (2, 8, 20, 50, 82 и 126 для нейтронов), и в особенности дважды магические ядра, выделяются среди остальных ядер особой прочностью. Кроме того, для ядер, содержащих магическое число протонов, характерна сферическая симметрия распределения зарядов в невозбужденных состояниях, т. е. отсутствие квадрупольных

электрических моментов. В этом отношении магические ядра напоминают атомы инертных газов, характеризующиеся сферической симметрией, химической пассивностью и связанными с ней наибольшими энергиями ионизации. Однако зарядовые числа инертных атомов ($Z = 2, 10, 18, 36, 54, 86$), за исключением $Z = 2$, не совпадают с магическими числами для ядер.

2. Магические свойства ядер проявляются прежде всего в том, что энергия присоединения последнего нуклона (равная энергии его отделения) при образовании магического ядра аномально велика по сравнению с энергией присоединения последнего нуклона для предыдущих и последующих соседних нуклонов. Рассмотрим, например, ряд самых легких ядер (${}_1^1\mathrm{H}$, ${}_1^2\mathrm{H}$, ${}_2^3\mathrm{He}$, ${}_2^4\mathrm{He}$, каждое из которых получается из предыдущего присоединением одного нейтрона или протона. Оказывается, что в этом случае энергии присоединения одного нуклона к предыдущему ядру соответственно равны 2,2; 5,5; 20,6 МэВ, т. е. резко возрастают по мере приближения к дважды магическому ядру ${}_2^4\mathrm{He}$ ($Z = N = 2$). Однако при дальнейшем добавлении к ядру ${}_2^4\mathrm{He}$ нейтрона или протона, т. е. при образовании ядер ${}_2^5\mathrm{He}$ и ${}_3^5\mathrm{Li}$, в обоих случаях энергия присоединения отрицательна, так что оба ядра ${}_2^5\mathrm{He}$ и ${}_3^5\mathrm{Li}$ неустойчивы и в природе не встречаются (см. § 64, задача 5). Особая прочность дважды магического ядра ${}_2^4\mathrm{He}$ проявляется также в том, что такие ядра (α-частицы) испускаются при α-распаде.

Увеличение энергии присоединения последнего нуклона по мере приближения к магическому ядру имеет место и для других ядер, хотя для тяжелых ядер оно выражено и не так резко, как для легких. При иллюстрации такого увеличения необходимо принять во внимание повышенную прочность стабильных ядер с четными числами нуклонов определенного типа по сравнению с нечетными числами нуклонов того же типа. А так как все магические числа четные, то приходится сравнивать только ядра с четным числом протонов или с четным числом нейтронов. В качестве примера укажем, что энергии присоединения протона к четно-четным ядрам ${}_4^8\mathrm{Be}$ и ${}_6^{12}\mathrm{C}$ соответственно равны $-0{,}18$ и $+1{,}9$ МэВ, тогда как для магического ядра ${}_8^{16}\mathrm{O}$ энергия присоединения протона (в результате чего образуется ядро ${}_9^{17}\mathrm{F}$) минимальна и составляет 0,6 МэВ. Аналогичная картина наблюдается и в случае присоединения нейтрона к тем же ядрам. Эти факты и свидетельствуют о повышенной прочности магического ядра ${}_8^{16}\mathrm{O}$ по сравнению с соседними четно-четными ядрами.

В случае тяжелых стабильных ядер, когда числа Z и N значительно отличаются друг от друга, приходится сравнивать ядра с четными числами протонов при неизменном числе нейтронов или ядра с четными числами нейтронов при неизменном числе протонов. Так, при переходе от $Z = 80$ к магическому числу $Z = 82$ энергия присоединения протона при всех N резко падает, что свидетельствует о повышенной стабильности ядра с магическим числом протонов $Z = 82$. Аналогичное явление имеет место в окрестности магического числа $N = 126$. Здесь также

энергия присоединения нейтрона к ядру для всех Z резко уменьшается при переходе от ядра с $N = 124$ к магическому ядру с $N = 126$.

3. Магические свойства ядер проявляются и в относительной распространенности химических элементов. Большей распространенностью обладают стабильные ядра с магическими числами Z или N по сравнению с ядрами с соседними четными значениями этих чисел. Например, для $Z = 20$ ($_{20}$Ca) имеются пять стабильных изотопов, для которых число нейтронов N четное. Как уже указывалось выше, сравнивать следует ядра, у которых N или Z отличаются от магического числа на два, так как четно-четные ядра обладают большей стабильностью, а потому и большей распространенностью в природе. При $Z = 18$ и при $Z = 22$ число стабильных ядер с четным N равно трем. Особенность области около $Z = 20$ проявляется даже и у ядер с нечетным массовым числом, так как при $N = 19$ или $N = 21$ не существует ни одного стабильного ядра. Существующий в природе изотоп калия $^{40}_{19}$K ($Z = 19$, $N = 21$) радиоактивен, хотя и обладает очень большим периодом полураспада ($1{,}4 \cdot 10^9$ лет). Магические свойства проявляются и для нейтронов при $N = 20$. Существуют пять изотопов при $N = 20$, два при $N = 18$ и два при $N = 22$.

Аргументация приведенного типа для случая $N = 50$ не очень убедительна, но существуют другие, не менее убедительные аргументы. При $N = 82$ приведенная аргументация вновь становится весьма убедительной. Имеется семь стабильных ядер с $N = 82$, три с $N = 80$ и два с $N = 84$. Олово ($Z = 50$) имеет наибольшее число стабильных изотопов (10), причем три из них — с нечетным A. Но, как и для случая $N = 50$, это не может считаться достаточно убедительной аргументацией, так как кадмий ($Z = 48$) и теллур ($Z = 52$) имеют каждый восемь изотопов.

Последним стабильным ядром с $Z = N$ является дважды магическое ядро $^{40}_{20}$Ca. Его содержание среди естественной смеси изотопов кальция составляет 97 %. Предшествующее четно-четное ядро с $Z = N$ ($^{36}_{18}$Ar) в естественной смеси изотопов аргона составляет 0,34 %, а следующий четно-четный $^{44}_{22}$Ti среди изотопов титана совсем не встречается.

Среди различных изотопов данного элемента обычно самым распространенным является изотоп со средним значением массового числа A. Исключение составляют случаи, когда среди изотопов элемента имеются изотопы, содержащие магические числа нейтронов $N = 50$ или $N = 82$. В этих случаях, как правило, самыми распространенными являются изотопы с $N = 50$ и $N = 82$, независимо от значения массового числа A.

4. Существует общее правило, что если наиболее тяжелый (или наиболее легкий) стабильный изотоп элемента с четным Z имеет N нейтронов, то наиболее тяжелый (или наиболее легкий) стабильный изотоп элемента с зарядовым числом $Z + 2$ имеет не меньше $N + 2$ нейтронов. Однако существуют пять исключений из этого правила, в которых ядра указанного типа имеют одно и то же число нейтронов. Сюда относятся

три пары наиболее тяжелых из сравниваемых изотопов:

$$^{48}_{20}\mathrm{Ca} - {}^{50}_{22}\mathrm{Ti}, \quad {}^{86}_{36}\mathrm{Kr} - {}^{88}_{38}\mathrm{Sr}, \quad {}^{136}_{54}\mathrm{Xe} - {}^{138}_{56}\mathrm{Ba}$$

и две пары наиболее легких:

$$^{90}_{40}\mathrm{Zr} - {}^{92}_{42}\mathrm{Mo}, \quad {}^{142}_{60}\mathrm{Nd} - {}^{144}_{62}\mathrm{Sm}.$$

У обоих ядер первой пары $N = 28$, у двух пар $N = 50$, а у двух остальных $N = 82$. Существование таких исключений указывает на более слабую связь для 29-, 51- и 83-го нейтронов по сравнению с обычной, а следовательно, более сильную связь для 28-, 50- и 82-го нейтронов. Но последние два числа как раз и являются магическими, а число 28 также часто относят к магическим для нейтронов.

5. Отметим, наконец некоторые закономерности α- и β-распада.

1) Все три радиоактивных семейства, встречающихся в природе, заканчиваются свинцом $_{82}\mathrm{Pb}$, т. е. элементом с магическим числом $Z = 82$. Ядра с $Z < 82$ из числа встречающихся в природе, как правило, стабильны.

2) Энергия испускаемых α-частиц увеличивается с ростом Z, как и должно быть согласно полуэмпирической формуле Вейцзеккера. Исключение составляют ядра с $Z = 84$, $N = 84$, $N = 128$, которые после α-распада переходят в магические ядра с $Z = 82$, $N = 82$, $N = 126$ и испускают при этом α-частицы с большей энергией по сравнению с соседними ядрами. Изотопы же полония $^{210}_{84}\mathrm{Po}$ и $^{212}_{84}\mathrm{Po}$ испускают также и длиннопробежные α-частицы.

3) Аналогичные результаты справедливы для максимальной энергии электронов, возникающих при β-распаде. Такая энергия достигает наибольшей величины при β-переходах с образованием магических ядер и наименьшей величины при β-распадах самих магических ядер.

6. Можно было бы привести и другие аргументы, свидетельствующие о существовании магических ядер. Но мы не будем делать этого. Периодическую повторяемость свойств атомных ядер и многие другие экспериментальные факты, связанные с существованием магических ядер, и пытается объяснить оболочечная модель ядра. Она строится по аналогии с теорией периодической системы химических элементов. Последняя, как известно, предполагает, что имеет смысл говорить не только о состоянии электронной оболочки в целом, но и о состоянии *каждого электрона в отдельности.* Каждый электрон независимо движется в *самосогласованном центрально-симметричном* поле, как если бы между электронами не существовало никакого взаимодействия. На самом деле взаимодействие есть, и отнюдь не слабое. Но оно эффективно учитывается посредством самосогласованного поля. Это поле образуется кулоновским полем ядра и усредненным электрическим полем всех электронов за исключением рассматриваемого. Этим и обеспечивается кажущаяся независимость движения электронов, при которой сохраняется энергия, а в силу центральной симметрии поля — и момент импульса отдельного электрона при его орбитальном движении. Это в свою очередь дает возможность характеризовать состояние каждого

электрона квантовыми числами, что и делается при построении периодической системы химических элементов. Хотя такой способ учета взаимодействия электронов в атоме и представляется логически не совсем последовательным, но при построении теории периодической системы химических элементов он приводит к правильным результатам. Причина этого уже была указана в § 47.

Казалось бы, что для ядра теория, построенная по аналогичной схеме, невозможна. Во-первых, потому, что у ядра нет силового центра для формирования центрально-симметричного самосогласованного поля, в котором двигались бы нуклоны. Во-вторых, в отличие от атомных оболочек, где электроны расположены далеко друг от друга, в ядре нуклоны упакованы очень плотно (концентрация нуклонов в ядре $n \approx 10^{38}$ см$^{-3}$), а ядерные силы, действующие между ними, очень велики. По этим причинам средняя длина свободного пробега нуклона в ядре от столкновения до столкновения, если ее оценивать классически, порядка размеров самого ядра и даже меньше. При таких условиях, казалось бы, не имеет смысла говорить о регулярном независимом орбитальном движении нуклонов в ядре.

Однако как уже отмечалось в предыдущем параграфе, необходимо принять во внимание следующие обстоятельства. В невозбужденном ядре нуклоны занимают все энергетически самые низкие состояния, а принцип Паули запрещает двум одинаковым нуклонам находиться в одном и том же квантовом состоянии (поскольку спин нуклона равен $1/2$). При столкновении двух нуклонов один из них должен терять энергию и перейти в энергетически более низкое состояние. А такой процесс невозможен, поскольку в невозбужденном ядре все такие состояния уже заняты. Если же потери энергии нет, то нуклоны просто обмениваются местами, а это, как также отмечалось в предыдущем параграфе, совсем не меняет состояния ядра, как если бы вообще никакого столкновения не было. В результате если ядро находится в невозбужденном состоянии, то эффективно все это проявляется так, как будто бы средняя длина свободного пробега нуклона в ядре стала больше и даже во много раз превосходила размеры самого ядра. В таком случае в нулевом приближении можно говорить как бы о независимом движении нуклонов в ядре. (Это справедливо и для квазичастиц.) Сильное же взаимодействие между нуклонами, а также малый радиус действия ядерных сил позволяют надеяться ввести центрально-симметричное нуклонное поле, в котором в нулевом приближении независимо и движутся нуклоны.

7. Шаткость приведенных соображений, конечно, не обеспечивает надежности теории, которая строится на их основе. Речь идет о рассматриваемой здесь оболочечной модели ядра. Тем не менее такая модель была создана. В ней реальные силы, действующие между нуклонами, заменяются самосогласованным полем, в котором независимо друг от друга и движутся нуклоны. Конечно, оболочечная модель ядра по своей обоснованности и результатам значительно уступает теории периодической системы химических элементов. Причина этого ясна из изложенного выше. Другая причина состоит в том, что в отличие от

атома, где действуют детально изученные электрические силы, силы между нуклонами в ядре изучены еще очень мало. Поэтому самосогласованное поле в ядре мы вынуждены подбирать эмпирически, руководствуясь при этом только тем, чтобы возможно лучше согласовать результаты оболочечной модели с экспериментальными фактами. Приблизительное соответствие результатов оболочечной модели опытным фактам и должно рассматриваться как ее истинное обоснование.

С учетом электрических взаимодействий между нуклонами самосогласованные поля должны подбираться различно для протонов и нейтронов. Однако как уже отмечалось, ядерные силы между протонами одинаковы с ядерными силами между нейтронами. Протоны отличаются от нейтронов дополнительными силами электрического отталкивания. А последние на малых расстояниях, на которых проявляются ядерные силы, в сотни раз слабее ядерных. Электрические силы, благодаря их дальнодействующему характеру, могут быть заметными только для самых тяжелых ядер. Поэтому в первом приближении можно учитывать только ядерные силы, пренебрегая электрическими. В этом приближении самосогласованные поля для протонов и нейтронов одинаковы. Это проявляется в том, что, как показывает опыт, магические числа протонов совпадают с магическими числами нейтронов. Поэтому в дальнейшем говорится о *нуклонном самосогласованном поле*, причем под нуклоном с равным основанием можно подразумевать как протон, так и нейтрон.

8. Замена реальных сил самосогласованным полем, одинаковым для всех нуклонов ядра, сводит задачу многих тел к задаче об *одной частице*, движущейся в этом самосогласованном поле. Соответствующий этому полю нуклонный потенциал, как уже сказано выше, подбирается эмпирически. Прежде всего для сферических ядер самосогласованный нуклонный потенциал $V(r)$ должен быть сферически-симметричным. А из-за малости радиуса действия ядерных сил $V(r)$ приближенно должен меняться в зависимости от расстояния r до центра ядра аналогично плотности ядерного вещества (см. § 65). Для средних и тяжелых ядер принимается, что $V(r)$ приблизительно постоянен внутри ядра, а в окрестности границы последнего очень круто (но все же плавно) спадает практически до нуля. Для легких ядер самосогласованный потенциал по форме напоминает потенциал трехмерного осциллятора, т. е. внутри ядра $V(r)$ представляется приближенной формулой $V(r) = (1/2)kr^2$ с постоянным k. Разумеется, вне ядра $V(r) = 0$. Приближенный вид обычно применяемого эмпирического самосогласованного нуклонного потенциала представлен кривой на рис. 141.

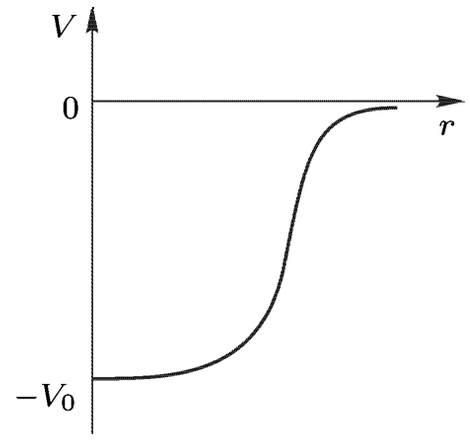


Рис. 141

После выбора самосогласованного нуклонного потенциала задача сводится к решению одночастичного уравнения Шредингера. Посмотрим сначала, к каким магическим числам приводит предположение, что $V(r)$ представляется потенциалом трехмерного гармонического осциллятора, т. е. внутри ядра $V = (1/2)kr^2$, а вне ядра $V = 0$. Представляя V в форме $V = (1/2)k(x^2 + y^2 + z^2)$, видим, что в уравнении Шредингера переменные x, y, z разделяются, так что надо решить три одинаковых уравнения Шредингера для одномерного гармонического осциллятора, отличающиеся одно от другого только обозначениями независимых переменных. Энергия одномерного гармонического осциллятора в стационарном состоянии равна $\mathscr{E}_x = (n_x + 1/2)\hbar\omega$, и аналогично для y и z. Полная энергия трехмерного осциллятора представится в виде $\mathscr{E} = \mathscr{E}_x + \mathscr{E}_y + \mathscr{E}_z = (n+3/2)\hbar\omega$, где $n = n_x + n_y + n_z$, причем все квантовые числа n_x, n_y, n_z могут принимать только целые положительные значения и нуль ($n = 0$). Самое существенное для нашей задачи состоит в том, что уровни энергии трехмерного осциллятора вырождены. Кратность вырождения (без спина) равна $(1/2)(n + 1)(n + 2)$. В частности, при $n = 1$ вырождение трехкратное. Одному и тому же значению n соответствуют различные состояния, отличающиеся одно от другого значениями квантовых чисел n_x, n_y, n_z (а также проекцией спина, которая может быть равна либо $+1/2$, либо $-1/2$). Возможные состояния (без учета спина) приведены в табл. 11 для n, равных 0, 1, 2. Таблицу легко продолжить и для бóльших значений n.

Таблица 11

n	n_x	n_y	n_z	Число состояний в оболочке
0	0	0	0	1
1	1	0	0	3
	0	1	0	
	0	0	1	
2	2	0	0	6
	0	2	0	
	0	0	2	
	1	1	0	
	0	1	1	
	1	0	1	

Естественно принять, что каждому значению n соответствует определенная оболочка ядра. Если учесть спин, то число состояний надо удвоить. Поэтому в модели трехмерного гармонического осциллятора числа нуклонов в последовательных оболочках должны были бы иметь значения 2, 6, 12, 20, 30, 42. Соответствующие магические числа получаются прибавлением каждого из этих чисел к сумме предыдущих. Таким образом, при рассматриваемой форме потенциала магические

числа должны были бы иметь значения 2, 8, 20, 40, 70, 112. Первые три числа совпадают с экспериментально установленными. Но остальные числа отличаются от экспериментально найденных значений магических чисел 50, 82, 126.

9. Указанное расхождение естественно отнести за счет неправильного выбора формы самосогласованного нуклонного потенциала и неучета ls-связи. При другом выборе этого потенциала разделение переменных в декартовых координатах при решении уравнения Шредингера уже не получится. Но разделение переменных останется в сферической системе координат, независимо от формы самосогласованного нуклонного потенциала, лишь бы только последний был сферически-симметричным. Возможные стационарные состояния нуклона в этом случае можно характеризовать совершенно так же, как это делается в случае одноэлектронных атомов, а именно четырьмя квантовыми числами: главным квантовым числом n, орбитальным квантовым числом l, квантовым числом j полного момента, слагающегося из орбитального и спинового, и квантовым числом m_j, которое определяет проекцию полного момента на избранное направление. Возможность такого описания состояния нуклона в ядре обеспечивается именно сферически-симметричной формой самосогласованного нуклонного потенциала, в силу которой для каждого нуклона имеет место сохранение орбитального и полного моментов импульса.

Главное квантовое число n может принимать целочисленные значения $1, 2, 3, \ldots$ Оно на единицу больше числа узлов радиальной волновой функции ядра. В частности, при $n = 1$ радиальная волновая функция совсем не имеет узлов. Следовательно, в случае ядра главное квантовое число имеет иной смысл, чем в атомной физике. (В случае атома квантовое число, определяющее число узлов радиальной волновой функции, называется радиальным и обозначается через n_r, а главное квантовое число определяется выражением $n = n_r + l + 1$.) Как и в атомной спектроскопии, состояния с $l = 0, 1, 2, \ldots$ обозначаются соответственно через s, p, d, f и далее по алфавиту. Число j при заданном $l \neq 0$ может принимать два значения, $l + 1/2$ и $l - 1/2$, в зависимости от ориентации спина нуклона: по орбитальному моменту или против него. При $l = 0$ возможно только единственное значение $j = 1/2$. При заданном j квантовое число m_j может принимать все значения: $-j, -(j-1), \ldots$ $\ldots, +(j-1), +j$, т. е. всего $2j + 1$ значений.

Для обозначения стационарных состояний нуклона применяется символика, аналогичная той, какая употребляется в атомной физике. На первом месте в качестве коэффициента ставится главное квантовое число n, затем (для определения l) идет одна из букв s, p, d, $\ldots$ с индексом внизу, который обозначает квантовое число j. Например, символ $2d_{5/2}$ означает состояние, у которого $n = 2$, $l = 2$, $j = 5/2$. Заметим, что ввиду сферической симметрии самосогласованного нуклонного потенциала энергия нуклона от квантового числа m_j не зависит.

10. В невозмущенном состоянии ядра с числом протонов Z и числом нейтронов N, согласно принципу Паули, должны быть заполнены все Z

энергетически самых низких протонных состояний и все N также энергетически самых низких нейтронных состояний. Поскольку энергия не зависит от m_j, каждый энергетический уровень характеризуется квантовыми числами n, l и j. Энергетические уровни нуклона, как требует опыт, надо сгруппировать в *оболочки.* Оболочками называются совокупности близких по энергии уровней, разделенных энергетическими интервалами, значительно превышающими расстояния между энергетическими уровнями внутри самих оболочек. Ядра с полностью застроенными протонной или нейтронной оболочками являются *магическими.* Действительно, на ядрах с полностью застроенными нуклонными оболочками должно происходить резкое увеличение энергии отделения нуклона от ядра. А это как раз то свойство, которое выделяет магические ядра среди остальных. Несовпадение ядерных магических чисел с порядковыми номерами атомов благородных газов, у которых также застроены электронные оболочки, связано с отличием самосогласованного поля ядра от соответствующего поля электронной оболочки атома.

Нуклонный самосогласованный потенциал надо подобрать так, чтобы получилась оболочечная структура, соответствующая опытным данным. При этом недостаточно ограничиться зависимостью самосогласованного поля от радиуса r. Надо еще учесть спин-орбитальное взаимодействие. На это в 1949 г. независимо друг от друга обратили внимание Мария Гёпперт-Майер (1906–1972) и Х. Йенсен (1907–1973), разработавшие наиболее удачный вариант оболочечной модели ядра. Спин нуклона может быть направлен либо по орбитальному моменту, либо против него. В первом случае энергетические уровни спускаются ниже, во втором поднимаются выше. Этот сдвиг надо подобрать так, чтобы получилась правильная последовательность наблюдаемых магических чисел, что и было сделано Гёпперт-Майер и Йенсеном. Спин-орбитальное взаимодействие математически учитывается выбором гамильтониана $\widehat{\mathscr{H}}$ в уравнении Шредингера $\widehat{\mathscr{H}}\psi = \mathscr{E}\psi$. Этот гамильтониан выбирают в виде

$$\widehat{\mathscr{H}} = V(r) + U(r)\mathbf{sl}, \tag{78.1}$$

где $\mathbf{s}$ — вектор спина нуклона, $\mathbf{l}$ — его орбитальный момент. Как уже указывалось выше, векторы $\mathbf{s}$ и $\mathbf{l}$ могут быть либо параллельны, либо антипараллельны. Потенциал $V(r)$ схематически представлен на рис. 141. Он имеет вид ямы с практически плоским дном, круто поднимается вблизи ее границы, а затем быстро и плавно обращается в нуль; $U(r)$ — центрально-симметричный потенциал, более слабый, чем $V(r)$. По аналогии с атомом обычно полагают

$$U(r) = b\frac{1}{r}\frac{\partial V}{\partial r}, \tag{78.2}$$

где постоянная b называется *постоянной спин-орбитального взаимодействия.* На основе нуклонного самосогласованного потенциала вида (78.2) и была разработана оболочечная модель ядра.

11. Расположение энергетических уровней нуклона, каким оно получается в результате решения уравнения Шредингера с простейшим эмпирически подобранным гамильтонианом вида (78.1), представлено в табл. 12. В каждой строке приведены состояния нуклонов, входящих

Таблица 12

Оболочки	Состояния	Число нуклонов в оболочке	Число нуклонов в ядре
I	$1s_{1/2}$	2	2
II	$1p_{3/2}1p_{1/2}$	$4+2=6$	8
III	$1d_{5/2}1d_{3/2}2s_{1/2}$	$6+4+2=12$	20
IV	$1f_{7/2}1f_{5/2}2p_{3/2}2p_{1/2}1g_{3/2}$	$8+6+4+2+10=30$	50
V	$1g_{7/2}2d_{5/2}1h_{11/2}2d_{3/2}3s_{1/2}$	$8+6+12+4+2=32$	82
VI	$2f_{7/2}1h_{9/2}1i_{13/2}3p_{3/2}2f_{5/2}3p_{1/2}$	$8+10+14+4+6+2=44$	126

в определенную оболочку. Энергии состояния (отрицательные) растут слева направо. Нумерация оболочек начинается с оболочки I и растет для последующих оболочек. В предпоследнем столбце указаны числа нуклонов в каждом состоянии (определяемые числом j) и в каждой оболочке, а в последнем — полное число нуклонов (протонов и нейтронов в отдельности) в ядре, заканчивающемся застроенной оболочкой. Расположение энергетических уровней и их группирование в оболочки приведено также на рис. 142. Отсчет энергии на рис. 142 ведется от дна потенциальной ямы. Таблица и рисунок относятся к любому типу нуклонов: как к протонам, так и к нейтронам. Обращаем внимание на расщепление уровня с определенным l на два подуровня с $j=l+1/2$ и $j=l-1/2$, обусловленное спин-орбитальным взаимодействием. Это расщепление растет с увеличением квантового числа l. Уже при $l=3$ расщепление $1f$-состояния на $1f_{7/2}$ и $1f_{5/2}$ столь значительно, что обнаруживается повышенная стабильность ядра с числом нуклонов 28. Поэтому иногда при рассмотрении некоторых свойств ядер число 28 относят к магическим, хотя на нем и не оканчивается заполнение протонной или нейтронной оболочки. Из-за известной неопределенности в выборе эмпирических потенциалов $V(r)$ и $U(r)$ в гамильтониане (78.1) распределения квантовых состояний по энергетическим уровням в различных литературных источниках слегка отличаются друг от друга. Здесь приведено одно из возможных распределений.

Успехи оболочечной модели ядра при объяснении магических чисел явились исторически первым и самым важным аргументом в пользу признания этой модели. Но оболочечная модель объясняет и некоторые другие факты. Сюда относятся, например, предсказания спинов и четностей ядер, а также вычисление их магнитных моментов. На этих вопросах мы останавливаться не будем. Многие факты остаются необъясненными и в оболочечной модели. Это вполне естественно,

если иметь в виду эмпирический характер модели и недостаточную обоснованность ее исходных положений.

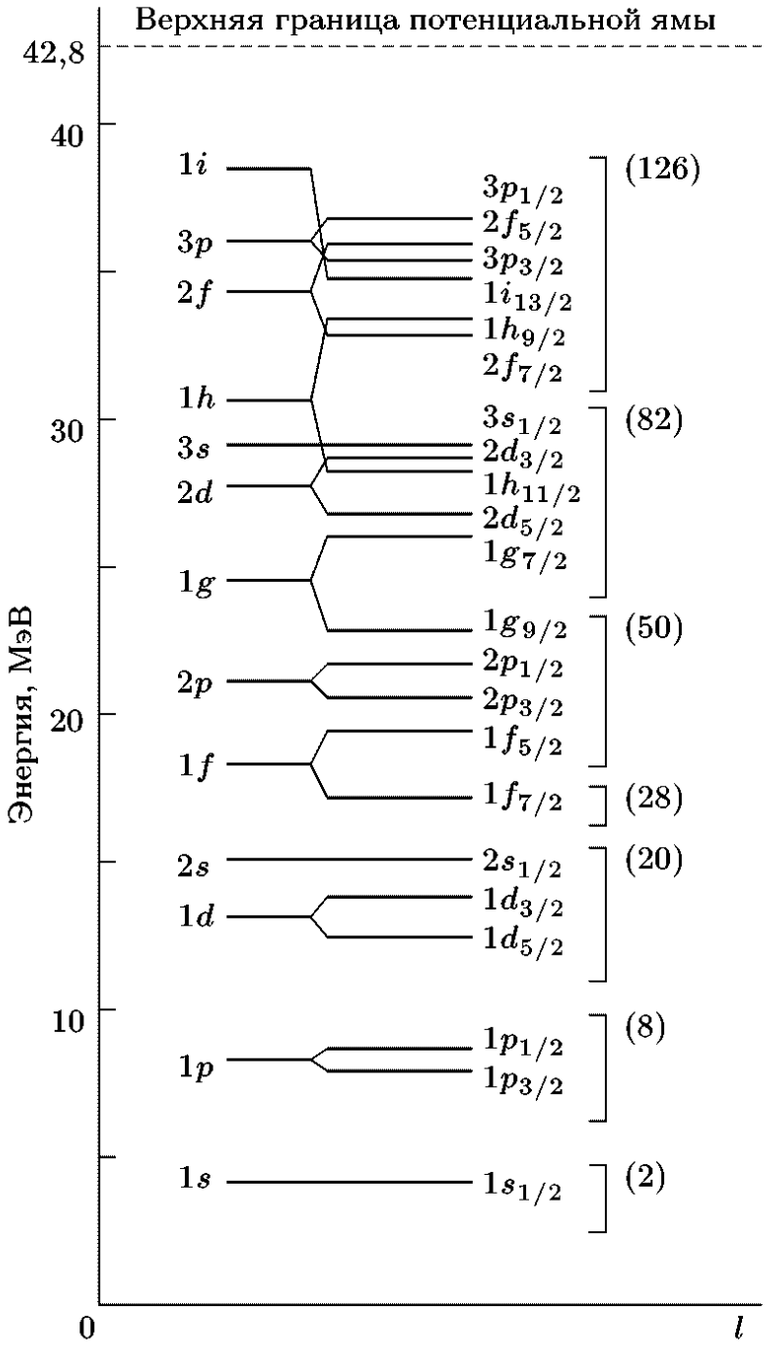


Рис. 142

Таблица 12 заканчивается шестой оболочкой для нейтронов. Вопрос о существовании оболочек более высокого порядка остается открытым, так как эти оболочки относятся к таким трансурановым ядрам, которые еще не получены и неизвестно, будут ли они вообще получены. Кроме того, надо иметь в виду, что с увеличением массовых чисел увеличивается число энергетических уровней, которые должны сформировать оболочку, а также уменьшается просвет между соседними оболочками. Поэтому при достаточно больших массовых числах, если даже соответствующие трансурановые ядра будут получены, само представление об оболочечной структуре ядра может потерять смысл. Для оболочки низкого порядка значения магических чисел при разумных выборах

самосогласованного потенциала почти не зависят от его формы. Для высших же оболочек предсказание значений магических чисел не вполне однозначно, так как с увеличением номера оболочки порядок ее заполнения становится более чувствительным к тонким деталям формы протонного и нейтронного самосогласованных потенциалов. Все же делаются попытки предсказать значения магических чисел гипотетических VI-й протонной и VII-й нейтронной оболочек. Наиболее вероятным значением магического числа для VI-й протонной оболочки считается $Z = 114$, а для VII-й нейтронной $N = 284$. Магическое число $Z = 114$ не совпадает с магическим числом $N = 126$ для того же VI-го номера нейтронной оболочки. Это связано с наличием кулоновского потенциала в случае протонов. Ожидают, что время жизни трансуранового элемента с $Z = 114$, если он будет получен, окажется на много порядков больше времен жизни соседних трансурановых элементов. Такой элемент, возможно, будет практически вести себя как стабильный и найдет научно-технические применения. Этот вопрос остается открытым.

12. Закончим эту главу краткой характеристикой так называемой обобщенной модели ядра, начала которой были заложены Рейнуотером (р. 1917) и которая была разработана главным образом Оге Бором (р. 1922) и Б. Моттельсоном (р. 1926). В обобщенной модели предполагается, что ядро состоит из внутренней устойчивой части — остова, образованного нуклонами заполненных оболочек, и внешних нуклонов, движущихся в поле, создаваемом остовом. Движение остова описывается коллективной моделью. Однако под влиянием наружных нуклонов остов может колебаться, изменяя свою форму. За счет этих колебаний изменяется и поле остова, в котором движутся наружные нуклоны. Согласно одночастичной оболочечной модели квадрупольный электрический момент ядра полностью определяется состоянием движения наружного нуклона. Поэтому в этой модели он не может по модулю превышать примерно 10^{-23} см^2 для всех ядер с нечетным числом протонов и должен обращаться в нуль для всех ядер с нечетным числом нейтронов. На самом деле имеются ядра с нечетным числом нейтронов, у которых квадрупольный электрический момент в десятки раз превышает одночастичное значение для одного протона. Например, у ядер ^{235}U и ^{233}U с числами нейтронов, равными соответственно 143 и 141, квадрупольные моменты равны $9 \cdot 10^{-21}$ и $14 \cdot 10^{-21}$ см^2. Очень велики квадрупольные электрические моменты и у многих ядер с нечетным числом протонов. Обобщенная модель объяснила большие квадрупольные моменты некоторых ядер сильной деформацией остова, вызываемой внешними нуклонами ядер. В результате этого остов становится несферическим, принимая форму вытянутого, сплюснутого или трехосного эллипсоида. Обобщенная модель позволила провести классификацию уровней энергии ядра — ввести понятие одночастичных (связанных с возбуждением наружных нуклонов) и коллективных (вращательных и колебательных, связанных с возбуждением остова) уровней ядра, определить энергии уровней, спин и четность.

Глава XI

ПРОХОЖДЕНИЕ ЗАРЯЖЕННЫХ ЧАСТИЦ И ГАММА-КВАНТОВ ЧЕРЕЗ ВЕЩЕСТВО

§ 79. Введение

В настоящей главе рассматривается прохождение через вещество заряженных частиц, γ- и рентгеновских квантов высоких энергий, т. е. таких энергий, которые на несколько или много порядков превосходят средний ионизационный потенциал электрона в электронной оболочке атома. Наибольший практический интерес для ядерной физики представляет интервал энергий от нескольких килоэлектронвольт до около 10 МэВ.

Несмотря на необычайную сложность процессов, связанных с прохождением рассматриваемых частиц через вещество, эти процессы поддаются сравнительно точным расчетам или, во всяком случае, оценкам. Это связано, во-первых, с тем, что основную роль при прохождении заряженных частиц, γ- и рентгеновских квантов через вещество играют хорошо изученные электромагнитные взаимодействия. Роль ядерных взаимодействий в большинстве случаев относительно невелика из-за короткодействующего характера ядерных сил, а также из-за того, что электронов в веществе значительно больше, чем атомных ядер. Поэтому рассматриваемые расчеты скорее относятся к атомной, а не к ядерной физике. Они важны для ядерной физики только потому, что именно она имеет дело с высокими энергиями заряженных частиц и γ-квантов. Вторая причина связана с тем, что высокая энергия проходящих через вещество заряженных частиц и γ-квантов позволяет часто пренебрегать энергией связи электронов в электронных оболочках атомов, рассматривая эти электроны как свободные. Даже в тех случаях, когда точный расчет процесса прохождения через вещество практически невыполним из-за его сложности и необходимые величины приходится находить опытным путем, качественное уяснение явлений позволяет определить, какие именно постоянные надо находить эмпирически.

По характеру механизма прохождения через вещество все частицы разделяются на: 1) *легкие заряженные частицы* (электроны и позитроны); 2) *тяжелые заряженные частицы* (к ним относятся все частицы кроме электронов и позитронов) и 3) *γ-кванты* (и кванты жестких рентгеновских лучей).

Нейтроны взаимодействуют только с атомными ядрами посредством ядерных сил, а поэтому их прохождение через вещество будет

рассмотрено в гл. XIV. Нейтрино подвергаются только слабым взаимодействиям и могут свободно проходить в веществе астрономические расстояния.

Знание закономерностей прохождения через вещество заряженных частиц и γ-квантов необходимо для понимания действия приборов ядерной физики, применяемых для регистрации и изучения свойств таких частиц, а также для расчета защиты от ядерных излучений при научных исследованиях, в атомной энергетике и при прочих применениях ядерной физики.

§ 80. Прохождение тяжелых заряженных частиц через вещество

1. Тяжелая заряженная частица массы M и высокой энергии взаимодействует с электрическими полями электронов и атомных ядер. Она либо *ионизует*, либо *возбуждает атомы.* Осуществляется также и чисто ядерное взаимодействие частицы с *атомным ядром.* За счет этих процессов энергия частицы уменьшается и ее движение замедляется. Если частица заряжена положительно, то в результате замедления она начинает энергично захватывать электроны, отбирая их от атомов окружающей среды. В результате она превращается в *ион* или *нейтральный атом* и приходит в тепловое равновесие с окружающей средой. Такова же судьба и быстрой отрицательной частицы. Регулярное движение частицы через среду прекращается — ее путь обрывается. Но процессы, происходящие с частицей в самом конце ее пути, здесь не рассматриваются. Не рассматриваются также ядерные превращения, которые может претерпевать движущаяся частица при столкновениях с атомными ядрами среды, так как из-за короткого действия ядерных сил такие превращения осуществляются гораздо реже, чем процессы, вызываемые кулоновскими силами. Заметим только, что для адронов высоких энергий заметную роль играют и ядерные взаимодействия.

В этом параграфе предполагается, что основную роль в замедлении частицы играют процессы *ионизации* и *возбуждения электронных оболочек атома.* Все они получили собирательное название *ионизационных потерь.* Только такие процессы и учитываются в настоящем параграфе [1]. Из-за дальнодействующего характера кулоновских сил частица взаимодействует сразу со многими электронами атомных оболочек, которые в свою очередь воздействуют на частицу. Это воздействие носит случайный, хаотический характер, так что путь частицы в веществе практически прямолинеен. Прямолинейность пути связана также с большой массой тяжелой частицы по сравнению с массой

[1] В § 80 и 81 совершенно не затронут *процесс многократного рассеяния* частиц, в основе которого лежит резерфордовское рассеяние частиц на ядрах. Следует иметь в виду, что этот процесс приводит к заметному искривлению следов даже тяжелых частиц, а *угол многократного рассеяния* используется для определения характеристик частиц, оставивших след.

легкого электрона, вследствие чего при каждом взаимодействии с электроном она отклоняется очень мало и теряет очень небольшую долю от первоначальной энергии.

Основной интерес представляют средние ионизационные потери энергии частицы $-d\mathscr{E}/dx$, отнесенные к единице пути, а также ее полный пробег R в веществе. Приближенное нахождение зависимости этих величин от характеристик частицы и среды и является целью настоящего параграфа. Рассмотрим решение этой задачи в предположении справедливости классической механики, а затем качественно учтем влияние квантовых эффектов. Последовательный квантовый расчет выходит за рамки этой книги.

2. Сначала рассчитаем потери энергии, вносимые отдельным электроном, а затем просуммируем эти потери по всем электронам среды. Таким образом, расчет будем проводить в приближении *парных столкновений*, т. е. будем считать, что взаимодействие каждого электрона с рассматриваемой частицей происходит так, как если бы других электронов не было. А поскольку энергия частицы предполагается высокой, электрон, с которым она взаимодействует, можно считать *свободным*. Более того, можно предполагать, что этот электрон покоится. Оправданием этого может служить следующее замечание: электрон входит в состав атомов и молекул и в среднем перемещается с ними с тепловыми скоростями. Сама же движущаяся частица имеет скорость, близкую к скорости света, или отличается от нее примерно на порядок.

Только после ионизации электрон теряет связь с молекулой или атомом и начинает быстро набирать скорость, а потому предположение о неподвижности электрона может и не совсем выполняться. Но процесс ионизации происходит на малых расстояниях от движущейся частицы, так что ускорение электрона совершается кратковременно, и можно думать, что оно не играет существенной роли. Саму частицу, как уже было выяснено выше, при расчете можно считать движущейся прямолинейно с постоянной скоростью v. Зарядовое число движущейся частицы будем обозначать малой буквой z, оставляя большую букву Z для обозначения зарядового числа атомных ядер окружающей среды.

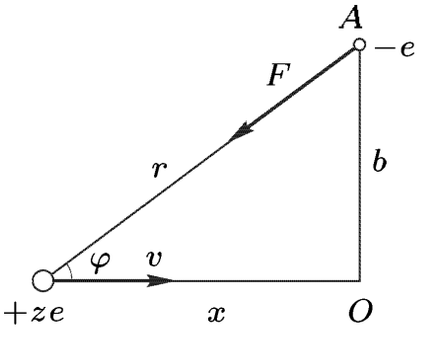


Рис. 143

Частица с зарядом ze, движущаяся мимо электрона A в направлении оси x (рис. 143), притягивает электрон с силой $F = ze^2/r^2$. Последняя за время dt сообщает ему импульс $F\,dt$. Продольная составляющая этого импульса не имеет значения, так как при переходе частицы через точку O она меняет знак. В результате приращение продольной составляющей будет компенсировано ее убыванием. Только поперечная составляющая импульса электрона представляет интерес в нашей задаче. Обозначим поперечную составляющую импульса просто через p. Тогда $dp = -F\sin\varphi\,dt$,

или

$$dp = -\frac{F \sin\varphi}{v}\, dx,$$

где dx — путь, пройденный частицей за время dt. Но $x = b\operatorname{ctg}\varphi$, $r = b/\sin\varphi$, а b в нашем приближении предполагается постоянным. Таким образом, приняв за независимую переменную угол φ, получим

$$dp = \frac{ze^2 \sin\varphi}{bv}\, d\varphi.$$

Полный поперечный импульс, полученный электроном, найдется интегрированием по φ в пределах от 0 до π. Таким путем находим

$$p = \frac{2ze^2}{bv}. \tag{80.1}$$

Электрон получит энергию $p^2/2m$, и такую же энергию потеряет частица (m — масса электрона).

3. Допустим теперь, что частица пересекает бесконечный плоскопараллельный слой вещества толщиной dx, в единице объема которого содержится n электронов. В части этого слоя, ограниченной цилиндрическими поверхностями с радиусами b и $b + db$, находится $dN = 2\pi n b\, db\, dx$ электронов. Если, как было предположено выше, электроны действуют независимо друг от друга, то взаимодействие частицы с dN электронами вызовет потерю ее энергии на величину $-dNp^2/2m$. Полная потеря энергии частицы на единице пути будет, таким образом,

$$-\frac{d\mathscr{E}}{dx} = \frac{4\pi n z^2 e^4}{mv^2} \int \frac{db}{b}, \tag{80.2}$$

где интегрирование распространено на всю область, заполненную электронами, существенно влияющими на торможение частицы. Строго говоря, этот способ выражения не совсем точен, так как он предполагает, что в этой области взаимодействие частицы с электронами происходит именно по той схеме, которая применялась при вычислении. Но это далеко не так. Такая схема заведомо неприменима при слишком больших и слишком малых значениях параметра b, а при промежуточных значениях применима только приближенно. Тем не менее, сознательно идя на потерю математической строгости, мы примем эту схему при промежуточных значениях, поскольку здесь она физически оправдана. Нельзя только производить интегрирование в пределах от $b = 0$ до $b = +\infty$, так как это приводит к расходящемуся интегралу, что физически означает мгновенное торможение частицы, а это бессмысленно. Поэтому интегрирование в формуле (80.2) следует производить в пределах от некоторого минимального значения $b = b_{\text{мин}}$ до некоторого максимального значения $b = b_{\text{макс}}$. Определение этих пределов представляет наиболее трудную часть задачи, которая вряд ли может быть решена с полной математической строгостью и достаточной физической ясностью. К счастью, в подавляющем большинстве случаев достаточно ограничиться сравнительно грубыми физически

оправданными оценками. Приведем одну из наиболее простых таких оценок.

4. Выясним прежде всего, почему необходимо ограничить верхний предел в интеграле (80.2). Это ограничение связано с *квантовыми свойствами атомов среды.* Для возбуждения атома внешнее воздействие должно быть достаточно сильным. Оно должно быть в состоянии перевести атом с одного энергетического уровня на другой. В противном случае атом возбуждаться не будет. Такой атом не влияет на замедление движущейся частицы и не вносит никакого вклада в интеграл (80.2). Следующая элементарная оценка позволяет уяснить суть дела. Движущаяся частица эффективно воздействует на электрон в течение времени $\tau \sim b/v$. Кулоновская сила, действующая на электрон, $F \sim ze^2/b^2$. Импульс, приобретаемый электроном, пропорционален $F\tau \sim ze^2/bv$, т. е. он тем меньше, чем больше b. Если b превышает некоторую величину $b_{\text{макс}}$, то соответствующий электрон не должен приниматься во внимание. Но если электрон рассматривается в течение времени τ, то его энергия не строго определена, и эта неопределенность $\Delta\mathscr{E}$ ограничена соотношением $\tau \cdot \Delta\mathscr{E} \approx \hbar$. Ориентировочно атом будет возбуждаться только тогда, когда $\Delta\mathscr{E}$ не меньше $\overline{I}$, где $\overline{I}$ — средний ионизационный потенциал атома. Полагая $\Delta\mathscr{E} = \overline{I}$, получаем оценку $\tau \approx \hbar/\overline{I}$ для времени эффективного взаимодействия электрона с рассматриваемой частицей. За это время частица проходит расстояние $b = v\hbar/\overline{I}$. Эту величину и можно принять в качестве грубого приближения для верхнего предела b:

$$b_{\text{макс}} = \hbar v/\overline{I} \quad \text{(нерелятив.)}. \tag{80.3}$$

Для средней энергии ионизации атома обычно принимают эмпирически установленную формулу

$$\overline{I} = 13{,}5Z \text{ эВ}. \tag{80.4}$$

Формула (80.3) получена в нерелятивистском приближении, что и отмечено в скобках. Когда частица движется с релятивистской скоростью, в эту формулу следует ввести поправку. Дело в том, что при ее выводе использовался закон Кулона для электрического поля точечного заряда. При релятивистских скоростях движущегося заряда электрическое поле его изменяется. Электрические силовые линии движущегося точечного заряда по-прежнему остаются прямолинейными, но вся картина силовых линий сжимается в направлении движения. Это показано на схематическом рис. 144. Кроме того, продольное поле, направленное вдоль линии движения частицы, уменьшается в $1/(1-\beta^2)$ раз, а поперечное экваториальное поле увеличивается в $1/\sqrt{1-\beta^2}$ раз. Первый эффект приводит

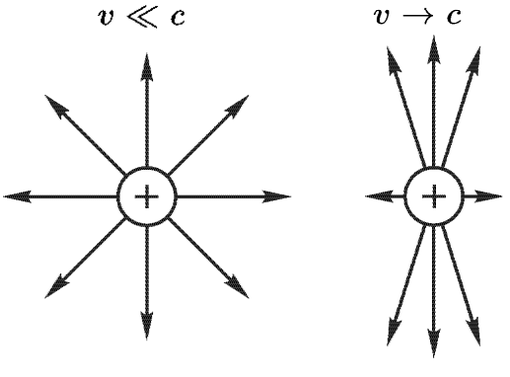


Рис. 144

к уменьшению эффективного времени взаимодействия частицы с электроном в $1/(1-\beta^2)$ раз, а второй эффект — к увеличению поперечной напряженности электрического поля в $1/\sqrt{1-\beta^2}$ раз. В результате прежняя величина $F\tau$ приобретает множитель $(1-\beta^2)/\sqrt{1-\beta^2} = \sqrt{1-\beta^2}$, а вместо выражения (80.1) получается

$$p = \frac{2ze^2}{bc}\sqrt{1-\beta^2}, \tag{80.5}$$

причем скорость частицы v мы заменили на c, поскольку ее движение теперь релятивистское. Что касается электрона, то мы по-прежнему предполагаем, что его движение, возникающее после столкновения с частицей, нерелятивистское. Поэтому приобретаемая им кинетическая энергия определяется прежним выражением $p^2/2m$. Таким образом, переход к релятивистскому случаю производится формальной заменой в формуле (80.1) величины b на $b\sqrt{1-\beta^2}$. В результате верхний предел $b_{\text{макс}}$ в рассматриваемом случае увеличивается в $1/\sqrt{1-\beta^2}$ раз, т. е.

$$b_{\text{макс}} = \frac{\hbar v}{\overline{I}}\frac{1}{\sqrt{1-\beta^2}} \quad \text{(релятив.)}. \tag{80.6}$$

5. Определим теперь нижний предел интеграла в формуле (80.2). При классическом рассмотрении скорость, сообщаемая электрону при лобовом столкновении с тяжелой частицей, не может превышать $2v$. Поэтому энергия, передаваемая электрону, не может превосходить $(1/2)m(2v)^2 = 2mv^2$. Значит, формула (80.1) может иметь смысл только при условии

$$\frac{1}{2m}\left(\frac{2ze^2}{bv}\right)^2 < 2mv^2, \text{ т. е. при } b > \frac{ze^2}{mv^2}.$$

Поэтому в качестве нижнего предела интеграла с классической точки зрения естественно принять выражение

$$b_{\text{мин}}^{\text{кл}} = \frac{ze^2}{mv^2} \quad \text{(нерелятив.)}. \tag{80.7}$$

К иному выражению приводит квантовое рассмотрение. Согласно соотношению неопределенностей импульс частицы $p = mv/\sqrt{1-\beta^2}$ и ее расстояние b до электрона должны удовлетворять условию $bp \gtrsim \hbar$. Поэтому с квантовой точки зрения в качестве нижнего предела естественно принять выражение

$$b_{\text{мин}}^{\text{кв}} = \frac{\hbar\sqrt{1-\beta^2}}{mv} \quad \text{(релятив.)}. \tag{80.8}$$

Из двух выражений (80.7) и (80.8) следует выбирать наибольшее. Сравнение этих выражений в нерелятивистском приближении приводит к результату

$$\frac{b_{\text{мин}}^{\text{кв}}}{b_{\text{мин}}^{\text{кл}}} = \frac{\hbar v}{ze^2} = \frac{\hbar c}{e^2}\frac{\beta}{z} = 137\frac{\beta}{z}. \tag{80.9}$$

В релятивистском приближении это отношение еще больше. Значит, ограничения, накладываемые квантовой механикой, начинают сказываться раньше. Поэтому следует выбрать квантовое выражение (80.8). В результате путем комбинации формул (80.2), (80.6) и (80.8) получается формула Бора:

$$-\frac{d\mathscr{E}}{dx} = \frac{4\pi n z^2 e^4}{mv^2} \ln \frac{b_{\text{макс}}}{b_{\text{мин}}} = \frac{4\pi n z^2 e^4}{mv^2} \ln \frac{mv^2}{\overline{I}(1-\beta^2)}. \tag{80.10}$$

Не следует слишком смущаться грубостью оценок пределов в $b_{\text{мин}}$ и $b_{\text{макс}}$, которые были произведены при выводе формулы (80.10), так как в нее входит логарифм отношения этих пределов, который слабо зависит от погрешностей, вносимых при оценках $b_{\text{мин}}$ и $b_{\text{макс}}$. Существует несколько более точных выражений для $-d\mathscr{E}/dx$. Ограничимся приведением простейших из них:

$$-\frac{d\mathscr{E}}{dx} = \frac{4\pi n z^2 e^4}{mv^2} \left[\ln \frac{2mv^2}{\overline{I}(1-\beta^2)} - \beta^2\right]. \tag{80.11}$$

Для протонов с энергией 1 МэВ в воздухе при нормальных температуре и давлении логарифмический член в последней формуле равен приблизительно 9.

Последовательная квантовая теория ионизационных потерь энергии заряженных частиц в веществе была разработана Бете и Блохом. На этом вопросе мы останавливаться не можем, поскольку изложение теории Бете–Блоха требует знания математического аппарата квантовой механики.

6. Формула Бора, по крайней мере качественно, а отчасти и количественно, позволяет понять, какими величинами определяется торможение тяжелых заряженных частиц за счет ионизационных потерь в веществе в широком диапазоне энергий частицы (от 1 МэВ до десятков и сотен гигаэлектронвольт).

Как видно из (80.10) или (80.11), основные потери определяются зарядом и скоростью частицы, числом электронов в единице объема среды и средним ионизационным потенциалом $\overline{I}$ атомов среды. Зависимость от $\overline{I}$ логарифмическая, а потому слабая. Зависимость от n сводится к зависимости от плотности среды ρ посредством формулы

$$n = \frac{Z\rho N_A}{A}, \tag{80.12}$$

где N_A — постоянная Авогадро, A — атомная масса, Z — порядковый номер атомов среды. Следовательно, величина $-d\mathscr{E}/d(\rho x)$ примерно одинакова для всех веществ. Величину ρx обычно вводят в качестве меры толщины вещества вместо линейной толщины x. Так поступают, например, рассчитывая толщину необходимой защиты от радиоактивных излучений, хотя в этих случаях в основном требуется защита от γ-квантов и нейтронов, а не от потоков заряженных частиц. Потери сильно зависят от скорости частицы — они тем больше, чем меньше скорость частицы. Вот почему толщина треков тяжелых заряженных

частиц в камере Вильсона или в фотоэмульсии значительно возрастает к их концу. При увеличении скорости частицы логарифмический член в (80.10) или (80.11) сначала убывает. Но при приближении скорости к релятивистскому пределу, т. е. при $v \to c$, убывание сменяется возрастанием, так как числитель $2mv^2$ становится практически постоянным, а знаменатель $1 - \beta^2$ приближается к нулю. В результате при $v \to c$ потери энергии $-d\mathscr{E}/dx$ проходят через минимум, который расположен примерно около $\mathscr{E} = 2Mc^2$. Это чисто релятивистский эффект.

Как видно из формулы (80.10) или (80.11) при заданных скорости и заряде частицы потери не зависят от ее массы M. Поэтому в случае протонов и пионов, например, потери одинаковы, если только эти частицы движутся с одинаковыми скоростями. Если же в нерелятивистском случае в формулу (80.10) ввести кинетическую энергию частицы $\mathscr{E} = Mv^2/2$, то получится

$$-\frac{d\mathscr{E}}{dx} = \frac{2\pi n z^2 e^4 M}{\mathscr{E} m} \ln \frac{2\mathscr{E} m}{M\bar{I}}. \tag{80.13}$$

Отсюда видно, что в нерелятивистском случае при одних и тех же заряде и массе частицы потери с логарифмической точностью пропорциональны массе частицы M. Поэтому треки у более тяжелой частицы жирнее и короче, чем у легкой. Наконец, квадратичная зависимость от z проявляется в сильном торможении α- и многозарядных частиц в веществе.

При очень малых и очень больших скоростях частицы формулы (80.10) и (80.11) дают завышенные значения для потерь энергии частицы.

7. При малых скоростях начинает сказываться захват электронов движущейся частицей. Такой захват в какой-то мере эквивалентен уменьшению числа z, а это приводит к меньшим потерям энергии по сравнению с тем, что дает формула Бора. Особенно сильно захват происходит в случае движения многократно заряженных положительных ионов, т. е. атомов, потерявших много электронов. Впрочем, иногда вместо захвата наблюдается и потеря электронов. Благодаря захвату электронов при уменьшении скорости частицы кривая потерь не уходит в бесконечность, как это было бы согласно формуле (80.10), а достигает максимума, после чего начинает постепенно снижаться.

При очень больших скоростях проявляется влияние поляризации среды, вызываемой электрическим полем частицы. Оно ослабляет или, как говорят, экранирует поле частицы, что уменьшает потери энергии последней. При нерелятивистских скоростях радиус экранирования (дебаевский радиус, см. т. III, § 121) превышает размеры атома. В этих случаях экранировка может проявиться лишь на расстояниях, больших $b_{\text{макс}}$, где ионизационные потери так и не возникают. Но в ультрарелятивистских случаях электрическое поле частицы сильно сплющено в направлении движения, растянуто в поперечном направлении и становится сильно неоднородным. В результате поляризация среды начинает сказываться уже на сравнительно малых расстояниях. Влияние

поляризации и в особенности захвата электронов среды трудно поддается теоретическому расчету. Эти эффекты учитываются эмпирически, и результаты выражаются в виде кривых пробег — энергия.

8. Расстояние, проходимое в веществе частицей до ее полной остановки, т. е. до того момента, когда она приходит в тепловое равновесие с окружающей средой, называется *пробегом.* Для вычисления пробега R замечаем, что на пути dx кинетическая энергия частицы $\mathscr{E} = Mv^2/2$ меняется на величину $d\mathscr{E}$, так что $dx = (dx/d\mathscr{E})\,d\mathscr{E} = (dx/d\mathscr{E})Mv\,dv$. Подставляя сюда вместо $d\mathscr{E}/dx$ выражение (80.10), приходим к дифференциальному уравнению

$$dx = -\frac{Mmv^3\,dv}{4\pi nz^2e^4\ln[mv^2/\overline{I}(1-\beta^2)]},$$

интегрирование которого дает

$$R = \frac{M}{z^2}\,f(v_0), \tag{80.14}$$

где v_0 — начальная скорость движения частицы, а функция f определяется интегралом

$$f(v_0) = -\int\limits_{v_0}^{0} \frac{mv^3}{4\pi ne^4\ln[mv^2/\overline{I}(1-\beta^2)]}\,dv. \tag{80.15}$$

Существенно, что эта функция для заданной среды одинакова для всех частиц. Если пренебречь слабой логарифмической зависимостью от скорости частицы, то

$$R \approx \frac{M}{z^2}\,v_0^4. \tag{80.16}$$

Однако как мы видели, применимость формулы (80.10) ограничена эффектами захвата электронов среды. Уточненную формулу для R можно получить из следующих простых соображений. Разделим весь путь движения частицы на две части: на часть, где захвата электронов практически не происходит и применима формула (80.10), и на оставшуюся часть, где существенную роль играют захваты. К первой части применимо выражение (80.14). Длина второй части пути от начальной скорости v_0 не зависит, т. е. является некоторой постоянной C. Значение этой постоянной различно для разных частиц и сред, в которых они движутся. Таким путем для полного пробега получается приближенная формула

$$R = \frac{M}{z^2}\,f(v_0) + C. \tag{80.17}$$

Для α-частицы в воздухе при комнатной температуре и нормальном давлении опыт дает $C = 0{,}2$ см. В алюминии пробег протона с энергией 5 МэВ равен 0,06 мм, а с энергией 10 МэВ — 0,17 мм.

Формула (80.17) справедлива при условии $R \ll \lambda_{\text{яд}}$, где $\lambda_{\text{яд}}$ — длина пробега относительного ядерного столкновения. Это условие не выполняется для адронов высоких энергий.

§ 81. Прохождение легких заряженных частиц через вещество

1. Благодаря малой массе при каждом столкновении движущейся легкой частицы (электрона или позитрона) изменение ее импульса относительно велико. Поэтому путь легкой частицы в среде не прямолинейный, а извилистый. Если пучок частиц направить на однородную среду, то он ведет себя по-разному в зависимости от того, состоит ли он из тяжелых частиц или из легких. В случае тяжелых частиц интенсивность пучка остается постоянной, если пройденный им путь x меньше длины пробега R. В очень же тонком слое вблизи границы $x = R$ частицы выбывают из пучка, и он резко обрывается. В случае же пучка из легких частиц интенсивность пучка убывает плавно и непрерывно на всем его протяжении. Поэтому об определенном пробеге R легкой частицы говорить не приходится. Можно ввести понятие *максимального* (или *экстраполированного*) *пробега* и *среднего пробега*. Максимальным пробегом называется минимальная толщина слоя вещества, в котором задерживаются все частицы. Он, очевидно, совпадает с полной длиной криволинейного пути, проходимого в веществе отдельной частицей. Чтобы получить средний пробег, надо взять длину прямолинейного пути, проходимого частицей в веществе до того, как она выбывает из пучка, и этот путь усреднить по всем частицам пучка.

Вторая особенность в поведении легких частиц состоит в том, что при изменении импульса в результате столкновения электрон (или позитрон) излучает. Поэтому помимо ионизационных, появляются *радиационные потери*, т. е. потери энергии на излучение фотонов.

Наконец, в-третьих, при движении электрона в среде проявляются *квантовые обменные эффекты*, наблюдающиеся во всякой системе тождественных частиц. Такие эффекты, разумеется, не возникают при движении позитрона в среде, поскольку электрон и позитрон — не тождественные частицы. Зато в этом случае возможен *процесс аннигиляции* позитрона с электроном. Впрочем, роль процессов аннигиляции, а также эффектов обмена относительно невелика. Поэтому торможение электрона и позитрона в среде происходит практически одинаково. Ниже для конкретности имеется в виду торможение электронов, так как позитронные пучки применяются в эксперименте значительно реже.

2. Качественно механизм ионизационных потерь в случае легких частиц такой же, что и в случае других заряженных частиц. Поэтому для электронов применима прежняя формула (80.2) с той только разницей, что из-за малости массы электрона и квантовомеханических эффектов обмена пределы интеграла $b_{\text{мин}}$ и $b_{\text{макс}}$ должны определяться иначе. С учетом этих и некоторых других факторов Бете (г. 1906) получил

следующую формулу для ионизационных потерь электронов:

$$-\frac{d\mathscr{E}}{dx}=\frac{2\pi ne^4}{mv^2}\left[\ln\frac{mv^2\mathscr{E}}{2\overline{I}(1-\beta^2)}-\ln 2(2\sqrt{1-\beta^2}-1+\beta^2)+\right.$$
$$\left.+1-\beta^2+\frac{1}{8}(1-\sqrt{1-\beta^2})\right], \quad (81.1)$$

где $\overline{I}$ — средний ионизационный потенциал атомов поглотителя, даваемый прежней приближенной формулой (80.5), а $\mathscr{E}$ — релятивистская кинетическая энергия электрона:

$$\mathscr{E}=\frac{mc^2}{\sqrt{1-\beta^2}}-mc^2. \quad (81.2)$$

В нерелятивистском пределе $\beta\to 0$:

$$-\frac{d\mathscr{E}}{dx}=\frac{4\pi ne^4}{mv^2}\ln\frac{mv^2}{2\overline{I}} \quad \text{(нерелятив.)}. \quad (81.3)$$

В ультрарелятивистском случае

$$-\frac{d\mathscr{E}}{dx}=\frac{2\pi ne^4}{mc^2}\left[\ln\frac{\mathscr{E}^2}{2\overline{I}^2\sqrt{1-\beta^2}}+\frac{1}{8}\right] \quad \text{(ультрарелятив.)}. \quad (81.4)$$

Ввиду малости массы электрона все три формулы (81.1), (81.3) и (81.4) находят практическое применение. Например, для электрона $mc^2=0{,}511$ МэВ, а потому электроны с энергиями в несколько мегаэлектронвольт уже являются ультрарелятивистскими.

3. При сравнении ионизационных потерь тяжелой и легкой частиц главное внимание следует обратить на то, что в формулах (80.10) и (80.11), с одной стороны, и (81.1), (81.3), (81.4) — с другой, определяющим является множитель перед логарифмом, так как логарифм медленно меняется с изменением параметров, характеризующих движение частиц. А этот множитель в случае движения однократно заряженных частиц фактически одинаков во всех формулах. Поэтому при одинаковых скоростях движения ионизационные потери тяжелой и легкой частиц примерно одинаковы. Это и понятно. Ионизационные потери возникают из-за воздействия электрического поля движущейся частицы на электроны среды. А эти поля́ совершенно одинаковы в случае тяжелой и легкой частиц, если только одинаковы их заряды и скорости движения.

Не так обстоит дело, когда сравниваются ионизационные потери легкой и тяжелой однозарядных частиц одинаковой энергии. В том случае, когда движение обеих частиц нерелятивистское, скорости частиц находятся в обратном отношении квадратных корней из их масс. Благодаря этому тяжелая частица более длительно эффективно воздействует на каждый электрон среды и поэтому быстрее теряет энергию. В этом

случае, как мы видели в предыдущем параграфе, ионизационные потери энергии пропорциональны массе частицы. Например, ионизационные потери протона примерно в 2000 раз превосходят ионизационные потери электрона той же энергии.

Более интересен случай, когда электрон ультрарелятивистский, но протон той же энергии еще может считаться нерелятивистским. В этом случае электрическое поле движущегося протона сферически-симметрично, тогда как у электрона оно сильно сплющено в направлении движения и растянуто в поперечном направлении. За счет этого ионизационные потери энергии электрона сильно возрастают. Сравним, например, электрон и протон с кинетической энергией $\mathscr{E} = 5$ МэВ. При такой энергии электрон уже может считаться ультрарелятивистским, тогда как протон остается нерелятивистским. Так как энергия покоя электрона $\mathscr{E}_0 = 0{,}5$ МэВ, а кинетическая энергия практически совпадает с полной, то $\mathscr{E} \approx \mathscr{E}_0/\sqrt{1-\beta^2}$, так что $1/\sqrt{1-\beta^2} \approx 10$. Сравнивая формулы (80.13) и (81.4), получаем

$$\frac{(d\mathscr{E}/dx)_\text{p}}{(d\mathscr{E}/dx)_\text{e}} = \frac{Mc^2}{\mathscr{E}}\,\frac{\ln A}{\ln B},$$

где

$$A = \frac{2\mathscr{E}m}{M\overline{I}} = \frac{2\cdot 5\cdot 10^6}{2000\cdot 10} \approx 5\cdot 10^2$$

(мы положили $\overline{I} = 10$ эВ),

$$B = \frac{\mathscr{E}^2}{2\overline{I}^2\sqrt{1-\beta^2}} = \frac{5^2\cdot 10^{12}\cdot 10}{2\cdot 10^2} \approx 2{,}5\cdot 10^{12},$$

т. е.

$$\ln A = 2\ln 10 + \ln 5 = 2\cdot 2{,}3 + 1{,}6 = 6{,}2,$$

$$\ln B = 12\ln 10 + \ln 2{,}5 = 12\cdot 2{,}3 + 0{,}9 = 28{,}5.$$

Таким образом, отношение логарифмов равно всего около $1/5$, а множитель при логарифмах $Mc^2/\mathscr{E} = 938/5 \approx 200$. Ионизационные потери протона превышают ионизационные потери электрона приблизительно в 40 раз.

Наконец, рассмотрим случай, когда обе частицы, тяжелая и легкая, — ультрарелятивистские и обладают одной и той же кинетической энергией, которую в рассматриваемом случае можно считать равной полной энергии: $\mathscr{E} = mc^2/\sqrt{1-\beta_\text{e}^2} = Mc^2/\sqrt{1-\beta_\text{p}^2}$, где β_e — отношение v/c для легкой частицы, а β_p — для тяжелой. Таким образом,

$$\frac{1}{\sqrt{1-\beta_\text{e}^2}} = \frac{\mathscr{E}}{mc^2}, \quad \frac{1}{\sqrt{1-\beta_\text{p}^2}} = \frac{\mathscr{E}}{Mc^2}.$$

Для тяжелой частицы (протона) пользуемся формулой (80.10), полагая в ней $v = c$, для легкой — формулой (81.4) и получаем

$$\frac{(d\mathscr{E}/dx)_{\text{p}}}{(d\mathscr{E}/dx)_{\text{e}}} = 2\frac{\ln A}{\ln B},$$

где

$$A = \frac{mc^2}{\overline{I}(1-\beta_{\text{p}}^2)} = \frac{m}{M}\,\frac{\mathscr{E}^2}{\overline{I}Mc^2},$$

$$B = \frac{\mathscr{E}^2}{2\overline{I}^2\sqrt{1-\beta_{\text{e}}^2}} = \frac{\mathscr{E}^3}{2\overline{I}^2mc^2}.$$

Возьмем числовой пример: $\mathscr{E} = 10$ ГэВ $= 10^{10}$ эВ, $\overline{I} = 10$ эВ, $m/M = 1/2000$, $Mc^2 = 1$ ГэВ $= 10^9$ эВ, $mc^2 = 0{,}5$ МэВ $= 0{,}5 \cdot 10^6$ эВ. Тогда $\ln A = 15{,}4$, $\ln B = 50{,}6$ и

$$\frac{(d\mathscr{E}/dx)_{\text{p}}}{(d\mathscr{E}/dx)_{\text{e}}} \approx 0{,}6.$$

Ионизационные потери ультрарелятивистского электрона в этом случае даже больше (примерно в два раза), чем ультрарелятивистского протона той же энергии. Причина этого в том, что по сравнению с электрическим полем неподвижной частицы электрическое поле ультрарелятивистского электрона изменяется более значительно (сплющивается сильнее в направлении движения и расширяется в поперечном направлении), чем электрическое поле ультрарелятивистского протона той же энергии ($\beta_{\text{e}} > \beta_{\text{p}}$).

Отличие в поведении заряженных частиц различных энергий проявляется, например, при их регистрации. Так, протон с энергией 5 МэВ оставляет в ядерной фотоэмульсии отчетливый след, а электрон с той же энергией практически незаметен. Ультрарелятивистские же частицы (например, в пузырьковой камере) трудно отличить друг от друга по оставляемым ими трекам, так как треки всех заряженных ультрарелятивистских частиц имеют практически одинаковую толщину.

4. Ускоренно движущаяся заряженная частица, как известно, испускает электромагнитные волны. В частности, это происходит при ее столкновениях с частицами вещества, через которые она проходит. Возникающее электромагнитное излучение называется *тормозным*, а потери энергии частицы на тормозное излучение — *радиационными*. Примером тормозного излучения может служить непрерывный рентгеновский спектр, возникающий при торможении электронов на антикатоде рентгеновской трубки. Торможение электронов высоких энергий используется в электронных ускорителях для получения пучков γ-лучей. В т. III, § 141 показано, что интенсивность тормозного излучения (т. е. электромагнитная энергия, испускаемая частицей в единицу времени) в нерелятивистском неквантовом приближении определяется выражением

$$w = \frac{2}{3}\,\frac{z^2e^2}{c^3}\,\dot{\mathbf{v}}^2, \tag{81.5}$$

где ze — заряд частицы, а $\dot{\mathbf{v}}$ — ее ускорение. Ускорение равно $\dot{\mathbf{v}} = \mathbf{F}/m$, где $\mathbf{F}$ — сила, действующая на частицу, а m — ее масса. Отсюда следует, что практически все радиационное торможение приходится на излучение электронов, так как излучение протона при равных действующих силах в $(m_\text{p}/m_\text{e})^2 = 1836^2 \approx 3{,}4 \cdot 10^6$ раз слабее, чем у электрона. Ионизационные потери энергии движущегося электрона обусловлены столкновениями его с электронами атомных оболочек. Они в основном пропорциональны числу электронов Z в атоме среды. Радиационные потери, напротив, в основном обусловлены столкновениями движущегося электрона с атомными ядрами среды. Они пропорциональны квадрату кулоновской силы притяжения между движущимся электроном и ядром. Эта сила в свою очередь пропорциональна Ze, а потому радиационные потери должны возрастать пропорционально второй, а не первой степени Z. Этот вывод остается справедливым и в последовательной релятивистской квантовой теории радиационного торможения, развитой Бете и Гайтлером (1904–1981).

5. Тормозное излучение, возникающее в каждом индивидуальном акте столкновения электрона с атомом, существенно зависит от степени экранирования электрического поля ядра атомными электронами. С классической точки зрения эта зависимость определяется соотношением между прицельным расстоянием налетающего электрона b и «радиусом ядра» a. Если $b/a \ll 1$, то экранирование несущественно, а при $b/a \gg 1$ экранирование полное. Все же основное значение имеет торможение электрона электрическим полем ядра. В пренебрежении экранированием энергия, теряемая электроном на радиационное торможение при прохождении одного и того же пути b в веществе, пропорциональна числу ядер, мимо которых пролетает электрон. Иными словами, эта энергия пропорциональна плотности вещества ρ и проходимому электроном пути dx. Поэтому радиационные потери энергии электрона определяются выражением

$$-\left(\frac{d\mathscr{E}}{dx}\right)_\text{рад} = \frac{\mathscr{E}}{l_r}, \tag{81.6}$$

где постоянная l_r называется *радиационной длиной.* Как уже говорилось в предыдущем параграфе (п. 6), при рассмотрении процессов поглощения вместо истинной толщины x вводят ее произведение на плотность вещества ρx (называя эту величину также толщиной). Во избежание недоразумений радиационную длину, понимаемую в таком смысле, мы будем обозначать большой буквой L_r. В таблицах обычно дают значения L_r в граммах на квадратный сантиметр. Не приводя теоретических выражений для L_r, к которым приводит теория Бете и Гайтлера ограничимся приведением числовых значений L_r для некоторых веществ (см. табл. 13). Из этой таблицы находим, например, что для сухого воздуха при температуре 18 °C и нормальном давлении ($\rho = 0{,}001213$ г/см^3)

$$l_r = \frac{36{,}5}{0{,}00121} = 30\,200 \text{ см} = 302 \text{ м}.$$

Согласно формуле (81.6) радиационные потери линейно растут с энергией, тогда как ионизационные потери при высоких энергиях

Таблица 13

Радиационные длины и критические энергии для различных веществ

Вещество	$\mathscr{E}_{\text{кр}}$, МэВ	L_r, г/см^2	Вещество	$\mathscr{E}_{\text{кр}}$, МэВ	L_r, г/см^2
H	340	58	Al	47	23,9
He	220	83	Fe	24	13,8
C	103	42,5	Cu	21,5	12,8
Воздух	83	36,5	Pb	6,9	5,8

меняются с энергией логарифмически, т. е. от энергии практически не зависят. Для сравнения можно пользоваться приближенным соотношением

$$\frac{(d\mathscr{E}/dx)_{\text{рад}}}{(d\mathscr{E}/dx)_{\text{ион}}} \approx \frac{Z\mathscr{E}}{800}, \tag{81.7}$$

где энергия $\mathscr{E}$ измеряется в мегаэлектронвольтах. Из формулы видно, что при $\mathscr{E} > 800/Z$ радиационные потери превышают ионизационные. Энергия $\mathscr{E}_{\text{кр}}$, при которой радиационные потери становятся равными ионизационным, называется *критической.* Для этой энергии приближенная формула (81.7) в мегаэлектронвольтах дает $\mathscr{E}_{\text{кр}} \approx 800/Z$. При очень высоких энергиях ионизационными потерями можно пренебречь и уравнение (81.6) проинтегрировать. Тогда получится

$$\mathscr{E} = \mathscr{E}_0 e^{-x/l_r}. \tag{81.8}$$

§ 82. Прохождение гамма-квантов через вещество

1. Ориентировочно к γ-излучению относят электромагнитные волны, длина которых значительно меньше межатомных расстояний, т. е. $\lambda \ll 10^{-8}$ см. В современных ускорителях получаются γ-кванты с энергией $\mathscr{E} \sim 20$ ГэВ, т. е. с длиной волны $\lambda = 2\pi\hbar c/\mathscr{E} \approx 6 \cdot 10^{-15}$ см $=$ $= 0{,}06$ Фм. Для практических приложений наибольший интерес представляет область от десятков килоэлектронвольт до 200–300 МэВ.

Теория прохождения γ-квантов в веществе [1]) есть проблема квантовой электродинамики, а потому здесь мы не можем ей касаться. Отметим только, что пучок γ-квантов поглощается веществом за счет электромагнитных взаимодействий. Однако по сравнению с заряженными частицами γ-кванты не имеют электрического заряда. По этой причине они не подвержены влиянию дальнодействующих кулоновских сил. Взаимодействие γ-кванта с электроном ограничено областью, линейные размеры которой порядка комптоновской длины волны электрона,

[1]) Изложенное в этом параграфе в равной мере относится к рентгеновскому излучению эквивалентной энергии.

т. е. порядка 10^{-11} см. Поэтому, проходя через вещество, γ-кванты сравнительно редко сталкиваются с электронами и атомными ядрами. Зато эти столкновения, как правило, сопровождаются резкими изменениями направления движения γ-квантов, что выводит их из пучка. Вторая особенность γ-квантов состоит в том, что они, как безмассовые частицы, могут двигаться только со скоростью c. Они не могут замедляться, а могут только либо поглощаться, либо отклоняться в сторону, либо порождать пары частица — античастица.

Таким образом, γ-кванты выбывают из пучка, как правило, в результате единичных актов столкновения с электронами или атомными ядрами вещества, через которое они проходят. Для γ-квантов нельзя ввести понятие пробега аналогично тому, как это делается для тяжелых заряженных частиц, испытывающих ионизационное торможение в веществе. Число γ-квантов, выбывающих из моноэнергетического пучка при прохождении слоя вещества толщиной dx, пропорционально dx и интенсивности пучка, падающего на этот слой. Поэтому с расстоянием x интенсивность параллельного моноэнергетического пучка γ-квантов должна убывать экспоненциально:

$$I(x) = I(0)e^{-n\sigma x}, \tag{82.1}$$

где σ — полное эффективное сечение ослабления (поглощения и рассеяния) γ-квантов на атоме, а n — число атомов поглотителя в единице объема. Величина $\tau = n\sigma$ называется *линейным коэффициентом поглощения γ-квантов.* Вместо нее удобнее пользоваться массовым коэффициентом поглощения $\mu = \tau/\rho$. Если расстояние x выражать в граммах на квадратный сантиметр, то формулу (82.1) можно переписать в виде

$$I(x) = I(0)e^{-\mu x}. \tag{82.2}$$

Основными процессами, выводящими γ-кванты из параллельного пучка, являются *фотоэффект*, *эффект Комптона* и *рождение электронно-позитронных пар.*

2. Фотоэффект уже рассматривался в § 2. Однако там речь шла преимущественно о фотоэффекте с поверхности металлов. Здесь же нас интересует главным образом фотоэффект на атомах среды, в которой распространяется пучок γ-квантов.

Свободный электрон принципиально не может поглощать и испускать γ-квант, так как в противном случае было бы нарушено одновременное выполнение законов сохранения энергии и импульса (см. § 1, п. 5). При фотоэффекте же на электроне атома вся энергия и импульс падающего γ-кванта передаются электрону и атомному остатку. Поэтому электрон должен быть обязательно связан. Кинетическая энергия $\mathscr{E}_{\mathrm{e}}$, которую получает освободившийся электрон, определяется соотношением

$$\mathscr{E}_{\mathrm{e}} = \mathscr{E}_{\gamma} - I_i, \tag{82.3}$$

где I_i — энергия ионизации той оболочки, с которой был вырван электрон (под i надо понимать $K, L_{\mathrm{I}}, L_{\mathrm{II}}, L_{\mathrm{III}}, \ldots$). Энергией отдачи,

которую приобретает атом, точнее, образовавшийся ион, ввиду ее малости мы пренебрегаем.

Ясно, что фотоэффект с i-й оболочки невозможен, если $\mathscr{E}_\gamma < I_i$, так как по своему смыслу кинетическая энергия $\mathscr{E}_e$ существенно положительна. Если же $\mathscr{E}_\gamma > I_i$, то естественно ожидать, что с возрастанием $\mathscr{E}_\gamma$ вероятность фотоэффекта должна убывать, так как при этом электрон становится все менее и менее связанным и его поведение должно приближаться к поведению свободного электрона. Опыт и теория оправдывают это ожидание.

На схематическом рис. 145 показан ход эффективного сечения фотоэффекта на атоме σ_Φ в зависимости от энергии падающего γ-кванта $\mathscr{E}_\gamma$. При очень больших $\mathscr{E}_\gamma$ сечение σ_Φ мало. С уменьшением $\mathscr{E}_\gamma$ оно возрастает, но при $\mathscr{E}_\gamma = I_K$ внезапно резко падает. Это падение происходит потому, что при $\mathscr{E}_\gamma < I_K$ вырывание электрона с K-слоя атома становится невозможным. После этого падения с дальнейшим уменьшением $\mathscr{E}_\gamma$ снова начинается возрастание σ_Φ, пока не наступит новое резкое падение на L-слое.

Рис. 145

Поскольку этот слой состоит из трех близко расположенных оболочек $L_\mathrm{I}, L_\mathrm{II}, L_\mathrm{III}$, на кривой $\sigma_\Phi = \sigma_\Phi(\mathscr{E}_\gamma)$ появляются три зубца, изображенные на рис. 145. После прохождения слоя L, при дальнейшем уменьшении $\mathscr{E}_\gamma$ опять происходит увеличение σ_Φ, пока не будет достигнут слой M и не произойдет соответствующее ему резкое падение σ_Φ, и т. д. При $\mathscr{E}_\gamma < 0{,}2$ МэВ эффективное сечение фотоэффекта падает с ростом энергии приблизительно как $\mathscr{E}_\gamma^{-7/2}$, а при $\mathscr{E}_\gamma > 0{,}5$ МэВ — приблизительно как $\mathscr{E}_\gamma^{-1}$. При этом, когда $\mathscr{E}_\gamma > I_K$, основную часть (около 80 %) в сечение σ_Φ вносит вырывание электронов из K-слоя (для $Z > 60$). Вероятность фотоэффекта растет с ростом Z приблизительно пропорционально Z^n, где n заключено между 4 и 5. Такая сильная зависимость опять-таки объясняется тем, что в легких элементах электроны в атомах связаны слабее, чем в тяжелых. Поэтому фотоэффект особенно существен при взаимодействии γ-излучения с тяжелыми атомами.

С описанным ходом эффективного сечения γ-излучения мы уже встречались в § 48, когда говорили о поглощении рентгеновских лучей в веществе.

В области энергий порядка энергий связи электронов в атомных оболочках сечение фотоэффекта очень велико по сравнению с сечением при более высоких энергиях. Например, для алюминия

$\sigma_\text{ф} \approx 6 \cdot 10^{-18}$ см^2 при $\mathscr{E}_\gamma = 1$ кэВ и $\sigma_\text{ф} \approx 6 \cdot 10^{-25}$ см^2 при $\mathscr{E}_\gamma = 0{,}1$ МэВ. Таким образом, фотоэффект является преобладающим механизмом поглощения при низких энергиях γ-излучения, а при высоких энергиях его роль становится ничтожной.

3. С увеличением энергии γ-квантов фотоэлектрическое поглощение отходит на задний план. Оно уступает место *эффекту Комптона*, разобранному в § 3. Эффект Комптона начинает играть существенную роль, когда энергия γ-квантов начинает превосходить энергию связи электрона в атоме. Когда энергия связи электрона в атоме мала по сравнению с энергией γ-кванта, электрон может считаться свободным, как мы и полагали в § 3. Ослабление пучка γ-квантов в веществе, обусловленное явлением Комптона, приводит и к рассеянию γ-излучения (при этом оно выбывает из параллельного пучка) и к частичному уменьшению энергии γ-излучения (т. е. к поглощению) за счет передачи части энергии комптоновским электронам отдачи.

Сечение рассеяния мягких γ-квантов ($\hbar\omega/m_\text{e}c^2 \ll 1$) на электроне определяется классической формулой Томсона

$$\sigma_\text{T} = \frac{8\pi}{3} r_\text{e}^2 = 0{,}665 \cdot 10^{-24} \text{ см}^2, \tag{82.4}$$

где r_e — «классический радиус электрона», т. е.

$$r_\text{e} = \frac{e^2}{m_\text{e}c^2} = 2{,}82 \cdot 10^{-13} \text{ см}. \tag{82.5}$$

Формула Томсона была выведена в § 10 (п. 2). Томсоновское, или классическое, рассеяние когерентно, т. е. происходит без изменения длины волны. Если условие $\hbar\omega/m_\text{e}c^2 \ll 1$ не выполняется, то формула Томсона не справедлива. В этом случае надо пользоваться формулой Клейна–Нишины–Тамма (10.4), которую дает квантовая электродинамика. В отличие от томсоновского рассеяния, комптоновское рассеяние (при энергии γ-квантов, большей энергии связи электрона в атоме) *некогерентно* и происходит с уменьшением длины волны. Это уменьшение, как было выяснено в § 3, одинаково для всех веществ.

Заметим еще, что вероятность комптоновского рассеяния на ядрах пренебрежимо мала, так как в этом случае роль «классического радиуса электрона» r_e играет величина $Z^2e^2/M_\text{яд}c^2$, а она в тысячи, десятки или сотни тысяч раз меньше, чем у электрона.

4. Гамма-кванты, если их энергия достаточно велика, взаимодействуют с веществом также посредством *образования пар электрон — позитрон.* Не будем останавливаться на истории предсказания существования позитрона, которое теоретически было сделано Дираком, так как его первое толкование (позитрон — «дырка» в состояниях электрона с отрицательной энергией) пришлось оставить. Напомним только, что позитрон является античастицей по отношению к электрону. Он обладает той же массой, тем же спином, теми же по величине, но противоположными по знаку зарядом и магнитным моментом, что

и электрон. Позднейшее развитие физики элементарных частиц, показало, что каждая элементарная частица, как правило, имеет свою античастицу.

Электрон и позитрон, сталкиваясь друг с другом, могут «исчезать», превращаясь в электромагнитное излучение. Этот процесс называется (не вполне удачно) *аннигиляцией.* При аннигиляции свободных электрона и позитрона не может появиться только один γ-квант, так как в противном случае нарушалось бы одновременное выполнение законов сохранения энергии и импульса. Это наиболее очевидно, если электрон и позитрон до столкновения находились в состоянии покоя. В этом случае суммарный импульс до столкновения равен нулю, тогда как импульс образовавшегося γ-кванта отличен от нуля. Но образование одного γ-кванта невозможно и в том случае, когда до столкновения электрон и позитрон двигались с различными скоростями.

Чтобы убедиться в этом, достаточно рассмотреть процесс аннигиляции в системе центра масс. В такой системе суммарный импульс до столкновения также равен нулю, а импульс образовавшегося γ-кванта по-прежнему отличен от нуля, так что заключение остается в силе. Но оно остается в силе и при рассмотрении процесса аннигиляции в любой системе отсчета, так как число образовавшихся γ-квантов от выбора системы отсчета не зависит. Таким образом, при аннигиляции электрона и позитрона должны возникать по меньшей мере два γ-кванта. Рассуждение, проведенное в обратном порядке, показывает, что свободно распространяющийся γ-квант не может породить пару, т. е. превратиться в электрон и позитрон. Но процесс образования пар может осуществляться и действительно осуществляется в электрическом поле атомного ядра. Как показывает квантово-механический расчет, согласующийся с опытом, превращение γ-квантов в электрон–позитронные пары происходит не внутри ядра, а около него в пределах области с линейными размерами порядка комптоновской длины волны электрона. Ядро воспринимает импульс отдачи, обеспечивая тем самым выполнение закона сохранения энергии — импульса, причем передача импульса отдачи ядру происходит посредством его кулоновского поля. Гамма-кванты могут рождать электрон–позитронные пары и в кулоновском поле электрона. (Возможно также рождение пар при столкновении двух γ-квантов.) Однако с наибольшей вероятностью происходит рождение пар γ-квантами в кулоновском поле ядра. Поскольку масса γ-кванта равна нулю, превратиться в электрон–позитронную пару он может только тогда, когда его энергия $\mathscr{E}_\gamma$ больше суммы энергий покоя электрона и позитрона, т. е. $2mc^2 \approx 1{,}02$ МэВ. Сечение $\sigma_{\text{пар}}$ рождения пары равно нулю, если $\mathscr{E}_\gamma < 2mc^2$. Таков действительно порог рождения пары, если оно происходит в электрическом поле тяжелой частицы — атомного ядра, так как тяжелая частица уносит малую энергию. Если же пара рождается при столкновениях γ-квантов с электроном, то электрон получает энергию того же порядка, что и частицы пары. Поэтому в этом случае рождение пары возможно только при энергии γ-кванта,

существенно превышающей $2m_e c^2$. В области энергий порядка $5m_e c^2 < \mathscr{E}_\gamma < 50m_e c^2$ теоретические расчеты в квантовой электродинамике для эффективного сечения $\sigma_{\text{пар}}$ образования пары на атомном ядре приводят к соотношению

$$\sigma_{\text{пар}} \sim Z^2 \ln \frac{\hbar\omega}{m_e c^2}. \tag{82.6}$$

По модулю сечение $\sigma_{\text{пар}}$ того же порядка, что и сечение тормозного излучения. При очень высоких энергиях величина $\ln(\hbar\omega/m_e c^2)$ заменяется постоянной из-за экранирования поля ядра электронами атомной оболочки.

Таким образом, выше порога рождения пар сечение $\sigma_{\text{пар}}$ постепенно возрастает, а при очень высоких энергиях (порядка $1000m_e c^2$) практически стремится к постоянному пределу:

$$\sigma_{\text{пар}} \approx 0{,}08 Z^2 r_e^2. \tag{82.7}$$

Наоборот, сечения фото- и комптон-эффектов при высоких энергиях γ-квантов спадают практически до нуля. При увеличении энергии рождение пар становится сначала основным, а при дальнейшем росте энергии практически единственным механизмом поглощения γ-излучения в веществе.

5. Если ограничиться только тремя основными механизмами ослабления, рассмотренными выше, то при определении полного линейного коэффициента ослабления τ γ-квантов в веществе надо принять во внимание, что в случае фотоэффекта и рождения пар рассеивающими центрами являются атомы, а в случае эффекта Комптона — электроны, которых в Z раз больше, чем атомов. Поэтому

$$\tau = n\sigma_{\text{ф}} + nZ\sigma_K^e + n\sigma_{\text{пар}}, \tag{82.8}$$

где n — число атомов в единице объема вещества, а $\sigma_{\text{ф}}, \sigma_K^e$ и $\sigma_{\text{пар}}$ — эффективные сечения фотоэффекта на атоме, эффекта Комптона на электроне и рождения пары на атомном ядре. Первое слагаемое в (82.8) преобладает при низких энергиях, второе — при средних (несколько мегаэлектронвольт), а третье — при высоких. Поэтому τ имеет минимум в области, где влияние комптоновского рассеяния наибольшее. Такой минимум особенно резко выражен для тяжелых элементов. В качестве примера на рис. 146 приведены кривые для свинца, которые наглядно показывают относительную роль всех трех рассмотренных нами механизмов ослабления в различных областях энергий γ-квантов.

6. Подводя итог последних трех параграфов, заметим, что заряженная частица, пролетая в воздухе, образует в среднем одну пару ионов противоположного знака на 33 эВ потерь. Например, α-частица с энергией 5 МэВ образует в воздухе $5 \cdot 10^6 : 33 \approx 150\,000$ пар ионов. Ионизационная способность заряженной частицы в других газовых средах примерно такая же, как и в воздухе. Гамма-кванты при прохождении через вещество передают свою энергию в основном электронам и, следовательно, вызывают ионизацию среды.

Приведем пример, иллюстрирующий это утверждение. Пусть электроны с энергией $\mathscr{E}_e = 1$ МэВ и γ-излучение той же энергии падают на поверхность алюминия. Экстраполированный пробег электро-

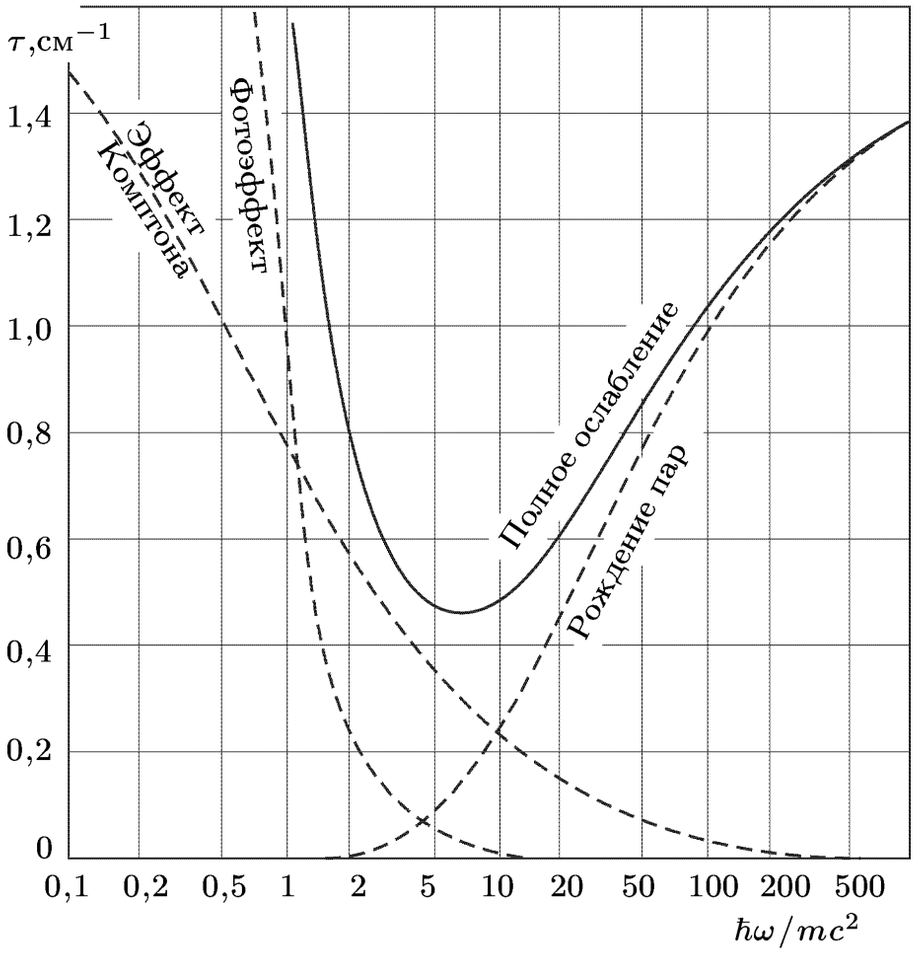


Рис. 146

на такой энергии может быть приближенно рассчитан по формуле $R\,(r/\text{см}^2) \approx 0{,}5\mathscr{E}_e\,(\text{МэВ}) - 0{,}1$.

При $\mathscr{E}_e = 1$ МэВ значение $R \approx 0{,}4r/\text{см}^2 \approx 0{,}15$ см. Средняя длина свободного пробега γ-лучей в веществе $\lambda = \dfrac{1}{n_0\sigma}$ (из формулы (82.1)) видно, что на пути $x = \lambda$ интенсивность γ-излучения уменьшается в e раз). Для γ-квантов с $\varepsilon = 1$ МэВ значения $n_0\sigma = 0{,}165$ см$^{-1}$ и $\lambda = 6$ см. Но в отличие от заряженных частиц проникающая способность γ-квантов очень велика. При облучении заряженными частицами ионизуется лишь только тонкий поверхностный слой, а при облучении γ-квантами — вся толща вещества.

§ 83. Другие проявления взаимодействия ядерных частиц с веществом

1. Заряженная частица, проходя через вещество, испытывает торможение из-за кулоновского взаимодействия не только с электронами, но и с *атомными ядрами*. Это — упругие столкновения с передачей энергии. Потери энергии частицы из-за этого эффекта выражаются

формулой типа (80.2):

$$-\left(\frac{d\mathscr{E}}{dx}\right)_{\text{яд.кул}} = \frac{4\pi z^2 Z^2 e^4 N}{M_{\text{яд}} v^2} \ln \frac{b_{\text{макс}}}{b_{\text{мин}}} = \frac{m_e}{m_p} \frac{Z}{A} \left(-\frac{d\mathscr{E}}{dx}\right)_{\text{эл}}, \qquad (83.1)$$

где $N = n/Z$ — число ядер в единице объема, а m_p — масса протона. При качественном сравнении этой формулы с (80.2) логарифмический множитель не играет роли. Существен только множитель $(m_e/m_p)Z/A \approx 1/4000$, стоящий перед логарифмом. Он появляется из-за различия масс и зарядов электрона и атомного ядра. Благодаря этому множителю потери энергии частиц при кулоновском торможении на атомных ядрах в тысячи раз меньше, чем на электронах. Однако из-за относительно большой массы ядра кулоновское рассеяние на ядрах может происходить на большие углы (и даже назад) уже в результате единичных актов рассеяния. Это как раз тот эффект, который привел Резерфорда к заключению о существовании атомного ядра (см. § 9).

2. Все частицы, за исключением фотонов, нейтрино, электронов, позитронов и мюонов, способны вступать в сильные (ядерные) взаимодействия. Такие частицы называются *адронами.*

Сильные взаимодействия положительно заряженных адронов с атомными ядрами начинаются с энергий 20–30 МэВ (в случае отрицательно заряженных адронов нет кулоновского барьера). Так как они проявляются на расстояниях порядка размеров атомного ядра, то частица сталкивается с ядром при сильных взаимодействиях примерно в $Z(R_{\text{ат}}/R_{\text{яд}})^2 \approx 10^{10} Z$ раз реже, чем с электронами при кулоновских взаимодействиях. Поэтому при движении в веществе частица испытывает ионизационное торможение, но ядерные столкновения претерпевает крайне редко. Это обстоятельство делает возможным рассматривать действие приборов для регистрации заряженных ядерных частиц (например, камеры Вильсона или пузырьковой камеры) без учета ядерных столкновений, а с учетом только ионизационного торможения, при котором, как известно, из-за малости массы электрона путь тяжелой частицы (трек) остается прямолинейным. Зато при каждом ядерном столкновении частица либо резко отклоняется в сторону (рассеивается на большой угол), либо поглощается, либо порождает новые частицы. Эти акты регистрируются по резкому излому трека, изменению его толщины и длины, по появлению новых треков, исходящих из одной точки («звезда»). Но при расчете радиационной защиты для релятивистских ускорителей и космических кораблей учет ядерных столкновений необходим.

3. Если скорость заряженной частицы превышает фазовую скорость света в рассматриваемой среде, то возникает *излучение Вавилова–Черенкова* (см. § 6, а также т. IV, § 38). Появляются потери энергии частицы на это излучение. Разумеется, они содержатся в найденных нами ранее ионизационных потерях и имеют тот же порядок, что и радиационные потери. По углу θ, под которым распространяется черенковское излучение, можно определить скорость частицы, что и делается в черенковских счетчиках. Если при этом известен импульс частицы

(по кривизне траектории в магнитном поле), то можно определить сорт частицы (по релятивистской ионизации частицы могут быть неразличимы). Поэтому черенковские счетчики — один из обязательных элементов установки, предназначенной для изучения релятивистских частиц.

4. Позитроны при прохождении через вещество в дополнение к ионизационным и радиационным потерям испытывают еще *аннигиляционные потери* за счет *двухфотонной аннигиляции* с электронами вещества:

$$e^+ + e^- \to \gamma + \gamma.$$

Сечение этого процесса при высоких энергиях меньше сечения ионизации. Но для медленных позитронов оно является определяющим.

5. Поглощаясь ядром, γ-кванты могут вызывать *ядерный фотоэффект*, т. е. выбивать из ядра нуклоны (обычно нейтроны), а также расщеплять атомное ядро. Но эти процессы практически не играют роли в поглощении γ-излучения. Порог ядерного фотоэффекта соответствует энергии связи нуклона в ядре, т. е. лежит в области энергий 6–10 МэВ. Эффективные сечения указанных процессов, как правило, возрастают с увеличением атомного номера Z. Если энергия γ-кванта во много раз превышает среднюю энергию связи нуклона, то возможно фоторасщепление ядра с вылетом нескольких нуклонов (нейтронов и протонов). При энергиях γ-квантов, превышающих $2m_\mu c^2 = 212$ МэВ (m_μ — масса мюона), в кулоновском поле ядра начинается процесс рождения мюонных пар ($\mu^+\mu^-$), аналогичный процессу рождения электрон–позитронных пар. При $\mathscr{E}_\gamma > m_\pi c^2 \approx 140$ МэВ (m_π — масса π-мезона) при взаимодействии с ядрами начинается фотогенерация π-мезонов. Поглощение γ-излучения за счет перечисленных процессов при высоких энергиях пренебрежимо мало по сравнению с поглощением их из-за рождения электрон–позитронных пар в кулоновском поле ядра.

6. Тормозное излучение электронов сопровождается возникновением мощных потоков γ-квантов, испускаемых преимущественно вперед. От таких потоков требуется усиленная защита, так как проникающая способность γ-квантов значительно превосходит проникающую способность электронов.

Электрон, позитрон или γ-квант, если их энергия достигает 1 ГэВ или выше, распространяясь в веществе, порождают *электрон–позитронные ливни*. Это явление заключается в следующем. Первичная частица, например электрон, тормозясь в электрическом поле ядра, испускает γ-квант высокой энергии. Этот γ-квант рождает электрон–позитронную пару в электрическом поле другого ядра. Электрон-позитронные пары в свою очередь порождают тормозные γ-кванты и т. д. Так возникает поток частиц, летящих практически в направлении первичной частицы, так как все эти частицы релятивистские. Этот поток и называется *ливнем*. В веществе поток частиц ливня после

своего возникновения сначала резко усиливается, но после прохождения некоторого расстояния начинает уменьшаться. Когда энергия отдельных частиц ливня уменьшается настолько, что ионизационные потери начинают преобладать над радиационными, ливень прекращается. Подобные множественные процессы образования частиц, но более разнообразные по составу вызываются и тяжелыми заряженными частицами (протонами, $\pi^{\pm}$-мезонами и пр.). Сначала они наблюдались в земной атмосфере и вызывались частицами космических лучей высоких энергий (см. § 103, п. 12). Каскады таких частиц, порождаемые первичными частицами с энергией $\mathscr{E} \gtrsim 10^5$ ГэВ, содержат 10^6–10^9 частиц и называются *широкими атмосферными ливнями.* Отдельный ливень покрывает площадь земной поверхности в несколько квадратных километров. С появлением ускорителей на высокие энергии основные исследования множественных процессов стали производиться на ускорителях.

7. Налетающие частицы достаточно высоких энергий при неупругих столкновениях с атомными ядрами могут частично разрушать их, например выбивать протоны, нейтроны или вызывать другие ядерные превращения. В результате образуются новые атомные ядра и новые изотопы химических элементов. Они, как правило, радиоактивны, так что в веществе возникает *наведенная радиоактивность.* Реакции выбивания протонов или нейтронов из ядра и прочие ядерные реакции, производимые электронами и γ-квантами, сильно эндотермичны и имеют порог около 10 МэВ. Но даже выше этого порога из-за слабости электромагнитных взаимодействий сечения этих процессов очень малы — на несколько порядков меньше площади эффективного сечения самого атомного ядра. Проникновению протонов и α-частиц в ядро препятствует *кулоновский потенциальный барьер*, особенно высокий в случае тяжелых ядер. Поэтому протоны и α-частицы могут создать заметную наведенную активность лишь при сравнительно высоких энергиях (во всяком случае, больше примерно 10 МэВ). Заметим еще, что α- и β-частицы, а также γ-кванты, возникающие в результате радиоактивных распадов ядер, обладают энергией всего в несколько мегаэлектронвольт. Такие радиоактивные излучения, как правило, создать дополнительную наведенную радиоактивность не могут.

Глава XII

ИСТОЧНИКИ И МЕТОДЫ РЕГИСТРАЦИИ ЯДЕРНЫХ ЧАСТИЦ

§ 84. Ускорители

1. Единственный способ изучения ядер и элементарных частиц, за исключением получения некоторых статических характеристик ядер (спины, магнитные и электрические квадрупольные моменты), состоит в осуществлении столкновений одних частиц с другими и регистрации последствий столкновений. Сначала для этой цели (в опытах Резерфорда и др.) использовались α- и β-частицы, возникающие в результате распада естественных радиоактивных ядер, а также частицы высоких энергий, содержащиеся в космических лучах. С изобретением ускорителей основные исследования рассматриваемого типа осуществляются с заряженными частицами, ускоряемыми для этой цели в электрических и магнитных полях.

Ускорители — очень сложные установки. Их устройство и работа — это вопрос не ядерной физики, а физической электротехники. Но роль ускорителей в ядерной физике и в особенности в физике элементарных частиц настолько велика (можно сказать — решающая), что понимание принципов их работы абсолютно необходимо. Краткое изложение этих принципов и является целью настоящего параграфа. При этом мы оставим в стороне многие технические и даже физические детали, хотя они и весьма существенны для работы ускорителей. Помимо физических применений, ускорители начинают все больше и больше использоваться за пределами физики (химия, биофизика, геофизика) и в прикладных целях (стерилизация продуктов, дефектоскопия, лучевая терапия и т. п.). Однако рассматривать подобные применения мы не будем.

При осуществлении столкновений более тяжелые частицы обычно покоятся и называются *частицами мишени*, а более легкие налетают на них в виде пучка *ускоренных частиц*. В ускорителях на встречных пучках частицы движутся навстречу друг другу, так что их деление на частицы мишени и частицы пучка теряет смысл. Кроме того, мишенями могут служить только частицы и ядра, которые входят в состав макроскопических тел и живут достаточно долго (не менее нескольких минут), а также протон и электрон. Из всех остальных частиц и ядер изготовить мишени не удается. В результате у нас нет, например, прямых экспериментальных данных о рассеянии нейтронов на нейтронах, тогда как рассеяние нейтронов на протонах и в особенности протонов на протонах экспериментально исследовано с большой полнотой.

Что касается ускоряемых частиц, то в большинстве ускорителей получаются пучки ускоренных протонов или электронов. В ускорителях получают также пучки дейтронов и α-частиц. Имеется небольшое количество ускорителей тяжелых ионов: многократно заряженных ионов углерода, азота, кислорода и пр. Такие ускорители применяются главным образом для синтеза сверхтяжелых трансурановых элементов (см. § 94), а также для поисков сверхтяжелого состояния ядерной материи. Созданы источники для ускорения позитронов и антипротонов. Энергии ускоренных заряженных частиц меняются в пределах от нескольких мегаэлектронвольт до сотен гигаэлектронвольт, причем верхний предел определяется не принципиальными трудностями, а уровнем развития ускорительной техники. Этот предел постоянно повышается примерно на порядок за десятилетие.

2. Первым ускорителем, получившим практические применения в ядерной физике (начиная с 30-х годов) является *электростатический генератор Ван-де-Граафа* (1901–1967). Принцип работы этого генератора уже был описан нами (см. т. III, § 11). Из внутренней области полого металлического шара, заряжаемого до очень высокого потенциала, выходит многосекционная ускорительная вакуумная трубка, в которой и происходит ускорение заряженных частиц. При работе генератора повышение напряжения на ускорительной трубке происходит до тех пор, пока ток утечки не сравняется с зарядным током, т. е. током, возникающим из-за движения заряженной транспортерной ленты, или пока не возникнет электрический пробой. Достижимый верхний предел напряжения на трубке ограничен напряжением пробоя между шаром и окружающими предметами. Обычные генераторы Ван-де-Граафа позволяют получать напряжения до 2–5 МВ, усовершенствованные — до 15–20 МВ.

Простым приемом эффективное напряжение генератора Ван-де-Граафа удается повысить в два раза. Для этого применяются две ускорительные трубки, расположенные вертикально одна на продолжении другой. Высоковольтный электрод (положительный) помещается между ними. Там же помещается тонкая металлическая фольга. На верхнем конце верхней трубки располагается источник ионов, на нижнем конце нижней трубки находится выпускное отверстие. Источник ионов и выпускное отверстие заземляются, а потому находятся при нулевом потенциале. Из источника должны выходить отрицательные ионы, т. е. частицы, содержащие избыток электронов. Чаще всего ими являются ионы водорода или дейтерия. Их получают, впуская водород или дейтерий в область, где горит электрический разряд. Там молекулы газа диссоциируют, а атомы ионизуются. Таким путем обычно получаются положительные ионы, но определенная часть ионов несет и отрицательный заряд. В последнем случае при ионизации атома водорода ион представляет собой протон, вокруг которого вращаются два электрона, слабо с ним связанные. Образовавшийся отрицательный ион направляется к положительному высоковольтному электроду и ускоряется до определенного потенциала. В конце пути он пронизывает металлическую фольгу, которая его «обдирает», т. е. лишает

обоих электронов. Здесь отрицательный ион превращается в положительный и продолжает двигаться в том же направлении к выпускному отверстию, уже отталкиваясь от высоковольтного электрода. В результате перезаряженная частица, подходя к выпускному отверстию, дополнительно приобретает такую же энергию, какую она получила на первой половине пути. Ускоритель, работающий по этому принципу, называется *перезарядным ускорителем* или *тандем-генератором.* По сравнению с обычным генератором Ван-де-Граафа он обладает тем важным преимуществом, что в нем источник газа находится не под высоким потенциалом, а при потенциале земли, а это сильно упрощает его питание и обслуживание.

Недостатком генератора Ван-де-Граафа является жесткое ограничение энергии пучка сверху. Но этот генератор обладает и важными преимуществами. Главнейшие из них — высокая степень монохроматичности пучка (до 10^6, т. е. выше, чем на любом другом ускорителе) и легкость регулирования его энергии. Благодаря этому эффективное сечение рассеяния протона на протоне (например, при низких энергиях) измерено на генераторе Ван-де-Граафа с точностью, которая недостижима при измерении любых других сечений в ядерной физике. Вот почему генератор Ван-де-Граафа до сих пор широко применяется в исследованиях при низких энергиях. Важным достоинством генератора Ван-де-Граафа является также возможность получения больших токов в пучке при высоких КПД. Ток в пучке в генераторах Ван-де-Граафа доходит до нескольких сот микроампер.

3. Перейдем к рассмотрению *линейных ускорителей.* Они относятся к классу *резонансных ускорителей*, так как в них для ускорения заряженных частиц применяются высокочастотные переменные поля, частота которых строго согласована со скоростью движения ускоряемой частицы. Простейшим из таких ускорителей является *ускоритель Видерое* (р. 1902), в котором полые цилиндрические электроды (называемые *дрейфовыми* или *пролетными трубками*) постоянно возрастающей длины располагаются друг за другом, как это указано на рис. 147. Трубки через одну соединены с одним полюсом генератора переменного напряжения, остальные трубки соединены с другим полюсом. Ускоритель является импульсным, т. е. дает не непрерывный ток ускоренных частиц, а ускоряет отдельные сгустки частиц (банчи). На верхней части рис. 147 показаны знаки зарядов на трубках и направления электрических полей в зазорах между ними в некоторый момент времени t. То же сделано на нижней части рис. 147, но через половину периода переменного напряжения генератора.

Пусть, например, положительная частица (протон) попадает в первый зазор между трубками в направлении электрического поля. Тогда в этом зазоре скорость частицы увеличится. После этого она попадает в канал цилиндрической дрейфовой трубки и будет двигаться по инерции, так как внутрь металлической трубки электрическое поле не проникает. Длина дрейфовой трубки такова, что, когда частица подходит ко второму зазору между трубками, направление электрического поля меняется на противоположное, так что в этом зазоре опять происходит

ее ускорение. То же самое происходит в третьем и во всех последующих зазорах между дрейфовыми трубками.

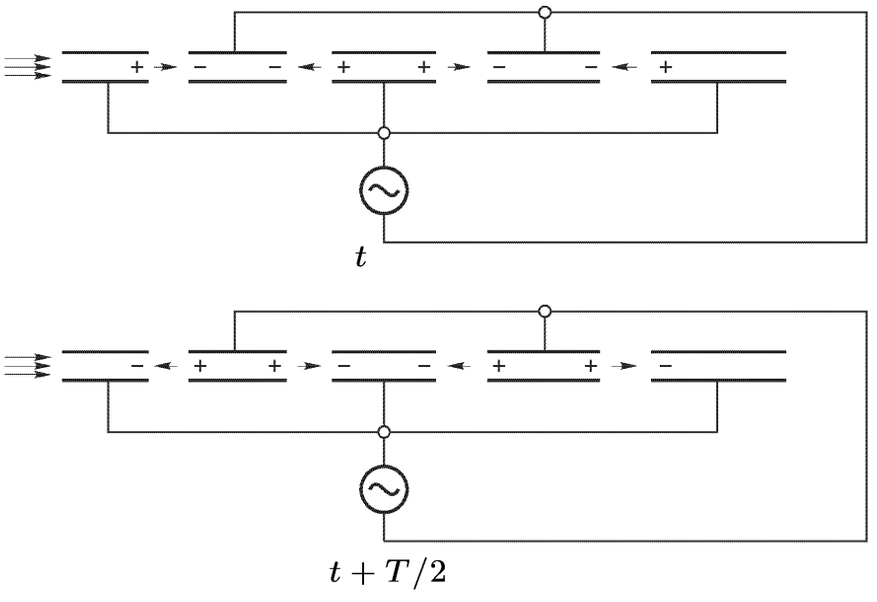


Рис. 147

Таким образом, в каждом зазоре между дрейфовыми трубками частица движется ускоренно. Для этого необходимо, чтобы длины последовательных трубок непрерывно возрастали в соответствии с ростом скорости, достигнутой частицей в процессе ускорения. Чтобы дрейфовые трубки получились не слишком длинными, надо применять генераторы высокой частоты.

В приведенном рассуждении молчаливо предполагалось, что фазы ускоряющих напряжений между соседними трубками меняются *синхронно*. На самом деле строгий синхронизм нарушается из-за конечной скорости распространения электромагнитных возмущений. При малых скоростях частицы это обстоятельство практически не играет роли. Но когда скорость частицы приближается к скорости света, нарушение синхронизма становится большим, и это обстоятельство должно быть принято во внимание. Поэтому линейные ускорители Видерое могут давать ускоренные частицы (протоны) сравнительно низких энергий (до 10 МэВ). В настоящее время они применяются редко и только на первых стадиях ускорения, пока скорость частицы невелика по сравнению со скоростью света.

4. Альварец (р. 1911) предложил и практически осуществил *линейный резонансный ускоритель*, в котором дрейфовые трубки не присоединяются к генератору высокого напряжения. Ускоритель Альвареца представляет собой цилиндрическую трубку (объемный резонатор), в которой возбуждается стоячая электромагнитная волна электрического типа, т. е. такая волна, электрический вектор которой направлен параллельно оси трубы. Электрическое поле в таком резонаторе имеет вид $E = A(r)\cos kx\cos\omega t$, где координата x отсчитывается вдоль трубы, а амплитуда $A(r)$ зависит от расстояния r до оси резонатора. Частота ω не может быть произвольной, так как в силу граничных условий на стенках трубы в ней можно возбудить стоячие волны только

с вполне определенными характеристиками. Понятно, что в ускорителях Видерое подобное ограничение на частоту генератора не накладывается. В узлах *1, 2, 3*, ... электрическое поле обращается в нуль, как это видно из рис. 148. Через половину периода T электрическое поле всякий раз меняет направление на противоположное.

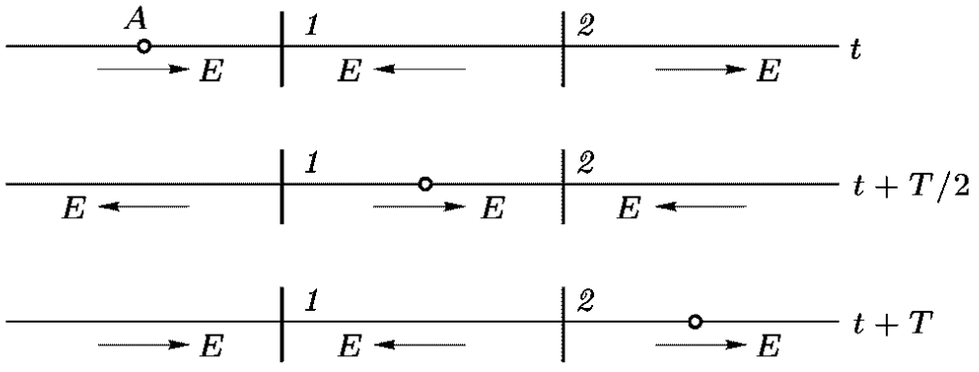


Рис. 148

Пусть в таком поле положительно заряженная частица, например p, движется с постоянной скоростью v и притом так, что в точке A она оказывается в тот момент времени, когда электрическое поле максимально. Тогда частица будет ускоряться, точнее, увеличивать свою энергию. Пусть через четверть периода она окажется в узле *1*, где электрическое поле обращается в нуль. В этот момент поле E всюду меняет направление на противоположное, так что оно будет ускорять частицу и между узлами *1* и *2*. К узлу *2* частица должна подходить в момент, когда снова происходит изменение направления электрического поля. Поэтому за узлом *2* ее движение снова будет ускоренным. То же справедливо и для движения между узлами *2* и *3* и т. д. Чтобы частица всюду двигалась ускоренно, ее скорость v должна определяться условием

$$v = \frac{\lambda}{2} : \frac{T}{2} = \frac{\lambda}{T} = v_{\text{фаз}}, \tag{84.1}$$

где $v_{\text{фаз}}$ — фазовая скорость электромагнитной волны вдоль трубы. Но фазовая скорость в трубе (не заполненной веществом) всегда превышает скорость света в вакууме c. Поэтому частица со скоростью $v_{\text{фаз}}$ двигаться не может. При реальном движении она должна будет последовательно проходить и через ускоряющие и через замедляющие участки. Альварец преодолел эту трудность, поместив на замедляющих участках дрейфовые трубки и тем самым обратив в нуль электрическое поле на этих участках. Эти трубки ни к каким источникам высокого напряжения присоединять не надо — они сами заряжаются колеблющимся электромагнитным полем.

Конечно, электромагнитное поле в трубе, нагруженной дрейфовыми трубками, не имеет столь же простой формы, что в их отсутствие. Поле вытесняется из участков, занятых трубками, в промежутки между ними. Но это обстоятельство не играет роли. Существенно только, чтобы частица всегда попадала в ускоряющие промежутки между трубками и двигалась в отсутствие поля внутри самих трубок. Для

этого длины дрейфовых трубок должны определенным образом возрастать со скоростью частицы. Именно, через дрейфовые трубки частица должна проходить за те полупериоды, когда в ненагруженном резонаторе электрическое поле ее тормозило бы. Поэтому длина дрейфовой трубки связана со скоростью движущейся частицы соотношением

$$l = v\frac{T}{2} = \frac{v}{c}\,\frac{cT}{2} = \frac{v}{c}\,\frac{\lambda}{2}. \tag{84.2}$$

Такое соотношение справедливо и в случае неравномерного движения ускоряемой частицы. Только в этом случае под v надо понимать скорость частицы в момент прохождения ее через рассматриваемую дрейфовую трубку. То же соотношение справедливо и в случае ускорителя Видерое.

5. Действие полого электромагнитного резонатора как ускорителя можно объяснить и с другой точки зрения. Для этого представим стоячую волну в резонаторе в виде суперпозиции двух бегущих волн одинаковых частот, распространяющихся навстречу друг другу:

$$A\cos kx\cos\omega t = \frac{A}{2}\cos(\omega t - kx) + \frac{A}{2}\cos(\omega t + kx).$$

Первая волна называется *попутной*, вторая — *встречной.* Пусть частица движется с фазовой скоростью $v = \omega/k$, все время находясь на гребне попутной волны $(A/2)\cos(\omega t - kx)$. Тогда эта волна будет непрерывно ускорять частицу. Встречная же волна будет оказывать на частицу ускоряющее и замедляющее действия, с большой частотой сменяющие друг друга. В результате возникнут небольшие, но быстрые колебания около некоторого усредненного плавного движения частицы, не играющие существенной роли. Конечно, для объяснения таких колебаний нет необходимости разлагать стоячую волну на сумму двух бегущих. Достаточно заметить, что отнюдь не все частицы пучка подходят к узлам электрического поля точно тогда, когда происходит изменение направлений поля. Это приводит к отставанию таких частиц от частицы, приходящей к узлам в моменты таких изменений, к ухудшению монохроматичности пучка и к явлениям колебательного характера.

Для ускорения частиц можно использовать только одну попутную волну, устранив вредное влияние встречной волны. Такой ускоритель называется *ускорителем с бегущей волной.* Для реального ускорения частиц требуется еще замедление попутной волны — снижение ее фазовой скорости до величины, меньшей скорости света в вакууме. Для этого применяются «нагруженные» волноводы, заполненные, например, перегородками с достаточно большими отверстиями. Этим можно также достигнуть того, чтобы фазовая скорость волны, а с ней и скорость частицы медленно возрастали вдоль волновода. Такое возрастание скорости волны не требуется только в случае ультрарелятивистских электронов, когда их скорости практически уже достигли релятивистского предела — скорости света в вакууме. Но и волноводы с медленно нарастающей фазовой скоростью можно использовать для

ускорения только уже достаточно быстрых частиц, так как большого замедления волн (по сравнению с вакуумом) в таких волноводах получить не удается. В ускорителе Альвареца труба (объемный резонатор) закрывается с концов проводящими крышками, от которых электромагнитные волны отражаются, в результате чего образуются стоячие волны. В ускорителях с бегущей волной отражение от внешнего конца надо устранить. Для этого с заднего конца к трубе присоединяется резонатор определенных размеров, в который уходит падающая волна, выделяя джоулево тепло.

Большой линейный ускоритель с бегущей волной для ускорения электронов до 1,8 ГэВ сооружен в 1964 г. в Физико-техническом институте АН УССР в Харькове. Его длина 240 м. Самый большой в мире линейный ускоритель с бегущей волной, ускоряющий электроны до 22,3 ГэВ, построен в Стэнфорде (США). Его длина 3,05 км. Для замедления электромагнитных волн в волноводе устанавливают специальные диафрагмы, делящие его на отдельные ячейки. Последние можно рассматривать как линейные полые резонаторы, в которых бегущая волна возбуждает электромагнитное поле. Из-за неизбежных потерь в стенках амплитуда волны непрерывно падает вдоль волновода. Для ее поддержания применяются специальные генераторы расположенные по всей длине ускорителя. В Стэнфордском линейном ускорителе таких генераторов более 80 тысяч.

Линейные ускорители стоят очень дорого. В мощных протонных ускорителях приращение энергии протона составляет 1–1,5 МэВ на каждый метр ускорителя, а в электронных — 10 МэВ. Предельно допустимая энергия пучка частиц определяется почти исключительно стоимостью ускоряющих резонаторов и связанных с ними радиочастотных источников энергии.

В настоящее время на основе Стэнфордского линейного ускорителя, энергия которого увеличена до 50 ГэВ, заканчивается сооружение установки для встречных электрон–позитронных пучков. Электроны и позитроны будут ускоряться в линейном ускорителе, а затем их траектории будут разводиться по разным окружностям, в месте пересечения которых будут возникать встречные столкновения.

В Институте ядерной физики СО АН СССР разработан проект линейного ускорителя (ВЛЭП), позволяющего получать встречные e^-- и e^+-пучки при темпе ускорения 100 МэВ/м. Это позволит при сравнительно малых размерах ускорителей получать встречные пучки e^- и e^+ с энергией 500 и даже 1000 ГэВ в каждом пучке. При таком темпе ускорения в линейном ускорителе можно будет ускорять до 1000 ГэВ даже нестабильные частицы, например $\pi^{\pm}$-мезоны. Создание таких высоких градиентов полей потребовало новой технологии (вакуум, чистота поверхностей и т. д.).

Важным достоинством линейных ускорителей электронов является то, что из-за прямолинейности траектории ускоренных частиц они практически свободны от электромагнитного излучения. Линейные ускорители обладают высокой интенсивностью, могут иметь большое

практическое значение, например, для выработки ядерного топлива (^{239}Pu, ^{233}U) из ^{238}U, ^{232}Th (см. § 95, п. 11).

6. В *циклических ускорителях* частица не проходит однократно через длинный ряд ускоряющих резонаторов, а многократно периодически возвращается к одним и тем же ускоряющим промежуткам. Таким путем достигается значительное снижение размеров и стоимости ускорителя. Первый циклический ускоритель был построен в 1930 г. Лоуренсом (1901–1958) и назван *циклотроном*. Чтобы понять идею циклотрона, напомним, что в однородном постоянном магнитном поле $\mathbf{H}$ заряженная частица, если ее скорость перпендикулярна к $\mathbf{H}$, равномерно вращается по окружности с циклической частотой

$$\Omega = \frac{eH}{mc}, \tag{84.3}$$

т. е. с периодом обращения $T = 2\pi/\Omega$. Существенно, что при нерелятивистском движении период T *не зависит* от скорости движения частицы. Это и используется в циклотроне.

Циклотрон представляет собой сплющенную цилиндрическую металлическую коробку, в которой вдоль диаметра имеется прорезь, разделяющая коробку на две половины A и B, называемые *дуантами* (рис. 149). Коробка помещается в постоянное однородное магнитное поле между противоположными полюсами магнита (электромагнита). Между дуантами A и B прикладывается электрическое напряжение от высокочастотного генератора с частотой Ω (84.3). Источником ионов (положительных) служит небольшая дуга, горящая в центре циклотрона между накаленным катодом и стенкой полости, служащей анодом. Непрерывной откачкой в дуантах поддерживается давление в 10^{-4}–10^{-5} мм рт. ст., тогда как в центре полости, где горит дуга, давление примерно в 100 раз выше. Положительный ион, выйдя из дуги, движется в электрическом поле зазора к отрицательному дуанту и набирает энергию. Внутри дуанта он движется равномерно, описывая полуокружность. Через промежуток времени $T/2$ он подходит к зазору между дуантами с противоположной стороны от центра. К этому моменту электрическое поле меняет направление на противоположное, так что ион снова попадает в ускоряющее поле и снова ускоряется. После этого в течение того же промежутка времени $T/2$ ион с приобретенной скоростью равномерно движется по полуокружности уже во втором дуанте. Через промежуток времени $T/2$ ион опять подходит к зазору между дуантами (когда электрическое поле изменит направление) и снова будет ускоряться. И так продолжается дальше. Ион движется по раскручивающейся спирали, периодически испытывая ускорение через время $T/2$. На последнем витке спирали включается отклоняющее электрическое поле, выводящее пучок наружу. Напряженность ускоряющего поля лимитируется возможностями пробоя. Она обычно не превышает 100 кВ.

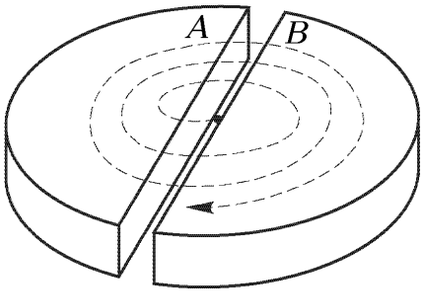


Рис. 149

Энергия, приобретаемая частицей при выходе из ускорителя, от ускоряющего поля не зависит. Она определяется лишь напряженностью магнитного поля H и радиусом R цилиндрической области, в которой оно создано. Действительно, скорость, приобретаемая частицей, $v = \Omega R$, а кинетическая энергия

$$\mathscr{E} = \frac{1}{2} m\Omega^2 R^2 = \frac{1}{2} \frac{(eHR)^2}{mc^2}. \tag{84.4}$$

Так, при $H = 15$ кГс и $R = 0{,}4$ м при ускорении протонов ($mc^2 = 938$ МэВ) получаем

$$\mathscr{E} = \frac{1}{2} \cdot \frac{(4{,}8 \cdot 10^{-10} \cdot 1{,}5 \cdot 10^4 \cdot 40)^2}{938 \cdot 10^6 \cdot 1{,}6 \cdot 10^{-12}} \text{ эрг} = \frac{1}{2} \frac{(4{,}8 \cdot 1{,}5 \cdot 4 \cdot 10^{-5})^2}{938 \cdot 1{,}6 \cdot 10^{-6} \cdot 1{,}6 \cdot 10^{-12}} \text{ эВ} =$$
$$= 17 \cdot 10^6 \text{ эВ} = 17 \text{ МэВ}.$$

Если амплитудное напряжение между дуантами равно 100 кВ, то при каждом полном обороте энергия протона увеличивается на 200 кэВ, так как при этом он дважды пересекает ускоряющий зазор между дуантами. Поэтому для набора энергии 17 МэВ протону требуется совершить $17 \cdot 10^6 : 2 \cdot 10^5 = 85$ оборотов.

Для работы циклотрона существенно, чтобы частица все время двигалась в центральной плоскости циклотрона или, во всяком случае, возвращалась к этой плоскости при малых отклонениях от нее в ту или другую сторону, а не попадала на один из дуантов. Таким свойством реальные циклотроны действительно обладают — оно называется *фокусировкой.* Фокусировка обеспечивается неоднородностью магнитного поля и отчасти электрического поля в зазорах между дуантами. Фокусирующим действием магнитное поле обладает потому, что оно убывает от центра к периферии. Благодаря этому магнитные силовые линии обращены выпуклостями наружу, т. е. имеют бочкообразную форму.

На рис. 150 изображена одна из магнитных силовых линий AB, причем предполагается, что удерживающее магнитное поле направлено вверх. Рассмотрим ради определенности положительную частицу. Пусть она сместилась из средней плоскости CD в положение M. В этом положении удерживающее магнитное поле вынуждает частицу двигаться в направлении от читателя за плоскость рисунка. Следовательно, сила Лоренца F имеет составляющую, направленную вниз. Эта сила и вернет отклонившуюся частицу снова в среднюю плоскость CD. Если бы магнитное поле достаточно быстро возрастало от центра к периферии, то магнитные силовые линии были бы обращены выпуклостью к центру. Возникла бы сила Лоренца с составляющей, направленной вверх, которая отклонила бы частицу от сред-

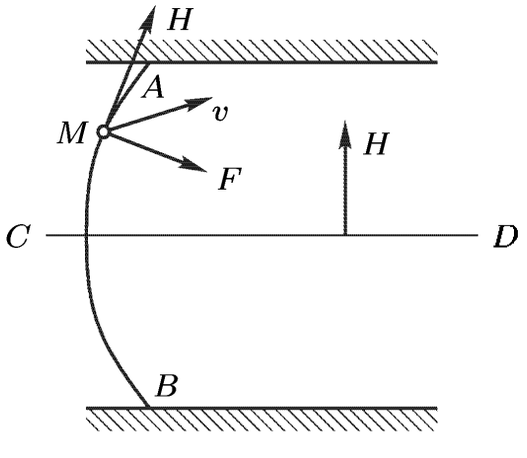


Рис. 150

ней плоскости CD еще дальше. В конце концов частица попала бы на верхнюю плоскость дуанта.

Существенный недостаток циклотрона состоит в том, что он позволяет ускорять частицы *только до нерелятивистских энергий.* Лишь в этом случае имеет место синхронизм колебаний электрического поля между дуантами с обращением частицы по окружности. При релятивистских движениях формула (84.3) сохраняется, но под m следует понимать не массу покоя, а релятивистскую массу частицы. Релятивистская же масса, а с ней и период обращения частицы T возрастают по мере ускорения частицы, что приводит к нарушению синхронизма.

В силу сказанного эффективно ускорять протоны на циклотроне можно лишь до энергий 20–25 МэВ, а ускорять электроны практически невозможно. Но в области низких энергий циклотрон, давая токи до 1 мА, по интенсивности значительно превосходит все другие ускорители. По этой причине циклотроны до сих пор широко используются для изучения ядерных реакций при низких энергиях, а также для промышленного получения тех (преимущественно нейтронодефицитных) изотопов, которые не получаются в ядерных реакторах. Циклотроны применяются также для получения пучков дейтронов, α-частиц, многократно ионизованных тяжелых ионов. Тяжелые ионы удается ускорять до энергий в несколько сотен мегаэлектронвольт, так как из-за большой массы релятивистские поправки для них становятся существенными при бóльших энергиях, чем для протонов.

7. *Фазотрон* (*синхроциклотрон*) позволяет ускорять частицы и до релятивистских энергий. Он отличается от циклотрона тем, что в нем для ускорения применяются электрические поля медленно меняющейся частоты, так что частота поля равна частоте обращения частицы, которая уменьшается из-за релятивистского увеличения массы. Такое поле ускоряет частицу на каждом витке пути, несмотря на релятивистское изменение массы. Фазотрон работает только в импульсном режиме — в каждый момент времени в камере ускоряется только один сгусток частиц.

Одним из основных условий, необходимых для работы фазотрона и большинства других современных резонансных ускорителей на релятивистские энергии является *автофазировка*, открытая В. И. Векслером (1907–1966) в 1944–1945 гг. и почти одновременно с ним Мак-Милланом (р. 1907). Дело в том, что полного совпадения частот ускоряющего поля и ускоряющейся частицы, как до сих пор предполагалось для упрощения рассуждений, в циклических резонансных ускорителях никогда не бывает.

На самом деле эти частоты несколько отличаются друг от друга. Да и сами частицы из-за некоторого разброса скоростей обращаются с несколько различными частотами. Возникает вопрос, не приведет ли это обстоятельство к потери устойчивости продольного движения сгустка частиц? Если бы это было так, то сгусток расплывался бы в продольном направлении, распадался, и работа ускорителя сделалась бы невозможной.

Для выяснения принципа автофазировки представим напряженность электрического поля E в ускоряющих промежутках в зависимости от времени t. Поле E изображается синусообразной кривой, но ее частота медленно меняется во времени. Существенно заметить, что время T, затрачиваемое частицей при переходе от одного ускоряющего промежутка к соседнему, зависит от ее энергии. Эта зависимость определяется двумя противоположно действующими факторами. С увеличением энергии частицы увеличивается ее скорость, что уменьшает T, но зато удлиняется путь между двумя последовательными ускорениями.

Результирующий эффект зависит от устройства ускорителя. Так, в линейном резонансном ускорителе второй фактор не действует и время T с увеличением энергии уменьшается. В других ускорителях соотношение может быть обратным. Примем для конкретности, что с увеличением энергии время T удлиняется. Противоположное допущение на ход рассуждения не влияет и, по существу, не отражается на окончательном выводе.

Пусть две частицы, мало отличающиеся по скоростям, одновременно выходят из какого-либо ускоряющего промежутка. При подходе к следующему ускоряющему промежутку более быстрая частица *1* опередит более медленную *1′* (рис. 151 вверху). Допустим, кроме того, что частицы проходят через ускоряющий промежуток тогда, когда электрическое поле, пройдя через максимум, начинает ослабевать. Тогда при нашем прежнем предположении в новом ускоряющем промежутке частица *1′* получит большее приращение скорости, чем частица *1*. При движении к следующему ускоряющему промежутку частица *1′* начнет ликвидировать свое отставание. Может даже случиться, что частица *1′* обгонит частицу *1*, как это изображено на нижнем рис. 151. Если это так, то в новом ускоряющем промежутке частица *1′* получит уже меньшее приращение скорости, чем частица *1*. При дальнейшем движении опережение частицы *1′* начнет уменьшаться и может оказаться, что при подходе к следующему ускоряющему промежутку она вновь окажется отстающей. Таким образом, при движении двух близких частиц расстояние между ними будет попеременно то увеличиваться, то уменьшаться.

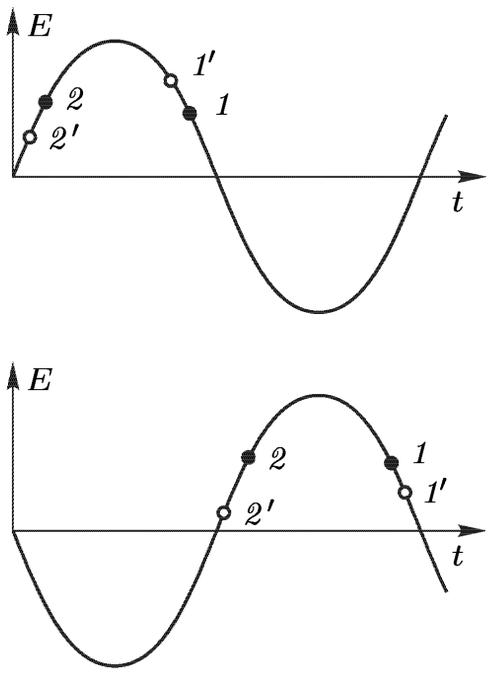


Рис. 151

В частности, изменение частоты поля можно подобрать так, чтобы частица *1* была резонансной, т. е. проходила через все ускоряющие участки в те моменты времени, когда фаза ускоряющего поля возвращается к одному и тому же значению. В этом случае возникнут малые колебания близких частиц вокруг положения резонансной частицы *1*, т. е. образуется *устойчивый сгусток* ускоряющихся частиц.

Из изложенного ясно, что так будет происходить в том случае, когда сгусток ускоряемых частиц проходит через ускоряющие промежутки в те моменты времени, когда ускоряющее поле уже прошло через максимум и начинает убывать. Если же оно еще не достигло максимума, то отставание частицы $2'$ от 2 (рис. 151 внизу) из-за приобретения дополнительной энергии в ускоряющих промежутках в дальнейшем будет увеличиваться. Частицы будут удаляться друг от друга, т. е. сгусток получится неустойчивым и развалится. Положение меняется, когда с возрастанием энергии частицы время T уменьшается. В этом случае устойчивый сгусток частиц образуется при условии, что он проходит через ускоряющие промежутки еще до того, как ускоряющее поле достигнет максимума. Так или иначе, устойчивый сгусток частиц образуется либо справа, либо слева от максимума электрического поля. Там же, но не в самом максимуме находится и соответствующая резонансная частица. Автоматическое образование устойчивого сгустка частиц в резонансных ускорителях и называется автофазировкой.

После изложения принципа автофазировки сообщим некоторые данные о фазотронах. Эти ускорители используются для ускорения тяжелых частиц — протонов, дейтронов, α-частиц. Они ускоряют частицы до энергий 1 ГэВ, давая в секунду от нескольких десятков до нескольких сотен импульсов. В одном импульсе содержится 10^9–10^{10} частиц. Интенсивность пучка в фазотроне намного меньше, чем в циклотроне, но все же довольно велика — порядка 2 мкА [1]. Ускоряющее напряжение равно 10–30 кВ. Поэтому в фазотроне на 700 МэВ частица должна совершить примерно 10^5 оборотов. Максимальная энергия, достигаемая на фазотроне, определяется не физическими, а экономическими соображениями — главным образом стоимостью магнита, который изготовляется из высококачественного трансформаторного железа и является наиболее дорогой частью установки. Дело в том, что в фазотроне частица раскручивается, начиная от центра к периферии. Поэтому магнитное поле должно быть создано во всем объеме камеры, в которой происходит это раскручивание. Для этого магнит должен быть снабжен полюсными наконечниками большого размера. При любой скорости импульс частицы определяется соотношением

$$p = \frac{eHr}{c}. \tag{84.5}$$

Площадь полюса возрастает как площадь последнего витка спирали, т. е. как квадрат импульса частицы. Примерно так же, а значит очень быстро, возрастает и стоимость магнита. Этим определяется область изменения энергии, в которой используется фазотрон. При энергиях от 25 до сотен мегаэлектронвольт фазотронный метод ускорения протонов, дейтронов и α-частиц в настоящее время является основным.

[1]) Существует ускоритель SIN (Швейцария) с интенсивностью $I = 200$ мкА (ее предполагают повысить до 2 мА). Интенсивность ускорителя в ОИЯИ (Дубна) после реконструкции доведена до 20 мкА.

8. К ускорителям, в которых используется постоянное во времени магнитное поле, относится *микротрон*, применяющийся для ускорения электронов. Идея микротрона была высказана В.И. Векслером еще в 1944 г., но первая действующая экспериментальная установка осуществлена только в 1948 г. в Канаде. В отличие от циклотрона и фазотрона источник ускоряемых электронов в микротроне помещается не в центре, а на краю области магнитного поля. Там же помещается полый ускоряющий резонатор, при прохождении через который энергия электрона всякий раз увеличивается на энергию покоя электрона $m_0c^2 = 0{,}511$ МэВ (или на величину, ей кратную). После этого электрон, описав окружность в магнитном поле, возвращается в ускоряющий промежуток, где его энергия снова возрастает на m_0c^2. В результате n-кратного прохождения через ускоряющий промежуток релятивистская масса электрона сделается равной $m_n = (n+1)m_0$, а частота обращения по окружности

$$\omega_n = \frac{eH}{m_n c} = \frac{eH}{(n+1)m_0 c} = \frac{\omega}{n+1}.$$

Таким образом, частота обращения уменьшается в $n+1$ раз по сравнению с частотой нерелятивистской частицы, а время обращения во столько же раз увеличивается. Именно по этой причине электрон проходит через ускоряющий промежуток всякий раз, когда электрическое поле находится в фазе ускорения.

Полная энергия электрона после n-кратного прохождения через ускоряющий промежуток определяется формулой $\mathscr{E}_n^2 - (pc)^2 = \mathscr{E}_0^2$, а потому

$$pc = \sqrt{n^2 + 2n}\,\mathscr{E}_0.$$

Отсюда на основании формулы (84.5) заключаем, что с возрастанием n радиусы траекторий электрона в магнитном поле микротрона возрастают как $\sqrt{n^2+2n}$. Последовательные круговые траектории, описываемые электроном в микротроне, схематически показаны на рис. 152.

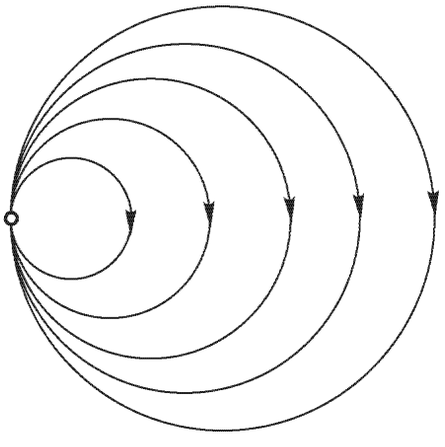


Рис. 152

В микротроне, как и в других ускорителях релятивистских частиц, осуществляется автофазировка. Она приводит к тому, что ускоряется не только резонансный электрон, проходящий в микротроне через ускоряющий резонатор в момент максимума электрического поля, но и ближайшие электроны, совершающие около него малые колебания. Амплитуда этих колебаний определяется областью устойчивости и приводит к небольшим колебаниям энергии электрона вокруг среднего значения, т.е. к нарушению монохроматичности. Но монохроматичность ускоренного пучка остается все же достаточно высокой. В этом отношении микротрон уступает только электростатическому генератору

Ван-де-Граафа. Зато он позволяет при монохроматичности, достаточной для постановки многих опытов, получать мощные пучки электронов такой энергии, которая для электростатических генераторов недостижима.

Большинство микротронов работают на длине волны $\lambda = 10$ см. Напряженность магнитного поля обычно невелика — порядка 1000 Гс. Диаметр наконечников магнита и камеры определяется длиной последней n-й орбиты: $D_n = \lambda n/\pi$. Число орбит обычно составляет 10–20. Микротрон, как и циклотрон, может работать и в непрерывном и в импульсном режимах. При энергии 5 МэВ ток в импульсе составляет 1–2 мА. Средний ток обычно в тысячу раз меньше.

Предельная энергия, достижимая на микротронах, оценивается в 50–100 МэВ. Дальнейшее повышение энергии требует выполнения весьма жестких допусков на магнитное поле для обеспечения устойчивости. Существующие микротроны позволяют ускорять электроны до 30 МэВ. При этом интенсивность ускоренного пучка резко падает с ростом энергии. Так, микротрон на 13 МэВ дает ток в импульсе 100 мА, а микротрон на 30 МэВ — всего лишь 0,05 мА.

Микротроны применяются только для ускорения электронов, так как уже в случае таких легких частиц требуется создание на резонаторах напряжений свыше $0{,}5 \cdot 10^6$ В, что вызывает серьезные технические трудности.

Если поместить у плоского резонатора тяжелую мишень (толщина которой порядка радиационной длины), то при торможении электронов возникнут γ-кванты высокой энергии. Они в свою очередь порождают пары электрон–позитрон. Образовавшиеся позитроны будут ускоряться тем же микротроном наряду с электронами, но двигаться при этом будут в противоположном направлении.

9. Ускорение электронов можно производить вихревым электрическим полем, возникающим при нарастании аксиально-симметричного магнитного поля во времени. Такой процесс осуществляется в *бетатроне.* В этом ускорителе электрон вращается по стационарной орбите, т. е. по окружности постоянного радиуса r, а потому никакого высокочастотного электрического поля и соблюдения синхронизма не требуется.

Найдем необходимое условие существования такой стационарной орбиты. Импульс электрона p возрастает по модулю в соответствии с уравнением $dp/dt = eE$, где e — заряд электрона по модулю, а E — напряженность вихревого электрического поля на стационарной орбите. Она определяется законом индукции

$$2\pi r E = \frac{1}{c}\frac{d\Phi}{dt},$$

а магнитный поток, пронизывающий площадь, ограниченную стационарной орбитой, $\Phi = \pi r^2 \overline{H}$, где $\overline{H}$ — средняя напряженность магнитного поля на этой площади. Таким образом,

$$\frac{dp}{dt} = \frac{e}{2\pi r c}\frac{d\Phi}{dt} = \frac{e}{2c}\frac{d\overline{H}}{dt}.$$

С другой стороны, в силу (84.5) при постоянном радиусе r

$$\frac{dp}{dt} = \frac{e}{c}\frac{dH}{dt},$$

где H — напряженность магнитного поля на стационарной орбите. Путем сравнения последних двух формул и интегрирования получаем

$$H = \frac{\overline{H}}{2}, \tag{84.6}$$

причем мы приняли во внимание, что при $t = 0$ $H = \overline{H} = 0$, что имеет место только при нарастании магнитного поля во времени. Что напряженность магнитного поля должна расти, а не уменьшаться с течением времени, это видно и без всяких вычислений. Ведь ускоритель должен увеличивать энергию электрона. А при увеличении энергии для удержания электрона на прежней орбите требуется более сильное магнитное поле.

Итак, для существования стационарной круговой орбиты электрона необходимо, чтобы напряженность магнитного поля на орбите была вдвое меньше средней напряженности того же поля на площади, ограниченной этой орбитой, причем магнитное поле должно нарастать во времени, начиная от нуля. Этот результат и лежит в основе устройства и действия бетатрона.

В бетатроне электроны разгоняются до ультрарелятивистских энергий, а потому в силу (84.5) окончательная кинетическая энергия ускоренного электрона определяется формулой

$$\mathscr{E}_{\text{кин}} = \mathscr{E}_{\text{полн}} = pc = eHr = \frac{e\overline{H}r}{2}, \tag{84.7}$$

причем здесь H и $\overline{H}$ означают напряженности удерживающего и среднего магнитных полей в конце времени ускорения. Мы видим, что энергия $\mathscr{E}_{\text{кин}}$ определяется лишь значениями H и $\overline{H}$ и радиусом стационарной орбиты, но не зависит от того, сколько оборотов сделал электрон при движении по стационарной орбите. Пусть, например, обмотка электромагнита питается синусоидальным переменным током с частотой $\nu = 50$ Гц, который создает среднее магнитное поле с амплитудой $\overline{H} = 10^4$ Гс. Бетатрон ускоряет электроны только в промежутки нарастания магнитного поля по величине. Примем ориентировочно, что длительность такого промежутка $\tau = T/4 = 1/4\nu \approx 1/200$ с, и допустим, что диаметр стационарной орбиты $D = 1$ м. Тогда максимальная кинетическая энергия электрона будет

$$\mathscr{E}_{\text{кин}} = \frac{1}{2} \cdot 4{,}8 \cdot 10^{-10} \cdot 10^4 \cdot \frac{1}{2} \cdot 10^2 = 1{,}2 \cdot 10^{-4} \text{ эрг} = 75 \text{ МэВ}.$$

На прохождение стационарной орбиты ультрарелятивистский электрон затрачивает время $t = 2\pi r/c = 3{,}14 \cdot 10^{-8}$ с, так что за все время

ускорения он совершает

$$\frac{\tau}{t} = \left(\frac{1}{2}\cdot 10^{-2}\right) : (3{,}14\cdot 10^{-8}) \approx 1{,}6\cdot 10^{5} \text{ оборотов.}$$

Устойчивость движения по стационарной орбите в вертикальном направлении обеспечивается бочкообразной формой магнитных силовых линий между полюсными наконечниками магнита бетатрона. На стационарной орбите центробежная сила $mv^2/r = pv/r$ уравновешивается лоренцевой силой eHv/c. Поэтому для устойчивости движения в радиальном направлении поле H должно убывать с радиусом медленнее, чем $1/r$. Только тогда сила Лоренца будет превосходить центробежную силу и сможет вернуть удалившуюся частицу на стационарную орбиту. Частица же, отклонившаяся в сторону центра, также вернется на стационарную орбиту, так как в этом случае, наоборот, центробежная сила будет превышать лоренцеву.

Бетатроны обычно применяются для ускорения электронов до энергий 1–50 МэВ, причем средний ток не превышает 10^{-2} мкА при 10^9–10^{10} частиц в импульсе. Вначале изготовлялись бетатроны и на более высокие энергии, вплоть до 240 МэВ. Однако это оказалось нецелесообразным, так как в бетатронах необходимо создавать магнитное поле по всей площади, находящейся внутри траектории электрона. Этот недостаток в бетатронах еще более существен, чем в фазотронах, так как среднее магнитное поле в бетатронах должно быть вдвое больше поля, необходимого для удержания электрона на стационарной орбите. Кроме того, при энергиях от 100 МэВ и выше режим ускорения электронов в бетатронах существенно ухудшается из-за электромагнитного излучения.

К недостаткам бетатронов относится трудность и даже практическая невозможность вывода пучка из камеры. Бетатроны часто применяются для получения γ-квантов высоких энергий. В таких случаях ускоренные электроны направляются на специальную мишень, расположенную в камере. Это достигается с помощью обмотки, создающей магнитное поле, нарушающее условие (84.6). В результате радиус орбиты увеличивается или уменьшается, и электроны попадают на мишень, где и генерируются γ-кванты.

10. Импульс частицы связан с ее полной энергией $\mathscr{E}$ соотношением $p = (\mathscr{E}/c^2)v$ или $pc = \mathscr{E}\beta$, где $\beta = v/c$. С другой стороны, $\mathscr{E}^2 - (pc)^2 = \mathscr{E}_0^2$. Из этих двух соотношений получается

$$\beta^2 = \frac{\mathscr{E}^2 - \mathscr{E}_0^2}{\mathscr{E}^2}, \tag{84.8}$$

при $\beta \sim 1$

$$1 - \beta = \frac{\mathscr{E}_0^2}{2\mathscr{E}^2}. \tag{84.8а}$$

Для электрона $\mathscr{E}_0 = 0{,}511$ МэВ. При релятивистской энергии электрона $\mathscr{E} = 10$ МэВ из формулы (84.8) находим $\beta = 1 - 0{,}000013$.

Таким образом, если энергия электрона больше примерно 10 МэВ, то его скорость практически постоянна и не отличается от скорости света c. Поэтому и период обращения электрона по круговой орбите практически постоянен. Это используется в *синхротронах* — циклических кольцевых резонансных ускорителях электронов с орбитой почти постоянного радиуса, в которых частота ускоряющего электрического поля постоянна, а напряженность удерживающего магнитного поля изменяется во времени. В синхротроне магнитное поле надо создавать только вдоль ускоряющего кольца, а не в его середине, что существенно уменьшает массу магнита и уменьшает его стоимость. В синхротроне могут ускоряться только ультрарелятивистские частицы. Это осложняет инжекцию (впуск) частиц в синхротрон. В крупных синхротронах применяется инжекция уже предварительно ускоренных (до энергий 1–50 МэВ) электронов, в меньших — бетатронная инжекция. Ускоритель сначала работает как бетатрон до достижения ультрарелятивистских энергий, а затем переходит на синхротронный режим. Выпуск ускоренных частиц из синхротрона также осложнен из-за постоянства радиуса орбиты. Тем не менее он осуществляется в большинстве современных синхротронов. Часто пучок ускоренных электронов не выпускается из камеры, а направляется на расположенную в ней мишень, где генерируются тормозные γ-кванты, используемые в различных исследованиях. В синхротронах высоких энергий число частиц в импульсе составляет примерно 10^{10}, а число импульсов в секунду — несколько десятков, так что получается средний ток около 0,1 мкА.

Ультрарелятивистские электроны, движущиеся в синхротроне по круговым орбитам (из-за наличия у них нормальных ускорений), являются мощными источниками электромагнитного излучения. Излучаемая энергия за один оборот электрона возрастает пропорционально четвертой степени энергии самого электрона при заданном радиусе (т. е. заданном H) [1]. Потеря энергии на излучение приводит к затуханию колебаний электронов около равновесной орбиты, а квантовый дискретный характер излучения — к их раскачке. Трудности создания мощных ускоряющих устройств, компенсирующих потери на излучение, ограничивают предельно достижимые энергии. В синхротронах достигнуты максимальные энергии 5–10 ГэВ. При меньших энергиях более экономичны бетатроны и микротроны, а при больших — линейные резонансные ускорители. Правда, существуют проекты синхротронов на 100–150 ГэВ.

11. Для ускорения тяжелых частиц (протонов или ионов) до максимальных энергий применяются *синхрофазотроны.* Синхрофазотрон — это циклический резонансный ускоритель тяжелых частиц, в котором меняются во времени как магнитное поле, так и частота ускоряющего

[1] Синхротронное излучение можно уменьшить, взяв больший R (меньшее H). В ЦЕРНе сооружается ускоритель ЛЭП со встречными e^+e^--пучками по 50 ГэВ в каждом (в дальнейшем по 100 ГэВ). Длина его окружности около 30 км. Потери на синхронное излучение достигают десятков мегаэлектронвольт.

электрического поля, и притом так, что радиус равновесной орбиты остается почти постоянным. Изменять частоту электрического поля в синхрофазотроне необходимо потому, что протоны с энергией 1 ГэВ еще не являются достаточно ультрарелятивистскими, так что период обращения их по орбите постоянного радиуса меняется с энергией ($\beta = 0{,}767$ при $\mathscr{E}_{\text{кин}} = 1$ ГэВ). Движение частиц происходит в кольцевой вакуумной камере, помещенной в магнитное поле системы магнитов, расположенных в определенном порядке по кольцу. Магнитное поле синхронно меняется с энергией частиц. В прямолинейных промежутках между магнитами (служащих для размещения ускоряющих электродов, а также устройств ввода и вывода пучка) магнитное поле спадает до нуля. Синхрофазотроны на очень высокие энергии построены по *многоступенчатому принципу*: линейный ускоритель (*инжектор*) впускает частицы в малый синхрофазотрон (*бустер*), где они ускоряются до промежуточной энергии, а затем поступают в большой синхрофазотрон для дальнейшего ускорения. В синхрофазотронах меньших энергий инжекция частиц производится непосредственно из линейного ускорителя.

Интенсивность ускоренных пучков в синхрофазотронах относительно низка, особенно при высоких энергиях. Так, синхрофазотрон в Дубне на 10 ГэВ за импульс дает около 10^{12} протонов (7,5 импульсов в минуту). Он позволяет ускорять не только протоны, но и атомные ядра до энергии 10 ГэВ на один протонный заряд. Например, полностью ионизованный атом изотопа углерода $^{12}_{6}\mathrm{C}$ ускоряется на этом ускорителе до энергии 60 ГэВ, т. е. до 5 ГэВ на нуклон (интенсивность 10^4 ядер $^{12}_{6}\mathrm{C}$ в импульсе). В 1967 г. в Серпухове был запущен синхрофазотрон на 76 ГэВ. Радиус ускорительного кольца в нем 236,14 м, средний ток $2 \cdot 10^{-12}$ мкА (10^{12} частиц в импульсе, 8 импульсов в минуту). Прирост энергии за один оборот 190 кэВ, так что за полный цикл ускорения частица совершает в ускорителе $(76 \cdot 10^9) : (190 \cdot 10^3) = 400\,000$ оборотов. До 1972 г. Серпуховский синхрофазотрон был самым большим ускорителем протонов в мире. К 1980 г. максимальная энергия, достижимая на синхрофазотроне, достигла 500 ГэВ (Батавия, США и ЦЕРН), проектируются синхрофазотроны на несколько тысяч гигаэлектронвольт. Предельная достижимая энергия ограничена в первую очередь технико-экономическими возможностями (размерами установки и ее стоимостью). Минимальная энергия, для получения которой применяются синхрофазотроны, равна около 1 ГэВ, для меньших энергий целесообразнее применять фазотроны.

На основе ускорителя на 500 ГэВ (ЦЕРН) были созданы встречные $\bar{\mathrm{p}}\mathrm{p}$-пучки по 200 и 310 ГэВ, на которых удалось открыть $\mathrm{W}^{\pm}$-, Z^0-бозоны. В 1985 г. в Батавии запущен ускоритель (со сверхпроводящими магнитами) на энергию 1000 ГэВ и вводятся в строй встречные пучки по 800 ГэВ. В Серпухове сооружается ускорительно-накопительный комплекс, основу которого составит протонный ускоритель со сверхпроводящими магнитами на 3–3,5 ТэВ с организацией встречных пучков $\bar{\mathrm{p}}\mathrm{p}$ и pp. Длина окружности основного ускорителя равна примерно

20 км, а в качестве бустера будет использоваться синхрофазотрон на 76 ГэВ. В ФРГ сооружается ускоритель HERA со встречными ер-пучками ($\mathscr{E}_e = 20$–30 ГэВ, $\mathscr{E}_p = 800$ ГэВ). Создается проект протонного ускорителя (США) на 20 ТэВ со встречными пучками.

12. Существенное значение для повышения энергии ускоряемых частиц в синхрофазотроне и других ускорителях на высокие энергии имело применение *жесткой* (иначе называемой *сильной*) *фокусировки*, предложенной в 1950 г. Н. Кристофилосом (р. 1917) и независимо от него в 1952 г. Э. Курантом, Х. Снайдером (р. 1913) и М. Ливингстоном (р. 1917). Идея жесткой фокусировки была уже изложена нами в т. IV, § 12. Она основана на том, что две линзы — собирающая и рассеивающая — с одинаковыми фокусными расстояниями всегда образуют собирающую систему, если первая линза рассеивающая. Если же первая линза собирающая, то это свойство сохраняется, когда расстояние между линзами меньше фокусного расстояния одной линзы. Результат был получен нами для тонких линз, но он (в несущественно уточненной форме) остается верным и для толстых линз.

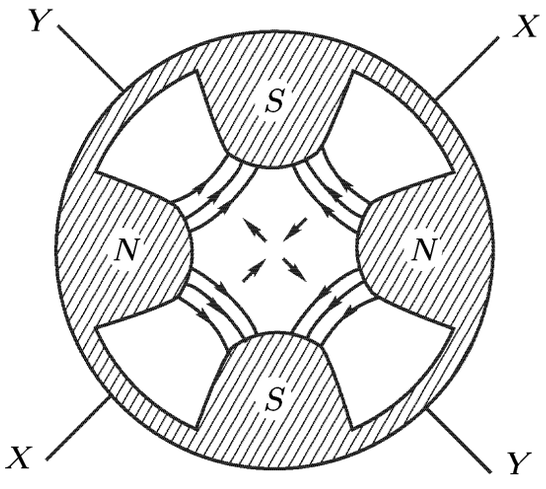


Рис. 153

В синхрофазотроне жесткая фокусировка осуществляется магнитными *квадрупольными линзами*. Одна из таких линз схематически изображена на рис. 153. Четыре магнитных наконечника обращены друг к другу противоположными полюсами, так что в центре квадрупольной линзы магнитное поле равно нулю. На рисунке показаны магнитные силовые линии. Предполагается, что положительная частица движется в направлении к читателю. Маленькие стрелки показывают направление лоренцевых сил, действующих на частицы со стороны магнитного поля квадрупольной линзы. Из рисунка видно, что в направлении оси X лоренцева сила стремится приблизить частицу к оси линзы (т. е. фокусировать), а в направлении оси Y — отдалить (т. е. дефокусировать).

Таким образом, квадрупольная магнитная линза действует как совокупность двух магнитных линз, одна из которых фокусирует частицы в направлении оси X, а другая — в направлении оси Y. Возьмем теперь две квадрупольных линзы и поставим их друг за другом, повернув одну из них относительно другой на 90°. Тогда в направлении оси X одна линза будет фокусировать, а другая — дефокусировать. В целом обе линзы вместе в направлении оси X будут фокусировать. То же самое относится и к направлению оси Y. Отсюда следует, что рассматриваемая система двух квадрупольных линз будет фокусировать и в любом поперечном направлении.

В синхрофазотронах с жесткой фокусировкой, разумеется, создается и изменяющееся во времени вертикальное магнитное поле, удерживающее частицы на равновесной круговой орбите. Но на это поле накладывается еще магнитное поле квадрупольных линз. Последние устанавливаются одна за другой таким образом, что линзы, скажем, с нечетными номерами фокусируют в вертикальном направлении, а с четными — дефокусируют в том же направлении. Наоборот, в горизонтальном направлении линзы с нечетными номерами дефокусируют, а с четными — фокусируют. Поэтому в каком бы направлении ни отклонилась частица от равновесной траектории, она последовательно встретит на своем пути пары квадрупольных линз, причем каждая из таких пар будет фокусировать. В результате пучок частиц сужается и можно получать ускоренные пучки большой интенсивности. Частицы в пучке совершают малые колебания около равновесной траектории, частота которых в несколько раз или десятков раз больше частоты обращения их по равновесной орбите. По этой причине фокусировка с помощью квадрупольных магнитных линз и называется жесткой или сильной. Жесткая фокусировка применяется не только в синхрофазотронах, но и во всех крупных ускорителях, в частности в Серпуховском синхрофазотроне, некоторые параметры которого были приведены выше.

Жесткая фокусировка позволяет уменьшить поперечные размеры камеры и, следовательно, массу магнита. Например, масса магнита ускорителя в Дубне на 10 ГэВ без жесткой фокусировки равна 36 000 тонн. Магнит же ускорителя в Серпухове на 76 ГэВ с применением жесткой фокусировки имеет меньшую массу — 20 000 тонн. Уменьшением массы магнита, а с ним и индуктивности можно добиться более частого повторения импульсов.

13. Циклотрон (см. п. 6), оставляя магнитное поле постоянным во времени, можно модернизировать так, чтобы он был пригоден и для ускорения *релятивистских частиц.* Для этого надо применять постоянное магнитное поле $H(r)$, зависящее определенным образом от радиуса r. Именно, согласно (84.3), магнитное поле должно меняться с радиусом так же, как и релятивистская масса частицы, т. е. $H = H_0/\sqrt{1-\beta^2}$. Тогда частота обращения Ω частицы будет оставаться одной и той же, равной частоте ускоряющего поля. Так как $\beta = v/c = \Omega r/c$, то указанному требованию удовлетворяет магнитное поле вида

$$H = \frac{H_0}{\sqrt{1-(\Omega r/c)^2}} \tag{84.9}$$

при постоянной Ω. Ускорители, использующие этот принцип, называются *изохронными циклотронами.* Они работают в непрерывном режиме и дают возможность получать столь же сильные токи (0,1–1 мА), что и обыкновенные циклотроны.

Изохронные циклотроны имеют огромные магниты, а потому довольно дороги. Казалось бы, что такие ускорители работать не могут, так как в них магнитное поле возрастает с радиусом (см. п. 6). Однако возникающую из-за этого вертикальную неустойчивость удается

компенсировать азимутальной неоднородностью магнитного поля довольно сложной формы. Правда, расчет показывает, что это можно сделать до энергий не выше 1 ГэВ. Большинство изохронных циклотронов проектируется и строится для ускорения протонов до 50–100 МэВ. Изохронные циклотроны применяются для изучения редких процессов в реакциях с протонами с энергией 50–1000 МэВ. Серьезными конкурентами их являются линейные ускорители тяжелых частиц.

14. Обычно ускоренный пучок частиц направляется из ускорителя на неподвижную мишень. При столкновении с частицей-мишенью общий импульс обеих частиц не меняется. Поэтому не меняется и та часть энергии, которая связана с движением центра масс как целого. Эта часть энергии не участвует ни в каких внутренних превращениях сталкивающихся частиц, и в этом отношении она является бесполезной. В системе центра масс эта бесполезная кинетическая энергия равна нулю и вся энергия является полезной, т. е. может быть использована для внутренних превращений сталкивающихся частиц.

В т. IV (§ 111, задача 1) был рассмотрен простейший случай столкновения двух частиц одинаковых масс, движущихся навстречу друг другу. Был поставлен и решен вопрос, какой энергией $\mathscr{E}_{\text{лаб}}$ должна обладать одна из этих частиц в лабораторной системе отсчета, в которой вторая частица (мишень) покоится, чтобы при столкновении получился такой же полезный эффект (в смысле возможности внутренних превращений частиц), что и в системе центра масс. Будем понимать под $\mathscr{E}_{\text{лаб}}$ полную (релятивистскую) энергию одной движущейся частицы, а под $\mathscr{E}_0$ — ее энергию покоя. В указанной задаче была получена формула

$$\mathscr{E}_{\text{лаб}} = 2\frac{\mathscr{E}_{\text{цм}}^2}{\mathscr{E}_0} - \mathscr{E}_0, \tag{84.10}$$

где $\mathscr{E}_{\text{цм}}$ — релятивистская энергия одной частицы в системе центра масс. Соответствующая кинетическая энергия частицы

$$\mathscr{E}_{\text{лаб}}^{\text{кин}} = \mathscr{E}_{\text{лаб}} - \mathscr{E}_0 = 2\left(\frac{\mathscr{E}_{\text{цм}}^2}{\mathscr{E}_0} - \mathscr{E}_0\right). \tag{84.11}$$

В нерелятивистском случае $\mathscr{E}_{\text{цм}} = \mathscr{E}_0 + \mathscr{E}_{\text{цм}}^{\text{кин}}$. Подставляя это значение в предыдущую формулу и пренебрегая квадратами $\mathscr{E}_{\text{цм}}^{\text{кин}}$, получим

$$\mathscr{E}_{\text{лаб}}^{\text{кин}} = 4\mathscr{E}_{\text{цм}}^{\text{кин}} = 2(2\mathscr{E}_{\text{цм}}^{\text{кин}}).$$

Таким образом, при неподвижной мишени для получения того же эффекта требуется вдвое большая кинетическая энергия, чем в системе центра масс (так как в этой системе полная кинетическая энергия обеих частиц равна $2\mathscr{E}_{\text{цм}}^{\text{кин}}$). Этот результат был уже получен в т. I, § 26.

Более интересен ультрарелятивистский случай. В этом случае в формуле (84.11) собственной энергией частицы можно пренебречь. Тогда связь между кинетическими энергиями частицы в лабораторной

системе и системе центра масс становится не линейной, а квадратичной:

$$\mathscr{E}_{\text{лаб}}^{\text{кин}} = \frac{2(\mathscr{E}_{\text{цм}}^{\text{кин}})^2}{\mathscr{E}_0}. \tag{84.12}$$

Здесь можно получить большой выигрыш энергии, осуществляя столкновение встречных частиц. Например, для протона ($\mathscr{E}_0 = 0{,}938$ ГэВ) при энергии $\mathscr{E}_{\text{цм}}^{\text{кин}} = 5$ ГэВ получается $\mathscr{E}_{\text{лаб}}^{\text{кин}} = 53$ ГэВ. *Ускоритель на встречных пучках* при энергии 5 ГэВ дает такой же эффект, как и ускоритель с неподвижной мишенью на энергию 53 ГэВ. Еще больший эффект получается в случае легких частиц, например электронов. Для электронов $\mathscr{E}_0 = 0{,}511 \cdot 10^{-3}$ ГэВ, так что при том же значении $\mathscr{E}_{\text{цм}}^{\text{кин}}$ получается $\mathscr{E}_{\text{лаб}}^{\text{кин}} \approx 10^5$ ГэВ, что уже недостижимо ни на каких реальных ускорителях с неподвижной мишенью.

Изложенного достаточно, чтобы уяснить принцип действия систем на встречных пучках. Наибольшее распространение получили устройства, в которых применяются электрон-электронные (e^-e^-), электрон–позитронные (e^-e^+) и протон-протонные (pp) пучки. Существенным недостатком систем на встречных пучках является малая интенсивность пучков. Для увеличения интенсивности пучков до процесса соударения заряженных частиц производится их накопление в специальных *накопительных кольцах*, чтобы циркулирующий в них ток был не меньше десятков ампер. Накопительное кольцо — это кольцевая вакуумная камера, помещенная в магнитное поле. Ускоренные заряженные частицы поступают в нее из синхротрона, синхрофазотрона или линейного ускорителя. Магнитное поле, как правило, создается секторами, разделенными прямолинейными промежутками (в которых магнитного поля нет) для областей пересечения пучков частиц и для компенсации синхротронного излучения и других целей.

Если заряды сталкивающихся частиц одинаковы, то система должна содержать два накопительных кольца, в которых частицы движутся в противоположных направлениях. Если же эти заряды противоположны, то достаточно только одного накопительного кольца, так как в одном и том же магнитном поле обеспечивается движение частиц с разными знаками заряда в противоположных направлениях и их ускорение в одном и том же электрическом поле. Встречные пучки в настоящее время могут быть использованы только для стабильных частиц. Ускорители с фиксированной мишенью являются источниками различных вторичных частиц: π-, K-мезонов, Λ-, Σ-гиперонов, нейтрино, μ-мезонов и т. д.

§ 85. Источники нейтронов и других нейтральных частиц

1. Нейтроны, как и все электрически нейтральные частицы, нельзя ускорять и фокусировать электромагнитными полями. Такие частицы образуются только в результате ядерных реакций. Ниже описываются принципы действия некоторых нейтронных источников и приводятся

для общей ориентировки некоторые их характеристики. Источники нейтронов можно разделить на три группы: 1) источники, в которых нейтроны создаются радиоактивными излучениями; 2) источники, в которых они создаются частицами, вылетающими из ускорителей; 3) ядерные реакторы. Во всех типах источников нейтроны, как правило, получаются быстрыми. Взаимодействие нейтронов с ядрами особенно интенсивно в случае медленных нейтронов (сечение взаимодействия пропорционально $1/v$). Полученные в источниках нейтроны используются либо сразу, либо после предварительного замедления. Особое значение имеют источники тепловых нейтронов (с энергиями порядка 1/40 эВ).

2. В так называемых (α, n)-источниках α-частицы от радиоактивного препарата облучают мишень, в результате чего и возникают нейтроны. Помимо энергии получающихся нейтронов, источник характеризуется *выходом* нейтронов и *интенсивностью*. Выходом называется среднее число нейтронов, возникающих при одном акте взаимодействия α-частицы с ядром мишени (или при активности 1 Ки, т. е. при $3{,}7 \cdot 10^{10}$ α-распадов). Интенсивность источника характеризуется числом нейтронов, возникающих при тех же условиях в одну секунду.

В самых первых источниках нейтронов использовалась реакция

$$^{9}_{4}\mathrm{Be} + \alpha \to {}^{12}_{6}\mathrm{C} + \mathrm{n} + 5{,}6 \text{ МэВ}, \qquad (85.1)$$

на которой Боте и Беккером (см. § 92) впервые наблюдалось нейтронное излучение. Дальнейшие опыты показали, что нейтроны образуются также при бомбардировке α-частицами изотопов элементов Li, B, N, F, Na, Mg, Al, ... Однако по интенсивности излучения реакция с бериллием значительно превосходит все остальные, поэтому эта реакция в течение долгого времени и использовалась при изготовлении источников нейтронов. Источник представляет собой герметическую ампулу, в которой помещена смесь порошка бериллия с α-активным препаратом, например полонием-210, испускающим α-частицы с энергией 5,298 МэВ. Альфа-частицы не могут проходить через стенки ампулы из-за их ничтожных пробегов. Нейтроны же, образующиеся в результате реакции (85.1), свободно выходят. Энергетический спектр нейтронов полоний-бериллиевого источника непрерывен и простирается примерно от 0,5 до 10 МэВ, интенсивность создаваемого нейтронного излучения около $0{,}8 \cdot 10^{-4}$ нейтронов в секунду на одну α-частицу, испускаемую полонием ($3 \cdot 10^{6}$ нейтронов в секунду на 1 Ки полония). Достоинством полоний-бериллиевого источника является незначительный фон нежелательного γ-излучения, недостатком — малый период полураспада (140 дней). В качестве источника α-частиц наряду с полонием применяются также радий, радон, плутоний. Радий-бериллиевый источник характеризуется практически неограниченным сроком действия (период полураспада 1600 лет) и в шесть раз большей интенсивностью нейтронного излучения, но у него очень велик фон мешающего γ-излучения. Плутоний-бериллиевый источник характеризуется малым фоном γ-излучения, большим сроком службы (период полураспада $2{,}3 \cdot 10^{4}$ лет)

и дает интенсивность около $0{,}46 \cdot 10^{-4}$ нейтронов в секунду на одну α-частицу, испускаемую плутонием.

Альфа-частицы испускаются не только при радиоактивном α-распаде материнского вещества, но и при α-распаде всех его α-активных дочерних продуктов. В равновесном состоянии α-активность всех α-излучателей одинакова. Но их способность к образованию нейтронов не одинакова. Она тем выше, чем меньше время жизни короткоживущих продуктов распада, так как такие продукты излучают более энергичные α-частицы. Поэтому если источник содержал вначале, например, чистый радий, то по мере накопления α-радиоактивных продуктов распада его способность излучать нейтроны возрастает в несколько раз и при насыщении стремится к постоянному пределу. Основной недостаток всех источников с использованием α-частиц — большой разброс по энергиям вылетающих нейтронов.

Относительно монохроматические нейтроны с энергиями 0,1–1 МэВ получают облучением γ-квантами ядер дейтерия и бериллия. При этом происходят реакции

$$\gamma + \mathrm{d} \to \mathrm{p} + \mathrm{n}, \quad {}^{9}_{4}\mathrm{Be} + \gamma \to {}^{8}_{4}\mathrm{Be} + \mathrm{n} \to \alpha + \alpha + \mathrm{n}, \tag{85.2}$$

идущие с поглощением энергии. Гамма-кванты создаются каким-либо радиоактивным ядром (${}^{24}_{11}\mathrm{Na}$, ${}^{72}_{31}\mathrm{Ga}$, ${}^{124}_{51}\mathrm{Sb}$ и пр.). Энергия этих γ-квантов не превышает нескольких мегаэлектронвольт. Поэтому в качестве мишеней и используются только легкие ядра d и ${}^{9}_{4}\mathrm{Be}$, так как у них энергия отделения нейтрона аномально низка (2,23 и 1,67 МэВ соответственно). Интенсивность γ-нейтронных источников примерно на два порядка ниже, чем α-нейтронных. Зато γ-нейтронные источники дают более монохроматические нейтроны. Действительно, энергия γ-кванта $\mathscr{E}_\gamma = p_\gamma c$, а кинетическая энергия α-частицы $\mathscr{E}_\alpha = p_\alpha^2/2m_\alpha$. Если эти энергии одинаковы, то

$$p_\gamma c = \frac{p_\alpha^2}{2m_\alpha},$$

откуда

$$\frac{p_\gamma}{p_\alpha} = \frac{p_\alpha}{2m_\alpha c} = \frac{v_\alpha}{2c},$$

где v_α — скорость α-частицы. Таким образом, при энергии несколько мегаэлектронвольт импульс γ-кванта примерно на два порядка меньше импульса α-частицы той же энергии. С этим и связана большая монохроматичность γ-радиоактивных источников нейтронов по сравнению с α-радиоактивными источниками. Комбинацией различных γ-излучателей либо с дейтерием, либо с бериллием и получают сравнительно монохроматические нейтроны с различными энергиями от 0,12 до 0,87 МэВ. Существенным недостатком этих источников наряду с низкой интенсивностью является высокий фон γ-излучения и малое время жизни (период полураспада от нескольких часов до нескольких десятков часов).

Альфа- и гамма-нейтронные источники находят применения в прикладных исследованиях (особенно в полевых условиях) как небольшие

лабораторные источники нейтронов, а также для калибровки нейтронных детекторов.

3. Получение нейтронов с помощью ускорителей отличается от способа, изложенного выше, в том отношении, что вместо α-частиц или γ-излучения естественных радиоактивных излучений применяются ускоренные заряженные частицы, при облучении которыми различных мишеней получаются нейтроны. Таким путем можно получать относительно моноэнергетические нейтроны самых разнообразных энергий. Дело в том, что при фиксированной энергии частиц в пучке и конкретном нейтронном канале реакции [1]) энергия получающегося нейтрона однозначно определяется углом его вылета и энергией реакции. Степень немоноэнергетичности образующихся нейтронов зависит от немонохроматичности исходного пучка заряженных частиц, их замедления в мишени и существования нескольких нейтронных каналов реакции. Для ослабления влияния этих факторов применяют тонкие мишени, в которых используемый нейтронный канал является единственным или по меньшей мере доминирующим.

В качестве бомбардирующих заряженных частиц обычно применяют протоны, дейтроны, α-частицы и пр. Например, для получения монохроматических нейтронов низких энергий используются реакции (p, n), т. е. такие реакции, в которых ядро-мишень бомбардируется протонами, а в результате реакции возникают нейтрон и другое ядро. Наиболее часто применяется реакция

$$^{7}_{3}\mathrm{Li} + \mathrm{p} \to {}^{7}_{4}\mathrm{Be} + \mathrm{n} - 1{,}6 \text{ МэВ}, \tag{85.3}$$

с помощью которой получаются моноэнергетические нейтроны с энергиями от 30 до 500 кэВ (в зависимости от энергии протонов). Удешевление производства трития сделало доступным использование реакции

$$\mathrm{p} + \mathrm{t} \to {}^{3}_{2}\mathrm{He} + \mathrm{n} - 0{,}735 \text{ МэВ}, \tag{85.4}$$

достоинством которой является не столько низкий порог, сколько отсутствие возбужденных состояний у ядра $^{3}_{2}$He. С помощью реакции (85.4) получают моноэнергетические нейтроны с энергиями от 0,06 до 3 МэВ. Недостатком реакции (85.4) является фон жесткого γ-излучения, возникающий за счет параллельно идущей реакции

$$\mathrm{p} + \mathrm{t} \to \alpha + \gamma. \tag{85.5}$$

Ограничиваясь приведенными примерами, заметим только, что применяется много реакций, в которых, в частности, получаются нейтроны относительно высоких энергий.

4. Наиболее мощным источником нейтронов является ядерный реактор, принцип действия которого описывается в § 95. В современных исследовательских реакторах плотность потока нейтронов в активной зоне и замедлителе ориентировочно составляет 10^{15} с$^{-1} \cdot$ см$^{-2}$. Спектр реакторных нейтронов, и в этом его недостаток, не моноэнергетичен,

[1]) Определение канала реакции дается в § 87, п. 3, а порога реакции — в § 88, п. 2.

а заполняет широкую непрерывную область энергий. Но суммарная интенсивность нейтронного потока настолько велика, что из него можно выделять сравнительно мощные пучки нейтронов с высокой степенью моноэнергетичности.

Мощным источником нейтронов может служить интенсивный протонный или дейтронный ускоритель, пучок которого направляется на мишень из тяжелых элементов, в которой происходит деление ядер.

Источниками нейтронов являются также и некоторые термоядерные реакции (см. § 98). Об использовании нейтронов в ядерной энергетике подробно говорится в § 95, 98.

5. Существенным недостатком всех источников дейтронов является сплошной спектр или низкая степень моноэнергетичности. Одна из причин этого — рассеяние нейтронов в самом источнике (а при высоких энергиях и обилие нейтронных каналов). Между тем для исследования взаимодействия нейтронов с ядрами крайне важно иметь нейтронные пучки высокой степени моноэнергетичности, чтобы, например, отделить друг от друга узкие и частые резонансы в сечениях взаимодействия нейтронов с ядрами. Для выделения из непрерывного спектра источников монохроматических пучков применяются разные методы, основанные на том, что нейтроны различных энергий обладают и различными скоростями.

Один из таких методов применяется тогда, когда источник немоноэнергетических нейтронов импульсный. Нейтроны от такого источника летят в трубе длиной в сотни метров. За время полета нейтронный сгусток разделяется по скоростям. В конце трубы ставится заслонка (прерыватель), синхронно открывающаяся лишь в моменты пролета нейтронов определенной скорости. В результате из трубы выходят почти моноэнергетические нейтроны, энергия которых заключена в узких пределах.

Другой монохроматор нейтронов, в принципе, действует так же, как и монохроматор для атомных пучков, описанный в § 18 (п. 8, рис. 37). Для него вращающиеся диски должны быть изготовлены из материала, непроницаемого для нейтронов (например, из кадмия, задерживающего тепловые нейтроны, или из специальных сплавов для задержания нейтронов, имеющих другие скорости). Как и предыдущий монохроматор, этот монохроматор является механическим. Механические монохроматоры эффективны для получения нейтронов с энергиями от тысячных долей до нескольких электронвольт.

6. В заключение коротко остановимся на получении других нейтральных частиц. К ним относятся прежде всего нейтральные частицы, участвующие в сильных взаимодействиях: π^0, η, K^0, $\overline{K}^0$, Λ, Σ^0, $\overline{n}$, $\overline{\Lambda}$, $\overline{\Sigma}^0$ и пр., а также нейтрино и антинейтрино всех сортов: ν_e, $\overline{\nu}_e$, ν_μ, $\overline{\nu}_\mu$, ν_τ, $\overline{\nu}_\tau$, которые участвуют только в слабых взаимодействиях. Сильновзаимодействующие частицы в заметных количествах возникают при бомбардировке мишени пучком заряженных частиц из ускорителя очень высокой энергии. Из-за релятивистских эффектов подавляющая

часть всех возникающих частиц летит вперед. Пучок заряженных частиц можно отклонить сильным магнитным полем и таким образом выделить пучок, состоящий из разных нейтральных частиц различной энергии. При энергиях первичного пучка до 700–800 МэВ в основном образуются нейтроны. При энергиях выше нескольких гигаэлектронвольт появляются нейтральные каоны и нейтральные гипероны и т. д. Нейтральные пионы образовать пучка не могут из-за слишком короткого времени жизни ($0{,}75 \cdot 10^{-16}$ с).

Нейтрино и антинейтрино подвержены только слабым взаимодействиям, а потому при столкновениях любых частиц с любыми мишенями рождаются в ничтожных количествах. Эти частицы рождаются при распадах. При β^--распаде рождается электронное антинейтрино $\overline{\nu}_e$, при β^+-распаде — электронное нейтрино ν_e. Мощным источником электронных антинейтрино является ядерный реактор, электронного нейтрино — Солнце. Мюонные нейтрино ν_μ и антинейтрино $\overline{\nu}_\mu$ возникают в современных протонных ускорителях: при столкновении первичного пучка с мишенью сначала возникают заряженные пионы, а затем идут следующие превращения:

$$\pi^+ \to \mu^+ + \nu_\mu, \quad \pi^- \to \mu^- + \overline{\nu}_\mu. \tag{85.6}$$

Из-за закона сохранения импульса при релятивистских скоростях все продукты распада в виде пучков летят в основном вперед. Если на пути пучков поставить достаточно толстую бетонную или железную стену, то все частицы будут поглощены, за исключением ν_μ и $\overline{\nu}_\mu$, которые беспрепятственно пройдут через нее (см. также § 74).

§ 86. Детекторы частиц

1. *Детекторами* называются приборы, служащие для регистрации частиц. Они разделяются на *счетчики* или *электронные детекторы*, вырабатывающие электрический импульс, когда в объем детектора попадает ионизующая заряженная частица, и *трековые детекторы*, позволяющие не только зарегистрировать факт и момент прохождения заряженной частицы, но и оставить след (или трек) частицы, воспроизводящий ее траекторию. Это дает более полную информацию о движении частицы, процессах столкновения ее с другими частицами и пр. Созданы приборы с использованием электронных детекторов, позволяющие определять траектории большого числа заряженных частиц.

Незаряженные частицы сами ионизацию не производят. Но в этом случае ионизация, а с ней и соответствующие треки создаются *вторичными заряженными частицами*, возникающими при прохождении нейтральных частиц через вещество. Так действуют, например, γ-кванты, вырывающие электроны из атомов и молекул или стенок счетчика, а также нейтроны, вызывающие ядерные превращения с образованием заряженных частиц.

Счетчики регистрируют прохождение частицы через определенное место пространства с погрешностью до 1–0,1 см, а время прохождения — с погрешностью до 10^4–10^9 с. В некоторых счетчиках определяются и другие характеристики частицы: энергия, заряд, масса, скорость и пр.

К счетчикам относятся *импульсные ионизационные камеры, пропорциональные счетчики, счетчики Гейгера–Мюллера, сцинтилляционные, черенковские* и *полупроводниковые счетчики.*

В трековых детекторах след, оставленный заряженной частицей, тем или иным способом фиксируется. Еще совсем недавно (лет 10–20 тому назад) единственным способом регистрации треков было их фотографирование, обычно в двух проекциях. Обработка таких фотографий — весьма трудоемкий процесс. Сейчас все большее распространение получает так называемый безфильмовый метод, не использующий фотографию. В этом методе параметры трека непосредственно передаются на ЭВМ для последующей полной обработки. Это на несколько порядков сократило время обработки и увеличило объем материала, который можно обработать.

К трековым детекторам относятся *камеры Вильсона, пузырьковые камеры, толстослойные фотоэмульсии, широкозазорные искровые* и *стримерные, пропорциональные* и *дрейфовые камеры.*

Искровая и пропорциональная камеры действуют как совокупности большого числа мелких счетчиков. Эти камеры напоминают счетчики в том отношении, что информация в них выдается мгновенно, без последующей обработки. Но они напоминают и трековый детектор, так как по действию многих счетчиков можно установить и трек частицы.

2. Основные характеристики детекторов.

1) *Эффективность* — отношение числа частиц, зарегистрированных детектором, к общему числу прошедших через него частиц (в долях от единицы или в процентах).

2) *Пространственное разрешение* — погрешность, с которой детектор может фиксировать положение частицы в пространстве.

3) *Временное разрешение* (разрешающее время) — минимальный промежуток времени между прохождениями двух частиц через детектор, когда сигналы от них еще не накладываются друг на друга. При меньших временах получится наложение сигналов, и две частицы будут регистрироваться детектором как одна частица.

4) *Время восстановления* (мертвое время) — время, за которое детектор, зарегистрировавший одну частицу, успевает вернуться в исходное состояние, чтобы быть готовым для регистрации следующей частицы. Частицы, прошедшие через детектор за это время, не регистрируются. Время восстановления является мерой инерционности детектора. Оно ограничивает максимальную интенсивность пучка частиц, при которой частицы еще могут регистрироваться детектором.

В табл. 14 приведены ориентировочные характеристики различных детекторов.

3. Наиболее обширную группу электронных детекторов составляют *ионизационные детекторы*, основанные на ионизации молекул

Таблица 14

Характеристики различных детекторов

Детектор	Пространственное разрешение, см	Временное разрешение, с	Время восстановления, с
Ионизационная камера	~ 1	10^{-6}	10^{-4}
Пропорциональный счетчик	~ 1	10^{-7}	10^{-5}
Счетчик Гейгера–Мюллера	~ 1	$>10^{-6}$	10^{-4}
Сцинтилляционный счетчик	~ 1	10^{-9}	10^{-8}
Полупроводниковый детектор	< 1	10^{-9}	10^{-9}
Фотоядерные эмульсии	10^{-4}	—	—
Камера Вильсона	10^{-1}	10^{-1}	10^{-2}
Пузырьковая камера	10^{-2}	10^{-3}	1
Искровая камера	10^{-2}	10^{-6}	10^{-3}
Пропорциональная камера	10^{-2}	10^{-7}	10^{-5}
Стримерная камера	10^{-2}	10^{-6}	10^{-1}

и атомов, производимой первичными или вторичными заряженными частицами. Одним из старейших детекторов этой группы является *ионизационная камера*, применявшаяся еще в первых опытах Резерфорда. Ионизационная камера представляет собой электрический конденсатор (плоский, цилиндрический или сферический), заполненный газом, между электродами которого приложена разность потенциалов.

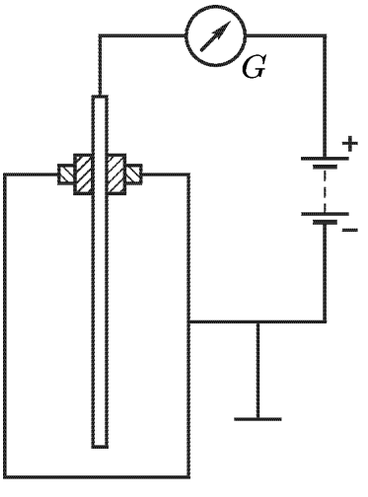


Рис. 154

На рис. 154 приведена схема ионизационной камеры с цилиндрическим конденсатором. Ионизующая частица, попадая в пространство между электродами, образует на своем пути электроны и ионы, которые перемещаются в электрическом поле и собираются на электродах. В цепи камеры появляется электрический ток, регистрируемый гальванометром G. Напряжение в камере подбирается таким (100–1000 В), чтобы все образовавшиеся ионы доходили до электродов, не успев рекомбинировать, но при этом не разгонялись бы настолько сильно, чтобы производить вторичную ионизацию. Поэтому в камере измеряется полная ионизация, производимая частицей, если только ее пробег целиком умещается в камере. Полная же ионизация согласно § 80 определяет и энергию частицы. Ионизационные камеры обычных типов пригодны лишь для регистрации короткопробежных частиц, т. е. тяжелых нерелятивистских частиц, так как пробеги электронов и релятивистских частиц в камере не умещаются.

Ионизационные камеры бывают *токовыми* и *импульсными*. Токовые ионизационные камеры дают сведения об общем количестве ионов, образовавшихся в 1 с. Соответствующие токи обычно малы (10^{-10}–10^{-15} А) и требуют усиления для регистрации. В импульсных ионизационных камерах регистрируются и измеряются импульсы напряжения на высокоомном сопротивлении R, которое должно быть включено в цепь вместо гальванометра G (рис. 154). Это напряжение появляется при прохождении через R ионизационного тока, вызванного движением частицы. Амплитуда и длительность импульса зависят от времени релаксации $\tau = RC$, где C — емкость конденсатора. Время релаксации τ должно быть велико по сравнению с временем собирания образовавшихся электронов на электродах, иначе электронный импульс рассосется уже во время его образования. Размеры ионизационных камер обычно порядка нескольких сантиметров. Скорость электронов при ионизации атома в тысячи раз превосходит скорость ионов. По этой причине электроны собираются на электродах примерно во столько же раз быстрее, чем ионы. Подбором сопротивления R можно добиться того, чтобы импульсы ионизационной камеры соответствовали бы сбору только электронов, а не ионов. Таким путем удается уменьшить длительность импульса до 1 мкс.

Рассмотрим в качестве примера прохождение α-частицы с энергией 5 МэВ через импульсную ионизационную камеру, наполненную аргоном. Средняя энергия ионизации атома аргона 25 эВ. Поэтому до замедления α-частица производит $5 \cdot 10^6 : 25 = 2 \cdot 10^5$ пар ионов. В результате через камеру проходит количество электричества $Q = 2 \cdot 10^5 e = 2 \cdot 10^5 \cdot 1{,}6 \cdot 10^{-19} = 3{,}2 \cdot 10^{-14}$ Кл. Если емкость конденсатора $C = 10^{-11}$ Ф, то ему соответствует импульс напряжения $V = Q/C = 3{,}2 \cdot 10^{-3}$ В. Так как сопротивление R высокоомное, то этот импульс напряжения почти весь приходится на R. Это весьма слабый импульс, поэтому он подается на вход усилителя. Наполнение камеры другим газом существенно ничего не меняет, так как средняя энергия, которая идет на образование одной пары ионов, для всех газов одного порядка (25–35 эВ). С другой стороны, увеличение импульса напряжения V за счет уменьшения емкости конденсатора C может производиться также в ограниченных пределах, так как время релаксации $\tau = RC$, как мы видели, нельзя брать меньше определенной величины. Для составления ориентировочного представления о порядках встречающихся величин заметим, что при напряженности электрического поля 500 В/см в воздухе при нормальных условиях скорость дрейфа электронов равна примерно 10^6, а ионов — только 10^3 см/с. При размерах камеры порядка нескольких сантиметров время собирания на электродах для электронов составляет 10^{-6}, а для ионов 10^{-3} с.

Ионизационные камеры применяются до сих пор, в частности в дозиметрии. Они просты, характеризуются высокой эффективностью регистрации, позволяют, как мы видели, оценивать энергию частицы и т. д. Однако амплитуда электрического сигнала в ионизационной

камере мала, требует усиления и поэтому делает аппаратуру чувствительной к помехам и шумам [1].

4. Недостатки ионизационной камеры в значительной степени устранены в *пропорциональном счетчике.* Это разновидность газонаполненного счетчика, в котором обычно используются коаксиальные электроды: катодом служит цилиндр, а анодом тонкая нить (толщиной 10–100 мкм), натянутая по оси цилиндра. Пропорциональный счетчик, как правило, наполняют инертным газом с небольшой примесью многоатомных газов. При попадании в счетчик ионизующей частицы большая часть образовавшихся электронов дрейфует в электрическом поле к аноду без размножения. Но вблизи анода, на расстояниях, сравнимых с диаметром нити, электрическое поле E настолько велико, что электроны вызывают вторичную ударную ионизацию — возникает несамостоятельный электрический разряд (см. т. III, § 112). В результате на анод приходит электронная лавина и в ряде случаев сигнал, порождаемый первичной частицей, получается достаточно сильным, так что для его регистрации не требуется усиления. В этом отношении пропорциональный счетчик отличается от ионизационной камеры. Отношение полного числа собранных электронов к их первоначальному числу называется *коэффициентом газового усиления* M. Этот коэффициент тем больше, чем больше отношение $E/\mathscr{P}$ ($\mathscr{P}$ — давление газа; в формировании импульса участвуют и ионы). При полном торможении частицы в пропорциональном счетчике сигнал, как и в ионизационной камере, пропорционален энергии частицы, затраченной на ионизацию. Дело в том, что вторичная ионизация, вызываемая электронами, происходит в очень малой области порядка 10^{-2} см вокруг нити. Объем этой области ничтожен по сравнению с полным объемом счетчика. Практически первичная ионизация всегда происходит вне этой области. Поэтому любой первичный электрон вызывает вблизи нити электронную лавину одной и той же мощности. А так как число первичных электронов, возникающих при ионизации атомов, производимой частицей, пропорционально ее энергии, то отсюда и следует вывод, сформулированный выше. Этот вывод и дает основание называть счетчик пропорциональным.

Типичные характеристики пропорциональных счетчиков: коэффициент газового усиления $M \approx 10^3$–10^4, но иногда может достигать и 10^6; амплитуда импульса 10^{-2} В при электрической емкости самого счетчика около 20 пФ; развитие лавины происходит за время 10^{-9}–10^{-8} с, однако момент появления сигнала на выходе счетчика зависит от места прохождения ионизующей частицы, т. е. от времени дрейфа первичных электронов до анода. Пропорциональные счетчики обладают практически стопроцентной эффективностью по отношению к заряженным частицам. Эти счетчики дешевы и просты в обращении.

---

[1] Импульсные ионизационные камеры используются для регистрации сильноионизирующих частиц: α-частиц, ядер-осколков и др. Для регистрации фона γ-излучения (в дозиметрии) используются токовые ионизационные камеры.

Пропорциональные счетчики используются для регистрации всех видов частиц: α-частиц, электронов, осколков деления, атомных ядер и пр., а также всех сильно взаимодействующих нейтральных частиц (по порождаемым ими заряженным частицам). Эти счетчики сыграли важную роль в развитии ядерной физики довоенного времени, так как они наряду с ионизационной камерой были практически единственными электронными спектрометрическими детекторами.

Главным недостатком пропорциональных счетчиков является сильная зависимость амплитуды импульса от состава газовой смеси и приложенного напряжения, а также недостаточно высокое временное разрешение (порядка 10^{-7} с). При работе с полным импульсом разрешающее время определяется временем дрейфа положительных ионов к катоду. Это время по порядку величины равняется 10^{-4} с. Однако его можно значительно уменьшить, использовав то обстоятельство, что скорость нарастания импульса наибольшая в первые моменты отхода ионов от нити. Это дает возможность работать на начальном участке импульса и достичь разрешающего времени 10^{-6}–10^{-8} с при не очень малом выходном импульсе (10^{-2} В). Пропорциональные счетчики применяются главным образом для регистрации и измерения энергии частиц с не слишком большими энергиями.

Заметим еще, что в связи с появлением ЭВМ в конце 60-х годов, в физике частиц высоких энергий начала применяться *пропорциональная камера.* Не вдаваясь ни в какие подробности, заметим, что она состоит из большого числа (10^3–10^4) небольших пропорциональных счетчиков, расположенных в одной плоскости и обычно в одном газовом объеме. С ее помощью можно воспроизводить траекторию частицы. Таким образом, она одновременно выполняет функции быстродействующего спектрографа и трекового детектора.

5. Если повышать напряжение на электродах пропорционального счетчика, то, начиная с некоторого напряжения, импульс становится не зависящим от энергии первичной частицы, так как прибор начинает работать в режиме самостоятельного разряда, но этот импульс возрастает с повышением напряжения. Такая область напряжений называется *областью Гейгера.* Прибор, работающий в указанном режиме, носит общее название *счетчика Гейгера.* В сущности, он представляет собой газоразрядный промежуток с сильно неоднородным полем. В первоначальных вариантах счетчика, разработанных самим Гейгером, сильно неоднородное электрическое поле возникало вблизи металлического острия, служившего анодом. В более поздних вариантах, предложенных совместно с Мюллером, анод был заменен тонкой проволочкой, натянутой вдоль оси цилиндрического конденсатора. В этой конструкции прибор называют *счетчиком Гейгера–Мюллера.* Один из видов счетчика Гейгера–Мюллера уже был описан нами в § 8 (п. 2, рис. 15).

В счетчике Гейгера–Мюллера электроды заключены в герметический резервуар, наполненный газом при давлении 100–200 мм рт. ст. К электродам прикладывается напряжение в несколько сотен вольт.

При попадании в счетчик ионизующей частицы в газе возникают свободные электроны, которые начинают двигаться к нити-аноду. Так зарождается электронная лавина. Около анода, где напряженность электрического поля очень велика, электроны, образовавшиеся в результате первичной ионизации, ускоряются настолько, что сами начинают ионизовать газ, усиливая ранее образовавшуюся лавину. Вновь возникшие электроны сами могут набрать достаточную скорость, чтобы в свою очередь вызвать новую ионизацию и еще больше усилить лавину электронов. В результате возникает самостоятельный — коронный — разряд, распространяющийся вдоль нити. Разряд обрывается включением большого сопротивления $R \approx 10^8$–10^9 Ом (несамогасящийся счетчик, рис. 155) либо использованием в счетчике газовой смеси, состоящей из инертного газа с примесью небольшого количества паров спирта (или другого многоатомного газа) и галоидов (самогасящийся счетчик).

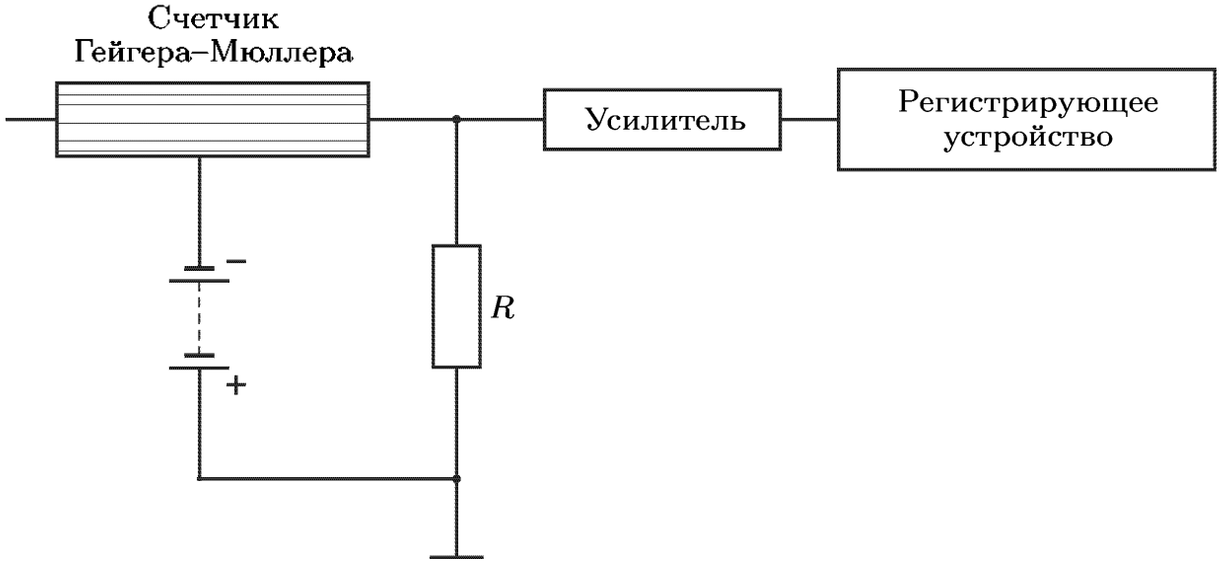


Рис. 155

В последнем случае механизм гашения разряда заключается в следующем. В отсутствие паров спирта фотоны достаточной энергии могут выбивать электроны с катода (см. т. III, § 115), а эти последние, направляясь к аноду, будут усиливать электронные лавины. Пары́ же спирта делают этот процесс невозможным, так как молекулы спирта сильно поглощают фотоны, возбуждаются и диссоциируют при этом, но практически не испускают электронов. Положительные ионы благородного газа (аргона), возникающие при первичной или последующей ионизации, направляются к катоду. Но на своем пути они сталкиваются с молекулами спирта и нейтрализуются, так как потенциал ионизации спирта (11,7 В) ниже потенциала ионизации аргона (15,7 В), а потому при столкновении энергетически выгодно, чтобы электрон перешел от молекулы спирта к иону аргона. Возбужденные атомы аргона испускают фотоны, которые в отсутствие молекул спирта вызывают фотоэффект с поверхностей электродов счетчика, способствующих развитию разряда. Но молекулы спирта поглощают фотоны и при этом диссоциируют; фотоэффекта не происходит, и разряд обрывается.

Электрические импульсы во внешней цепи, возникающие при вспышках разряда в счетчике Гейгера–Мюллера, усиливаются и регистрируются (см. рис. 155). Гамма-кванты регистрируются по вторичным заряженным частицам — фотоэлектронам, комптоновским электронам, электрон–позитронным парам; нейтроны — по ядрам отдачи и продуктам ядерных реакций, возникающих в газе счетчика.

Счетчики Гейгера–Мюллера характеризуются высокой эффективностью и большой амплитудой сигнала (2–40 В). Эффективность регистрации заряженных частиц счетчиками Гейгера–Мюллера близка к 100 %. Эффективность всех газоразрядных счетчиков по отношению к регистрации γ-квантов не превышает 1–3 %. Недостатки счетчиков Гейгера–Мюллера: невозможность измерять энергию частицы, невысокое временное разрешение (10^{-6} с), большое время восстановления их чувствительности (10^{-4}–10^{-3} с). Последнее определяется временем дрейфа к катоду положительных ионов, возникающих при ионизации частиц газа.

6. Ионизационные газовые детекторы сыграли фундаментальную роль на раннем этапе развития ядерной физики. Они применялись для регистрации естественных радиоактивных излучений и космических лучей. Однако с появлением ускорителей, создающих интенсивные пучки заряженных частиц (10^6–10^7 $с^{-1}$) высоких энергий, ионизационные газовые детекторы оказались слишком медленными и были дополнены *сцинтилляционными* и *черенковскими* счетчиками. Эти счетчики оказались удобными и при исследовании космических лучей.

Сцинтилляционный счетчик состоит из вещества, способного люминесцировать под действием заряженных частиц (сцинтиллятора), и фотоумножителя (ФЭУ) — прибора в высшей степени чувствительного и быстродействующего (см. т. III, § 103). Заряженная частица, попадая в сцинтиллятор, возбуждает атомы или молекулы. Возвращаясь в основное состояние, они испускают фотоны. Последние, попадая на катод ФЭУ, выбивают электроны, в результате чего на катоде возникает электрический импульс, который далее усиливается и регистрируется.

Благодаря высокой чувствительности и быстроте действия сцинтилляционный метод детектирования получил широкое распространение. Высокая чувствительность достигается тем, что в ФЭУ устанавливается до 10–20 эмиттеров (*динодов*) из специально обработанного материала. В среднем каждый электрон, попадая на динод, выбивает из него от 3 до 10 новых электронов. С последнего динода усиленный поток электронов попадает на анод, создавая электрический импульс, регистрируемый радиотехническими методами. В результате достигается большой коэффициент усиления (10^6–10^8), малое время восстановления (около 10^{-8} с) при высокой его стабильности, высокое временное разрешение (около 10^{-9} с). Замечательной особенностью ФЭУ является хорошо соблюдаемая линейность усиления. Вторая, не менее важная причина распространенности метода состоит в том, что в качестве сцинтилляторов стали использовать вещества, прозрачные по отношению к собственному излучению сцинтиллятора. Благодаря

этому эффективным в детекторе становится весь объем сцинтиллятора, а не только его поверхностный слой.

Доля энергии регистрируемой частицы, которая превращается в световую энергию, называется *конверсионной эффективностью* η. Наибольшими значениями η обладают кристаллы NaI, активированные Tl, кристаллы антрацена и ZnS. Для этих сцинтилляторов длина волны в максимуме лежит в видимой области спектра вблизи его фиолетового конца. Интенсивность свечения после прохождения частицы изменяется во времени экспоненциально: $I \sim e^{-t/\tau}$, где τ — среднее время высвечивания, определяемое временем жизни молекул или атомов на возбужденных уровнях. Для большинства сцинтилляторов $\tau \approx 10^{-9}$–10^{-5} с. Чем меньше τ, тем более быстродействующим будет счетчик. Наименьшими временами τ обладают пластмассы, активированные терфинилом.

7. Использование прозрачных жидкостей в качестве сцинтиллирующих веществ открыло возможность применять сцинтилляторы громадных размеров. Изготовить сцинтиллятор таких размеров из кристалла затруднительно и даже невозможно, так как кристалл надо вырастить однородным. В качестве примера укажем, что именно с помощью сцинтилляционных счетчиков с жидким сцинтиллятором Рейнесу и Коуэну в 1956 г. впервые удалось зарегистрировать взаимодействие антинейтрино с веществом (см. § 74, п. 15). Для регистрации был использован сцинтиллятор, содержащий 5000 л воды с примесью люминесцирующего вещества и соли кадмия. Источником антинейтрино служил ядерный реактор, дающий плотность потока антинейтрино 10^{13} с$^{-1}\cdot$см$^{-2}$. Сцинтиллятор был помещен глубоко под землей для максимального ослабления фона космического излучения и отделен мощной защитой от реактора. Исследовалась реакция

$$\overline{\nu}_{\mathrm{e}} + \mathrm{p} \to \mathrm{n} + \mathrm{e}^{+}, \tag{86.1}$$

в которой антинейтрино взаимодействовали с протонами воды сцинтиллятора. Позитрон аннигилировал с электроном вещества с испусканием двух γ-квантов, давая первую вспышку. Нейтрон за несколько микросекунд замедлялся до надтепловых скоростей, после чего захватывался кадмием. Ядро кадмия при захвате нейтрона превращалось в новое ядро с энергией возбуждения 9,1 МэВ. При переходе в нормальное состояние оно испускало каскад γ-квантов, дававших вторую вспышку. Обе вспышки регистрировались схемой запаздывающих совпадений, что позволяло уверенно отделить нужные события от фоновых излучений. В среднем регистрировались три события в час, причем на проведение опыта было затрачено около полугода. Для сечения реакции (86.1) получилось $\sigma \approx (11 \pm 4)\cdot 10^{-44}$ см^2.

8. Черенковские счетчики используют явление *излучения Вавилова–Черенкова* (см. § 6, а также т. IV, § 38). Такое излучение возникает, когда заряженная частица движется в среде со скоростью v, превышающей фазовую скорость света в этой среде c/n (n — показатель преломления). В черенковском счетчике эта среда называется *радиатором*. Радиатор должен быть прозрачен для черенковского излучения

и иметь низкий уровень сцинтилляций, создающих фоновые сигналы. В качестве радиаторов для регистрации частиц со сравнительно небольшими скоростями применяют органическое стекло ($n = 1{,}5$), свинцовое стекло ($n = 1{,}5$) и воду ($n = 1{,}33$). Оптическая система счетчика фокусирует свет от частицы на катоде ФЭУ, превращающего световой сигнал в электрический. Общее количество света, попадающее на фотокатод черенковского счетчика, как правило, в несколько раз меньше, чем в сцинтилляционном счетчике. Но оно все же достаточно для регистрации после преобразования в ФЭУ в электрический сигнал и последующего усиления последнего. Именно фотоумножитель сделал возможным в черенковском счетчике уверенно регистрировать отдельные частицы. Так как излучение света в черенковских счетчиках происходит мгновенно, то их разрешающее время достигает 10^{-9} с.

Излучение Вавилова–Черенкова характеризуется резкой направленностью. Все оно концентрируется в тонком слое вблизи поверхности конуса, образующие которого составляют угол

$$\theta = \arccos \frac{c}{nv} = \arccos \frac{1}{\beta n} \tag{86.2}$$

с направлением движения частицы. Эта особенность излучения и используется в черенковских счетчиках. Если скорость частицы v близка к скорости света, то используются *газовые радиаторы.* В этом случае из-за малого отличия n от единицы угол θ, а с ним и интенсивность излучения очень малы. Для получения интенсивности вспышки, достаточной для регистрации, приходится увеличивать длину черенковского счетчика до 10 м и больше. Зато газовый черенковский счетчик позволяет плавно менять n путем изменения давления газа.

Черенковские счетчики делятся на *пороговые*, *дифференциальные* и *полного поглощения.* Пороговые счетчики должны регистрировать все частицы со скоростями больше некоторой (пороговой) скорости. Дифференциальные счетчики регистрируют частицы, скорости которых заключены в определенном интервале: $v_1 < v < v_2$. Это достигается оптической системой, выделяющей лучи, распространяющиеся под определенными углами между θ_1 и θ_2 к направлению движения частицы. Линза или сферическое зеркало с фокусным расстоянием f, поставленные на пути этих лучей, фокусируют их на кольцевую щелевую диафрагму с радиусами $R_1 = f\theta_1$ и $R_2 = f\theta_2$ (углы θ_1 и θ_2 малы), а затем на фотоумножитель. В результате свет будет зарегистрирован только от частиц, излучающих в направлении выделенного интервала углов и имеющих скорости в соответствующем ему интервале скоростей. Таким путем удается выделить частицы, которые по величине β отличаются всего на 10^{-6} от других частиц.

Черенковский счетчик полного поглощения предназначен для регистрации и спектрометрии электронов и γ-квантов. Он содержит блок радиаторов большой толщины, в котором электрон или γ-квант образует электрон-фотонную лавину и теряет всю (или почти всю) свою энергию. Как правило, радиатор изготовляют из стекла с большим содержанием свинца. При толщине радиатора 40 см в нем практически

полностью тормозятся электроны с энергией до 10 ГэВ. Количество света, излученного в черенковских счетчиках такого типа, пропорционально энергии первичного электрона или γ-кванта.

Для регистрации заряженных частиц с энергией порядка 10^2–10^3 ГэВ используются детекторы, в которых вспышки света возникают при прохождении регистрируемой частицы через границу двух сред с резко различными свойствами (*переходное излучение*). Интенсивность света, излучаемого при этом, пропорциональна энергии частицы, но значительно меньше, чем в случае черенковского излучения. Поэтому детекторы на переходном излучении делают многослойными, содержащими сотни слоев газ — твердое тело.

9. К концу 50-х годов после преодоления различных технических трудностей появились *полупроводниковые счетчики.* Они используют переходное излучение в рентгеновской области и имеют большое будущее в физике ТэВ-области для идентификации частиц. Схема такого счетчика показана на рис. 156. Принцип работы полупроводникового счетчика тот же, что и ионизационной камеры, только вместо газа применяется полупроводник. Это имеет то преимущество, что в твердом теле на одном и том же отрезке пути заряженная частица отдает в сотни раз больше энергии, чем в газе. Проходя через полупроводник, частица вызывает переход электронов из валентной зоны в зону проводимости, порождая пару электрон-дырка (см. т. III, § 100). Энергия W, необходимая для рождения такой пары, составляет 2,9 эВ в германии и 3,5 эВ в кремнии. Если энергия частицы равна $\mathscr{E}$, то на своем пути в полупроводнике она создает $\mathscr{E}/W$ электронно-дырочных пар. Приложенное электрическое поле оттягивает отрицательные заряды к поверхности, заряженной положительно, а положительные заряды — к поверхности, заряженной отрицательно. В результате во внешней цепи появляется электрический сигнал, который усиливается и регистрируется. Чтобы в отсутствие регистрируемой частицы через полупроводник не проходил ток, поверхности полупроводника обрабатывают сильно легированными донорными и акцепторными примесями (см. т. III, § 108). При этом отрица-

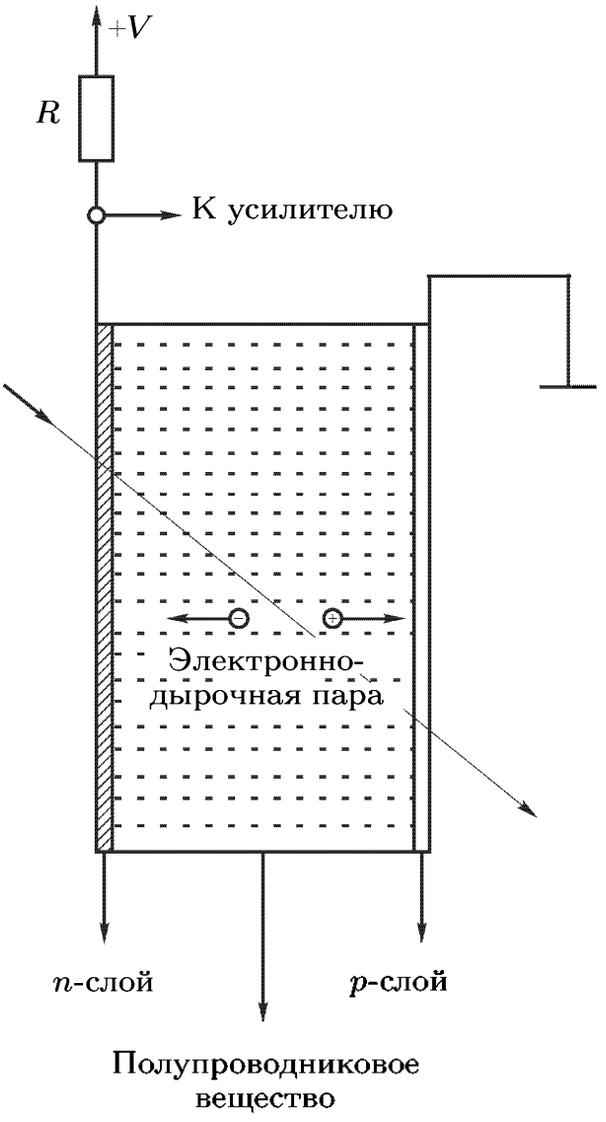


Рис. 156

тельный n-слой присоединяют к положительному, а положительный p-слой — к отрицательному полюсу внешнего источника. При комнатной температуре тепловое возбуждение может приводить к нежелательным импульсам тока в приборе, и поэтому полупроводниковые детекторы нередко охлаждают до температуры жидкого азота.

Размеры полупроводниковых счетчиков невелики — несколько кубических сантиметров. Большими считаются полупроводниковые счетчики, объем которых доходит до 100 $см^3$. Малые размеры рабочей области позволяют довести разрешающее время до 10^{-7} с. В области низких энергий (электроны до 2 МэВ, протоны до 20 МэВ) полупроводниковые счетчики обладают практически 100 %-ной эффективностью, малым разрешающим временем и превосходят счетчики других типов по компактности и точности измерения энергии.

10. Старейшим трековым детектором является *камера Вильсона* (1869–1959), изобретенная ее автором в 1912 г. Принцип действия этой камеры был подробно рассмотрен в т. III (§ 33, задача 4). Он основан на конденсации пересыщенных паров воды или спирта на ионах, образованных при движении вдоль траектории заряженной частицы. Камера в нужный момент освещается импульсным источником света и фотографируется несколькими фотоаппаратами для получения стереоскопических изображений треков. Природа и свойства частицы устанавливаются по длине ее пробега, толщине трека и его искривлению в магнитном поле. Измеряя кривизну трека в сильном магнитном поле, можно определить импульс частицы и знак ее заряда. Этот метод был предложен в 1923 г. независимо друг от друга П. Л. Капицей (1894–1984) и Д. В. Скобельциным (р. 1892).

Камера Вильсона сыграла выдающуюся роль в ядерной физике и физике космических лучей. На протяжении нескольких десятилетий она была единственным трековым детектором заряженных частиц. Но в 50-е и 60-е годы она утратила свое значение, уступив место *пузырьковой* и *искровой камерам.* В настоящее время камера Вильсона используется главным образом как эффектный лекционный демонстрационный прибор.

Родственной с камерой Вильсона по принципу действия является *диффузионная камера.* В камере Вильсона пересыщение создается адиабатическим расширением газа (содержащего пары воды или спирта), а в диффузионной камере оно достигается за счет непрерывного потока пара от горячей поверхности у крышки камеры к более холодной поверхности у ее дна. В отличие от камеры Вильсона, в диффузионной камере пересыщение существует постоянно в некотором горизонтальном слое пара, в котором и получаются треки ионизующих частиц. Поэтому диффузионная камера чувствительна к ионизующим частицам непрерывно.

11. Важнейшим трековым детектором заряженных частиц является *пузырьковая камера*, изобретенная Глезером (р. 1926) в 1954 г. Принцип действия этой камеры был описан в т. II (§ 119). Он основан на том, что можно получать жидкость с температурой выше точки кипения.

Такая перегретая жидкость нестабильна и через короткое время вскипает. Наиболее употребляемыми рабочими жидкостями служат жидкие водород, дейтерий, смесь неона с водородом — в криогенных камерах, а также пропан (C_3H_8), фреон и Xe (обычно в смеси с пропаном) — в тяжеложидкостных камерах. Перегревание жидкости производится быстрым понижением давления от начального значения $\mathscr{P}_н > \mathscr{P}_0$ до $\mathscr{P} < \mathscr{P}_0$ ($\mathscr{P}_0$ — равновесное давление при температуре T). Понижение (снятие) давления достигается либо быстрым перемещением поршня в жидководородных камерах, либо посредством использования упругой мембраны (в пропановой и фреоновой камерах). Давление в камере снимается за 5–15 мс, после чего жидкость оказывается перегретой, т. е. чувствительной к регистрации заряженных частиц. В момент максимальной чувствительности частицы впрыскиваются в камеру.

Заряженная частица при прохождении через перегретую жидкость приводит к образованию центров кипения вдоль своей траектории. По современным представлениям причиной образования таких центров являются δ-электроны, т. е. вторичные электроны, испускаемые атомами среды при соударении с ними ионизующей частицы. Эти электроны поглощаются небольшими объемами жидкости, что приводит к нагреванию последних, возникновению зародышей кипения, а в дальнейшем — к образованию и росту пузырьков. За время порядка 0,5–3 мс пузырьки достигают размера 50–300 мкм и при освещении их импульсным источником света могут быть сфотографированы несколькими фотоаппаратами.

Пузырьковые камеры обычно применяются для регистрации актов взаимодействия частиц высоких энергий с ядрами атомов жидкости, наполняющей камеру, а также актов распада частиц. В первом случае жидкость в пузырьковой камере одновременно выполняет и роль мишени, и роль детектора частиц.

Искажения следов частиц в пузырьковой камере невелики и определяются главным образом многократным рассеянием.

Эффективность регистрации определяется в основном размерами пузырьковой камеры. Камеры обычно имеют размеры 1–2 м. Однако на ускорителях сверхвысоких энергий используются и очень большие пузырьковые камеры. Так, пузырьковая камера СКАТ в Институте физики высоких энергий (г. Серпухов), заполненная жидким фреоном, имеет размеры $4{,}5 \times 1{,}6 \times 1{,}5$ м, водородная камера на ускорителе Национальной ускорительной лаборатории им. Ферми (США) — 33 м^3.

Пузырьковые камеры сыграли важную роль в открытии и изучении странных частиц, нестабильных частиц-резонансов и т. д. Отметим некоторые их недостатки. Главный недостаток пузырьковой камеры — невозможность в процессе работы отбирать нужные события, при исследовании редких событий — это приводит к необходимости просматривать очень много фотографий. Нельзя различить релятивистские частицы. На очень высоких энергиях требуется сепаратирование пучка, а это трудно сделать. Даже в самые большие камеры нельзя одновременно запускать более 5–6 заряженных частиц. Поэтому процессы

с сечением меньше 1 микробарна практически недоступны (за исключением нейтрино, где проходит 10^{10} частиц), тогда как электронные приборы могут регистрировать до 10^7–10^8 частиц в секунду.

12. Самым дешевым трековым детектором заряженных частиц являются *толстослойные ядерные фотоэмульсии.* Они отличаются от обычных фотоэмульсий более высокой чувствительностью и большей толщиной (сотни микрометров, тогда как толщина обычных фотоэмульсий порядка 10 мкм). Пластинки, покрытые слоем чувствительной фотоэмульсии, ставятся на пути ионизующих частиц, после чего проявляются. В результате частица оставляет след, образованный зернами металлического серебра (размер зерна около 0,3 мкм). Этот след рассматривается в микроскоп с увеличением в 500–1000 раз. Измерение длины пробега при известной массе частицы дает ее энергию, а измерение плотности зерен — потери энергии на единицу пути, по которым можно судить о скорости частицы. Пластинку можно помещать в магнитное поле для измерения импульса частицы. Однако из-за короткой длины треков магнитное поле должно быть очень сильным. Главный недостаток метода — очень большая трудоемкость обработки пластинок, так как каждую пластинку приходится подолгу просматривать под сильным микроскопом.

В последние годы в связи с изучением частиц с тяжелыми кварками техника фотоэмульсий вновь возродилась (изучение частиц со временем жизни 10^{-12}–10^{-13} с).

Используются *гибридные установки* с электронными детекторами для идентификации и определения вершин [1]) в эмульсии. В гибридных установках используются также в качестве вершинных детекторов небольшие быстродействующие пузырьковые камеры высокого разрешения с лазерным и голографическим съемом информации. Разрешение — десятки микрометров.

13. В *искровой камере*, появившейся в конце 50-х годов, сочетается немедленность регистрации (т. е. отсутствие последующей обработки) с полнотой информации трековых детекторов. Принцип действия искровой камеры основан на возникновении искрового разряда в газе при попадании в него ионизующей частицы. Простейшая искровая камера состоит из двух плоскопараллельных электродов, пространство между которыми заполнено газом (чаще всего Ne, Ar или их смесью). Площадь пластин — от десятков квадратных сантиметров до нескольких квадратных метров. Одновременно с прохождением частицы или с некоторым запаздыванием (около 1 мкс) по сигналу системы детекторов (сцинтилляционных счетчиков, черенковских счетчиков и пр.), выделяющих искомое событие, на электроды искровой камеры подается короткий (порядка 10–100 нс) импульс высокого напряжения (5–20 кВ/см). Электроны, возникшие вдоль траекторий движения частицы при ионизации атомов газа, ускоряются электрическим полем и сами вызывают ионизацию и возбуждение атомов. В результате на очень

[1]) Вершина — точка в фотоэмульсии, где произошло взаимодействие.

коротком пути частицы возникают электрон-фотонные лавины, перерастающие либо в видимый глазу искровой разряд, либо в локальные светящиеся области небольшого объема (*стримеры*).

Узкозазорная искровая камера состоит из набора близко расположенных одинаковых плоскопараллельных электродов, отделенных друг от друга промежутками около 1 см. Электроды соединены через один. Половина электродов заземлена, а на другую в момент прохождения частицы подается короткий высоковольтный импульс напряжения порядка 1 кВ/см. Цепь искр, проскакивающих между пластинами в момент подачи этого импульса, воспроизводит траекторию частицы с точностью до долей миллиметра. Временное разрешение порядка 10^{-6} с, полное время восстановления порядка 10^{-3} с.

В широкозазорных искровых камерах расстояние между электродами 3–50 см. Электрон-фотонные лавины, развивающиеся в результате прохождения первичной частицы, сливаются в узкий светящийся канал вдоль траектории частицы. В этом режиме могут регистрироваться треки под углами не более 50° к направлению электрического поля в камере. Для наблюдения треков под большими углами вплоть до 90° используется так называемый *стримерный режим*. *Стримерная камера* есть разновидность искровой камеры, на которую подается импульс 20 кВ/см с более коротким фронтом и длительностью 10 нс. В этой камере разряд, вызванный импульсом высокого напряжения, обрывается на ранней стадии. В результате треки заряженных частиц, прошедших через камеру, выглядят как цепочки отдельных (несливающихся) стримеров, длиной в несколько миллиметров каждый.

Трековые искровые камеры и стримерные камеры характеризуются высокой эффективностью к одновременной регистрации многих частиц (ливней частиц), а также малой пространственной и угловой погрешностью определения траектории (около 10^{-3} рад).

Кроме фотографирования в искровых камерах широко применяют безфильмовые методы регистрации с использованием ЭВМ. Координаты искр записываются в память ЭВМ, где сразу же подвергаются математической обработке.

Глава XIII

ЯДЕРНЫЕ РЕАКЦИИ

§ 87. Терминология и определения

1. О *ядерных реакциях* уже неоднократно говорилось на протяжении нашего курса. Рассмотрим теперь относящиеся сюда вопросы более систематично, допуская повторения в тех немногих случаях, где это необходимо.

В самом широком смысле слова ядерной реакцией называется процесс, возникающий в результате взаимодействия нескольких (как правило двух) сложных атомных ядер или элементарных частиц. Под это определение подходит и упругое рассеяние частиц, например нуклона на нуклоне, т. е. такое рассеяние, при котором не образуются новые частицы и не происходит их возбуждения, а происходит лишь перераспределение энергии и импульса между ними. При неупругом рассеянии также не возникают новые частицы, но по крайней мере одна из них переходит в возбужденное состояние.

В настоящей главе рассматриваются ядерные реакции только в узком смысле. В таких реакциях среди исходных частиц обязательно имеется по крайней мере одно ядро. Оно сталкивается с другим ядром или элементарной частицей, в результате чего происходит ядерная реакция и образуются новые частицы.

Как правило, ядерные реакции идут под действием ядерных сил. Но могут быть и исключения из этого правила. Например, процесс расщепления ядра под действием γ-квантов высоких энергий или быстрых электронов есть ядерная реакция, но она происходит под действием электромагнитных, а не ядерных сил, так как ядерные силы на фотоны и электроны не действуют. Точно так же к ядерным реакциям в широком смысле относятся процессы, возникающие при столкновениях нейтрино с другими частицами, но они совершаются под действием слабых сил.

Ядерные реакции могут происходить в естественных условиях (например, в недрах звезд или в космических лучах). Но их изучение обычно производится в лабораторных условиях на экспериментальных установках, в которых заряженные частицы ускоряются с помощью ускорителей. В этих случаях более тяжелые частицы обычно покоятся и называются *частицами мишени*, а более легкие налетают на них в составе *ускоренного пучка*. В ускорителях на встречных пучках частицы движутся навстречу друг другу, так что их деление на частицы мишени и частицы пучка теряет смысл. Кроме того, мишень не может состоять из нейтронов и большинства элементарных частиц, так как не существует способов приготовления таких мишеней.

2. Ускорением достигается сближение частиц (двух ядер, ядра и нуклона и т. д.) на расстояния порядка 10^{-13} см, на которых начинают действовать ядерные силы. Энергия налетающей положительной частицы должна быть порядка или больше высоты кулоновского потенциального барьера ядра (для однозарядной частицы и средних ядер порядка 10 МэВ, а для легких ядер меньше). В этих случаях ядерная реакция, как правило, и осуществляется за счет ускорения бомбардирующей частицы на ускорителе. Энергия частицы может быть и меньше высоты кулоновского барьера, как это было в классических опытах Кокрофта (1897–1967) и Уолтона (р. 1903), которые в 1932 г. впервые осуществили искусственное расщепление ядер лития путем бомбардировки их ускоренными протонами (см. (87.1)). Но в их опытах проникновение протона в ядро лития происходило путем туннельного перехода через кулоновский потенциальный барьер. Вероятность такого процесса, разумеется, очень мала. Для отрицательно заряженных и нейтральных частиц кулоновский барьер не существует и ядерные реакции могут протекать даже при тепловых энергиях налетающих частиц.

3. Наиболее универсальная и наглядная запись ядерных реакций заимствована из химии. Слева пишется сумма исходных частиц, затем ставится стрелка, а за ней сумма конечных продуктов реакции. Например, запись

$$ {}_1^1\mathrm{p} + {}_3^7\mathrm{L} \to {}_0^1\mathrm{n} + {}_4^7\mathrm{Be} \qquad (87.1) $$

означает ядерную реакцию, возникающую при бомбардировке изотопа лития ${}_3^7\mathrm{Li}$ протонами, в результате которой возникают нейтрон и изотоп бериллия ${}_4^7\mathrm{Be}$. Запись

$$ \gamma + {}_4^9\mathrm{Be} \to 2\,{}_2^4\mathrm{He} + {}_0^1\mathrm{n} $$

означает реакцию расщепления ядра ${}_4^9\mathrm{Be}$ на две α-частицы и нейтрон. Такой способ записи ядерных реакций уже неоднократно применялся нами, хотя и без его расшифровки.

Ядерные реакции в узком смысле часто записывают в более короткой символической форме: A(a, bcd ...)B, где A — ядро мишени, a — бомбардирующая частица, bcd ... — испускаемые в ядерной реакции частицы, B — остаточное ядро (в скобках после запятой записывают более легкие продукты реакции, вне — более тяжелые). Например, реакция (87.1) может быть записана так: $^7\mathrm{L(p, n)}^7\mathrm{Be}$. Запись $^{40}\mathrm{Ca}(\gamma, \mathrm{pn})^{38}\mathrm{K}$ означает выбивание из ядра $^{40}\mathrm{Ca}$ протона и нейтрона под действием γ-кванта, в результате чего образуется ядро калия $^{38}\mathrm{K}$. Часто применяют и еще более короткую запись, указывая только легкие частицы и не указывая ядра, участвующие в реакции. Например, запись (p, n) означает выбивание нейтрона из какого-то ядра под действием протона.

Ядерная реакция часто может идти несколькими способами, например: $^{63}\mathrm{Cu(p, n)}^{63}\mathrm{Zn}$; $^{63}\mathrm{Cu(p, 2n)}^{62}\mathrm{Zn}$; $^{63}\mathrm{Cu(p, pn)}^{62}\mathrm{Cu}$; $^{63}\mathrm{Cu(p, p)}^{63}\mathrm{Cu}$; $^{63}\mathrm{Cu(p, p')}^{*}\ ^{63}\mathrm{Cu}$ (неупругое рассеяние протонов). Совокупность сталкивающихся частиц в определенном квантовом состоянии (например, p и ядро $^{63}\mathrm{Cu}$) называется *входным каналом ядерной реакции.* Частицы,

рождающиеся в результате ядерной реакции в определенных квантовых состояниях (например, p, n и ядро ^{62}Cu с определенными орбитальными моментами и проекциями спинов на выделенное направление) образуют *выходной канал ядерной реакции.*

4. Количественное описание ядерных реакций с квантовомеханической точки зрения может быть только *статистическим*, т. е. таким, в котором принципиально можно говорить лишь о вероятностях различных процессов, характеризующих ядерную реакцию. Например, реакция $a + A \to b + B$, в начальном и конечном состояниях которой имеются по две частицы, в этом смысле полностью характеризуется *дифференциальным эффективным сечением* рассеяния $d\sigma/d\Omega$ внутрь телесного угла $d\Omega = \sin\theta\, d\theta\, d\varphi$, где θ и φ — полярный и азимутальный углы вылета одной из частиц (обычно легкой), причем угол θ отсчитывается от направления движения налетающей частицы. Зависимость дифференциального эффективного сечения от углов θ и φ называется *угловым распределением частиц*, образующихся в реакции. *Интегральным* или *полным сечением*, которым характеризуется интенсивность реакции, называется дифференциальное эффективное сечение, проинтегрированное по всем значениям углов θ и φ:

$$\sigma = \int \frac{d\sigma}{d\Omega}\, d\Omega. \tag{87.2}$$

Если налетающая частица и частицы мишени не имеют спинов или ориентированы хаотически, то все процессы аксиально-симметричны относительно направления движения частицы. В частности, в этом случае $d\sigma/d\Omega$ зависит только от полярного угла θ и можно написать $d\Omega = 2\pi \sin\theta\, d\theta$. Подавляющее большинство экспериментальных исследований выполняется в *лабораторной системе координат* (ЛС), т. е. системе, в которой мишень покоится. В теоретических исследованиях удобнее *система центра масс* (СЦМ), или *система центра инерции* (СЦИ), в которой суммарный импульс сталкивающихся частиц равен нулю. Результаты, полученные в СЦМ, можно пересчитать в ЛС, на чем мы останавливаться не будем. В ядерных реакциях в узком смысле слова масса налетающей частицы обычно много меньше массы ядра. В этом случае при нерелятивистских движениях центр масс практически совпадает с центром тяжелой частицы, так что обе системы координат фактически совпадают. Но при сверхвысоких скоростях налетающей частицы (почти совпадающих со скоростью света), когда ее кинетическая энергия значительно превосходит сумму энергий покоя обеих сталкивающихся частиц, системы ЛС и СЦМ уже сильно отличаются одна от другой.

При более детальной количественной характеристике ядерной реакции, идущей по нескольким каналам, вводят дифференциальные и интегральные сечения для отдельных каналов реакций. Для множественных процессов важное значение имеют так называемые *инклюзивные сечения*, определяющие вероятность появления в данном столкновении какой-либо определенной частицы или группы частиц.

5. Допустим теперь, что спины частиц, участвующих в реакции, не равны нулю. Если они как-то упорядочены, то частицы называются *поляризованными.* Так, поляризация света связана с ориентацией спинов фотонов. Частица с ненулевой массой покоя и спином I имеет $2I + 1$ квантовых состояний, отвечающих различным значениям проекции спина на некоторое направление. Состояние частицы представляет собой суперпозицию этих состояний. Если коэффициенты суперпозиции полностью определены (*чистое состояние*), то говорят, что *частица полностью поляризована.* Если же они определены не полностью, а заданы некоторыми статистическими характеристиками (*смешанное состояние*), то говорят о *частичной поляризации.* В этом случае для характеристики состояния поляризации вводят *вектор поляризации*, модуль которого называется просто *поляризацией* и измеряется в процентах. Сам вектор поляризации равен среднему значению спина, деленному на максимальное значение его проекции. Если вектор поляризации не параллелен импульсу налетающей частицы, то угловое распределение частиц после реакции может оказаться не аксиально-симметричным, т. е. может зависеть от азимутального угла φ. Но и при отражении или рассеянии неполяризованного пучка частиц на неполяризованной мишени может возникнуть поляризация. Это явление, напоминающее поляризацию, возникающую при отражении естественного света от изотропных диэлектриков, используется для получения поляризованных частиц. Оказывается, что вектор поляризации частиц перпендикулярен к плоскости рассеяния.

6. Наглядно эффективное сечение можно рассматривать как *эффективную площадь мишени*, попадая в которую частица вызывает требуемую реакцию. Но из-за волновых и квантовых свойств частиц такое толкование имеет ограниченную область применимости. В частности, для квантовой частицы существует ненулевая вероятность пройти без отклонения в сторону через область, в которой на нее действуют силы. Тогда действительное полное эффективное сечение окажется меньше поперечного сечения области, в которой происходит взаимодействие. В этом случае по аналогии с оптикой ядро-мишень называют *частично прозрачным* или *серым.*

Из-за волновых свойств частиц невозможно и чисто неупругое рассеяние, при котором рассеивающее тело только поглощает все падающие на него частицы. В противном случае за телом появилась бы резкая тень, что невозможно. Тень обязательно размоется из-за дифракции. А это означает, что появятся частицы, отклонившиеся в сторону, т. е. поглощение будет сопровождаться еще упругим рассеянием. Это рассеяние ввиду своего происхождения называется *дифракционным.*

Сечения ядерных реакций σ зависят от энергии $\mathscr{E}$ вылетающей частицы, типа ядерной реакции, углов вылетов и ориентации спинов частиц — продуктов реакции. Величина σ колеблется в пределах 10^{-27}–10^{-21} см^2. Если длина волны де Бройля $\lambda\!\!\!^{-}$ налетающей частицы мала по сравнению с радиусом ядра R, то применимы методы геометрической оптики. В этом случае (для нерезонансных реакций и непрозрачных ядер) эффективное сечение реакции определяется геометрическим по-

перечным сечением ядра: $\sigma \approx \pi R^2$. На это соотношение, разумеется, нельзя смотреть как на точное. Действительно, сечение σ есть точная, статистически определенная величина, тогда как в определении радиуса ядра R содержится известный элемент неопределенности и условности. Это, в частности, проявляется в том, что с изменением скорости налетающей частицы меняется и эффективное сечение реакции. При малых скоростях, когда $\lambda\!\!\!^{-} \gg R$, методы геометрической оптики уже неприменимы. Величиной, определяющей сечение реакции, становится главным образом $\lambda\!\!\!^{-}$, а не R (см. § 88). Ведь обе величины R и $\lambda\!\!\!^{-}$, как имеющие размерность длины, входят в формулы, определяющие эффективное сечение реакции.

Для суждения о том, какая из этих двух величин является определяющей в том или ином конкретном случае, представляет интерес определить энергию $\mathscr{E}$ бомбардирующего нуклона, при которой $R = \lambda\!\!\!^{-}$. Для этого замечаем, что $\mathscr{E} = p^2/2m$, где m — масса нуклона, $p = \hbar/\lambda\!\!\!^{-}$, так что при условии $\lambda\!\!\!^{-} = R$ получим

$$p = \frac{\hbar}{R}, \quad \mathscr{E} = \frac{\hbar^2}{2mR^2}.$$

Воспользуемся теперь формулой (65.1). Тогда

$$\mathscr{E} = \frac{\hbar^2}{2mr_0^2 A^{2/3}} = \frac{(\hbar c)^2}{2\pi mc^2 r_0^2 A^{2/3}}$$

или после подстановки числового значения r_0 из формулы (65.2)

$$\mathscr{E} \approx 10 A^{-2/3} \text{ МэВ}. \tag{87.3}$$

7. В реальных физических опытах далеко не всегда удается непосредственно измерить само дифференциальное или интегральное эффективное сечение реакции. Непосредственно измеряемой величиной является *выход реакции*. Выходом ядерной реакции W называется доля частиц пучка, испытавших ядерное взаимодействие с частицами мишени. Если S — площадь поперечного сечения пучка, а I — плотность его потока, то на такую же площадь мишени ежесекундно падает $N = IS$ частиц. Из них в одну секунду в среднем реагирует $\Delta N = IS n\sigma$ частиц, где σ — эффективное сечение реакции частиц пучка, а n — концентрация ядер в мишени. Таким образом,

$$W = \Delta N/N = \sigma n. \tag{87.4}$$

§ 88. Законы сохранения в ядерных реакциях

1. При рассмотрении ядерных реакций, как и других процессов, обсуждающихся в ядерной физике, используются следующие точные законы сохранения:

1) *закон сохранения энергии*;
2) *закон сохранения импульса*;
3) *закон сохранения момента импульса*;
4) *закон сохранения электрического заряда*;

5) *закон сохранения барионного заряда*;

6) *закон сохранения лептонных зарядов.*

Кроме того, используются и другие законы сохранения, а именно:

7) при пренебрежении слабым взаимодействием — *закон сохранения четности волновой функции*;

8) при пренебрежении электромагнитным взаимодействием — *закон сохранения изотопического спина.* Здесь этот закон рассматриваться не будет.

В физике элементарных частиц к перечисленным законам сохранения добавляются некоторые другие законы (см. § 109). Но здесь, в ядерной физике, мы рассматривать их не будем.

Законы сохранения позволяют предсказать, какие из мысленно возможных реакций могут действительно осуществляться, а какие невозможны или, как говорят, «запрещены» в силу невыполнения одного или нескольких законов сохранения. В этом отношении применительно к ядерным реакциям законы сохранения играют особо важную роль.

2. Начнем с законов сохранения энергии и импульса. Для процесса столкновения двух частиц эти законы в релятивистской форме можно записать так:

$$\mathscr{E}_{1\,\text{рел}} + \mathscr{E}_{2\,\text{рел}} = \mathscr{E}'_{1\,\text{рел}} + \mathscr{E}'_{2\,\text{рел}} + \ldots + \mathscr{E}'_{n\,\text{рел}}, \tag{88.1}$$

$$\mathbf{p}_{1\,\text{рел}} + \mathbf{p}_{2\,\text{рел}} = \mathbf{p}'_{1\,\text{рел}} + \mathbf{p}'_{2\,\text{рел}} + \ldots + \mathbf{p}'_{n\,\text{рел}}, \tag{88.2}$$

где величины без штриха обозначают релятивистские энергии и импульсы частиц до столкновения, а штрихованными буквами обозначены те же величины для частиц, образовавшихся в результате столкновения. Разумеется, соотношения (88.1) и (88.2) написаны в любой инерциальной системе отсчета. В частности, в лабораторной системе отсчета, когда частица 2 до столкновения неподвижна, $\mathbf{p}_{2\text{рел}} \equiv \mathbf{p}_{2\text{нерел}} = 0$, а в системе центра масс $\mathbf{p}_{1\text{рел}} + \mathbf{p}_{2\text{рел}} = 0$.

В экспериментальных исследованиях под энергией частицы всегда понимают ее кинетическую энергию $\mathscr{E} \equiv \mathscr{E}_{\text{кин}} = \mathscr{E}_{\text{рел}} - mc^2$, где m — масса покоя частицы. Для безмассовых частиц $\mathscr{E}_{\text{кин}}$ и $\mathscr{E}_{\text{рел}}$ тождественно совпадают. В этом случае нерелятивистского приближения не существует. Вычитая из (88.1) соответствующие энергии покоя, получим

$$\mathscr{E}_1 + \mathscr{E}_2 = \mathscr{E}'_1 + \mathscr{E}'_2 + \ldots + \mathscr{E}'_n + Q, \tag{88.3}$$

где

$$Q = (m'_1 + m'_2 + \ldots + m'_n)c^2 - (m_1 + m_2)c^2. \tag{88.4}$$

Величина Q представляет собой энергию, выделяющуюся в результате реакции. Ее, как и в химии, часто включают в уравнение самой реакции. Например, реакция A(a, b)B в более полной форме записывается так:

$$\mathrm{a} + \mathrm{A} \to \mathrm{b} + \mathrm{B} + \mathrm{Q}. \tag{88.5}$$

Это равенство означает, что если покоящиеся частицы a и A вступают в реакцию друг с другом, то частицы b и B получаются не в состоянии покоя, а в состоянии, в котором их суммарная кинетическая энергия

равна Q. В химии величина Q называется *теплотой реакции.* В ядерной физике ее принято называть *энергией реакции.*

Реакция называется *экзотермической*, если $Q > 0$, т. е. если она идет с выделением энергии. При этом в соответствии с (88.4) под энергией всегда понимается кинетическая энергия. Примером экзотермической реакции может служить реакция (в СЦМ)

$$\mathrm{p} + {}^7\mathrm{Li} \to \alpha + \alpha + 17 \text{ МэВ},$$

впервые полученная искусственно Кокрофтом и Уолтоном. Если же $Q < 0$, то реакция идет с поглощением энергии и называется *эндотермической.* Очевидно, всякая реакция, обратная экзотермической, будет эндотермической. Так, эндотермична реакция

$$\alpha + \alpha \to \mathrm{p} + {}^7\mathrm{Li} - 17 \text{ МэВ}.$$

Если не вводить отрицательных энергий, а воспользоваться обозначением $Q_{\text{энд}} = -Q$, то равенство (88.5) можно переписать так:

$$\mathrm{a} + \mathrm{A} + Q_{\text{энд}} \to \mathrm{b} + \mathrm{B}, \tag{88.6}$$

где величина $Q_{\text{энд}}$ уже положительна. Это равенство означает, что для возможности реакции между частицами a и A им необходимо сообщить суммарную кинетическую энергию $Q_{\text{энд}}$, чтобы конечные продукты реакции b и B получились в состоянии покоя.

Положительную величину $Q_{\text{энд}}$ условимся называть *энергией эндотермической реакции.*

Экзотермическая реакция может идти при сколь угодно малой энергии сталкивающихся частиц. Напротив, эндотермическая реакция может идти только тогда, когда энергия сталкивающихся частиц превосходит некоторое минимальное значение. Это минимальное значение энергии, начиная с которого эндотермическая реакция может идти, называется *порогом реакции.* Существенно заметить, что порог реакции измеряется всегда в лабораторной системе координат, в которой частица-мишень покоится. Что касается величины $Q_{\text{энд}}$, то она связана только с относительным движением реагирующих частиц. Кинетическая энергия, связанная с движением центра масс системы частиц, сохраняется и в реакции не участвует. Напротив, она оказывает существенное влияние на величину $\mathscr{E}_{\text{пор}}$. Поэтому порог реакции $\mathscr{E}_{\text{пор}}$ вообще говоря, не совпадает с энергией эндотермической реакции Q.

Возьмем, например, две одинаковые частицы, движущиеся навстречу друг другу с одинаковыми, но противоположно направленными скоростями. В этом случае центр масс системы находится в покое. Он останется в покое и после столкновения. Пусть в результате столкновения частицы сливаются в одну частицу (неупругое столкновение). Она будет находиться в состоянии покоя. На образование этой частицы затрачивается энергия, равная сумме кинетических энергий сталкивающихся частиц. Предположим, что это — минимальная энергия, при которой слияние частиц возможно. Тогда реакция будет эндотермической и будет иметь порог. Пусть теперь одна частица покоится, а другая на нее налетает с той же относительной скоростью. Тогда реакция

и энергия эндотермической реакции останутся прежними, но не вся начальная кинетическая энергия пойдет на превращение. Останется еще кинетическая энергия движения центра масс. Таким образом, порог реакции больше энергии эндотермической реакции.

3. Найдем связь между порогом и энергией эндотермической реакции в общем случае, когда сталкивающиеся частицы, а также частицы, образующиеся в результате реакции, движутся с релятивистскими скоростями. Будем предполагать, что в лабораторной системе неподвижна частица 2 (мишень), а частица 1 движется с произвольной скоростью. Тогда $\mathbf{p}_{2\text{рел}} = 0$, $\mathbf{p}_{1\text{рел}} \neq 0$. В этой системе координат законы сохранения энергии и импульса запишутся так:

$$\mathbf{p}_{1\text{рел}} = \mathbf{p}'_{1\text{рел}} + \mathbf{p}'_{2\text{рел}} + \ldots + \mathbf{p}'_{n\text{рел}},$$

$$\mathscr{E}_{1\text{рел}} + m_2c^2 = \mathscr{E}'_{1\text{рел}} + \mathscr{E}'_{2\text{рел}} + \ldots + \mathscr{E}'_{n\text{рел}}.$$

Воспользуемся теперь тем обстоятельством, что в любой системе частиц с энергией $\mathscr{E}_{\text{рел}}$ и импульсом $\mathbf{p}_{\text{рел}}$ величина $\mathscr{E}^2_{\text{рел}} - (\mathbf{p}_{\text{рел}}c)^2$ инвариантна относительно выбора системы координат. Применим этот результат к нашей системе частиц. До столкновения возьмем указанный инвариант в лабораторной системе, а после столкновения — в системе центра масс. Так как в системе центра масс импульс равен нулю, то таким путем получим

$$(\mathscr{E}_{1\text{рел}} + m_2c^2)^2 - \mathbf{p}^2_{1\text{рел}}c^2 = (\mathscr{E}'_{1\text{рел}} + \mathscr{E}'_{2\text{рел}} + \ldots + \mathscr{E}'_{n\text{рел}})^2.$$

Напомним, что здесь $\mathscr{E}'_{1\text{рел}}, \mathscr{E}'_{2\text{рел}}, \ldots$ представляют полные (релятивистские) энергии образовавшихся частиц в системе центра масс. В общем случае частицы движутся относительно друг друга. Если же относительного движения частиц не возникает, то $\mathscr{E}_{1\text{рел}}$ будет минимальной релятивистской энергией частицы 1, при которой может начаться рассматриваемая реакция. Иными словами, $\mathscr{E}_{1\text{рел}}$ и будет порогом реакции, если в порог включить и энергию покоя частицы 1. Таким образом, значение порога найдется из предыдущего уравнения, если потребовать, чтобы в нем все величины $\mathscr{E}'_{1\text{рел}}, \mathscr{E}'_{2\text{рел}}, \ldots$ не содержали кинетических энергий, а являлись только энергиями покоя частиц, получающихся в результате реакции. Это дает

$$(\mathscr{E}_{1\text{рел}} + m_2c^2)^2 - \mathbf{p}^2_{1\text{рел}}c^2 = (m'_1 + m'_2 + \ldots + m'_n)^2c^4,$$

или, ввиду соотношения $\mathscr{E}^2_{1\text{рел}} - \mathbf{p}^2_{1\text{рел}}c^2 = \text{Inv} = (m_1c^2)^2$,

$$2\mathscr{E}_{1\text{рел}}m_2 + (m_1^2 + m_2^2)c^2 = (m'_1 + m'_2 + \ldots + m'_n)^2c^2.$$

Исключим теперь из величины $\mathscr{E}_{1\text{рел}}$ энергию покоя налетающей частицы m_1c^2. Тогда останется только кинетическая энергия этой частицы, которая и представляет собой пороговую энергию в обычном смысле

этого слова: $\mathscr{E}_{\text{пор}} \equiv \mathscr{E}_1 = \mathscr{E}_{1\text{рел}} - m_1c^2$. В результате простых преобразований получим

$$\mathscr{E}_{\text{пор}} = \frac{(m_1' + m_2' + \ldots + m_n' + m_1 + m_2)(m_1' + m_2' + \ldots - m_1 - m_2)c^2}{2m_2}. \tag{88.7}$$

Но

$$Q_{\text{энд}} = (m_1' + m_2' + \ldots + m_n' - m_1 - m_2)c^2, \tag{88.8}$$

так что

$$\mathscr{E}_{\text{пор}} = Q_{\text{энд}}\left(1 + \frac{m_1}{m_2} + \frac{Q_{\text{энд}}}{2m_2c^2}\right) \quad \text{(релятив.)}. \tag{88.9}$$

4. Эта формула упрощается в нерелятивистском приближении, когда $Q_{\text{энд}} \ll m_2c^2$. А в химии, хотя формула (88.8) и остается справедливой, проверить ее не удается из-за недостаточной точности измерения масс. В ядерной физике такая проверка не составляет особых затруднений, но все же в случае обычных ядерных реакций энергия $Q_{\text{энд}}$ мала по сравнению с энергией покоя мишени. Этим как раз и характеризуется нерелятивистский случай. Тогда квадратичным по $Q_{\text{энд}}$ членом в (88.9) можно пренебречь и получить

$$\mathscr{E}_{\text{пор}} = Q_{\text{энд}}\left(1 + \frac{m_1}{m_2}\right) \quad \text{(нерелятив.)}. \tag{88.10}$$

Разумеется, это приближение проще получить непосредственно из нерелятивистской механики, что и рекомендуется сделать читателю. Когда $m_1 \ll m_2$, то в нерелятивистском случае $\mathscr{E}_{\text{пор}} = Q_{\text{энд}}$. Это и понятно, так как в этом случае можно считать, что центр масс все время совпадает с центром второй частицы, т. е. остается неподвижным.

В ядерной физике обычно можно ограничиться нерелятивистской формулой (88.10). Но в процессах с участием ультрарелятивистских частиц, как правило, надо применять релятивистские формулы (88.7) и (88.9). Например, в случае расщепления атомного ядра под действием γ-кванта высокой энергии формула (88.9) переходит в

$$\mathscr{E}_{\text{пор}} = Q_{\text{энд}}\left(1 + \frac{Q_{\text{энд}}}{2m_2c^2}\right).$$

5. Рассмотрим теперь несколько примеров на применение релятивистских формул (88.7)–(88.9).

Пример 1. Рождение пары нуклон–антинуклон (например, нейтрон–антинейтрон) при столкновении двух протонов:

$$\mathrm{p} + \mathrm{p} \to \mathrm{p} + \mathrm{p} + \mathrm{n} + \bar{\mathrm{n}}. \tag{88.11}$$

В этом случае массы покоя всех частиц одинаковы (938 МэВ, если пренебречь различием масс протона и нейтрона) и формулы (88.7)

и (88.8) дают

$$\mathscr{E}_{\text{пор}} = \frac{6m_{\text{p}} \cdot 2m_{\text{p}}}{2m_{\text{p}}} c^2 = 6m_{\text{p}}c^2 = 5{,}63 \text{ ГэВ},$$

$$Q_{\text{энд}} = 2m_{\text{p}}c^2 = 1{,}88 \text{ ГэВ}.$$

Таким образом, порог этой эндотермической реакции в три раза превышает энергию $Q_{\text{энд}}$.

Пример 2. Пусть теперь в результате столкновения двух протонов рождается N нуклон–антинуклонных пар. Тогда

$$\mathscr{E}_{\text{пор}} = 2(N+2)Nm_{\text{p}}c^2, \quad Q_{\text{энд}} = 2Nm_{\text{p}}c^2 = \mathscr{E}_{\text{пор}}/(N+2).$$

Если, например, $N = 3$, то $\mathscr{E}_{\text{пор}} = 5Q_{\text{энд}}$, так что только одна пятая пороговой энергии участвует в реакции. Остальные четыре пятых уносятся с движением центра масс частиц.

Пример 3. Рождение пары электрон–позитрон. Два γ-кванта одинаковой энергии распространяются навстречу друг другу и взаимодействуют, в результате чего образуется пара электрон–позитрон:

$$\gamma + \gamma \to \text{e}^- + \text{e}^+. \tag{88.12}$$

В этом случае $m_1 = m_2 = 0$. Формулами (88.7) и (88.9) пользоваться нельзя, так как они предполагают, что частица 2 перед столкновением покоится, что для γ-кванта невозможно. Но формулой (88.8), конечно, пользоваться можно. Полагая в ней $m_1' = m_2' = m_{\text{e}}$, получим $Q_{\text{энд}} = 2m_{\text{e}}c^2$, что очевидно и без вычислений. Величина $Q_{\text{энд}}$ в рассматриваемом случае одновременно является и порогом реакции, если, конечно, порог определять в системе центра масс.

Пример 4. Фоторождение пиона на протоне:

$$\gamma + \text{p} \to \text{n} + \pi^+. \tag{88.13}$$

В этом случае

$$m_1 = 0, \quad m_2 = m_1' = m_{\text{p}} = 938 \text{ МэВ},$$

$$m_2' = m_\pi = 140 \text{ МэВ} = 0{,}149m_{\text{p}},$$

$$\mathscr{E}_{\text{пор}} = \frac{(m_{\text{p}} + m_\pi + m_{\text{p}})(m_{\text{p}} + m_\pi - m_{\text{p}})c^2}{2m_{\text{p}}} = 0{,}160m_{\text{p}}c^2 =$$

$$= 150 \text{ МэВ} = 1{,}07m_\pi c^2,$$

$$Q_{\text{энд}} = m_\pi c^2 = 140 \text{ МэВ}.$$

6. В ядерных реакциях действует закон сохранения электрического заряда, согласно которому алгебраическая сумма электрических зарядов частиц до реакции равна алгебраической сумме электрических зарядов частиц после реакции. Примером могут служить реакции (88.11)–(88.13). Наряду с этим законом в ядерной физике действует закон сохранения барионного заряда, аналогичный закону сохранения лептонного заряда (см. § 74, п. 10). В ядерных реакциях в узком смысле

при низких энергиях содержание этого закона сводится к тому, что суммарное число нуклонов не меняется в результате реакции. Но этот закон в расширенном смысле остается справедливым и при высоких энергиях, когда происходят превращения элементарных частиц (рождение античастиц). Хотя этот вопрос и выходит за рамки ядерной физики в узком смысле и рассматривается в физике элементарных частиц, мы считаем необходимым коротко остановиться на нем.

Под *барионами* понимают группу «тяжелых» элементарных частиц с полуцелым спином и массой не меньше массы протона. К ним относятся протон и нейтрон, гипероны, часть резонансов и «очарованных» частиц и, возможно, некоторые другие частицы. Как и у большинства элементарных частиц, у барионов существуют античастицы, называемые *антибарионами.* Они отличаются от барионов знаком некоторых характеристик (например, знаком электрического заряда и магнитного момента). Единственным стабильным барионом является протон (а антибарионом — антипротон). Остальные барионы *нестабильны* и путем последовательных распадов превращаются в протон и легкие частицы (например, в свободном состоянии нейтрон — нестабильная частица, но он становится стабильным в связанном состоянии — внутри стабильных ядер). Барионы участвуют во всех известных фундаментальных взаимодействиях: сильном, электромагнитном, слабом и гравитационном.

Во всех наблюдавшихся процессах разность между числами барионов и антибарионов оставалась постоянной. Этому результату можно придать форму закона сохранения, напоминающего закон сохранения электрического заряда. Для этого условились каждой частице приписывать определенный *барионный заряд.* Его условились считать равным $+1$ для бариона, -1 для антибариона и нулю для всех остальных частиц. Тогда сформулированный выше результат принимает форму закона сохранения барионного заряда, согласно которому суммарный барионный заряд системы частиц при всех процессах, совершающихся в ней, остается постоянным. Одно из проявлений этого закона состоит в том, что рождение всякого антибариона обязательно должно сопровождаться рождением дополнительного бариона.

До 70-х годов считалось, что закон сохранения барионного заряда строго выполняется для всех типов фундаментальных взаимодействий. Однако в связи с созданием различных моделей единой теории поля (так называемого «великого объединения» слабого, электромагнитного и сильного взаимодействий) строгая справедливость этого закона поставлена под сомнение. В частности, допускается возможность распада протона, например, по каналу

$$\mathrm{p} \to \mathrm{e}^+ + \pi^0$$

со временем жизни τ в различных моделях от 10^{30} до 10^{32} лет, что примерно в 10^{20}–10^{22} раз превосходит возраст наблюдаемой части Вселенной (согласно экспериментальным данным $\tau > 10^{32}$ лет). Это предсказание требует еще экспериментальной проверки, трудность которой

состоит в исключительно большом значении ожидаемого времени жизни протона.

7. Отметим теперь существенное обстоятельство, связанное с законом сохранения и квантованием момента импульса при низких энергиях сталкивающихся частиц. Орбитальный момент относительного количества движения двух сталкивающихся частиц может принимать только целочисленные значения $l = 0, 1, 2, \ldots$ (в единицах $\hbar$). Это обстоятельство в связи с ограниченным радиусом действия ядерных сил приводит к заключению, что реакция между частицами возможна практически только при небольших значениях l. Оценку максимального значения числа l строго можно получить на основе последовательной квантовой механики. Не имея возможности сделать это, воспользуемся полуклассическими представлениями в духе теории Бора, а затем дополним их качественными соображениями квантовой механики.

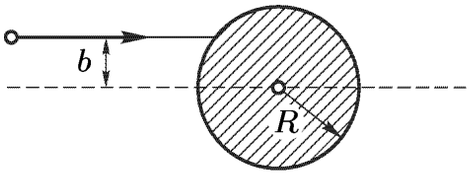


Рис. 157

Момент импульса частицы с импульсом p, налетающей на неподвижное ядро, равен pb, где b — прицельное расстояние (рис. 157). По классическим представлениям реакция может произойти только в тех случаях, когда b меньше радиуса действия ядерных сил: $b \leqslant R$, где R — радиус ядра, а потому $bp \leqslant Rp$ или $\hbar l \leqslant Rp$. Отсюда, вводя длину волны де Бройля, получаем

$$l \leqslant \frac{2\pi R}{\lambda} \approx \frac{R}{\lambda\!\!\!^{-}}. \tag{88.14}$$

Это и есть искомое ограничение. Оно существенно при больших значениях λ, т. е. при низких энергиях налетающей частицы.

С учетом волновых свойств частицы реакция в принципе возможна при любых значениях l, но вероятность реакции резко падает, если соотношение (88.14) не выполняется. Если полное сечение реакции представить в виде суммы $\sigma = \sum \sigma_l$, где σ_l — *парциальное сечение реакции*, т. е. сечение реакции, идущей при определенном значении l, то оказывается, что в случае электрически нейтральной частицы $\sigma_l \sim$ $\sim (2l+1)p^{4l} \sim (2l+1)\mathscr{E}^{2l}$, т. е. при уменьшении энергии частицы $\mathscr{E}$ сечение σ_l убывает тем быстрее, чем больше l. Для электрически заряженных частиц это убывание происходит еще быстрее. В этих случаях в СЦМ (но не в ЛС) угловое распределение вылетающих частиц сферически-симметрично, т. е. частицы разлетаются по всем направлениям с одинаковой вероятностью.

8. О лептонных зарядах и законах их сохранения уже говорилось (§ 74, п. 10). О четности состояния и законе сохранения четности было сказано в § 69. Напомним здесь, что закон сохранения четности выполняется в сильных и электромагнитных взаимодействиях и нарушается в слабых. Для ядерной реакции a+A $\to$ b+B закон сохранения четности требует выполнения равенства

$$P_a P_A (-1)^{l_a A} = P_b P_B (-1)^{l_b B}, \tag{88.15}$$

где P_a, P_A, P_b, P_B — внутренние четности взаимодействующих и образующихся частиц и ядер, а $l_a A$, $l_b B$ — орбитальные угловые моменты частиц a и b в относительном движении около ядер A и B.

При упругом рассеянии состояния ядра и бомбардирующей его частицы не изменяются. У них может произойти только переориентация спинов, при которой четность сохраняется. Но тогда из (88.15) следует, что $(-1)^{l_a A} = (-1)^{l_b B}$, т. е. l может изменяться только на четное число (в пределах, допускаемых законом сохранения момента импульса).

§ 89. Составное ядро

1. Многие ядерные реакции при невысоких энергиях проходят через стадию образования так называемого *составного*, или *промежуточного*, *ядра*. Представление о составном ядре было введено в физику в 1936 г. Нильсом Бором. Он руководствовался следующими соображениями. Частица, проникшая внутрь ядра, как правило, сильно взаимодействует с его нуклонами — настолько сильно, что ее энергия взаимодействия с отдельным нуклоном обычно того же порядка, что и кинетическая энергия самой частицы. Поэтому весьма вероятен захват частицы ядром. Частица застревает в ядре, причем из-за взаимодействия с нуклонами ядра ее энергия уменьшается настолько, что она длительное время не может покинуть ядро. Более того, частица, попавшая в ядро, вообще теряет свою индивидуальность и действует как система новых нуклонов, присоединившихся к прежним нуклонам ядра. Из-за множества столкновений между нуклонами в поведении новых и прежних нуклонов принципиально пропадает всякое различие. С другой стороны, если первоначальная энергия влетевшей частицы не слишком велика, в системе не окажется нуклонов столь большой энергии, чтобы один или несколько из них немедленно покинули ядро. Задерживаясь в ядре на некоторое время, нуклоны образуют систему частиц, которую в течение этого времени можно рассматривать как связанную. Эта система и есть составное, или промежуточное, ядро.

Составное ядро возникает в возбужденном состоянии и стремится потерять энергию возбуждения за счет какого-либо возможного для него процесса. В принципе оно не отличается от радиоактивного ядра. Один из возможных механизмов радиоактивного превращения состоит в том, что энергия захваченной частицы, беспорядочно распределившаяся между нуклонами составного ядра, в результате флуктуационных процессов вновь сконцентрируется на одной из частиц. Тогда такая частица и вылетит из ядра. Не обязательно, чтобы это была та же частица, которая влетела в ядро. Она может быть и другой: протон, нейтрон, α-частица и пр. Возможны несколько каналов радиоактивного распада составного ядра.

2. Необходимо особо подчеркнуть, что говорить о возбужденном составном ядре имеет смысл только тогда, когда оно существует достаточно долго. «Достаточно долго», конечно, надо понимать не в макроскопическом, а в ядерном масштабе. Например, чтобы нейтрон со

скоростью 10^9 см/с пролетел через ядро, не испытав столкновений, для него требуется время порядка $10^{-13} : 10^9 \approx 10^{-22}$ с. Это и есть «ядерное время», которым надо пользоваться для суждения о длительности процессов, происходящих в атомном ядре. Если составное ядро живет в возбужденном состоянии 10^{-14} с, то в ядерном масштабе это время должно считаться «очень большим». Из-за короткодействующего характера ядерных сил за это время влетевший нейтрон успеет претерпеть в ядре более $10^{-14} : 10^{-22} \approx 10^8$ столкновений, чего с избытком достаточно, чтобы его движение в ядре приобрело запутанный характер, совершенно не зависящий от того, с какой скоростью и в каком направлении нейтрон влетел в исходное ядро. Таким образом, возникает система нуклонов, поведение которой совершенно не зависит от истории ее образования. В ядерном масштабе времени такая система ведет себя так, как если бы она существовала бесконечно долго. А это и есть необходимое условие того, чтобы такую систему можно было рассматривать как составное ядро.

Наряду с описанным процессом образования и распада составного ядра возможен и конкурирующий процесс. Это процесс *радиационного захвата*. В этом процессе ядро переходит в возбужденное состояние также в результате захвата какой-то частицы. Но в основное состояние оно возвращается путем испускания γ-кванта. Последний процесс происходит под действием уже электромагнитных сил, т. е. в ядерном масштабе по-прежнему достаточно «медленно». Поэтому и при радиационном захвате может также образоваться составное ядро.

Заслуживает внимания тот парадоксальный с точки зрения классической физики факт, что силы кулоновского отталкивания между протонами и другими положительно заряженными частицами ядра (например, между α-частицами) не способствуют, а препятствуют выходу этих частиц из ядра. Об этом факте уже говорилось (§ 73, п. 12) в связи с влиянием центробежного барьера на α-распад. Объяснение его состоит в том, что силам отталкивания соответствует положительная энергия. Она увеличивает высоту, а с ней и ширину кулоновского потенциального барьера. Выход же протона и всякой положительно заряженной частицы из ядра есть *подбарьерный процесс*. Он тем менее вероятен, чем выше и шире потенциальный барьер. Особенно существенно это обстоятельство проявляется в случае средних и тяжелых ядер.

3. Таким образом, процесс столкновения частицы a с ядром A разбивается на два этапа. На первом этапе частица a сближается с ядром A. Завершением этапа является образование составного ядра C^* в возбужденном состоянии. Схематически этот процесс можно изобразить так:

$$a + A \to C^*. \tag{89.1}$$

На втором этапе происходит распад составного ядра по схеме

$$C^* \to b + B. \tag{89.2}$$

Ядерная реакция в целом изображается схемой

$$\mathrm{a} + \mathrm{A} \to \mathrm{C}^* \to \mathrm{b} + \mathrm{B}. \tag{89.3}$$

Эффективное сечение σ_{ab} реакции (89.3) найдется, если эффективное сечение σ_{C^*} образования составного ядра умножить на вероятность W_{b} распада этого ядра по каналу b:

$$\sigma_{\mathrm{ab}} = \sigma_{\mathrm{C}^*} W_{\mathrm{b}}. \tag{89.4}$$

Если то же составное ядро распадается по другому каналу b' с вероятностью $W_{\mathrm{b}'}$, то $\sigma_{\mathrm{ab}'} = \sigma_{\mathrm{C}^*} W_{\mathrm{b}'}$. Исключение σ_{C^*} дает

$$\frac{\sigma_{\mathrm{ab}}}{\sigma_{\mathrm{ab}'}} = \frac{W_{\mathrm{b}}}{W_{\mathrm{b}'}}. \tag{89.5}$$

Пусть теперь такое же составное ядро C^* с той же энергией возбуждения образуется в результате столкновения других частиц m и M: $\mathrm{m} + \mathrm{M} \to \mathrm{C}^*$. Ввиду тождественности этого ядра с прежним оно будет распадаться по тем же каналам с образованием тех же частиц b и b′ и с прежними вероятностями W_{b} и $W_{\mathrm{b}'}$, а потому $\sigma_{\mathrm{mb}}/\sigma_{\mathrm{mb}'} = W_{\mathrm{b}}/W_{\mathrm{b}'}$. Таким образом,

$$\frac{\sigma_{\mathrm{ab}}}{\sigma_{\mathrm{ab}'}} = \frac{\sigma_{\mathrm{mb}}}{\sigma_{\mathrm{mb}'}}. \tag{89.6}$$

Выполнение такого соотношения является подтверждением того, что обе реакции, рассмотренные выше, идут с образованием составного ядра. Если же соотношение (89.6) не выполняется, то схема с образованием составного ядра либо не выполняется, либо составное ядро еще не совсем успело сформироваться.

4. Другой характерной особенностью ядерных реакций, идущих через составное ядро, при выполнении определенных условий является *симметрия в системе центра масс* углового распределения разлетающихся частиц, которые образуются при распаде составного ядра. Предположим, что результирующий спин составного ядра равен нулю. Тогда частицы, образующиеся при распаде составного ядра, в системе центра масс будут с одинаковой вероятностью разлетаться вперед и назад. (Направление «вперед» задается скоростью налетающей частицы.) Про такое угловое распределение говорят, что оно обладает *симметрией «вперед-назад»*. Для обоснования высказанного утверждения заметим, что первоначальная система из-за наличия импульсов и угловых моментов у налетающей частицы и исходного ядра симметрией «вперед-назад» не обладает. Но такая симметрия появится в процессе формирования составного ядра. В самом деле, в системе центра масс результирующий импульс исходной системы равен нулю. При формировании составного ядра не только сохранится равенство нулю полного импульса, но и произойдет хаотическое распределение импульсов отдельных нуклонов по направлениям в пространстве. То же самое относится и к их спинам, так как по предположению спин составного ядра равен нулю. Остается только орбитальный момент импульса частиц. Но в системе центра масс он перпендикулярен к импульсу налетающей частицы. При формировании составного ядра он, конечно, сохранится,

но хаотически перераспределится между нуклонами составного ядра. Получится составное ядро, обладающее симметрией «вперед-назад». Естественно, что такая симметрия сохранится и при распаде составного ядра.

§ 90. Ядерные реакции, идущие через составное ядро

1. Энергетический спектр составного ядра — непрерывный. Действительно, пусть составное ядро образуется от слияния частицы a и другого ядра A. В начальном состоянии частица a и исходное ядро A бесконечно далеко удалены друг от друга. Движение такой системы двух частиц *инфинитно*, а потому ее полная энергия положительна и не должна квантоваться. Следовательно, она не будет квантоваться и после того, как частица проникнет в ядро, так как энергия сохраняется. Таким образом, энергия возбуждения составного ядра, вообще говоря, превышает энергию, которую необходимо затратить для удаления из ядра хотя бы одной частицы типа a.

Но если вероятность распада составного ядра достаточно мала, то имеет смысл говорить о почти стационарных, или *квазистационарных*, *состояниях* его, в которых оно длительное время совершает движение в ограниченной области пространства. Такое движение в течение ограниченного времени приближенно можно рассматривать как финитное. Время τ, в течение которого это движение допустимо, называется *временем жизни составного ядра.* Вероятность распада W ядра в единицу времени связана с τ соотношением $W = 1/\tau$. Для таких составных ядер приближенно можно говорить о *квазистационарных состояниях* и соответствующих им *квазистационарных уровнях энергии.* Но каждый квазистационарный уровень характеризуется не только энергией, но и определенной шириной Г, которая может быть определена с помощью соотношения неопределенностей

$$\Gamma = \frac{\hbar}{\tau}. \tag{90.1}$$

Таким образом, энергетический спектр составного ядра состоит из ряда дискретных полос конечной ширины. Строго говоря, такова же структура энергетического спектра и всякого радиоактивного ядра. Только радиоактивное ядро живет настолько долго, что вероятность его распада ничтожна, так что обычно нет надобности отличать квазистационарные уровни энергии от стационарных.

2. Применим изложенные соображения к вопросу об эффективном сечении ядерных реакций, предполагая для простоты, что налетающая частица электрически нейтральна. Главнейшими из таких частиц являются нейтроны. Для них, в отличие от положительно заряженных частиц, не существует кулоновского потенциального барьера, а потому они легко могут проникать в ядро и вызывать ядерные превращения. Подробнее ядерные реакции под действием нейтронов будут рассмотрены в гл. XIV. Сейчас же мы ограничимся кратким рассмотрением

вопроса о зависимости эффективного сечения реакции от скорости налетающего нейтрона. Последовательное и строгое решение этого вопроса (насколько в теории ядра вообще можно говорить о строгости) возможно лишь с использованием вычислительных методов квантовой механики, что в общем курсе физики сделать невозможно. Поэтому мы должны ограничиться лишь качественными, и нестрогими соображениями, а многие результаты привести вообще без обоснования.

3. Предположим, что кинетическая энергия нейтрона мала, так что энергия возникающего составного ядра значительно меньше энергии его первого возбужденного уровня. Этим исключаются резонансные явления в ядерных реакциях, о которых говорится ниже. Кроме того, предположим, что энергия нейтрона настолько мала, что длина волны де Бройля $\lambda\!\!\!^{-}$ значительно больше размеров ядра. Вероятность проникновения нейтрона в ядро, а с ней и среднее число актов распада N составного ядра в единицу времени пропорциональны $|\psi|^2$, где ψ — волна де Бройля для падающего нейтрона: $N \infty |\psi|^2$. С другой стороны, средний поток нейтронов на ядро $I \infty v|\psi|^2$, где v — средняя скорость нейтронов. Эти величины не зависят от длины волны де Бройля (а значит и от скорости v, так как последняя предполагается бесконечно большой по сравнению с размерами ядра. На основании определения эффективного сечения

$$\sigma = \frac{N}{I} \sim \frac{1}{v}. \tag{90.2}$$

Зависимость (90.2) носит название закона $1/v$. Ее происхождение легко понять и из других физических соображений: при уменьшении скорости нейтрона увеличивается время взаимодействия его с ядром, а это увеличивает вероятность захвата нейтрона ядром. Закон (90.2) имеет исключительно важное значение в ядерной энергетике и объясняет, почему в ядерных реакторах требуется замедление нейтронов (см. § 95).

4. Теперь рассмотрим случай, когда суммарная энергия $\mathscr{E}$ нейтрона и исходного ядра лежит в области расположения энергетических полос составного ядра. Согласно квантовой механике, если $\mathscr{E}$ равна энергии одного из квазистационарных уровней составного ядра, то вероятность образования последнего особенно велика. Сечение ядерных реакций при таких энергиях частиц резко возрастает, образуя так называемые *резонансные максимумы.* В этих случаях ядерные реакции называются *резонансными.* Вблизи резонансного уровня сечение реакции описывается зависимостью, напоминающей дисперсионную формулу оптики вблизи линии поглощения. Эта формула была получена в 1936 г. Брейтом (р. 1899) и Вигнером (р. 1902) и носит их имя.

Ограничимся случаем медленных нейтронов, когда достаточно учитывать лишь частицы с орбитальным моментом $l = 0$ (т. е. в s-состоянии). Кроме того, предположим, что на значение эффективного сечения оказывает заметное влияние только один резонансный уровень. Тогда для реакции A(n, b)B формула Брейта–Вигнера может быть

записана в виде

$$\sigma_{\mathrm{nb}} = \pi \lambda\!\!\!^{-2}_{\mathrm{n}} g \frac{\Gamma_{\mathrm{n}} \Gamma_{\mathrm{b}}}{(\mathscr{E} - \mathscr{E}_0)^2 + \Gamma^2/4}, \tag{90.3}$$

где $\mathscr{E}_0$ — энергия резонансного уровня, а $\lambda\!\!\!^{-}_{\mathrm{n}}$ — длина волны налетающего нейтрона. Величина Γ в знаменателе есть *полная ширина уровня*, равная сумме ширин уровней по всем возможным входным и выходным каналам реакции. В частном случае одного входного и одного выходного каналов $\Gamma = \Gamma_{\mathrm{n}} + \Gamma_{\mathrm{b}}$, где Γ_{b} соответствует поглощению нейтрона, т. е. испусканию частицы b, а Γ_{n} — упругому рассеянию нейтрона. Статистический весовой множитель g учитывает возможные ориентации моментов импульса частиц до столкновения и частиц, образовавшихся после столкновения. При этом предполагается, что нейтрон и исходное ядро A линейно поляризованы, т. е. их спины имеют определенные направления.

Для вычисления весового множителя g предположим, что налетающий нейтрон и ядро A не поляризованы, а их спины $\mathbf{I}_{\mathrm{n}}$ и $\mathbf{I}_{\mathrm{A}}$ ориентированы хаотически. Тогда существуют $2I_{\mathrm{n}} + 1$ возможных ориентаций нейтрона и $2I_{\mathrm{A}} + 1$ ориентаций ядра — всего $(2I_{\mathrm{n}} + 1)(2I_{\mathrm{A}} + 1)$ исходных состояний. В результате столкновения получится составное ядро со спином $\mathbf{I}$, которому соответствует $2I + 1$ возможных ориентаций. Если все эти ориентации равновероятны, то каждой паре линейно поляризованных нейтрона n и ядра A соответствует в среднем сечение, в g раз меньшее, чем в случае отсутствия поляризации. Таким образом,

$$g = \frac{2I + 1}{(2I_{\mathrm{n}} + 1)(2I_{\mathrm{A}} + 1)}. \tag{90.4}$$

Если числитель и знаменатель формулы (90.3) умножить и разделить на Γ, то эта формула приводится к виду (89.4), где

$$\sigma_{\mathrm{C}} = \pi \lambda\!\!\!^{-2}_{\mathrm{n}} g \frac{\Gamma_{\mathrm{n}} \Gamma}{(\mathscr{E} - \mathscr{E}_0)^2 + \Gamma^2/4}, \tag{90.5}$$

так как, очевидно, вероятность распада составного ядра по каналу b равна

$$W_{\mathrm{b}} = \frac{\Gamma_{\mathrm{b}}}{\Gamma}. \tag{90.6}$$

Следовательно, эффективное сечение образования составного ядра будет определяться выражением (90.5).

Так как спин нейтрона $I_{\mathrm{n}} = 1/2$, то $(2I_{\mathrm{n}} + 1) = 2$. В этом случае для спина составного ядра могут получиться только два значения: либо $I = I_{\mathrm{A}} + 1/2$, либо $I = I_{\mathrm{A}} - 1/2$.

При захвате медленного нейтрона возможны различные эффекты. Наиболее вероятным является испускание нейтрона (рассеяние) или γ-кванта (радиационный захват нейтрона). Возможны и другие процессы: например, ^{14}N при захвате нейтрона может испустить протон, а ^{10}B и ^{6}Li — α-частицу. Наиболее тяжелые ядра при захвате медленного нейтрона могут испытать деление. Как уже неоднократно говорилось, малая вероятность испускания заряженной частицы (α-частицы или

протона) в результате захвата медленного нейтрона связана с тем, что при вылете из ядра положительная частица должна преодолеть кулоновский потенциальный барьер. Поэтому вылет заряженных частиц при захвате медленных нейтронов в подавляющем большинстве случаев наблюдается лишь для самых легких ядер (^{10}B, ^{6}Li). Для большинства тяжелых ядер захват нейтронов сопровождается γ-излучением, а при захвате нейтронов легкими ядрами наиболее вероятным оказывается вылет из ядра нейтронов же, т. е. осуществляется реакция упругого рассеяния нейтронов. Резонансное сечение захвата тепловых нейтронов может в 10^5–10^6 раз превосходить πR^2.

5. Зависимость эффективного сечения σ_{nb} от скорости v налетающего нейтрона определяется не только резонансным знаменателем $(\mathscr{E}-\mathscr{E}_0)^2+\Gamma^2/4$ в формуле (90.3), но и числителем $\Gamma_{\mathrm{n}}\Gamma_{\mathrm{b}}$. Когда энергия $\mathscr{E}$ составного ядра очень близка к $\mathscr{E}_0$, основное значение имеет резонансный знаменатель. При $|\mathscr{E}-\mathscr{E}_0|<\Gamma/2$ слагаемым $(\mathscr{E}-\mathscr{E}_0)^2$ можно пренебречь. Тогда по формуле (90.3) $\sigma_{\mathrm{nb}}=4\pi\lambda\!\!\!^{-2}g\Gamma_{\mathrm{n}}\Gamma_{\mathrm{b}}/\Gamma^2$, т. е. в сечении σ_{nb} возникает резкий максимум. Такие резкие максимумы для медленных нейтронов могут во много раз (например, в тысячи) превосходить геометрическое сечение ядра πR^2. Напротив, когда $\mathscr{E}\ll\mathscr{E}_0$, что имеет место для очень медленных нейтронов, знаменатель в формуле (90.3) меняется со скоростью мало, и этим изменением можно пренебречь. Можно также не учитывать зависимость Γ_{b} от v, так как Γ_{b} определяется только вероятностью распада самого составного ядра. Зависимость эффективного сечения от v определяется только множителями $\lambda\!\!\!^{-2}_{\mathrm{n}}$ и Γ_{n}, а именно $\sigma_{\mathrm{nb}}\sim\lambda\!\!\!^{-2}_{\mathrm{n}}\Gamma_{\mathrm{n}}$. Но $\Gamma_{\mathrm{n}}=\hbar/\tau$, а время τ, в течение которого нейтрон проникает в поле ядра, пропорционально v, так что $\Gamma_{\mathrm{n}}\sim v$. Кроме того, $\lambda\!\!\!^{-}_{\mathrm{n}}\sim 1/v$. Поэтому $\sigma_{\mathrm{nb}}\sim(1/v)^2v\sim 1/v$. Это — закон (90.2), полученный выше другим путем. Как правило, он оправдывается для нейтронов, энергии которых лежат в интервале от 1/40 эВ (тепловые нейтроны) до 1 эВ. Впрочем, для легких ядер (He, Li, B), у которых первый резонансный уровень расположен высоко, зависимость (90.2) остается справедливой и при энергиях вплоть до нескольких сотен электронвольт.

6. Теперь рассмотрим качественно нерезонансные реакции, попрежнему идущие через составное ядро. Они возникают, например, тогда, когда налетающая частица сообщает ядру высокую энергию возбуждения, так что его энергетические уровни перекрываются. В этих случаях говорить об отдельных резонансных уровнях уже не приходится. Но представление о составном ядре можно дополнить статистическими соображениями. В результате получится *статистическая теория ядерных реакций*, или *модель испарения*. Частица, попавшая в ядро, быстро теряет свою энергию, передавая ее всем нуклонам ядра. Возникает равновесное внутреннее состояние ядра, аналогичное термодинамическому равновесию макроскопической системы. Можно даже для характеристики такого состояния ввести некоторую величину, подобную обычной термодинамической температуре. В течение некоторого времени — времени жизни составного ядра — энергии каждого

нуклона недостаточно для вылета из ядра, хотя ядро в целом сильно возбуждено. Наконец, по истечении некоторого времени возникает какая-то большая флуктуация, в результате которой один из нуклонов (обычно нейтрон) получает энергию, достаточную для «испарения» из ядра. Затем может испариться другой нуклон, и т. д.

Следует ожидать, что в модели испарения составного ядра угловое распределение частиц, образующихся после распада составного ядра, в системе центра масс должно быть сферически-симметричным, а не только обладать симметрией «вперед-назад». Это должно происходить потому, что сферически-симметрично само составное ядро, поскольку его симметрия устанавливается в процессе достижения термодинамического равновесия. На самом деле, как показывает опыт, угловое распределение, как правило, содержит анизотропную часть, не обладающую даже симметрией «вперед-назад»: обычно бóльшая часть частиц летит вперед. Следует ожидать также, что в модели испарения процентная доля частиц, образующихся при распаде составного ядра по определенному каналу, не должна зависеть от способа образования составного ядра, так как в процессе установления термодинамического равновесия ядро совершенно «забывает» о способе своего образования. Эта закономерность обычно выполняется тоже лишь приближенно. Наконец, распределение по энергиям вылетающих частиц не полностью согласуется с тем, что предсказывает модель испарения. Напомним еще, что и в случае резонансных ядерных реакций, идущих через составное ядро, наблюдаются некоторые расхождения с опытом. Это свидетельствует о том, что помимо реакций, идущих через составное ядро, имеются и другие существенные механизмы ядерных реакций.

§ 91. Дополнительные сведения о ядерных реакциях

1. Помимо ядерных реакций, идущих через составное ядро, существуют *прямые ядерные реакции.* В таких реакциях энергия, вносимая в ядро, передается преимущественно одному или небольшой группе нуклонов, которые и покидают ядро. Прямые ядерные реакции могут вызываться всевозможными налетающими частицами — от γ-квантов до многозарядных ионов во всем доступном диапазоне энергий (до нескольких гигаэлектронвольт и выше). Первые из таких реакций были открыты в начале 50-х годов. Это были *реакции срыва* (d, p) и (d, n), а также *реакции подхвата* (p, d) и (n, d) при взаимодействии дейтронов с легкими ядрами. В реакции срыва (d, p), например, налетающий дейтрон, взаимодействуя с одним или несколькими периферийными нуклонами ядра, лишается своего нейтрона, который остается в ядре, а протон продолжает лететь дальше. Напротив, в реакции подхвата (p, d) налетающий протон уносит один из периферийных нейтронов, так что из ядра вылетает дейтрон. Образующиеся в этих реакциях протон и дейтрон вылетают в основном вперед (т. е. в направлении пучка налетающих частиц).

Известны прямые ядерные реакции, в которых нуклон (или группа нуклонов) переходит от одного из сталкивающихся ядер к другому (*реакции передачи*), прямые реакции квазиупругого рассеяния, например (p, 2p), реакции выбивания из ядра дейтронов, например (p, pd), и т. д. Все прямые ядерные реакции характеризуются сильной угловой анизотропией вылета частиц и сравнительно слабой зависимостью сечения σ от энергии налетающей частицы. Ядро, остающееся в результате прямой ядерной реакции, находится, как правило, либо в основном, либо в слабовозбужденном состоянии.

Особенности прямых ядерных реакций можно понять, если предположить, что вылетающая из ядра частица получает энергию и импульс в результате взаимодействия непосредственно с налетающей частицей. Надо также предположить, что такие процессы происходят на периферии атомного ядра. Это уже предполагалось, когда выше говорилось о реакциях срыва и подхвата. На периферии ядра плотность нуклонов заметно меньше, чем в центре, так что нуклон (или группа нуклонов), получив энергию и импульс от влетевшей частицы, может со значительной вероятностью покинуть ядро. Толщина периферического слоя ядра порядка 1 ферми, а радиус тяжелого ядра — 10 ферми. Поэтому относительная вероятность прямой ядерной реакции для тяжелого ядра должна быть около 10 % (у легких ядер она больше), что подтверждается опытом.

2. При низких энергиях налетающих частиц основным механизмом ядерных реакций является образование составного ядра (за исключением ядерных реакций с дейтронами). Если энергия возбуждения составного ядра меньше энергии отделения от него нуклона или системы нуклонов, то единственным способом его распада является испускание γ-квантов (радиационный захват). Если же нуклоны или частицы, образованные из них, могут вылетать из ядра, то этот вылет может происходить еще до того, как сформируется равновесное составное ядро (так называемый *предравновесный распад*). При больших энергиях налетающих частиц преобладают прямые процессы.

Если энергия налетающей частицы превышает 100 МэВ, то представления Бора о составном ядре перестают оправдываться. Влетая в ядро и сталкиваясь с его нуклонами, частица высокой энергии не успевает потерять всю свою энергию, так как число столкновений ее с нуклонами может оказаться для этого недостаточным. Частица вылетает из ядра, потеряв лишь часть своей энергии. Ядерная реакция при больших энергиях обычно (но не всегда) проходит через две стадии. Сначала частица выбивает из ядра несколько быстрых нуклонов, энергия и угловое распределение которых зависят от энергии налетающей частицы и параметров столкновения. Часть вторичных нуклонов в течение некоторого времени совершает запутанное движение внутри ядра, в результате чего образуется составное ядро, сильно отличающееся от ядра-мишени. На второй стадии происходит распад составного ядра с вылетом из него нуклонов, групп нуклонов или элементарных частиц — пионов, каонов, гиперонов и пр. При очень высоких энергиях налетающих частиц (несколько сотен мегаэлектронвольт и выше) могут

происходить «взрывы» ядер, в результате которых ядро разваливается на несколько более мелких осколков. Такой взрыв оставляет в фотоэмульсиях или пузырьковых камерах треки, образующие многолучевую «звезду». Соответствующие процессы называются *процессами с образованием звезд.*

Время протекания прямых ядерных реакций — это время, необходимое частице, чтобы пролететь область пространства, заполненную ядром (порядка 10^{-22} с), тогда как среднее время жизни составного ядра много больше (порядка 10^{-14}–10^{-16} с).

3. Эффективное сечение прямой ядерной реакции $\sigma(\mathscr{E})$ зависит от энергии $\mathscr{E}$ монотонно, тогда как для реакций, идущих через составное ядро, как мы видели, наблюдаются резонансные максимумы. Резонансные максимумы энергии составного ядра налагаются друг на друга и энергетический спектр приобретает сплошной характер, когда энергия налетающей частицы достаточно велика (1 МэВ и выше). В этих случаях ход поперечного сечения реакции в зависимости от энергии достаточно удовлетворительно передается так называемой *оптической моделью ядра.* В оптической модели ядро рассматривается как сплошная среда, преломляющая и поглощающая волны де Бройля падающих на него частиц. Тем самым задача многих тел формально сводится к задаче о движении одной частицы в заданном силовом поле.

Свойства ядра в оптической модели по аналогии с оптикой характеризуются *показателем преломления.* Для учета поглощения волн де Бройля этот показатель считается комплексным и выражается через гамильтониан взаимодействия налетающей частицы с ядром, который также выбирается комплексным. Для согласования с опытными данными наряду с объемным поглощением ядра вводится еще *поверхностное поглощение.* Предельным частным случаем оптической модели является *модель черного ядра.* Она характеризуется тем, что всякая частица, упавшая на ядро, поглощается последним. Однако для согласования с опытом в случае больших энергий падающих частиц черное ядро приходится считать частично прозрачным.

Недостаток оптической модели состоит в том, что все параметры ядра — комплексный показатель преломления, комплексный гамильтониан, поверхностное поглощение, прозрачность черного ядра — подбираются эмпирически и вводятся искусственно для описания некоторых свойств ядер, а не получаются из какой-то последовательной физической теории ядра. Но этот недостаток присущ всякой искусственно вводимой модели, а не только оптической модели ядра.

4. Остановимся на некоторых особенностях ядерных реакций, отличающихся различной природой бомбардирующих частиц. Сюда относятся ядерные реакции под действием нейтронов, протонов, дейтронов, α-частиц, многозарядных (тяжелых) ионов, γ-квантов.

Реакции под действием нейтронов, протонов и α-частиц во многих отношениях сходны между собой. Причина этого — в одинаковости механизма этих реакций: в большинстве случаев все они идут с образованием и последующим распадом составного ядра. Различие же

между ними в основном связано с различием зарядов бомбардирующих частиц. Оно сказывается на проницаемости кулоновского барьера, когда бомбардирующая частица находится за пределами ядра, и становится мало существенным для последующего распада образовавшегося составного ядра, когда начинают действовать ядерные силы, определяющие механизм реакции.

Сечения ядерных реакций, обусловленные захватом α-частиц малой (около 1 кэВ) и средней (от 1 кэВ до 1 МэВ) энергий, ничтожно малы даже при взаимодействии с легкими ядрами и быстро возрастают с увеличением энергии. Под действием α-частиц идут преимущественно реакции типа (α, p) и (α, n). Альфа-частицы, возникающие при радиоактивных распадах, могут вызывать реакции только на легких ядрах ($A \lesssim 50$), для которых высота кулоновского потенциального барьера порядка 10 МэВ, причем вероятность реакции (α, n) больше, чем реакции (α, p), так как испусканию протонов препятствует кулоновский потенциальный барьер, а для нейтронов такого барьера нет. В случае тяжелых ядер ($A \gtrsim 100$) высота кулоновского барьера для α-частиц около 25 МэВ. Когда энергия α-частицы превышает это значение, реакции типа (α, p) и (α, n) идут практически с одинаковой вероятностью. При реакциях типа (α, p), как правило, образуются стабильные ядра, а при реакциях типа (α, n) — радиоактивные. Примером реакции типа (α, n) может служить реакция

$$^{9}_{4}\mathrm{Be} + {}^{4}_{2}\mathrm{He} \to {}^{12}_{6}\mathrm{C} + {}^{1}_{0}\mathrm{n} + 5{,}6\ \text{МэВ},$$

до настоящего времени широко используемая в качестве простейшего источника нейтронов (см. § 85).

Реакции, идущие с вылетом заряженных частиц, в сильной степени зависят от прозрачности кулоновского барьера. По этой причине поглощение медленных нейтронов приводит к испусканию либо γ-квантов (радиационный захват), либо нейтронов первоначальной энергии (упругое рассеяние). Исключение составляют только самые легкие ядра. Но и на самых легких ядрах при малых энергиях падающих частиц ядерные реакции с испусканием заряженных частиц наблюдаются очень редко. У самых тяжелых ядер (например, у $^{235}\mathrm{U}$) возможен также процесс деления. Таким образом, в области малых энергий основной вклад в полную ширину уровня Γ вносят радиационная Γ_γ и нейтронная Γ_n ширины. Остальные ширины пренебрежимо малы.

5. В соответствии с формулой (89.4) сечения радиационного захвата $\sigma_{\mathrm{n}\gamma}$ и упругого рассеяния σ_{nn} можно представить в виде

$$\sigma_{\mathrm{n}\gamma} = \frac{\sigma_\mathrm{n}\Gamma_\gamma}{\Gamma_\gamma + \Gamma_\mathrm{n}}, \tag{91.1}$$

$$\sigma_{\mathrm{nn}} = \frac{\sigma_\mathrm{n}\Gamma_\mathrm{n}}{\Gamma_\gamma + \Gamma_\mathrm{n}}, \tag{91.2}$$

где σ_{n} — сечение образования составного ядра при захвате нейтрона. Следовательно,

$$\frac{\sigma_{\text{n}\gamma}}{\sigma_{\text{nn}}} = \frac{\Gamma_\gamma}{\Gamma_{\text{n}}}. \tag{91.3}$$

У тяжелых ($A > 100$) ядер $\Gamma_\gamma > \Gamma_{\text{n}}$. Следовательно, поглощение медленных нейтронов тяжелыми ядрами в основном приводит к испусканию γ-квантов. Однако с увеличением энергии падающей частицы увеличивается и энергия возбуждения составного ядра. Это ведет к увеличению нейтронной ширины Γ_{n}. При энергии падающего нейтрона 1–10 кэВ Γ_{n} начинает превосходить Γ_γ. При переходе к еще большим энергиям падающей частицы упругое рассеяние нейтронов начинает преобладать над радиационным захватом даже у тяжелых ядер. Для относительно легких ядер ($A \approx 50$) и энергии падающих нейтронов порядка 100 эВ Γ_{n} достигает величины порядка нескольких электронвольт и оказывается больше Γ_γ.

Поглощение нейтронов ядрами $^{10}_{5}\text{B}$ в широком диапазоне энергий (вплоть до 10 кэВ) следует закону $1/v$. Сечение реакции $^{10}_{5}\text{B}(\text{n}, \alpha)^{7}_{3}\text{Li}$ примерно в 10^5 раз превосходит сечение радиационного захвата нейтрона. Столь большая вероятность испускания заряженной α-частицы связана с тем, что захват нейтрона ядром сопровождается выделением энергии, значительная часть которой (1,77 МэВ) уносится α-частицей. Поэтому кулоновский барьер не оказывает α-частице серьезного препятствия для вылета из ядра. Следует отметить, что рассматриваемая реакция включает в себя своеобразное деление составного ядра $^{11}_{5}\text{B}$ на два осколка ($^{4}_{2}\text{He}$ и $^{7}_{3}\text{Li}$).

6. При больших энергиях падающих нейтронов становится возможным неупругое рассеяние (n, n'), при котором конечное ядро получается не в основном, а в одном из возбужденных состояний. В этом случае оказывается, что Γ_{n} значительно превышает все остальные ширины $\Gamma_{\text{p}}, \Gamma_\gamma, \Gamma_\alpha$ и т. д., так что $\Gamma \approx \Gamma_{\text{n}}$. Поэтому сильно возбужденное ядро с подавляющей вероятностью переходит в основное состояние с испусканием нейтронов, а вероятности реакций (n, p), (n, α), ... незначительны. Если бы эти последние реакции шли через составное ядро, то следовало бы ожидать, что их сечения должны представляться выражениями

$$\sigma_{\text{np}} \approx \frac{\pi R^2 \Gamma_{\text{p}}}{\Gamma}, \quad \sigma_{\text{n}\alpha} \approx \frac{\pi R^2 \Gamma_\alpha}{\Gamma}.$$

А так как $\Gamma_{\text{p}} \ll \Gamma$, $\Gamma_\alpha \ll \Gamma$ и т. д., то сечения этих реакций должны были бы быть значительно меньше геометрического сечения πR^2. Этот вывод не подтверждается опытом. Измерения для большого числа ядер по неупругому рассеянию нейтронов с энергией 14 МэВ показали, что действительные сечения во много раз превышают те, которые следует ожидать согласно модели составного ядра. Значит, рассматриваемые реакции, по крайней мере в значительной части, идут не через составное ядро, а являются реакциями прямого взаимодействия. Это может

служить иллюстрацией того, что было сказано в начале настоящего параграфа.

7. Под действием протонов идут реакции (p, α), (p, n), (p, p), (p, γ) и с меньшей вероятностью — (p, d). Реакции (p, α) обычно экзотермические и идут преимущественно на легких ядрах, так как выход α-частиц из тяжелых ядер сильно затруднен кулоновским барьером. Напротив, реакции типа (p, n), как правило, эндотермические и обладают порогом 1–3 МэВ. В результате таких реакций заряд ядра увеличивается на единицу, а потому оно становится β^+- или K^+-активным. Примером могут служить реакции

$$^7_3\mathrm{Li} + {}^1_1\mathrm{p} \to {}^7_4\mathrm{Be} + {}^1_0\mathrm{n}, \quad {}^7_4\mathrm{Be} \to {}^7_3\mathrm{Li} + \mathrm{e}^+ + \nu.$$

Реакции (p, p) при энергии падающих протонов, превышающих высоту кулоновского барьера, идут примерно с той же вероятностью, что и реакции (p, n). Реакции типа (p, γ) обычно характеризуются меньшим выходом, так как Γ_γ меньше Γ_i для всех других частиц ($i = \alpha$, n, p); такие реакции приобретают большое значение только в тех случаях, когда вылет частицы i из ядра почему-либо затруднен. Реакции типа (p, d), как правило, эндотермические. Они встречаются реже остальных, так как дейтрон — слабо связанное ядро. Энергия связи дейтрона равна 2,23 МэВ, и нуклоны в дейтроне находятся на довольно большом расстоянии друг от друга (порядка $4 \cdot 10^{-13}$ см).

Таким образом, дейтрон представляет собой сравнительно «рыхлое» ядро, и эта рыхлость проявляется также в реакциях под действием дейтронов, когда энергия падающей частицы составляет несколько мегаэлектронвольт. Из-за относительно больших размеров дейтрона входящий в него нейтрон может проникнуть в ядро мишени, тогда как протон продолжает находиться еще вне поля действия ядерных сил, а со стороны ядра подвергается только действию кулоновских сил отталкивания и по этой причине так и не проникает в ядро. Именно это кулоновское отталкивание вызывает развал дейтрона и приводит к реакции срыва (d, p), происходящей под действием дейтронов. Существование срыва объясняет, почему реакция (d, p) является преобладающей над всеми реакциями, производимыми дейтронами. Понятен тот факт, почему при малых энергиях реакция (d, p) более вероятна, чем реакция (d, n). Также понятно, почему выход реакции под действием дейтронов гораздо больше, чем соответствующий выход под действием других частиц. Продукты реакции в системе центра масс летят преимущественно вперед, что указывает на то, что все реакции срыва идут не через составное ядро.

Важная особенность реакции срыва состоит в том, что в результате нее получаются изотопы ядер мишеней с малой энергией возбуждения, которая в ряде случаев меньше энергии связи по отношению к отделению нейтрона. Этим открывается практическая возможность в реакции срыва под действием дейтронов получать ядра в слабовозбужденных состояниях. Процесс срыва остается основным и в ядерных реакциях под действием дейтронов при высоких энергиях. Только в этих случаях кулоновское отталкивание протонов не играет существенной роли

и обе реакции срыва (d, p), и (d, n) становятся практически одинаково вероятными.

8. Остановимся теперь кратко на ядерных реакциях, идущих под действием γ-квантов. Такие реакции называются также *фотоядерными.* Они вызываются не ядерными, а *электромагнитными взаимодействиями.*

При малых энергиях γ-квантов последние могут испытывать при взаимодействии с ядрами только упругое рассеяние. При возрастании энергии γ-квантов возможны реакции типа (γ, n), (γ, p), $(\gamma, 2\mathrm{n})$, (γ, pn), (γ, α) и пр. Эти реакции аналогичны поглощению γ-квантов атомами, а потому все они называются *ядерным фотоэффектом.* В делящихся ядрах (см. § 93) с большой вероятностью идет реакция фотоделения ядра (γ, f). При энергиях, больших порога рождения мезонов, наряду с расщеплением ядра протекают процессы *фоторождения* (например, пионов).

Для того чтобы какая-либо из фотоядерных реакций могла идти, необходимо, чтобы энергия γ-кванта превосходила энергию освобождения соответствующей частицы или группы частиц из ядра. Например, для вылета нейтрона должно быть $\hbar\omega > \mathscr{E}_{\mathrm{n}}$, и аналогично для протона и α-частицы. Энергия γ-квантов, испускаемых естественными радиоактивными веществами, для этого недостаточна, за исключением только двух случаев, когда величина $\mathscr{E}_{\mathrm{n}}$ мала. Эти случаи осуществляются в реакциях

$$\gamma + {}^{2}_{1}\mathrm{H} \to {}^{1}_{0}\mathrm{n} + {}^{1}_{1}\mathrm{p}, \quad \gamma + {}^{9}_{4}\mathrm{Be} \to {}^{8}_{4}\mathrm{Be} + {}^{1}_{0}\mathrm{n}.$$

В других случаях используются γ-кванты, получаемые на ускорителях в реакции (p, γ) или при тормозном излучении электронов. Необходимо отметить, что в экспериментальном отношении исследование фотоядерных реакций (γ, n) и (γ, p) представляет значительно большие трудности по сравнению с исследованием реакций, вызываемых нейтронами и в особенности протонами и α-частицами, так как γ-кванты, получаемые на ускорителях, в высшей степени не монохроматичны.

Особенностью фотоядерных реакций является существование *гигантских резонансов* в сечениях поглощения γ-квантов — больших и широких максимумов в зависимости эффективных сечений $\sigma_{\gamma\mathrm{n}}$ и $\sigma_{\gamma\mathrm{p}}$ от энергии $\mathscr{E}_{\gamma}$. У легких ядер (^{12}C, ^{16}O) этот максимум расположен в области 20–25 МэВ, у средних и тяжелых — в области 13–18 МэВ. Ширина максимума около 3–4 МэВ. Основной вклад в полное сечение поглощения γ-квантов в области гигантского резонанса вносят реакции (γ, p) и (γ, n). Сечение поглощения γ-квантов ядрами даже в области гигантского резонанса равно $0{,}05 \cdot 10^{-24}$–$0{,}1 \cdot 10^{-24}$ см^2, т. е. на 1–2 порядка меньше сечения поглощения таких же γ-квантов, вызываемого атомными электронами. Таким образом, фотоядерные реакции слабо влияют на поглощение γ-излучения веществом.

Энергетическое и угловое распределение вылетающих частиц при ядерном фотоэффекте не согласуется с концепцией Бора о составном

ядре. Так, выход фотопротонов и фотонейтронов из ядра обычно больше ожидаемого по концепции Бора, угловое распределение вылетающих быстрых протонов и нейтронов анизотропно и переходит в изотропное для медленных. Эти факты говорят в пользу предположения, что поглощение γ-кванта происходит на поверхности ядра одним или несколькими нуклонами.

9. Наряду с ядерным фотоэффектом отметим еще *электроядерные процессы*, проявляющиеся при бомбардировке ядра заряженными частицами, например протонами или α-частицами. В этом случае идет конкуренция между процессами, происходящими под действием ядерных и электрических сил. Ядерные силы в области их действия примерно в 100–1000 раз сильнее электрических. Зато они очень быстро убывают с расстоянием и практически перестают действовать вне ядра. Электрические силы, напротив, слабо убывают с расстоянием, и поэтому именно они определяют поведение налетающей заряженной частицы вне ядра. Этим объясняется, почему α-частица, например, при высоких энергиях свободно проникает в ядро и вступает с ним в интенсивное ядерное взаимодействие. При высоких энергиях влияние электрических сил ничтожно. При низких энергиях, напротив, α-частица не может приблизиться к ядру вплотную и взаимодействует с ним только через электрическое поле. Граничной является энергия $\mathscr{E}_{\text{гр}}$ α-частицы, равная примерно кулоновской энергии при соприкосновении ядра с α-частицей, т. е.

$$\mathscr{E}_{\text{гр}} = \frac{2Ze^2}{R+r},$$

где $2e$ — заряд α-частицы, r — ее радиус, Ze — заряд ядра, R — его радиус. Для средних ядер $\mathscr{E}_{\text{гр}} \sim 10$ МэВ.

Кулоновское взаимодействие тяжелых заряженных частиц с ядрами вызывает не только упругое рассеяние этих частиц, но и может приводить к неупругому рассеянию *с кулоновским возбуждением ядра.* Хотя этот процесс приводит к возбуждению только низких уровней ядра, однако он представляет интерес в связи с тем, что позволяет измерять внутренние квадрупольные моменты ядер Q_0 (см. § 70).

Взаимодействие электронов с ядрами при всех энергиях происходит исключительно через электромагнитное поле (если пренебречь слабым взаимодействием), так как электроны в сильных взаимодействиях не участвуют. Особое значение имеет изучение упругого рассеяния электронов высоких энергий на ядрах и на отдельных нуклонах. Такой процесс поддается точному расчету и позволяет измерять распределение электрического заряда в ядрах и нуклонах.

Глава XIV

НЕЙТРОНЫ И ДЕЛЕНИЕ АТОМНЫХ ЯДЕР

§ 92. История открытия нейтрона

1. Нейтрон в ядерной физике и ее приложениях играет не меньшую роль, чем электрон в электронике. Хотя при изучении физики ядра мы все время имели дело с нейтронами, до сих пор ничего не было сказано об истории открытия этих частиц. А этот вопрос в учебнике, несомненно, представляет определенный интерес. Заполним теперь коротко этот пробел.

Первая искусственная ядерная реакция

$$^{14}_{7}\mathrm{N} + ^{4}_{2}\mathrm{He} \to ^{17}_{8}\mathrm{O} + ^{1}_{1}\mathrm{H} \qquad (92.1)$$

наблюдалась Резерфордом в 1919 г. Протоны в этой реакции регистрировались по сцинтилляциям, вызываемым ими на экране из сернистого цинка. После открытия реакции (92.1) Резерфордом, Чедвиком и др. началось энергичное изучение аналогичных ядерных реакций. Однако до 1931 г. не существовало ускорителей, пригодных для этой цели. Искусственные превращения атомных ядер пытались осуществлять путем облучения их α-частицами, особенно наиболее энергичными, испускаемыми $^{214}_{84}$Po (радиохимическое обозначение этого изотопа RaC′). Было установлено, что многие элементы, преимущественно легкие, у которых кулоновский барьер сравнительно невысок, при таком облучении испускали протоны. Однако среди самых легких элементов оказались и такие, например Be и Li, которые при облучении α-частицами протонов не испускали.

2. Был поставлен вопрос, не появляется ли при α-облучении различных веществ какое-то новое излучение, которое не дает сцинтилляций на экране из сернистого цинка. Для ответа на этот вопрос Боте (1891–1957) и Г. Беккер в конце 1930 г. заменили экран из сернистого цинка счетчиком Гейгера. В качестве источника α-частиц был взят $^{218}_{84}$Po, хотя энергия испускаемых им α-частиц и не очень велика (5,25 МэВ). Такой выбор был сделан потому, что $^{218}_{84}$Po испускает только α-частицы и не дает β- и γ-излучений, осложняющих опыт. Между источником α-частиц и счетчиком Гейгера вводилась пластинка из исследуемого вещества, например бериллия, которая и подвергалась облучению. За облучаемой пластинкой можно было вводить свинцовые пластины различной толщины для суждения о поглощении неизвестных лучей, если таковые будут обнаружены.

Опыт действительно показал, что при облучении некоторых веществ (бериллий, бор, литий) возникает какое-то излучение, вызывающее слабые разряды счетчика Гейгера. Особенно сильное излучение

давал бериллий. *Бериллиевые лучи*, как мы будем временно называть неизвестное излучение, оказались весьма проникающими. При прохождении через слой свинца толщиной 2 см интенсивность бериллиевого излучения уменьшалась всего на 13 %. На основании этого можно было заключить, что бериллиевые лучи должны состоять из электрически нейтральных частиц, так как заряженные частицы ионизуют среду и поэтому поглощаются значительно сильнее. Из нейтральных же частиц в 1930 г. были известны только фотоны. Поэтому Боте и Беккер вполне естественно предположили, что бериллиевые лучи состоят из γ-квантов, и притом очень жестких. Однако при более подробных исследованиях это предположение не подтвердилось.

Если бы бериллиевые лучи представляли собой γ-излучение, то по их поглощению в свинце, например, можно было бы оценить энергию соответствующего γ-кванта. Таким путем было найдено, что максимальная энергия γ-кванта должна быть около 7 МэВ. А это не согласовывалось с результатами, найденными из других опытов. Кроме того, поглощение γ-лучей тем больше, чем больше плотность поглотителя. Для бериллиевых лучей соотношение оказалось обратным.

3. И. Кюри и Ф. Жолио-Кюри в 1932 г. повторили опыты Боте и Беккера, заменив счетчик Гейгера ионизационной камерой. Помещая между источником бериллиевых лучей и ионизационной камерой листы парафина, они обнаружили, что из парафина исходили протоны, максимальный пробег которых в воздухе достигал 26 см. Такому пробегу соответствовала энергия протона $\mathscr{E} \approx 4{,}5$ МэВ. Появление протонов отдачи в этих опытах было непосредственно доказано путем наблюдения треков в камере Вильсона. Подобно парафину действуют также все водородсодержащие вещества.

Зная максимальную энергию протонов, на основе гипотезы Боте и Беккера о природе бериллиевого излучения супруги Кюри вычислили необходимую максимальную энергию γ-квантов. Расчет производился следующим образом.

Если протоны появляются под действием γ-квантов, то, очевидно, этот процесс есть комптон-эффект на протонах водорода, в котором их можно считать свободными. Поэтому для изменения длины волны λ при рассеянии γ-кванта на угол θ можно воспользоваться известной формулой

$$\lambda' - \lambda = \frac{h}{mc}(1 - \cos\theta),$$

в которой, однако, m означает массу протона, а не электрона (см. § 3). Максимальное изменение длины волны получается при рассеянии назад, т. е. при $\theta = \pi$. В этом случае $\lambda' - \lambda = 2h/mc$. Отсюда находим частоту ν', а затем и энергию выбитого протона:

$$\mathscr{E} = h\nu - h\nu' = \frac{h\nu}{1 + mc^2/(2h\nu)}. \tag{92.2}$$

Подставляя сюда $\mathscr{E} = 4{,}5$ МэВ, $mc^2 = 938$ МэВ, вычисляем $h\nu = 48$ МэВ. Такова должна быть энергия γ-квантов, чтобы при их

рассеянии получились протоны требуемой энергии. Сами γ-кванты должны возникать в бериллии под действием α-частиц, излучаемых полонием. Так как энергия последних всего 5,25 МэВ, то трудно допустить, чтобы они возбуждали γ-кванты столь высоких энергий. Кроме того, полученное значение 48 МэВ не согласуется с энергией 7 МэВ тех же γ-квантов, найденной по их поглощению.

4. Наиболее тщательные исследования природы бериллиевого излучения были произведены Чедвиком в 1932 г. Схема его установки принципиально не отличается от той, какой пользовались супруги Кюри. Она изображена на рис. 158. На подложку D наносили слой полония Po. Альфа-частицы, испускаемые ядрами Po, вступали в ядерную реакцию (α, n) с ядрами Be. Нейтроны на пути к ионизационной камере пропускались через слой парафина. Заметим, что ионизационная камера I прикрывалась листами алюминия различной толщины для определения пробегов, а следовательно и энергий протонов, исходивших из парафина. Для максимальной энергии протонов из парафина Чедвик получил $\mathscr{E} = 5{,}7$ МэВ. Если бы бериллиевые лучи состояли из γ-квантов, то при таком значении $\mathscr{E}$ максимальная энергия γ-кванта получилась бы $h\nu = 55$ МэВ. Но Чедвик исследовал рассеяние бериллиевых лучей не только парафином, но и другими веществами и рассчитал по формуле (92.2) соответствующие максимальные энергии γ-квантов. Из измерения пробегов ядер отдачи в азоте получилось $h\nu = 90$ МэВ, а в аргоне $h\nu = 150$ МэВ.

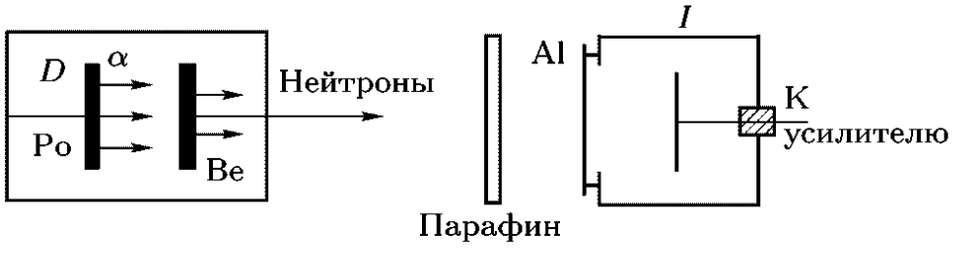


Рис. 158

Итак, гипотеза, отождествлявшая бериллиевые лучи с γ-квантами, приводила к противоречивым результатам. Для одной и той же величины различные методы давали значения 7, 55, 90, 150 МэВ. Это доказывало неправильность указанной гипотезы. Однако, как указал Чедвик, все противоречия устраняются, если предположить, что бериллиевые лучи образуются потоком не γ-квантов, а других *нейтральных частиц*, названных им *нейтронами*.

Возможность существования нейтронов предусматривалась еще Резерфордом в 1920 г. Он допускал возможность существования сложной частицы, состоящей из протона и электрона. Но в этой частице протон и электрон связаны между собой еще более прочно, чем в атоме водорода в основном состоянии. Существование подобной нейтральной частицы подтвердилось, но гипотезу о ее строении из протона и электрона пришлось оставить (см. § 63).

Однако гипотеза Резерфорда о строении нейтрона позволяет сразу понять, почему для выбивания наблюдаемых протонов из парафина требуются нейтроны несравненно меньшей энергии, чем энергия γ-квантов. Дело в том, что в соответствии с гипотезой Резерфорда масса нейтрона должна быть с большой точностью равна массе протона. Примем, что масса m протона и нейтрона одна и та же. При лобовом столкновении γ-кванта с энергией $h\nu$ с протоном последний получает импульс $mv = h\nu/c$. Отсюда

$$\frac{mv^2/2}{h\nu} = \frac{v}{2c}.$$

Но при равенстве масс протона и нейтрона величина $mv^2/2$ равна также энергии нейтрона. Из последней формулы видно, что эта энергия должна быть значительно меньше энергии соответствующего γ-кванта. Отметим, что в этом рассуждении принималось во внимание только предположение о равенстве масс протона и нейтрона, а гипотеза о строении последнего совсем не использовалась.

5. Чедвик произвел и первое, хотя и недостаточно точное, измерение массы нейтрона m_n. Он исходил из следующих соображений. Если нейтрон претерпевает лобовое столкновение с покоящимся ядром массой m, то на основании законов сохранения энергии и импульса в нерелятивистском приближении можно написать

$$m_\mathrm{n}v_{\mathrm{n}1}^2 = m_\mathrm{n}v_{\mathrm{n}2}^2 + mv^2, \quad m_\mathrm{n}v_{\mathrm{n}1} = m_\mathrm{n}v_{\mathrm{n}2} + mv,$$

где $v_{\mathrm{n}1}$ и $v_{\mathrm{n}2}$ — скорости нейтрона до и после столкновения, а v — скорость ядра отдачи. Отсюда

$$v = \frac{2m_\mathrm{n}}{m + m_\mathrm{n}}\, v_{\mathrm{n}1}.$$

Производились измерения максимальных скоростей ядер отдачи в водороде (v_H) и в азоте (v_N). Максимум скорости ядер отдачи свидетельствовал о том, что в обоих случаях скорости нейтронов были одинаковы, а именно максимальны. Поэтому полагая в предыдущем уравнении сначала $m = m_\mathrm{H}$, а затем $m = m_\mathrm{N}$, почленным делением получаем

$$\frac{v_\mathrm{H}}{v_\mathrm{N}} = \frac{m_\mathrm{N} + m_\mathrm{n}}{m_\mathrm{H} + m_\mathrm{n}} = \frac{m_\mathrm{N}/m_\mathrm{H} + m_\mathrm{n}/m_\mathrm{H}}{1 + m_\mathrm{n}/m_\mathrm{H}}. \tag{92.3}$$

Из своих измерений Чедвик нашел, что наиболее вероятные максимальные скорости отдачи ядер водорода (протонов) и азота равны $v_\mathrm{H} = 3{,}3 \cdot 10^9$ см/с и $v_\mathrm{N} = 4{,}7 \cdot 10^8$ см/с. Пользуясь этими значениями и полагая $m_\mathrm{N}/m_\mathrm{H} = 14$, из предыдущего соотношения находим $m_\mathrm{n}/m_\mathrm{H} = 1{,}15$. Учитывая довольно большую ошибку метода, Чедвик заключил, что $m_\mathrm{H} = m_\mathrm{n}$, т. е. масса нейтрона равна массе протона m_p. Точное значение массы нейтрона $m_\mathrm{n} = (939{,}5731 \pm 0{,}0027)$ МэВ было получено из баланса масс различных ядерных реакций с участием нейтронов. Масса нейтрона отличается от массы протона $m_\mathrm{p} = (938{,}2796 \pm 0{,}0027)$ МэВ на $m_\mathrm{n} - m_\mathrm{p} = (1{,}29343 \pm 0{,}00004)$ МэВ.

§ 93. Деление атомных ядер

1. Открытие деления атомных ядер — одно из важнейших фундаментальных открытий, получившее многочисленные научно-технические применения. Предыстория этого открытия начинается с 1934 г., когда Ферми со своими сотрудниками начали облучать нейтронами атомные ядра с целью получения новых химических элементов. Нейтроны не имеют электрического заряда — для них не существует кулоновского потенциального барьера. Поэтому им легче проникать внутрь ядра и вызывать различные ядерные превращения, чем заряженным частицам. При облучении самого тяжелого из естественных (встречающихся в природе) химических элементов — урана — Ферми надеялся получить еще более тяжелые химические элементы, называемые *трансуранами*. Эти элементы должны быть радиоактивными, так как в противном случае они встречались бы на Земле в естественном состоянии.

Стабильных изотопов у урана всего три: ^{238}U, ^{235}U и ^{234}U. Самым распространенным является ^{238}U. Относительная распространенность остальных двух изотопов весьма мала: ^{235}U составляет около $1/140$, а ^{234}U — всего $1/17\,000$ долю ^{238}U. Допустим, что облучению подвергается один из этих изотопов, например ^{238}U. При таком облучении возникает новый радиоактивный изотоп $^{239}_{92}$U. Этот изотоп перегружен нейтронами, а потому должен испытывать β^--распад, в результате чего возникает трансурановый элемент с $Z = 93$. При последующем β^--распаде должен возникать элемент с $Z = 94$, и т. д.

Ферми и его сотрудники действительно обнаружили, что при облучении урана нейтронами возникают новые радиоактивные ядра. Доказательством этого служило появление новых периодов полураспада. Особое внимание было уделено периодам в 13 и 90 мин, так как среди всех ядер с зарядовыми числами между $Z = 86$ и $Z = 92$ не было известно ни одного ядра, которое бы обладало таким периодом полураспада. Поэтому естественным было предположение Ферми, что новое β-излучение исходит от элемента с $Z = 93$ или от элементов с еще более высокими зарядовыми числами. Такое предположение как будто бы подтверждалось исследованиями химических свойств гипотетических элементов методами радиохимии, но это было признано неубедительным.

2. Систематические многолетние исследования О. Гана (1879–1968), Лизе Мейтнер (1878–1988) и Штрассмана (1902–1980) показали, что кривая ослабления интенсивности различных излучений выглядела различно в зависимости от времени начала наблюдения. Это указывало на то, что, по-видимому, какие-то радиоактивные вещества возникают не сразу, а спустя некоторое время после начала облучения нейтронами. Отсюда вытекало, что существует цепь последовательных радиоактивных превращений, подобная радиоактивным рядам естественных радиоактивных веществ. Периоды полураспада 13 и 90 мин, найденные Ферми, были уточнены и по измерениям Мейтнер оказались равными

16 и 59 мин соответственно. Кроме того, было обнаружено много периодов полураспада, в особенности более длительных, которые раньше не были замечены Ферми. Общее число таких периодов достигло девяти, а наибольший атомный номер гипотетических радиоактивных элементов возрос до 97. Интересно отметить также, что период полураспада 23 мин появлялся при облучении только медленными нейтронами. Лишь по этой причине его удалось отделить от близкого к нему периода 16 мин.

И. Кюри и П. Савич (р. 1909) помимо периодов, найденных Ганом, Мейтнер и Штрассманом, обнаружили еще период полураспада в 3,5 ч. Оказалось, что по своим химическим свойствам соответствующий ему элемент является аналогом редкоземельного элемента лантана. Этот факт определенно приводил к гипотезе о *делении ядра.* Однако Кюри и Савич не заметили такой возможности и приписали наблюдаемое радиоактивное излучение трансурановым элементам.

Между тем Ган и Штрассман продолжали систематические опыты по исследованию радиоактивных веществ, возникающих из урана и тория при облучении их нейтронами. (Лизе Мейтнер не могла принять участия в этих опытах. Она была еврейкой и австрийской подданной. После захвата Австрии гитлеровской Германией она в середине 1933 г. была вынуждена эмигрировать из Германии в Швецию. Но она непрерывно поддерживала активную идейную связь с лабораторией Гана в Берлине.) Ган и Штрассман были химиками — ведущими специалистами в области радиохимии. Главный итог их исследований состоит в том, что при облучении урана и тория нейтронами возникают изотопы химических элементов, лежащие где-то посередине периодической системы. К ним, в частности, относятся барий, а также лантан, наблюдавшийся в опытах Кюри и Савича.

3. Природа этих удивительных явлений была понята Л. Мейтнер и ее племянником О. Фришем (1904–1979) в середине января 1939 г. Они предположили, что возбужденное ядро урана или тория, получающееся после захвата нейтрона, делится на две более мелкие части приблизительно одинаковой массы, которые были названы *осколками.* Обнаруженные Кюри и Савичем, Ганом и Штрассманом радиоактивные изотопы, лежащие в середине периодической системы, и являются такими осколками ядер урана и тория. Поскольку относительная доля нейтронов в ядрах увеличивается с увеличением массы последних, эти осколки, получившиеся из тяжелых ядер, перегружены нейтронами. Поэтому в ядрах осколков происходят процессы превращения нейтронов в протоны путем испускания электронов. Такие же процессы происходят и в дочерних ядрах, возникающих в результате этих превращений. Этим и объясняются цепочки последовательных β^--превращений, наблюдавшиеся еще Ферми. Следует ожидать, и это было подтверждено последующими исследованиями, что ядро может делиться не только на два, но также на три и большее число осколков. Однако такие деления происходят значительно реже, чем деление на два осколка.

Предположение Мейтнер и Фриша сразу же было подтверждено опытами Фриша с атомами отдачи, которому удалось с убедительностью наблюдать взрывной характер деления ядер. В этих опытах малая ионизационная камера покрывалась изнутри окисью урана. Когда к этой камере подносили источник нейтронов (Rn + Be), то примерно раз 10–30 в минуту наблюдались сильные ионизационные толчки, которые после предварительного усиления фиксировались с помощью осциллографа. Такие толчки могли вызываться только атомными частицами, обладающими большими массами, зарядами и громадной энергией. Вся схема была рассчитана так, что импульсы, создаваемые α-частицами, были недостаточны для приведения в действие осциллографа. Это было видно из того, что толчки не наблюдались, когда не было урана или источника нейтронов. При окружении источника нейтронов слоем парафина число ионизационных толчков возрастало примерно в два раза. Это указывает на то, что для процесса деления ядер урана более эффективными являются *медленные нейтроны.*

После опытов Фриша такие же и аналогичные опыты, свидетельствующие о делении ядер, сразу же были поставлены во многих лабораториях. Не останавливаясь на этой стороне дела, перейдем к изложению основных закономерностей явления деления ядер и их простейшей теоретической интерпретации.

4. Ранее уже говорилось, что средняя энергия связи нуклона в тяжелых ядрах меньше, чем в ядрах середины периодической системы элементов (см. рис. 121). Допустим, что ядро урана ^{235}U после захвата нейтрона делится на два одинаковых осколка. Из рис. 121 видно, что средняя энергия связи нуклона в ядре урана равна 7,6 МэВ, тогда как в каждом из осколков она равна 8,5 МэВ. Таким образом, при делении ядра урана освобождается энергия 8,5–7,6 = 0,9 МэВ на каждый нуклон. Так как полное число нуклонов при делении не изменяется, то полная энергия, освобождающаяся при одном акте деления урана, составляет $0{,}9 \cdot 236 \approx 210$ МэВ. Эта громадная энергия проявляется главным образом в кинетической энергии образовавшихся осколков.

Оценку энергии, освобождающейся при делении, можно произвести несколько точнее, использовав приводимые в таблицах значения дефектов массы ядер, т. е. разностей $\Delta = M - A$ между массой ядра M и массовым числом A. Допустим, что исходным ядром является ядро ^{235}U. После захвата нейтрона образуется промежуточное ядро ^{236}U, которое и испытывает деление на два осколка. Предположим, что массовые числа этих осколков одинаковы, т. е. равны каждое 118. Из таблиц дефектов массы ядер находим, что для промежуточного ядра урана $\Delta = {}^{236}M - {}^{236}A \approx 42$ МэВ, а для каждого из осколков $\Delta = {}^{118}M - {}^{118}A \approx -86$ МэВ. Полная энергия, освобождаемая при делении, определяется разностью энергии промежуточного ядра и энергии двух образовавшихся осколков. Она составляет $42-(-2\cdot 86) = 214$ МэВ, что в пределах ошибок совпадает с предыдущим результатом.

По существу те же результаты получаются и из полуэмпирической формулы Вейцзеккера (64.6). В этой формуле, поскольку она получена

в предположении несжимаемости ядерного вещества, можно опустить первый член, так как его значение не меняется при делении ядра, а потому он не влияет на разность масс до и после деления. По той же причине можно опустить четвертый член, так как энергия симметрии также не меняется при делении. Отбросим еще последний член — энергию спаривания, так как ее изменением при делении можно пренебречь. С учетом этих упрощений энергию ядра можно представить формулой

$$\mathscr{E} = C_{\text{пов}} A^{2/3} + C_{\text{кул}} Z^2 A^{-1/3}. \qquad (93.1)$$

Знаки в формуле (64.6) изменены, так как эта формула дает работу, которую надо затратить, чтобы полностью расщепить ядро на составляющие его нуклоны. А эта работа, взятая с противоположным знаком, как раз и дает энергию ядра. Воспользуемся еще значениями коэффициентов (64.7), из которых нам нужны только два: $C_{\text{пов}} = 17{,}8$ МэВ и $C_{\text{кул}} = 0{,}710$ МэВ. Простой расчет дает для энергии урана $\mathscr{E}_{236} = 1653$ МэВ, а для каждого осколка $\mathscr{E}_{118} = 735$ МэВ. Искомая энергия, освобождаемая при делении, таким образом, равна $1653 - 2 \cdot 735 = 183$ МэВ, что практически совпадает с предыдущими результатами. (См. также задачу 3 к § 64.)

5. Простейшая теория деления ядра была разработана в 1939 г. Н. Бором и Д. Уилером (р. 1911) и независимо и даже несколько раньше их Я. И. Френкелем (1894–1952). В основу этой теории была положена капельная модель ядра. Конечно, капельная модель в состоянии описать только некоторые, простейшие черты такого сложного явления, какое представляет собой деление атомного ядра. Многие стороны этого явления остаются необъясненными и даже находятся в противоречии с теорией. В. М. Струтинский усовершенствовал капельную модель учетом оболочечных эффектов. Однако мы ограничимся изложением, в основном качественным, только простейшей — капельной — теории деления ядра.

Рассмотрим ядро ^{235}U, захватившее нейтрон, в результате чего образовалось составное ядро ^{236}U. Составное ядро ^{236}U в основном состоянии практически стабильно; хотя оно и α-радиоактивно, но период полураспада составляет $2{,}4 \cdot 10^7$ лет. Такое ядро может совершать внутренние колебания около своей равновесной формы, не подвергаясь делению. Однако в результате захвата нейтрона получается ядро ^{236}U в сильно возбужденном состоянии, и амплитуда колебаний может стать настолько большой, что ядро разделится на две части. Под действием кулоновских сил отталкивания эти части разлетятся со значительными скоростями. Ядерные силы, как силы притяжения, препятствуют этому разлету. Но при разлете превалируют электрические силы. В результате части разделившегося ядра и получают кинетическую энергию около 200 МэВ, о которой говорилось выше.

Если ядро уподобить жидкой капле, то различные стадии, через которые проходит форма ядра до и в результате деления, схематически представлены на рис. 159. Однако не вся энергия деления переходит в кинетическую энергию осколков. Сами осколки получаются в возбужденном состоянии, и их внутренняя энергия высвобождается в первую

очередь путем испарения нейтронов, а затем идет на испускание γ-лучей. Основными продуктами процесса деления являются, таким образом, два примерно равных ядра, несколько нейтронов и γ-квантов. Как уже указывалось в п. 3, ядра-осколки, как правило, оказываются β^--активными и распадаются с самыми разными периодами полураспада.

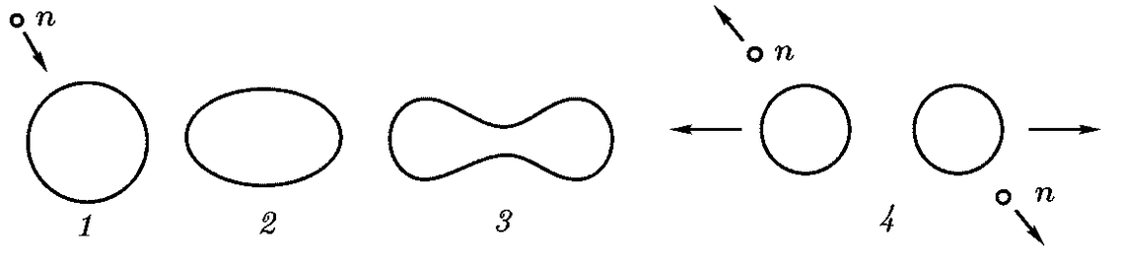


Рис. 159

6. При увеличении поверхности ядра энергия поверхностного натяжения возрастает, а энергия кулоновского отталкивания убывает. Изменениями объемной энергии, энергии симметрии и энергии спаривания можно пренебречь. Поверхностное натяжение стремится вернуть деформированное ядро в исходное состояние, а кулоновское отталкивание способствует его делению. Отношение кулоновской энергии к поверхностной пропорционально $Z^2A^{-1/3} : A^{2/3} = Z^2/A$. Поэтому отношение Z^2/A может служить основным параметром, определяющим способность ядра к делению. Чем больше значение этого параметра, тем легче ядро подвержено делению. Чтобы приближенно оценить критическое значение параметра Z^2/A, при котором ядро уже не может быть стабильным, рассмотрим устойчивость ядра относительно его малых деформаций.

При малых деформациях можно считать, что исходное сферическое ядро принимает форму вытянутого эллипсоида вращения с полуосями $a = R(1+\varepsilon)$ и $b = R/\sqrt{1+\varepsilon}$, где ε — малая величина, играющая роль параметра деформации. При таких значениях a и b объем ядра не меняется, что соответствует предположению о несжимаемости ядра. Нахождение поверхностной энергии сводится к геометрической задаче вычисления поверхности ядра, а вычисление кулоновской энергии — к задаче о поле заряженного эллипсоида, если предположить, что заряд ядра равномерно распределен по его объему. Расчет показывает, что при малых деформациях ядра суммарная энергия поверхностного натяжения и кулоновского отталкивания с точностью до членов ε^2 включительно меняется на величину

$$\Delta\mathscr{E} = \frac{1}{5}\left(2C_{\text{пов}}A^{2/3} - C_{\text{кул}}Z^2A^{-1/3}\right)\varepsilon^2, \qquad (93.2)$$

где $C_{\text{пов}}$ и $C_{\text{кул}}$ — коэффициенты, входящие в формулу Вейцзеккера (64.6). Конечно, это изменение происходит за счет внутренней энергии возбужденного ядра. Если $\Delta\mathscr{E} > 0$, то поверхностные силы будут превалировать над силами кулоновского отталкивания, стремясь вернуть ядро к исходной сферической форме, около которой оно будет

совершать колебания. Если же $\Delta\mathscr{E} < 0$, то преобладающими будут силы кулоновского отталкивания, способствующие делению.

Критическое значение параметра Z^2/A, больше которого ядро становится абсолютно неустойчивым к делению, определится из требования, чтобы обратилось в нуль выражение в круглых скобках формулы (93.2). Это дает

$$\left(\frac{Z^2}{A}\right)_{\text{крит}} = 2\,\frac{C_{\text{пов}}}{C_{\text{кул}}} \approx 50, \tag{93.3}$$

причем мы использовали значения коэффициентов $C_{\text{пов}}$ и $C_{\text{кул}}$ из (64.7). Конечно, на вычисленное значение критического параметра Z^2/A надо смотреть как на ориентировочное, поскольку оно получено в предположении справедливости капельной модели и малых деформаций ядра.

Формулой (93.3) еще не определяются в отдельности критические значения Z и A, больше которых ядро абсолютно неустойчиво по отношению к делению, так как в общем случае между Z и A не существует однозначной связи. Такая связь имеет место только в случае β-стабильных ядер и определится формулой (64.8а). С помощью этой формулы можно исключить из (93.3) зарядовое число Z и получить иррациональное уравнение только для A. Решая это уравнение, находим критическое значение массового числа $A_{\text{крит}} \approx 385$, а затем и критическое значение зарядового числа $Z_{\text{крит}} \approx 138$.

7. Следующие качественные соображения позволяют уяснить влияние на процесс деления и конечных деформаций ядра. Разделим мысленно исходное стабильное ядро на две части (два осколка) с зарядовыми числами Z_1 и Z_2. Если эти осколки разойдутся достаточно далеко на расстояние r, то энергия их кулоновского отталкивания будет $Z_1 Z_2 e^2/r$. При этом появится еще кинетическая энергия осколков около 200 МэВ. Она, разумеется, возникает за счет потенциальной энергии ядра в исходном состоянии, так что потенциальная энергия в исходном состоянии составит около 200 МэВ, т. е. равна энергии, освобождающейся при делении. Однако исходное ядро по предположению стабильно. Следовательно, на кривой потенциальной энергии $V = V(r)$ должен существовать максимум — потенциальный барьер, который должен быть преодолен, чтобы фактически произошло деление (рис. 160). Вместо одного максимума, в особенности когда атомные ядра не сферические, возможны и два максимума (двухгорбая кривая). Но от такой возможности при выяснении основных вопросов деления можно отвлечься. Для абсолютно неустойчивых ядер, у которых $Z^2/A > 50$, максимума на кривой $V = V(r)$ нет. Такие ядра, если бы они возникли, мгновенно претерпевали бы деление за времена порядка 10^{-23}–10^{-24} с.

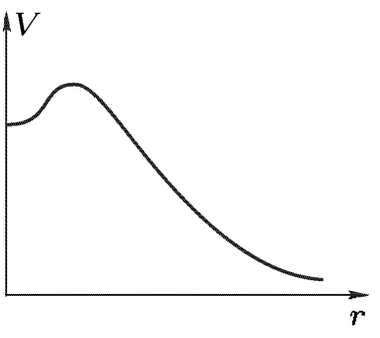


Рис. 160

Разность между максимальным значением потенциальной энергии $V(r)$ и ее значением при $r = 0$ для стабильных ядер называется *энергией активации*. По классическим представлениям для деления ядра необходимо сообщить ему энергию, не меньшую энергии активации. Эту энергию приносят нейтроны, при поглощении которых и образуются возбужденные составные ядра, удовлетворяющие указанному условию. Если же энергия возбуждения меньше энергии активации, то по классическим представлениям деление невозможно. Исследования показали, что ядра ^{235}U претерпевают деление после захвата любых, в том числе и медленных (тепловых) нейтронов. Для деления же ядер ^{238}U, захвативших нейтрон, требуются быстрые нейтроны с энергиями больше 1 МэВ.

8. Такое различие в поведении ядер ^{235}U и ^{238}U связано в первую очередь с эффектом спаривания нуклонов (см. § 64). В нечетном ядре ^{235}U неспаренный нуклон связан слабее остальных нуклонов, которые спарены. Допустим, что ядро ^{235}U поглощает свободный нейтрон, превращаясь в четно-четное промежуточное ядро ^{236}U. При захвате свободного нейтрона промежуточное ядро получается в возбужденном состоянии. Возбуждение еще более возрастает за счет того, что бывший неспаренный нуклон находит себе партнера и тем самым становится связанным сильнее. Если же свободный нейтрон захватывается ядром ^{238}U, то получается четно-нечетное промежуточное ядро ^{239}U. В нем захваченный нейтрон, не имея себе партнера, связан слабее, чем остальные нуклоны. Поэтому в процессе захвата нейтрона ядро ^{238}U возбуждается слабее, чем ядро ^{235}U. Для его деления требуется дополнительная энергия возбуждения. Некоторое влияние на способность ядер к делению, возможно, оказывает и то обстоятельство, что параметр Z^2/A для ядер ^{236}U (35,86) больше, чем для ядер ^{239}U (35,41).

Изложенное в основном справедливо и в отношении других четно-нечетных ядер. Особенно следует отметить ядра ^{239}Pu и ^{233}U, которые делятся медленными нейтронами, тогда как их четно-четные изотопы, как правило, требуют для деления быстрых нейтронов (^{240}Pu делится и на тепловых нейтронах). Это обстоятельство играет решающую роль в ядерной энергетике (см. § 95).

9. С квантовой точки зрения не обязательно, чтобы энергия возбуждения превосходила энергию активации ядра. Ядро может разделиться и в том случае, когда имеет место обратное соотношение. В этом случае деление может произойти путем *туннельного просачивания* через потенциальный барьер, как это имеет место при α-распаде. Однако вероятность этого процесса мала. Она тем меньше, чем выше потенциальный барьер. В частности, возможно и *спонтанное деление* ядра, происходящее при полном отсутствии внешнего возбуждения.

Спонтанное деление ядер урана впервые наблюдали К. А. Петржак и Г. Н. Флеров (р. 1913) в 1940 г. В их опытах была использована многослойная ионизационная камера с рабочей площадью пластин около 1000 см^2 и больше, на которую наносился уран. При поднесении к ионизационной камере источника нейтронов в осциллографе наблюдались

мощные ионизационные толчки, происходящие от вынужденного деления ядер урана. Эти предварительные опыты позволяли изучить характер ионизационных толчков, происходящих от деления ядер урана. При удалении источника нейтронов такие толчки полностью не прекращались, что свидетельствовало о спонтанном делении ядер. В начальных опытах Петржак и Флеров наблюдали в среднем примерно 6 толчков в час. Окончательные опыты были поставлены на одной из станций Московского метрополитена на глубине 50 м, чтобы как можно лучше защитить аппаратуру от мешающего фона космических лучей. В одном грамме урана за час самопроизвольно делится в среднем всего около 25 ядер урана. Пользуясь этим, легко рассчитать, что период полураспада урана по отношению к спонтанному делению составляет около 10^{16} лет. На рис. 161 приведена зависимость периода T спонтанного деления от

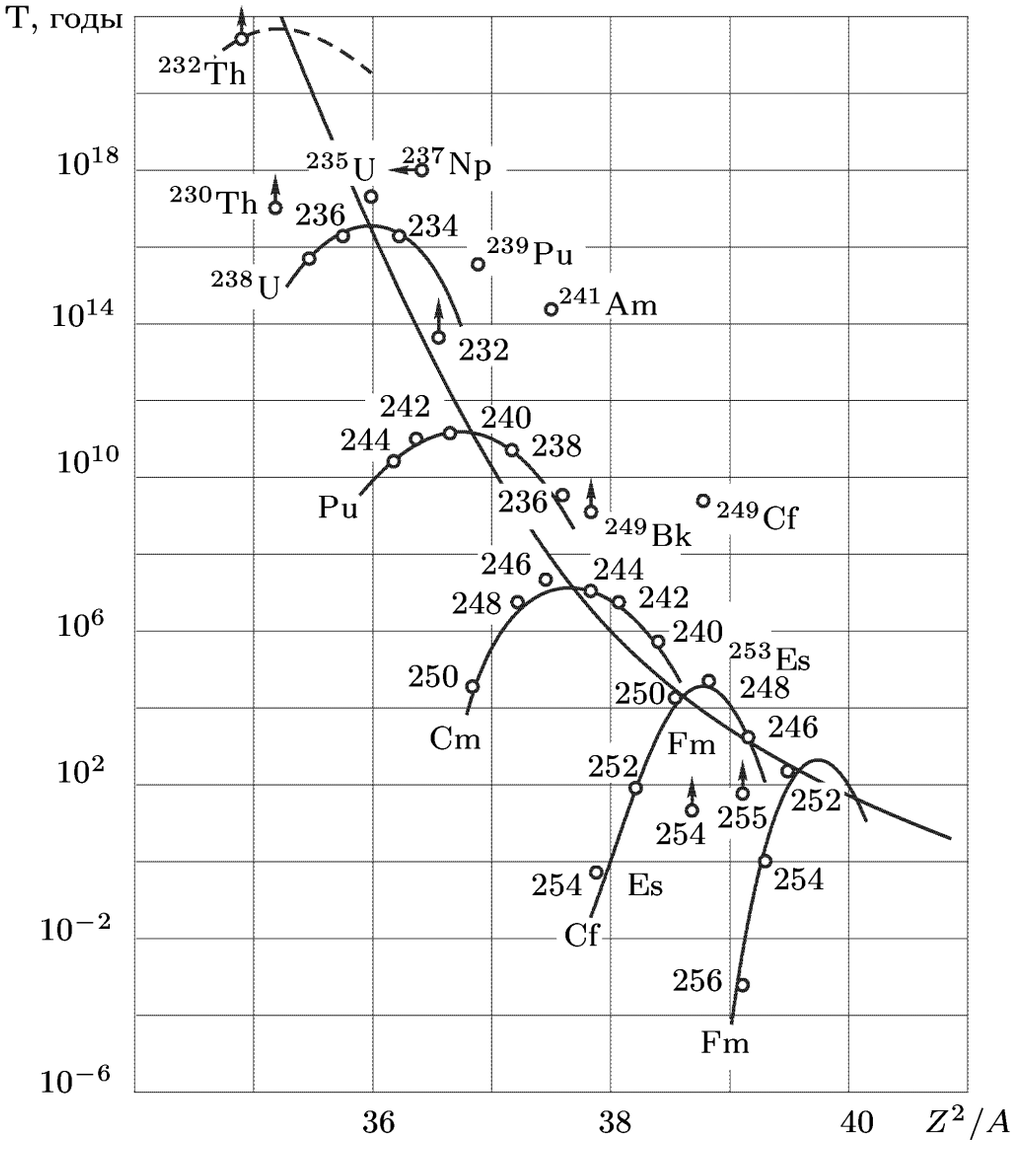


Рис. 161

параметра Z^2/A для различных тяжелых ядер. Светлые точки соответствуют экспериментальным данным, пересекающая рисунок сплошная кривая рассчитана по модели жидкой капли.

10. Как уже указывалось выше, ядро чаще всего делится на два осколка. Но отношение масс этих осколков может быть разным. При

делении тепловыми нейтронами осколки равных или близких масс почти не наблюдаются, хотя все числовые оценки выше мы делали в предположении равенства масс осколков, так как для этой цели такое предположение не существенно. На самом деле наиболее вероятно деление на осколки, один из которых примерно в полтора раза тяжелее другого. Это иллюстрируется рис. 162, на котором приведено количество осколков различных массовых чисел в процентах для ^{235}U

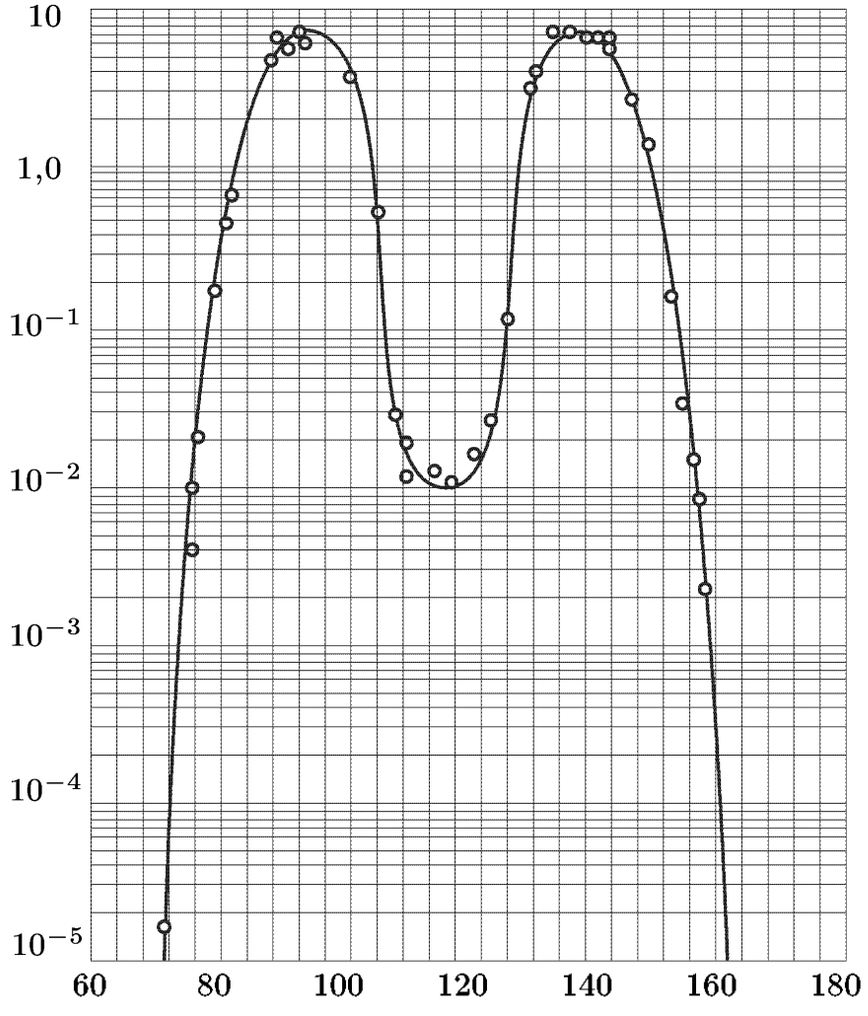


Рис. 162

при делении тепловыми нейтронами. Считается, что такая асимметрия деления объясняется влиянием ядерных нейтронных оболочек — ядру энергетически выгоднее делиться так, чтобы число нейтронов в каждом из осколков было близко к одному из магических чисел — 50 или 82.

11. Процесс деления ядра сопровождается вылетом *вторичных нейтронов*. Этого следует ожидать на основе следующих простых соображений. Отношение числа нейтронов N к числу протонов Z в ядре, вообще говоря, тем больше, чем тяжелее ядро. Для β-стабильных ядер это видно из рис. 120, а также из формулы (64.8а), если ее переписать в виде

$$\frac{N}{Z} = 0{,}97 + 0{,}0150\, A^{2/3}.$$

Допустим теперь, что тяжелое ядро разделилось на два осколка. Если бы при делении общее число нейтронов в системе не изменилось,

то отношения N/Z в обоих осколках были бы либо равны такому же отношению в исходном тяжелом ядре, либо в одном из них оно было бы больше, а в другом меньше. Значит, по крайней мере один из осколков получился бы нейтроноизбыточным и должен был бы освобождаться от лишних нейтронов.

Появление вторичных нейтронов можно понять и с точки зрения капельной модели ядра. В шейке гантелеобразного ядра (положение *3* на рис. 159) почти нет протонов, так как из-за кулоновского отталкивания они сосредоточиваются главным образом в удаленных — шарообразных — частях ядра. Поэтому при разрыве шейки в процессе деления из нее могут вылетать лишь нейтроны, и притом за времена порядка ядерных (10^{-22}–10^{-23} с), т. е. практически мгновенно. Такие вторичные нейтроны называются *мгновенными.*

Наряду с мгновенными при делении ядер появляются и *запаздывающие вторичные нейтроны.* Дело в том, что осколки деления ядер радиоактивны. При их β-распаде могут появляться нейтроноизбыточные ядра в сильно возбужденных состояниях. Возбуждение таких ядер снимается путем конкурирующих процессов испускания нейтронов и γ-квантов. Оба процесса происходят практически мгновенно. Первый из этих процессов и порождает запаздывающие нейтроны. Запаздывание понимается в смысле временного отставания момента вылета рассматриваемых нейтронов от момента испускания мгновенных нейтронов (появляющихся непосредственно при делении ядра). Оно определяется периодом полураспада изотопа-предшественника, при β-распаде которого и получилось возбужденное ядро, испустившее затем нейтрон.

Процесс возникновения запаздывающих нейтронов поясняется схемами на рис. 163 и 164. На рис. 163 показано ядро ^{87}Br, появившееся

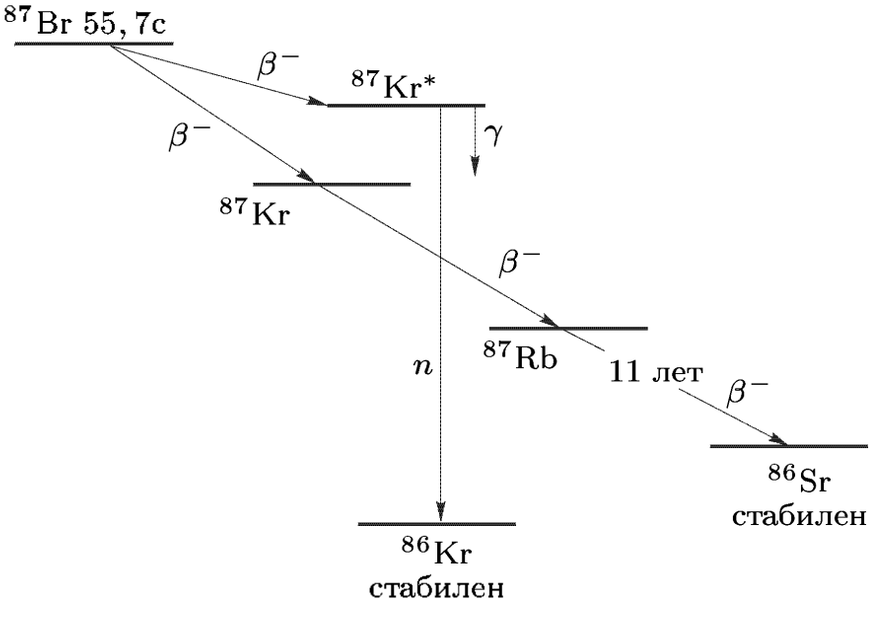


Рис. 163

в результате ядерного деления. Это ядро испытывает β^--распад с периодом полураспада 55,6 с. В результате таких распадов возникают ядра ^{87}Kr, бóльшая часть которых находится в невозбужденном состоянии

и лишь незначительная часть — в сильно возбужденном состоянии ^{87}Kr*. Невозбужденное ядро после двух последовательных β^--рас-

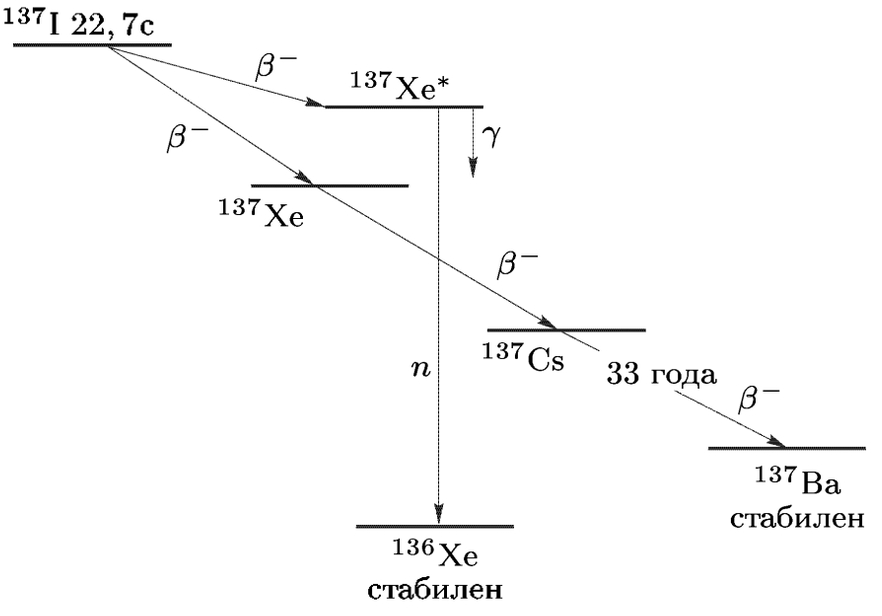


Рис. 164

падов переходит в стабильное ядро ^{87}Sr. Сильно же возбужденное ядро ^{87}Kr* почти мгновенно переходит в невозбужденное состояние ^{87}Kr (с испусканием γ-квантов) или же в стабильное состояние ^{86}Kr (с испусканием нейтрона). В последнем процессе и возникает запаздывающий нейтрон [1]. Аналогичная схема приведена и на рис. 164.

Таблица 15

Запаздывающие нейтроны

Период полураспада	Выход, % (к общему числу вторичных нейтронов)	Кинетическая энергия запаздывающих нейтронов, МэВ
55,7	0,021	0,25
22,7	0,140	0,56
6,2	0,125	0,43
2,3	0,253	0,62
0,6	0,074	0,42
0,2	0,027	—
—	—	—
Средние 8,97	0,64	0,508

Запаздывающие нейтроны делятся на несколько групп, различающихся временем запаздывания и энергией. В табл. 15 приведены ха-

[1] Следует отметить, что изотоп ^{87}Kr не случайно оказался «нейтроноактивным». В ядре этого изотопа 51 нейтрон, т. е. на один нейтрон больше магического числа 50; этот «лишний» нейтрон и покидает возбужденное ядро. У изотопа ^{137}Xe (рис. 164) число нейтронов равно 83, так что ситуация аналогична.

рактеристики запаздывающихся нейтронов, возникающих при делении ядер урана-235.

Свыше 99 % от всех выделяющихся при делении нейтронов освобождается мгновенно. На долю запаздывающих нейтронов приходится примерно 0,75 %. Несмотря на это, запаздывающие нейтроны играют важную роль в ядерной энергетике — они позволяют сделать управляемыми ядерные реакции, происходящие при делении в ядерных реакторах (см. § 95, п. 5).

12. В заключение приведем более точное примерное распределение энергии деления между различными способами ее освобождения при делении ^{235}U тепловыми нейтронами:

Кинетическая энергия осколков	166,2 ± 1,3 МэВ
Энергия нейтронов деления	4,8 ± 0,1 МэВ
Энергия мгновенных γ-квантов	8,0 ± 0,8 МэВ
Энергия β-частиц продуктов деления	7,0 ± 0,3 МэВ
Энергия γ-излучения продуктов деления	7,2 + 1,1 МэВ
Энергия антинейтрино продуктов деления	9,6 ± 0,5 МэВ
Всего	202,8 ± 0,4 МэВ

§ 94. Трансурановые элементы

1. Причину, по которой периодическая система химических элементов, встречающихся в природе в естественном состоянии, обрывается на уране $_{92}$U, надо искать в происхождении химических элементов и их последующих радиоактивных превращениях. В настоящее время можно считать доказанным, что все химические элементы Солнечной системы образовались свыше 5 млрд лет назад в результате гравитационного сжатия огромных масс галактического водорода, при котором сильно повысилась его температура. Чрезвычайно высокая температура и обеспечила синтез сложных ядер из более простых ядер водорода. При этом были синтезированы и многочисленные изотопы химических элементов с атомными номерами больше 92. Однако период их полураспада оказался значительно меньше времени существования Земли (около $4{,}5 \cdot 10^9$ лет). К настоящему времени все эти изотопы из-за радиоактивных превращений полностью распались и исчезли на Земле (см. п. 10). Самым тяжелым из сохранившихся оказался изотоп урана ^{238}U, период полураспада которого ($4{,}5 \cdot 10^9$ лет) того же порядка, что и возраст Земли.

2. Истинные границы периодической системы, за которыми существование более тяжелых элементов невозможно, казалось бы, должны определяться нестабильностью электронной оболочки атома. Дело в том, что самый глубокий электрон атомной оболочки не может приблизиться к центру ядра на расстояние, меньшее радиуса последнего. Такой электрон был бы сразу поглощен ядром, что привело бы

к уменьшению атомного номера элемента. Расстояние самого глубокого электрона атомной оболочки до центра атомного ядра с зарядовым числом Z по теории Бора равно r_B/Z, где r_B — боровский радиус. Радиус самого ядра определяется формулой (65.1). Таким образом, для возможности существования атома, во всяком случае, необходимо выполнение условия $r_B/Z \geqslant r_0 A^{1/3}$, или

$$Z A^{1/3} \leqslant \frac{r_B}{r_0} \approx 38\,000. \tag{94.1}$$

Исключение Z с помощью соотношения (64.8а) приводит к неравенству

$$\frac{A^{4/3}}{1{,}97 + 0{,}015 A^{2/3}} \lesssim 38\,000. \tag{94.1а}$$

Решение этого неравенства сводится к решению квадратного уравнения относительно $A^{2/3}$. В результате получается $A \lesssim 17700$ и, следовательно, $Z \lesssim 1460$. Итак, совершенно невозможно существование атомов с порядковым числом $Z \gtrsim 1460$.

Приведенная оценка дает слишком завышенное значение для верхней границы Z. Она, в сущности, основана на безусловно верном утверждении, что атом не может существовать, если ближайший к ядру электрон находится на уровне, с которого практически мгновенно происходит электронный захват, в результате которого зарядовое число Z (равное порядковому номеру элемента) уменьшается на единицу. А такой захват (из-за туннельного эффекта) с большой вероятностью происходит не только с расстояния до центра ядра, равного его радиусу, но и с расстояний, значительно больших. Поэтому стабильность электронной оболочки атома должна нарушаться уже при Z, значительно меньших найденного выше значения. Более подробное рассмотрение показывает, что такое нарушение должно начинаться при $Z = 180$–200. И все же периодическая система элементов обрывается еще раньше. Ее верхняя граница связана не с неустойчивостью электронной оболочки, а определяется *неустойчивостью самого ядра.*

3. Элементы с атомными номерами больше 92, как уже указывалось, называются *трансурановыми* («зауарановыми») или, короче, *трансуранами.* Все они получаются искусственно, т. е. в результате различных ядерных реакций, в которых тот или иной (тяжелый) элемент подвергается облучению или бомбардировке различными ядерными частицами. Наиболее важными из них являются нейтроны, поскольку они лишены электрического заряда и, следовательно, им не требуется преодолевать кулоновский потенциальный барьер. Но и заряженные частицы являются эффективными снарядами для получения трансурановых элементов, если только их энергия достаточно велика.

При облучении нейтронами изотопа урана ^{238}U получается промежуточное ядро ^{239}U. Обладая излишком нейтронов, оно испытывает β^--распад и превращается в ядро трансуранового элемента с $Z = 93$. Этот элемент был назван *нептунием* (Np) по аналогии с Солнечной

системой, в которой за планетой Уран следует планета Нептун. Претерпевая дальнейший β^--распад, нептуний превращается в наиболее важный искусственный трансуран с $Z = 94$, названный *плутонием* (Pu), опять-таки по аналогии с Солнечной системой, в которой за Нептуном следует планета Плутон.

Такая схема получения трансуранов с $Z = 93$ и $Z = 94$ была абсолютно ясна Ферми уже в 1934 г., когда он начал облучать нейтронами уран. Правда, в то время Ферми не удалось наблюдать Np и Pu из-за их больших периодов полураспада (2, 3 дня для $^{239}_{93}\mathrm{Np}$ и $2{,}4 \cdot 10^4$ лет для $^{239}_{94}\mathrm{Pu}$) и слабости радон-бериллиевого источника нейтронов, которым он тогда пользовался. В опытах Ферми в действительности наблюдались не радиоактивные излучения трансуранов, а излучения радиоактивных изотопов, получающихся при *делении ядер урана* (см. § 93). Только после создания ядерных реакторов, из которых исходят мощные потоки нейтронов (до 10^{15} нейтронов через 1 см^2 в 1 с), открылась возможность получать Np и Pu и притом в промышленных масштабах. Их получение можно представить следующей схемой:

$$^1_0\mathrm{n} + {}^{238}_{92}\mathrm{U} \to {}^{239}_{92}\mathrm{U} \xrightarrow{\beta^-} {}^{239}_{93}\mathrm{Np} \xrightarrow{\beta^-} {}^{239}_{94}\mathrm{Pu} \xrightarrow{\alpha} . \qquad (94.2)$$

Это, конечно, не единственная схема получения нептуния и плутония. Например, изотоп нептуния $^{238}_{93}\mathrm{Np}$ был получен в результате облучения урана $^{238}\mathrm{U}$ дейтронами с энергией 16 МэВ:

$$^2_1\mathrm{d} + {}^{238}_{92}\mathrm{U} \to {}^{238}_{93}\mathrm{Np} + 2\mathrm{n}. \qquad (94.3)$$

Всего в настоящее время известно не менее 12 изотопов нептуния с массовыми числами от 230 до 241. Изотоп нептуния $^{239}_{93}\mathrm{Np}$ представляет собой тяжелый ($\overline{\rho} = 19{,}5$ г/см^3) металл с серебристым блеском, медленно окисляющийся на воздухе. Температура плавления нептуния 640 °C.

В настоящее время известны не менее 15 изотопов плутония с массовыми числами от 232 до 246. Элемент плутоний — блестящий металл с голубоватым оттенком. Плотность его в разных модификациях изменяется от 15,9 до 19,8 г/см^3. Температура плавления плутония 640 °C. Коэффициент линейного расширения плутония отрицательный.

4. Путем облучения нейтронами плутония получается следующий трансуран с $Z = 95$. Он был получен в США и назван *америцием* (Am). Способ его получения в ядерном реакторе представляется следующей схемой:

$$^1_0\mathrm{n} + {}^{239}_{94}\mathrm{Pu} \to {}^{240}_{94}\mathrm{Pu},$$

$$^1_0\mathrm{n} + {}^{240}_{94}\mathrm{Pu} \to {}^{241}_{94}\mathrm{Pu} \xrightarrow{\beta^-} {}^{241}_{95}\mathrm{Am}.$$

Америций $^{241}_{95}\mathrm{Am}$ представляет собой серебристый металл с плотностью 11,7 г/см^3.

Облучение $^{241}_{94}\mathrm{Am}$ нейтронами с последующим β^--распадом промежуточного ядра приводит к образованию трансурана с $Z = 96$ по схеме

$$^1_0\mathrm{n} + {}^{241}_{95}\mathrm{Am} \to {}^{242}_{95}\mathrm{Am} \xrightarrow{\beta^-} {}^{242}_{96}\mathrm{Cm}. \qquad (94.4)$$

Он был назван *кюрием* (Cm) в честь Марии и Пьера Кюри. Это также серебристый металл с плотностью 7 г/см^3. Он α-радиоактивен с периодом полураспада 160 дней. Его α-радиоактивность настолько велика (10^{13} α-частиц в 1 мин на 1 мг), что она приводит к разогреванию, разбрызгиванию и даже разложению воды в водных растворах солей кюрия, а также к свечению их в темноте.

Не вдаваясь в подробности, заметим, что 97-й и 98-й элементы также могут быть получены путем длительного облучения нейтронами плутония $^{239}_{94}\mathrm{Pu}$ в ядерных реакторах. Они были названы соответственно *берклием* (Bk) и *калифорнием* (Cf). Эти названия происходят от города Беркли (США) и Калифорнийского университета в Беркли, где впервые были получены рассматриваемые трансураны. Существуют различные способы получения берклия и калифорния, равно как америция и кюрия. Однако в больших количествах все эти элементы получаются только реакторным способом. В настоящее время известно 11 изотопов америция с массовыми числами от 237 до 247, 13 изотопов кюрия ($238 \leqslant A \leqslant 252$), 9 изотопов берклия ($243 \leqslant A \leqslant 251$) и 16 изотопов калифорния ($240 \leqslant A \leqslant 255$).

5. Реакторный способ, хотя он и является единственным способом получения трансуранов в больших количествах, все же не всегда применим. Оказалось, что следующие за калифорнием ($Z = 98$) новые элементы вблизи ядерного реактора упорно не появляются. Причина этого в том, что не все ядра одного и того же изотопа после поглощения нейтрона претерпевают β^--распад. Большая часть ядер или испытывает деление, или распадается иным способом, не приводящим к образованию нового элемента с большим атомным номером. Эти процессы идут тем интенсивнее, чем тяжелее атомные ядра, и, наконец, полностью вытесняют β^--распад. Можно было бы думать, что указанные трудности имеют только технический характер и могут быть преодолены путем повышения мощности реактора. Но помимо технических трудностей есть и трудности принципиальные, когда в результате длинного ряда радиоактивных превращений опять возникает исходный элемент, подвергнутый облучению нейтронами. К этому, в частности, может привести α-распад какого-либо ядра. Примером может служить следующая цепочка превращений:

$$^{239}_{94}\mathrm{Pu} + {}^1_0\mathrm{n} \to {}^{240}_{94}\mathrm{Pu},$$

$$^{240}_{94}\mathrm{Pu} + {}^1_0\mathrm{n} \to {}^{241}_{94}\mathrm{Pu} \xrightarrow{\beta^-} {}^{241}_{95}\mathrm{Am},$$

$$^{241}_{95}\mathrm{Am} + {}^1_0\mathrm{n} \to {}^{242}_{95}\mathrm{Am} \xrightarrow{\beta^-} {}^{242}_{96}\mathrm{Cm} \xrightarrow{\alpha} {}^{238}_{94}\mathrm{Pu},$$

$$^{238}_{94}\mathrm{Pu} + {}^1_0\mathrm{n} \to {}^{239}_{94}\mathrm{Pu}.$$

Непреодолимое препятствие для реакторного способа возникает также, когда одно из звеньев в цепочке превращений выпадает. Такое препятствие возникает, когда в этом звене время жизни ядра очень мало, а также тогда, когда ядро, входящее в звено, полностью распадается нежелательным образом, например путем спонтанного деления.

6. Указанные затруднения связаны прежде всего с тем, что при облучении нейтронами от ядерных реакторов облучаемое ядро, как правило, поглощает только один или малое число нейтронов. Если увеличить поток облучающих нейтронов во много раз, то некоторые ядра, возможно, будут поглощать сразу несколько нейтронов и окажутся в состоянии перепрыгнуть через «запрещенную» область в реакторном методе. Таким путем могут образовываться новые трансурановые элементы или новые изотопы уже известных элементов. И действительно, в 1952 г. в США произвели испытание водородной бомбы, в которой происходит сгорание дейтериево-тритиевой смеси. Реакция происходит при температуре порядка ста миллионов градусов и длится всего доли секунды. В течение такого короткого времени через урановую оболочку бомбы проходит поток нейтронов, на много порядков превышающий поток нейтронов в ядерных реакторах за то же время. В радиоактивной пыли, образовавшейся при взрыве, были обнаружены ранее неизвестные изотопы плутония $^{244}\mathrm{Pu}$ и $^{246}\mathrm{Pu}$. Они образуются в результате захвата ядром $^{238}\mathrm{U}$ сразу 6 или 8 нейтронов с последующим β^--распадом, т. е.

$$^{238}_{92}\mathrm{U} + 6\,^{1}_{0}\mathrm{n} \to {}^{244}_{92}\mathrm{U} \xrightarrow{\beta^-} {}^{244}_{93}\mathrm{Np} \xrightarrow{\beta^-} {}^{244}_{94}\mathrm{Pu}, \tag{94.5}$$

$$^{238}_{92}\mathrm{U} + 8\,^{1}_{0}\mathrm{n} \to {}^{246}_{92}\mathrm{U} \xrightarrow{\beta^-} {}^{246}_{93}\mathrm{Np} \xrightarrow{\beta^-} {}^{246}_{94}\mathrm{Pu}. \tag{94.6}$$

Американские ученые запросили срочно доставить с атолла Бикини, где произошел взрыв водородной бомбы, около тонны кораллов. После весьма кропотливого труда из этих кораллов были выделены ничтожные количества 99-го и 100-го элементов. Они были названы соответственно *эйнштейнием* и *фермием* в честь Эйнштейна и Ферми. Обнаруженные изотопы эйнштейния ($^{253}_{99}\mathrm{Es}$) и фермия ($^{255}_{100}\mathrm{Fm}$) получаются в результате следующих процессов. При взрыве водородной бомбы некоторые ядра урана одновременно захватывают 15 или 17 нейтронов. Последующая цепь β^--распадов и приводит к образованию новых трансуранов:

$$^{238}_{92}\mathrm{U} + 15\,^{1}_{0}\mathrm{n} \to {}^{253}_{92}\mathrm{U} \xrightarrow{\beta^-} {}^{253}_{93}\mathrm{Np} \xrightarrow{\beta^-} \ldots \xrightarrow{\beta^-} {}^{253}_{99}\mathrm{Es}, \tag{94.7}$$

$$^{238}_{92}\mathrm{U} + 17\,^{1}_{0}\mathrm{n} \to {}^{255}_{92}\mathrm{U} \xrightarrow{\beta^-} {}^{255}_{93}\mathrm{Np} \xrightarrow{\beta^-} \ldots \xrightarrow{\beta^-} {}^{255}_{100}\mathrm{Fm}. \tag{94.8}$$

Всего в настоящее время известно 14 изотопов эйнштейния с массовыми числами от 243 до 256 и 16 изотопов фермия с массовыми числами от 242 до 258. Эйнштейний и фермий характеризуются еще меньшими периодами полураспада относительно процессов спонтанного деления

и испускания α-частиц, чем предшествующие трансураны. Самый долгоживущий изотоп эйнштейния $^{254}_{99}$Es имеет период α-распада около 480 дней, а самый долгоживущий изотоп фермия $^{257}_{100}$Fm — только 80 дней. У изотопа фермия $^{256}_{100}$Fm очень небольшой период полураспада относительно спонтанного деления — всего около 160 мин.

Можно было бы надеяться, что дальнейшие трансураны могут быть получены таким же способом, что и эйнштейний и фермий, если только увеличить мощность потока облучающих нейтронов. Но эти надежды не оправдались. В 60-х годах в США была произведена серия подземных ядерных взрывов. Поток нейтронов в наиболее мощном из них по своей интенсивности превосходил примерно в 10 раз соответствующий поток на атолле Бикини. Но трансураны с атомным номером выше атомного номера фермия не образовывались. Причиной этого является спонтанное деление ядер. В цепочке ядерных превращений появляется ядро, практически мгновенно испытывающее спонтанное деление и тем самым обрывающее эту цепочку.

Изложенные методы не позволяют получить ядра с $Z > 100$. Причины — недостаточная плотность нейтронных потоков, малая вероятность захвата большого числа нейтронов и (что наиболее важно) быстрый радиоактивный распад ядер с $Z > 100$.

7. Для получения трансуранов с $Z > 100$ пришлось обратиться к более старому методу, который применялся еще до открытия нейтрона. В этом методе мишень, содержащую атомные ядра одного элемента, подвергали бомбардировке атомными ядрами другого элемента. Происходило столкновение двух атомных ядер, в результате чего образовывалось новое атомное ядро. Оно могло оказаться промежуточным и испытать в дальнейшем радиоактивный распад с образованием других атомных ядер. В качестве мишени брали чаще всего уран или какой-либо из ранее полученных трансуранов, а в качестве бомбардирующего ядра — более легкие атомные ядра, например α-частицы.

Недостаток метода состоит в том, что оба сталкивающихся ядра заряжены положительно, а потому они отталкиваются друг от друга. Принято говорить, что между ядрами существует *кулоновский потенциальный барьер.* Чтобы при проникновении частицы-снаряда внутрь ядра-мишени между этими частицами произошла ядерная реакция, необходимо этот барьер преодолеть. Это сделать тем легче, чем больше энергия бомбардирующей частицы.

Естественные α-частицы, испускаемые радиоактивными элементами, обладают сравнительно небольшими энергиями (не превышающими 8 МэВ). Хотя они и могут преодолевать потенциальный барьер, но это происходит только в результате маловероятного квантовомеханического эффекта — туннельного просачивания через барьер. Поэтому-то вместо α-частиц, возникающих при распаде, применяются искусственные α-частицы, т. е. ускоренные ионы гелия $^{4}_{2}$He, а также ускоренные ионы других элементов.

В 1940–1950 гг. в распоряжении физиков имелись только маломощные ускорители, которые могли ускорять до небольших энергий

лишь простейшие ядра водорода — протоны ${}^{1}_{1}\mathrm{H}$, дейтроны ${}^{2}_{1}\mathrm{H}$, а также ядра гелия ${}^{4}_{2}\mathrm{He}$ (α-частицы). Для синтеза нового трансурана сначала получали в реакторе ядра самого тяжелого из известных трансуранов и из них приготовляли мишень, а затем бомбардировали ее протонами, дейтронами или α-частицами. В результате получался элемент, атомный номер которого на одну или две единицы был больше, чем у ядра мишени.

Примерами могут служить ядерные реакции

$$
{}^{2}_{1}\mathrm{H} + {}^{238}_{92}\mathrm{U} \to 2\,{}^{1}_{0}\mathrm{n} + {}^{238}_{93}\mathrm{Np} \xrightarrow[2{,}1\ \text{дня}]{\beta^-} {}^{238}_{94}\mathrm{Pu},
$$

$$
{}^{4}_{2}\mathrm{He} + {}^{239}_{94}\mathrm{Pu} \to {}^{242}_{96}\mathrm{Cm} + {}^{1}_{0}\mathrm{n},
$$

$$
{}^{4}_{2}\mathrm{He} + {}^{241}_{95}\mathrm{Am} \to {}^{243}_{97}\mathrm{Bk} + 2\,{}^{1}_{0}\mathrm{n},
$$

$$
{}^{4}_{2}\mathrm{He} + {}^{242}_{96}\mathrm{Cm} \to {}^{245}_{98}\mathrm{Cf} + {}^{1}_{0}\mathrm{n},
$$

с помощью которых впервые были синтезированы изотопы элементов 94, 96, 97, 98. Они были получены в США в 1940, 1944, 1949 и 1950 гг. соответственно. Отметим, что сначала был открыт кюрий, а потом америций. Впоследствии другие изотопы тех же элементов получались другими способами, в частности путем облучения ядер трансуранов и самого урана нейтронами. На образном языке технику эксперимента путем облучения трансуранов легкими заряженными частицами можно было бы назвать «легкой артиллерией».

Последним элементом, синтезированным с помощью «легкой артиллерии», был 101-й элемент, названный *менделевием* в честь великого русского химика Д. И. Менделеева. Его впервые удалось получить в Беркли (США) в 1955 г. в результате реакции

$$
{}^{4}_{2}\mathrm{He} + {}^{253}_{99}\mathrm{Es} \to {}^{256}_{101}\mathrm{Md} + {}^{1}_{0}\mathrm{n}. \tag{94.9}
$$

Мишенью служил невидимый глазом тонкий слой атомов изотопа эйнштейния-253 (всего около миллиарда атомов), нанесенный на тонкую (тоньше человеческого волоса) золотую фольгу. На ускорителе в Беркли эта фольга облучалась α-частицами с энергией 41 МэВ. Альфа-частица, проникая в ядро эйнштейния, увеличивала его заряд на 2 единицы. В результате образовывались нейтрон и ядро 101-го элемента, которые попадали на расположенную сзади вторую золотую фольгу и осаждались на ней. Со второй фольги в первом опыте химикам удалось собрать всего 17 атомов 101-го элемента. И все же химики определили некоторые радиоактивные и химические свойства этого элемента. Работой по синтезу 101-го элемента руководил А. Гиорсо (р. 1915).

8. Все трансурановые элементы от 93-го до 101-го включительно были впервые получены в США под руководством Г. Сиборга (р. 1912) и А. Гиорсо. Для получения трансуранов с большими атомными номерами от «легкой артиллерии» требовалось перейти к «артиллерии среднего калибра», что и было сделано. В качестве снарядов для бомбардировки мишеней стали использоваться ускоренные ионы углерода

$_6$C, азота $_7$N, кислорода $_8$O, неона $_{10}$N, аргона $_{18}$Ar. Существенно, чтобы бомбардирующие ионы были многозарядными, так как при прочих равных условиях энергия, приобретаемая ионом в циклическом ускорителе, пропорциональна квадрату заряда иона. В середине 60-х годов лишь физики Дубны (СССР) и Беркли (США) располагали ускорителями, пригодными для ускорения ионов аргона.

В дальнейшем для бомбардировки мишеней стали применять еще более массивные ионы:

$$_{26}\mathrm{Fe},\ _{27}\mathrm{Co},\ _{28}\mathrm{Ni},\ _{29}\mathrm{Cu},\ _{30}\mathrm{Zn},\ _{32}\mathrm{Ge},\ _{36}\mathrm{Kr},\ _{54}\mathrm{Xe}$$

и т. д. вплоть до $_{92}$U включительно. Короче говоря, в строй вступила «тяжелая артиллерия». Но чем больше заряд ядра иона, тем до большей кинетической энергии должен быть ускорен ион. Действительно, чтобы ядро ускоренного иона могло слиться с ядром мишени, кинетическая энергия иона должна превышать высоту кулоновского барьера между ядрами. А последняя пропорциональна зарядам обоих сталкивающихся ядер. Например, кулоновский барьер для цинка ($Z = 30$) в 15 раз выше кулоновского барьера для α-частицы ($Z = 2$). Вот почему для использования тяжелых ионов требуются ускорители на высокие энергии.

Для ускорения ионов при получении новых трансуранов в США пользовались линейными ускорителями различных типов, в том числе и перезарядными линейными ускорителями (тандемами), принцип действия которых такой же, по которому работает соответствующий тандем-генератор Ван-де-Граафа (см. § 84, п. 2). Линейными ускорителями пользовались и ученые Дармштадта (ФРГ) и во многих других лабораториях мира. В СССР в Объединенном институте ядерных исследований (Дубна) в Лаборатории ядерных реакций, руководимой Г. Н. Флеровым, для получения новых трансуранов и изучения их свойств применялись циклотроны. Поскольку ученым Дубны при изучении трансуранов принадлежит ведущее место в мире, мы остановимся только на краткой характеристике этих циклотронов. При этом мы совершенно не будем касаться способов выделения полученного трансуранового изотопа из смеси или соединения его с другими элементами, а также способов идентификации заряда и массового числа соответствующего ему ядра, хотя эти вопросы не менее трудны и важны, чем и получение самого изотопа.

9. В 1961 г. в Дубне для ускорения тяжелых ионов был построен циклотрон У-300, который для того времени был самым мощным в мире и специально приспособлен для синтеза новых трансурановых элементов. (Число 300 означает, что диаметр полюсов магнита в циклотроне составляет 300 см, точнее, 310 см. Аналогично, символы У-200 и У-400 означают циклотроны с диаметрами полюсов магнитов 200 и 400 см соответственно.)

Энергия, приобретаемая ионом в циклотроне, пропорциональна квадрату заряда иона q, квадрату радиуса r полюсов магнита и квадрату напряженности магнитного поля H (см. задачу 2 к этому параграфу). Масса магнита в циклотроне У-300 2200 т, мощность высоко-

частотного генератора, питающего циклотрона 500 кВт, напряжение между высокочастотными электродами-дуантами 260 кВ.

В 1966 г. в Дубне появился новый циклотрон У-200. Он стал одним из лучших в своем классе ускорителей (т. е. ускорителей с такими же диаметрами полюсов магнита) и оказался непревзойденным по интенсивности пучка легких ионов, а это очень важно для синтеза трансуранов. На строившихся до него циклотронах того же класса напряженность магнитного поля составляла, как правило, 14–16 кГс с вариацией амплитуды относительно этого значения около 4 кГс. На циклотроне У-200 удалось довести напряженность поля до 20 кГс, а амплитуду — до 6. В результате энергия ускоряемых частиц на циклотроне увеличилась вдвое, если учесть ее зависимость от магнитного поля, отмеченную выше. На циклотроне У-200 впервые были синтезированы изотопы 106-го и 107-го элементов, о чем подробнее будет сказано ниже.

С момента вступления в строй циклотрона У-200 появилась возможность дальше повышать энергию ускоряемых ионов комбинацией циклотрона У-300 с меньшим циклотроном У-200. Таким путем получался *перезарядный* или *тандем-циклотрон*, работающий по принципу, сходному с тем, по которому работает тандем-генератор Ван-де-Граафа (см. § 84, п. 2). В тандем-циклотроне ион сначала ускорялся циклотроном У-300, затем поступал в циклотрон У-200, пройдя предварительно через угольную пленку толщиной 0,2 мкм. В пленке частично «обдирается» электронная оболочка иона, так что положительный заряд иона увеличивается. Благодаря этому появляется возможность дальнейшего ускорения иона в циклотроне У-200, хотя он и меньше циклотрона У-300. Однако этот метод не получил распространения из-за своей сложности, а главным образом потому, что интенсивность ионных пучков на выходе получалась очень малой.

На циклотроне У-200, как уже отмечалось выше, были произведены важные исследования по синтезу трансурановых элементов. Но циклотрон У-200 главным образом был задуман как уменьшенная вдвое модель более мощного циклотрона У-400, который вступил в строй в конце декабря 1979 г. Циклотрон У-400 проектировался как специализированный ускоритель для получения ускоренных пучков ионов средней массы от неона ^{20}Ne до серебра ^{108}Ag. При его сооружении был учтен опыт строительства, эксплуатации и модернизации циклотронов У-300 и У-200, а также тандем-циклотрона. В результате У-400 оказался лучшим в мире в своем классе ускорителей. В частности, при той же массе электромагнита и том же потребляемом им токе питания, что и у ускорителя У-300, интенсивность ускоренных пучков на циклотроне У-400 возросла в сотни раз.

10. Обращаемся к вопросу о синтезе трансурановых элементов после получения 101-го элемента — менделевия. В 1956 г. к работам по синтезу 102-го элемента приступили ученые Нобелевского института физики в Стокгольме, Калифорнийского института в Беркли и Института атомной энергии в Москве.

Не прошло и года, как из Нобелевского института сообщили, что на стокгольмском циклотроне при облучении мишени из кюрия $^{244}_{96}$Cm

ионами углерода $^{13}_{6}$C обнаружены атомы 102-го элемента. Авторы работы предложили называть новый элемент *нобелием* ($_{102}$No) в честь учредителя Нобелевской премии. Однако через несколько месяцев ученые в Беркли в точности воспроизвели стокгольмские опыты и не нашли активности с описанными свойствами. Но и они сообщили данные о свойствах якобы 102-го элемента, впоследствии опровергнутые. Поэтому и их работа не может претендовать на открытие 102-го элемента.

Впервые элемент 102, точнее, его изотоп $^{254}102$ [1]) был с достоверностью синтезирован в Дубне в 1963 г. путем бомбардировки $^{238}_{92}$U ионами $^{22}_{10}$Ne и правильно по порядку величины был определен период полураспада синтезированного изотопа (уточненное значение этого периода 55 с). Эта работа и может претендовать на приоритет в получении 102-го элемента. Впоследствии данные, полученные в Дубне по всем изотопам 102-го элемента, были согласованы с данными, полученными в Беркли.

В 1961 г. из Беркли пришло сообщение о синтезе изотопа $^{257}103$. Элемент с $Z = 103$ был назван *лоуренсием* в честь изобретателя циклотрона Лоуренса. Однако данные, сообщенные об изотопе $^{257}103$, оказались ошибочными. Впоследствии сами авторы выразили сомнение в том, что массовое число обнаруженного ими изотопа равно 257, а не 258 или 259. В 1965 г. в Дубне с достоверностью был открыт изотоп $^{256}103$ путем бомбардировки америция атомами изотопа кислорода:

$$^{18}_{8}\mathrm{O} + {}^{243}_{95}\mathrm{Am} \to {}^{256}103 + 5\,{}^{1}_{0}\mathrm{n}. \qquad (94.10)$$

При этом были получены правильные данные о вновь открытом изотопе $^{256}103$. Поэтому можно считать, что элемент 103 был также впервые синтезирован в Дубне в 1965 г.

Элементы 102 и 103 не получили еще общепринятых названий. Ученые социалистических стран, работающие в Объединенном институте ядерных исследований в Дубне, где впервые с достоверностью были синтезированы указанные элементы, предложили называть элемент 102 *жолиотием* (Jl) в честь Фредерика Жолио-Кюри, а элемент 103 — *резерфордием* (Rf) в честь Резерфорда. Но и эти названия до сих пор еще не получили всеобщего признания. Споры о приоритете открытия элементов 102 и 103 и их названиях продолжаются до сих пор. По этой причине мы условно и сохраняем первоначальные названия — нобелий и лоуренсий.

В 1964 г. в Дубне на циклотроне У-300 был синтезирован 104-й элемент (точнее, изотоп $^{260}104$), названный *курчатовием* (Ku) в честь видного советского физика и организатора науки И. В. Курчатова (1903–1960). При синтезе этого элемента в качестве бомбардирующих частиц применялись ионы неона $^{22}_{10}$Ne с энергией 115 МэВ, а в качестве

[1]) Символ $^{254}102$ означает, что порядковый номер элемента равен 102, а его массовое число 254. Такая символика употребляется тогда, когда название элемента еще не установлено.

мишени использовался $^{242}_{94}\mathrm{Pu}$. Происходила ядерная реакция

$$^{22}_{10}\mathrm{Ne} + {}^{242}_{94}\mathrm{Pu} \to {}^{260}_{104}\mathrm{Ku} + 4\,{}^{1}_{0}\mathrm{n}, \qquad (94.11)$$

в результате которой и получился указанный изотоп курчатовия. Всего в настоящее время открыто восемь изотопов курчатовия.

11. Во всех рассмотренных выше способах получения трансурановых элементов путем бомбардировки мишеней заряженными ионами в качестве мишеней брался либо уран, либо ранее полученный трансуран. Последним элементом, найденным таким путем, был элемент 105, впервые синтезированный в Дубне в 1970 г. в результате реакции

$$^{22}_{10}\mathrm{Ne} + {}^{243}_{95}\mathrm{Am} \to {}^{261}105 + 4\,{}^{1}_{0}\mathrm{n}. \qquad (94.12)$$

Он был назван *нильсборием* (Ns) в честь Нильса Бора.

В 1973 г. Ю. Ц. Оганесян в Дубне предложил вместо тяжелых радиоактивных мишеней применять обычный свинец. Это предложение основывалось на следующих соображениях. Бомбардирующее ядро должно быть ускорено до громадной кинетической энергии, достаточной для преодоления кулоновского потенциального барьера между этим ядром и ядром мишени. Пусть в результате взаимодействия этих ядер сначала образуется компаунд-ядро. Оно получается в сильно возбужденном, или, как принято говорить, в сильно «нагретом» состоянии. В дальнейшем происходит «остывание» ядра.

С точки зрения капельной модели механизм остывания аналогичен тому, который описан в предыдущем параграфе (см. п. 11). Сильно возбужденное компаунд-ядро начинает совершать колебания. Если эти колебания становятся достаточно интенсивными, то ядро принимает форму гантели (см. рис. 159, положение *3*). Из шейки гантели, как было выяснено в указанном параграфе, начинают интенсивно испаряться мгновенные нейтроны. За счет этого испарения и происходит остывание компаунд-ядра. При испускании нейтрона ядро может разделиться на части, но может и не разделиться. Испарение одного нейтрона без деления ядра — процесс, больше чем в 100 раз менее вероятный по сравнению с испарением, которое сопровождается делением. При испускании двух нейтронов доля ядер, не испытавших деления, уменьшается больше чем в $100^2 = 10\,000$ раз, при испускании трех нейтронов — больше чем в $100^3 = 1\,000\,000$ раз, и т. д.

Таким образом, чем больше нейтронов испустит компаунд-ядро при образовании нового трансуранового элемента, тем менее вероятно появление последнего.

Допустим сначала, что мишенью служит малоустойчивое, или «рыхлое», ядро. При слиянии его с ядром-снарядом лишь относительно малая часть энергии последнего тратится на дополнительное «разрыхление» образовавшегося компаунд-ядра. Основная доля энергии ядра-снаряда расходуется на дальнейшее возбуждение уже «рыхлого» компаунд-ядра. Поэтому последнее получается сильно возбужденным, или «горячим». Такая ситуация имеет место, когда мишенью служит тяжелое радиоактивное ядро. Так, при использовании ионов C, O, Ne и мишеней из Pu, Cm, Cf образуются сильно возбужденные ядра

с энергией возбуждения 40–60 МэВ. Каждый испаряемый нейтрон способен унести из ядра энергию, в среднем равную 10–12 МэВ, так что для «остывания» компаунд-ядра и образования более устойчивого трансурана из компаунд-ядра должно вылететь до 5 нейтронов. Подтверждением этого могут служить ядерные реакции (94.10), (94.11), (94.12), рассмотренные выше. Таким образом, вероятность образования трансурана будет очень малой.

Допустим теперь, что в качестве мишени берется очень устойчивое, или «плотноупакованное», ядро. Тогда ядро-снаряд предварительно «разрыхляет» его, на что тратится значительная доля энергии. Лишь оставшаяся энергия ядра-снаряда идет на дальнейшее возбуждение образовавшегося компаунд-ядра. Последнее получается менее возбужденным, или «холодным».

Наиболее устойчивыми плотноупакованными ядрами являются магические и в особенности дважды магические ядра. У ядер с магическим числом протонов энергия, как функция числа N нейтронов, минимальна. Минимальна энергия и у ядер с магическим числом нейтронов, если ее рассматривать как функцию зарядового числа Z. Для дважды магических чисел энергия минимальна уже как функция обоих переменных Z и N. Ядро изотопа свинца $^{208}_{82}\mathrm{Pb}$ ($Z = 82, N = 126$) дважды магическое, все остальные ядра этого элемента магические ($Z = 82, N \neq 126$). Магическим является и ядро висмута $^{209}_{83}\mathrm{Bi}$. При использованиии мишеней из этих ядер компаунд-ядро получается сравнительна мало возбужденным. Так, при бомбардировке плотноупакованного ядра свинца ионами аргона, титана, хрома энергия возбуждения оказывается сравнительно низкой (всего 10–15 МэВ), так что для снятия возбуждения достаточно испускания уже одного-двух нейтронов. Примерами могут служить ядерные реакции (94.13), (94.14), приводимые ниже.

На основании изложенного можно было надеяться, что метод так называемого «холодного» или «магического» синтеза трансурановых элементов, предложенный Оганесяном, более эффективен, чем метод, в котором в качестве мишеней используются тяжелые радиоактивные элементы. Эти надежды оправдались.

12. Эффективность метода Оганесяна была испытана в Дубне прежде всего при облучении свинцовой мишени $_{82}\mathrm{Pb}$ ускоренными ионами $_{22}\mathrm{Ti}$. В результате этого сразу были открыты три новых изотопа курчатовия $_{104}\mathrm{Ku}$ (в зависимости от того, какие изотопы свинца и титана применялись на опыте). Несмотря на то что ускоренных ионов титана было в сотни раз меньше, чем ионов неона в реакции (94.11), на изучение свойств каждого изотопа курчатовия потребовалось всего несколько десятков часов работы циклотрона У-300, в то время как длительность первых опытов, в которых был синтезирован элемент 104, составляла свыше 10^4 ч.

Методом Оганесяна в Дубне в 1974 г. был открыт 106-й элемент. Он получался в реакции

$$^{54}_{24}\mathrm{Cr} + {}^{207}_{82}\mathrm{Pb} \to {}^{259}106 + 2\,{}^{1}_{0}\mathrm{n}, \qquad (94.13)$$

а в 1976 г. в реакции

$$^{54}_{24}\mathrm{Cr} + {}^{209}_{83}\mathrm{Bi} \to {}^{261}107 + 2\,{}^{1}_{0}\mathrm{n} \qquad (94.14)$$

был синтезирован 107-й элемент. Попутно были получены новые изотопы предшественников этих элементов по периодической системе Менделеева — от 100-го до 105-го.

Группа западно-германских экспериментаторов в Дармштадте на линейном ускорителе «Унилак» синтезировала методом Оганесяна новый изотоп 107-го элемента, воспроизвела и подтвердила многие другие результаты дубненских исследователей и получила ряд новых данных.

Предпринятые затем в Дубне и в Дармштадте попытки синтезировать 109-й элемент в реакции между висмутом ^{209}Bi и железом ^{58}Fe встретили затруднения. Выяснилось, что вероятность его образования в десятки раз меньше вероятности образования 107-го элемента. В 1982 г. в опыте западно-германских физиков, который длился двенадцать суток подряд, удалось зарегистрировать только одно событие, которое могло претендовать на распад 109-го элемента с атомной массой 266. Для дальнейшего продвижения вперед требовалось значительно повысить чувствительность опытов. В Дубне она была повышена в несколько десятков раз. Там же на рубеже 1982–1983 гг. в ходе 260-часового опыта было зафиксировано восемь событий, свидетельствовавших о рождении ядер 109-го элемента. Несмотря на столь малый эффект, можно было сделать первые общие суждения о свойствах этого элемента.

В начале 1984 г. в Дубне впервые были синтезированы два изотопа элемента 108 с массовыми числами 264 и 265. Они получались путем бомбардировки изотопов свинца $^{207}_{92}$Pb и $^{208}_{92}$Pb ионами железа ^{58}Fe. Изотоп с массовым числом 265 примерно в то же время был синтезирован и в Дармштадте. Весной 1984 г. в трех ядерных реакциях, проведенных на новом циклотроне У-400 (в реакциях свинца с железом и висмута с марганцем), были зарегистрированы 44 случая распада трех изотопов 108-го элемента с массовыми числами 263, 264 и 265. Все они чрезвычайно короткоживущи и распадаются не спонтанно, а испуская α-частицы. По спонтанному же делению они не менее живучи, чем 104-й и 106-й элементы, — вывод очень важный для теории и практики.

Одновременно три ядра 108-го элемента (точнее, изотопа $^{265}108$) были зарегистрированы в аналогичной ядерной реакции, проведенной на ускорителе «Унилак» группой западно-германских физиков. Располагая уникальным, пока единственным в мире чрезвычайно селективным масс-спектрографом, западно-германские физики установили важнейшие характеристики изотопа $^{265}108$: период полураспада — около 2 мс, энергия α-частиц — 10,4 МэВ.

В середине 1987 г. появилось сообщение о получении 110-го элемента периодической таблицы Менделеева. Ядро этого элемента было получено в Объединенном институте ядерных исследований в Дубне в результате бомбардировки тория и изотопов урана мощным пучком ускоренных ядер кальция и аргона. Всего было получено около 40 ядер 110-го элемента. Время жизни этих ядер оказалось порядка сотых долей

секунды. Для столь тяжелых элементов такое время жизни исключительно велико. Возможно, что это является указанием на существование острова стабильности сверхтяжелых элементов. Существование такого острова стабильности современная теория предсказывает в районе элементов с атомными номерами 110–114 (см. п. 13).

13. Трансурановые элементы испытывают все виды радиоактивного распада. Однако процессы β-распада относительно медленные; для ядер с $Z > 100$, имеющих короткие времена жизни относительно α-распада и спонтанного деления, роль β-распада невелика. С ростом Z конкуренция между спонтанным делением, α- и β-распадами проявляется все более отчетливо.

Нестабильность относительно спонтанного деления и определяет верхнюю границу периодической системы элементов. Период полураспада относительно спонтанного деления $^{238}_{92}\mathrm{U}$ равен 10^{16} лет, $^{239}_{94}\mathrm{Pu}$ — 10^{10} лет, $^{255}_{100}\mathrm{Fm}$ — 20 ч. Для $^{260}_{104}\mathrm{Ku}$ он снижается до 0,1 с, а для 106-го и 107-го элементов измеряется тысячными долями секунды.

Естественно, возникает вопрос, почему физики упорно работают над получением столь короткоживущих элементов и исследованием их свойств? К тому же такие элементы получаются в ничтожных количествах, иногда составляющих всего несколько атомов. Возможно, что в чисто научном отношении это и представляет интерес, но на непосредственное практическое применение рассматриваемых элементов вряд ли можно рассчитывать. Однако такое заключение в настоящее время является поспешным.

В 1966 г. В.М. Струтинский предложил метод расчета ядерных оболочек, который неплохо воспроизводил периодические изменения свойств изученных ядер. Экстраполируя найденные таким путем закономерности на неисследованную область, можно было попытаться предсказать свойства еще не полученных трансуранов. В последующее десятилетие физиками разных стран были произведены расчеты на самых сложных вычислительных машинах. Согласно этим расчетам ядра с числом протонов 114 или числом нейтронов 184 должны быть магическими, а ядро $^{298}114$ — дважды магическим.

Гипотетические элементы с числом протонов, близким к 114, и числом нейтронов около 184 называются *сверхэлементами*, а их ядра — *сверхядрами*. Есть основание ожидать, что сверхядра должны характеризоваться повышенной стабильностью, проявляющейся в относительно больших временах жизни. Поэтому совокупность значений Z и A сверхэлементов называют *островом стабильности*.

Основными типами радиоактивного распада сверхядер являются спонтанное деление и α-распад. По теории самым устойчивым по отношению к спонтанному делению должно быть дважды магическое ядро $^{298}114$. Однако время жизни по отношению к α-распаду у всех ядер сильно уменьшается с увеличением зарядового числа Z. Конкуренция между α-распадом и спонтанным делением должна по теории приводить к тому, что самым долгоживущим сверхэлементом будет изотоп $^{294}110$. Ожидается, что наиболее вероятное время жизни для него должно составлять 10^8–10^9 лет. Однако этот вывод основан на

теории ядра, а она далеко не совершенна. Удовлетворительная теория ядра еще не создана. Не исключено, что истинное время жизни во много раз больше или меньше приведенного выше, а самым долгоживущим ядром окажется одно из ядер с зарядовым числом между 108 и 126.

Предсказываемое теорией поведение гипотетических элементов в окрестности магического ядра $^{298}114$ характеризуется резким изменением устойчивости и времени жизни даже при незначительных изменениях хотя бы одного из чисел Z или N. Так, если у самого долгоживущего ядра $^{294}110$ изменить всего на 2–3 единицы число протонов или нейтронов, то время жизни по теории уменьшится примерно в миллионы раз. Такой эффект аналогичен тому, что наблюдается у обычных элементов в окрестности предшествующего дважды магического ядра свинца $^{208}_{82}\mathrm{Pb}$. Это ядро настолько устойчиво, что его радиоактивный распад никому еще не удавалось наблюдать. Но стоит к нему добавить всего один нейтрон, как время жизни полученного изотопа $^{209}_{82}\mathrm{Pb}$ из-за β^--распада снизится до 4,8 ч.

Таким образом, по теории при приближении к сверхэлементам и в их окрестности время жизни меняется нерегулярно: с возрастанием атомного номера оно то уменьшается, то снова увеличивается, достигая очень больших значений для некоторых изотопов на острове стабильности, так что они могут считаться практически стабильными. В этом отношении элементы в окрестности дважды магического ядра $^{298}114$ ведут себя аналогично элементам в окрестности предшествующего дважды магического ядра $^{208}_{82}\mathrm{Pb}$. У таллия ($Z = 81$) и следующих за ним свинца и висмута есть изотопы со столь большими временами жизни, что эти времена практически не удается измерить. За висмутом идут радиоактивные элементы: полоний ($Z = 84$), астат ($Z = 85$), радон ($Z = 86$) ... Все изотопы элементов от полония до актиния превращаются в дочерние за очень короткие времена. Зато периоды полураспада наиболее стабильных изотопов урана и тория — элементов с большими порядковыми номерами — равны соответственно 4,5 и 14 млрд лет и сравнимы с возрастом Земли. Нептуний, плутоний, америций, кюрий менее радиоактивны, чем полоний или радий, хотя их порядковые номера значительно выше.

14. Если гипотетический остров стабильности действительно существует и если бы удалось найти способ синтезирования долгоживущих сверхэлементов в макроскопических количествах, то научно-техническое значение такого открытия было бы трудно переоценить. Поэтому понятен интерес ученых к вопросу о сверхэлементах.

Вопрос о существовании сверхэлементов, в особенности стабильных и долгоживущих, может быть окончательно решен только экспериментально. Для этой цели можно воспользоваться уже рассмотренным нами методом синтеза трансурановых элементов путем слияния ядра-мишени с ядром-снарядом с последующим охлаждением образовавшегося компаунд-ядра в процессе испускания нейтронов. Помимо этого метода слияния есть и другой метод получения трансурановых элементов — метод деления ядер, предложенный Г.Н. Флеровым в 1965 г. В этом методе для бомбардировки урана используются ускоренные тяжелые

ионы (Xe, U). Ядро-снаряд на очень короткое время (порядка 10^{-21} с) слипается с ядром-мишенью с образованием сильно перегретого гантелеобразного ядра, быстро вращающегося вокруг его центра масс (см. задачу 4 к этому параграфу). В это время нуклоны в количестве до нескольких десятков имеют возможность перейти из одной половины гантелеобразного ядра в другую. Затем под действием громадной центробежной силы гантелеобразное ядро разрывается. Образовавшиеся осколки, будучи сильно перегретыми, сами испытывают деление. В результате деления образуются сотни изотопов самых разных элементов, в том числе и тяжелых.

И действительно, уже в первом опыте в Дубне на тандем-циклотроне были обнаружены нейтроноизбыточные ядра тяжелых трансуранов, которые ранее получались при термоядерных взрывах.

В дальнейшем подобные опыты неоднократно повторялись не только в Дубне, но и в других странах, в частности в ФРГ (Дармштадт), где был использован наиболее мощный в мире ускоритель ионов урана, причем мишенью служил также уран. Ученые ФРГ обнаружили, что ядерные «гантели» чаще всего делятся на два равных ядра урана. Более тяжелые и легкие осколки встречались реже. Чем больше различие в массах осколков, образовавшихся при делении ядерной «гантели», тем менее вероятно такое деление. Была надежда, что среди изотопов, образовавшихся при делении, появятся и изотопы сверхэлементов. Но эта надежда не оправдалась. С большим трудом удалось найти лишь ничтожные количества 102-го элемента.

15. Во многих лабораториях мира пытались найти ядерные реакции, в которых, по предположению, должны получаться магические ядра $^{298}114$, хотя бы и в ничтожных количествах. Примерами могут служить предполагаемые реакции

$$^{238}_{92}\mathrm{U} + {}^{238}_{92}\mathrm{U} \to {}^{476}184 \to {}^{166}_{70}\mathrm{Yb} + {}^{298}114 + 12\,{}^{1}_{0}\mathrm{n}, \tag{94.15}$$

$$^{136}_{54}\mathrm{Xe} + {}^{238}_{92}\mathrm{U} \to {}^{374}146 \to {}^{298}114 + {}^{72}_{32}\mathrm{Ge} + 4\,{}^{1}_{0}\mathrm{n}. \tag{94.16}$$

Однако практически осуществить все гипотетические реакции подобного рода пока никому не удалось. Таким образом, на вопрос о существовании острова стабильности сверхэлементов опыты пока не дают ответа.

16. Быть может, сверхэлементы или следы их существования надо искать в самой природе — на Земле или, лучше, в метеоритах? Не останавливаясь на подробностях, ограничимся только некоторыми, далеко не полными, принципиальными замечаниями по этому трудному вопросу.

Ядра урана и всякого трансуранового элемента испытывают спонтанное деление. Получающиеся осколки в момент своего образования обладают большой кинетической энергией — порядка 100 МэВ. Если деление произошло в кристаллической среде (минерале), то, двигаясь в ней, осколок до замедления производит множество разрушений — ведь атомы в решетке связаны относительно слабыми силами, для преодоления которых достаточно всего 2–3 эВ. Осколки оставляют в среде следы, или треки, в виде микроскопически узких каналов. Подвергнув минерал травлению, т. е. подействовав на него подходящей кислотой или щелочью, эти каналы можно расширить в сотни раз, так как вещество нарушенных участков кристалла гораздо легче

подвергается травлению, чем ненарушенных. Расширенные треки легко увидеть в обычный оптический микроскоп (с увеличением в 100 и более раз).

Продукты деления, накапливающиеся при делении, являются изотопами элементов, лежащих в середине периодической системы. Они могут быть выделены из исследуемого минерала, например, при нагревании его до 1000 °С и выше (разумеется, для этого не требуется производить травления минерала), а затем подвергнуты анализу с помощью масс-спектрографа. По изотопному составу того или иного элемента можно однозначно судить о том, какие ядра испытали спонтанное деление, в результате которого образовались выделенные элементы. Конечно, за время, прошедшее с момента образования изотопов, часть их могла рассеяться и уйти из исследуемого образца. Однако это обстоятельство не играет роли, так как для рассматриваемого метода существен лишь *относительный* изотопный состав, а не абсолютные количества образовавшихся изотопов. А относительный изотопный состав одного и того же элемента не меняется при их рассеянии.

В изложенном методе наиболее удобны изотопы ксенона, так как естественная распространенность этого элемента в земных и космических минералах очень мала — она в миллионы и миллиарды раз меньше, чем распространенность других химических элементов. Поэтому даже незначительные добавки ксенона, образующиеся при делении, легко зафиксировать на слабом фоне обычного ксенона, содержащегося, например, в атмосфере. У других элементов природный фон полностью маскирует такие добавки.

Для проблемы обнаружения следов спонтанного деления ядер сверхэлементов более подходящими являются минералы космического, а не земного происхождения. Дело в том, что на Земле такие следы не могут сохраняться столь длительное время, как это возможно в метеоритах. На Земле они исчезают под действием многочисленных разрушающих факторов. При исследовании рассматриваемой проблемы предпочтение было отдано *метеоритам-ахондритам*. Так называются метеориты, в состав которых входят *хондры* — затвердевшие капли силикатного вещества, возможно, образовавшегося еще в начальной стадии формирования Солнечной системы. Размеры хондр колеблются от долей миллиметра до сантиметров.

17. При исследовании кристаллов минералов, извлеченных из некоторых метеоритов-ахондритов, были обнаружены миллионы следов осколков деления каких-то ядер на каждом квадратном сантиметре. Однако сами ядра обнаружены не были. Естественно было предположить, что время жизни их недостаточно велико, так что за время существования Солнечной системы они полностью распались. Что же это за ядра?

При термоядерных взрывах, как уже говорилось в п. 6, обнаружен плутоний-244 — наиболее долгоживущий элемент из всех известных трансурановых элементов (период полураспада ^{244}Pu равен $8{,}3\cdot10^7$ лет). Состояние вещества, подобное тому, какое было при термоядерном взрыве, но уже в громадных космических масштабах существовало и при возникновении Солнечной системы — в период образования химических элементов. Несомненно, что в этот период образовывался и плутоний-244, и другие трансурановые элементы, но к настоящему времени все они полностью исчезли на Земле из-за сравнительно коротких времен жизни.

Следы спонтанного деления ядер, оставленные в метеоритах-ахондритах, не могут принадлежать ядрам урана-238 или тория-232, так как в противном случае эти долгоживущие элементы сохранились бы в метеоритах до настоящего времени, как это имеет место на Земле. Можно думать, что эти следы оставлены ядрами трансуранов. Однако все известные относительно короткоживущие трансураны должны быть сразу же исключены, поскольку они распались бы еще задолго до затвердевания метеоритов. Исключение

составляет только самый долгоживущий трансуран, а именно плутоний-244. Только он мог оставить рассматриваемые следы, а затем еще задолго до настоящего времени полностью исчезнуть из метеоритов. Это предположение подтвердилось экспериментально.

В метеоритах сохранились не только треки, оставленные осколками деления ядер, но и сами осколки в виде соответствующих атомов. Эти атомы частично можно было извлечь из метеоритов путем нагревания их до температуры 1000–1200 °С и масс-спектроскопически определить изотопный состав извлеченного газа. Оказалось, что в этом газе среди прочих изотопов содержатся и изотопы ксенона, представляющие для нас основной интерес. Как уже подчеркивалось, важно знать только относительное содержание изотопов ксенона, а не количество его, сохранившееся в метеорите.

С другой стороны, в потоке нейтронов, исходящем из ядерного реактора, синтезируется множество элементов, в том числе и ^{244}Pu. Несколько миллиграммов этого изотопа было помещено в кварцевую ампулу, из которой затем был выкачан воздух, после чего сама ампула была герметически запаяна. Из-за спонтанного деления ядер плутония в ампуле среди прочих продуктов образовывался и ксенон. Примерно через два года ампулу вскрыли, а образовавшийся ксенон был исследован с помощью масс-спектрометра. Хотя за это время путем деления распалась мизерная часть плутония-244, однако при высокой чувствительности современных приборов ее оказалось достаточно, чтобы исследовать изотопный состав полученного ксенона. Он оказался в точности таким же, как и в метеоритах-ахондритах. Это и доказывает, что следы деления, наблюдавшиеся в метеоритах, оставлены ядрами именно плутония-244.

18. Возникла идея, что в углистых метеоритах-хондритах надо искать следы деления ядер гипотетических сверхэлементов. (Углистыми называют метеориты, в состав которых входят углеродсодержащие минералы и соединения. В этих веществах в метеорите и содержатся хондры.) Их былое существование можно было бы обнаружить по изотопному составу ксенона, содержащегося в метеоритах. В основе такой идеи лежали следующие соображения. На основе квантовой механики с использованием ЭВМ было рассчитано строение электронных оболочек гипотетических сверхтяжелых атомов и тем самым определены их химические свойства. Оказалось, что наиболее устойчивые сверхэлементы должны быть химическими аналогами легколетучих элементов вроде свинца. А из всех метеоритов наиболее богаты такими элементами именно углистые метеориты.

Из углистых метеоритов-хондритов путем нагревания до 1000–1200 °С были выделены нейтроноизбыточные изотопы ^{136}Xe, ^{134}Xe, ^{132}Xe и ^{131}Xe. Казалось, что эти изотопы выделялись при делении ядер какого-то трансурана. Но все ядра известных трансуранов при делении не выделяют ксенон с таким же изотопным составом. Первоначально это считалось доказательством былого существования сверхэлемента, при делении ядер которого и получился ксенон наблюдаемого изотопного состава. Однако такое заключение вскоре было опровергнуто. Действительно, непосредственно при делении сначала получаются нейтроноизбыточные, но β^--неустойчивые ядра (см. § 93, п. 11). Претерпев β^--распад, они в конце концов переходят в стабильные ядра ксенона. Эти стабильные ядра также еще должны быть нейтроноизбыточными. Нейтронодефицитные ядра ксенона при делении получиться не могут. А между тем в метеоритах они содержатся. Оказалось, что чем больше нейтроноизбыточных стабильных ядер, тем больше и нейтронодефицитных стабильных ядер. Это не согласуется с предположением, что все стабильные изотопы ксенона в метеоритах получились в результате деления. Другие контрольные опыты также подтвердили такое заключение.

Таким образом, исследование изотопного состава ксенона в метеоритах не принесло доказательства существования в настоящее время или в отдаленном прошлом сверхэлементов. Другие методы, основанные на измерении длины треков в метеоритах и на Земле, также пока не дали окончательного ответа на вопрос о существовании сверхэлементов. Для этого требуются дальнейшие исследования.

ЗАДАЧИ

1. Для качественного уяснения влияния электронного захвата на стабильность электронной оболочки атома можно предположить, что захват электронов как бы мгновенно происходит с боровской орбиты с главным квантовым числом n. Исходя из такого представления, оценить наименьшее значение числа Z, выше которого стабильность электронной оболочки атома становится невозможной, предполагая, что $n = 1, 2, 3, 4, 5$.

Решение. Такое представление эквивалентно предположению, что внутри сферы с радиусом n-й боровской орбиты электронов нет, а размер ядра (при том же заряде) как бы увеличен до размеров этой сферы. Но радиус n-й боровской орбиты пропорционален n^2. Тем самым задача свелась к задаче, разобранной в п. 2, если только в уравнении (94.1) r_0 заменить на $r_0 n^2$. Иначе говоря, надо оперировать с неравенством

$$\frac{A^{4/3}}{1{,}97 + 0{,}015A^{2/3}} \lesssim \frac{38\,000}{n^2}. \tag{94.1б}$$

Из него находим, что для стабильности электронной оболочки необходимо выполнение условий:

$$Z \lesssim 1460 \quad (n = 1), \quad Z \lesssim 370 \quad (n = 3),$$

$$Z \lesssim 630 \quad (n = 2), \quad Z \lesssim 250 \quad (n = 4),$$

$$Z \lesssim 185 \quad (n = 5).$$

2. Определить в нерелятивистском приближении кинетическую энергию, приобретаемую зарядом q массой m, если радиус орбиты, по которой он вращается, равен r, а магнитное поле H однородно.

Ответ. $\mathscr{E} = q^2R^2H^2/(2mc^2)$.

3. Ядро ускоренного иона ^{238}U сталкивается с покоящимся ядром того же урана и на короткое время сливается с ним в гантелеобразное вращающееся ядро. Оценить минимальную угловую скорость вращения ω полученного гантелеобразного ядра.

Решение. Момент импульса вращающегося гантелеобразного ядра определяется формулой $L^2 = l(l+1)h^2$, в которой следует положить $l = 1$. Таким образом, $L = \sqrt{2}\hbar$. С другой стороны, $L = J\omega$, где J — момент инерции гантелеобразного ядра относительно его центра масс. Для оценки J заменим гантель двумя соприкасающимися ядрами урана. Момент инерции сплошного шара относительно касательной к нему равен $(7/5)mr^2$, а двух шаров относительно общей касательной — $(14/5)mr^2$. Радиус ядра урана $r = 8{,}7 \cdot 10^{-13}$ см, $m = 4 \cdot 10^{-22}$ г, $J = 8{,}5 \cdot 10^{-46}$ г $\cdot$ см^2. Следовательно,

$$\omega = \frac{\sqrt{2} \cdot 1{,}05 \cdot 10^{-27}}{8{,}5 \cdot 10^{-46}} = 1{,}7 \cdot 10^{18}\ \text{с}^{-1}.$$

4. В урановых рудах обнаружены ничтожные количества $^{239}_{94}\mathrm{Pu}$. Такой плутоний образуется из $^{238}_{92}\mathrm{U}$ под действием нейтронных потоков, всегда имеющихся в руде. Нейтронные потоки в свою очередь возникают и поддерживаются при спонтанном делении ядер урана или под действием космических лучей.

Оценить долю атомов $^{239}_{94}\mathrm{Pu}$, отнесенную к числу атомов руды, предполагая, что руда состоит из чистого урана-238 и что плутоний образуется только под действием нейтронов от спонтанного деления ядер урана. Период полураспада урана-238 относительно спонтанного деления $T^{\mathrm{U}}_{1/2} \approx 10^{16}$ лет, плутония $T^{\mathrm{Pu}}_{1/2} \approx 2{,}4 \cdot 10^4$ лет.

Р е ш е н и е. Предположим, что каждый нейтрон, возникающий при спонтанном делении ядра урана, поглощается ядром того же урана с образованием плутония. Как при всяком радиоактивном равновесии, искомая доля плутония $N_{\mathrm{Pu}}/N_{\mathrm{U}}$ найдется из требования, чтобы число распадающихся в единицу времени при спонтанном делении ядер урана было равно числу распадающихся за то же время ядер плутония. Это дает

$$N_{\mathrm{Pu}}/N_{\mathrm{U}} = T^{\mathrm{U}}_{1/2}/T^{\mathrm{Pu}}_{1/2} \approx 10^{-12}.$$

В действительно встречающихся рудах отношение числа атомов плутония к числу атомов руды примерно на два порядка меньше приведенного значения. Причина расхождения заключается в том, что $N_{\mathrm{U}} < N_{\text{руды}}$, а главное в том, что большая часть нейтронов, возникающих при спонтанном делении ядер урана, поглощается ядрами других веществ руды, а также уходит наружу. Кроме того, не всякий нейтрон, поглощенный ядром урана, вызывает деление.

§ 95. Цепная реакция и ядерные реакторы

1. При делении тяжелых ядер освобождается энергия, равная в среднем около 200 МэВ на каждое делящееся ядро (см. § 93). Эта величина в сотни миллионов и миллиарды раз превосходит энергию, освобождающуюся в единичном акте химических реакций (последняя величина в лучшем случае составляет несколько электронвольт, а обычно — десятые и даже сотые доли электронвольта). Энергия, освобождающаяся при делении атомных ядер, называется *ядерной* или *атомной энергией.* Ее получение в больших макроскопических количествах производится в *ядерных реакторах.* Теория ядерных реакторов, включающая наряду с физическими множество технических проблем, весьма обширна и сложна. Из этой теории в общем курсе физики, понятно, можно затронуть только немногие принципиальные вопросы чисто физического порядка.

Вещества, используемые в ядерных реакторах для осуществления ядерных реакций деления, называются *ядерным топливом.* В природе в естественном состоянии встречается только одно ядерное топливо — уран. Естественный уран содержит примерно 0,7 % изотопа $^{235}\mathrm{U}$. Собственно, изотоп $^{235}\mathrm{U}$ и является *ядерным горючим*, поддерживающим реакцию ядерного деления. В естественном уране почти 99,3 % составляет изотоп $^{238}\mathrm{U}$. Он, а также $^{232}\mathrm{Th}$ являются сырьевыми материалами, из которых искусственно получаются делящиеся изотопы $^{239}\mathrm{Pu}$ и $^{233}\mathrm{U}$, являющиеся также ядерным горючим. Но эти изотопы в естественном

состоянии в природе не встречаются. Плутоний-239 получается в реакции (94.2), а уран-233 — в реакции

$$^{232}_{90}\mathrm{Th}(\mathrm{n},\gamma)^{233}_{90}\mathrm{Th} \xrightarrow[22{,}4\text{ мин}]{\beta^-} {}^{233}_{91}\mathrm{Pa} \xrightarrow[27{,}4\text{ дня}]{\beta^-} {}^{233}_{92}\mathrm{U} \xrightarrow[1{,}6\cdot 10^5\text{ лет}]{\alpha} . \qquad (95.1)$$

Высшие трансураны — америций, кюрий, берклий, калифорний, ввиду трудности и дороговизны их получения, не могут быть использованы для освобождения ядерной энергии в больших количествах.

Основную роль в ядерной энергетике играют ядра изотопов $^{235}_{92}\mathrm{U}$, $^{239}_{94}\mathrm{Pu}$, $^{233}_{92}\mathrm{U}$, содержащие нечетное число нейтронов, так как они делятся нейтронами любых энергий, в том числе и тепловыми. Под *тепловыми* понимают нейтроны, энергия которых порядка kT. При комнатной температуре (300 K) $kT \approx 4{,}14 \cdot 10^{-14}$ эрг $\approx 0{,}025$ эВ. При таких малых энергиях сечение деления пропорционально $1/v$, где v — скорость нейтрона. Если энергия нейтрона $\mathscr{E} \gtrsim 1$ МэВ, то такие нейтроны называются *быстрыми*. Ядра $^{238}\mathrm{U}$ и $^{232}\mathrm{Th}$ медленными нейтронами не делятся, а делятся только быстрыми нейтронами.

Важнейшей величиной, характеризующей делящееся ядро, является среднее число вторичных нейтронов ν, вылетающих при делении одного ядра. Другая величина η характеризует уже не отдельное ядро, а однородную среду из одинаковых ядер, причем размеры среды достаточно велики (в пределе — бесконечно велики). Это есть среднее число вторичных нейтронов, приходящееся на единичный акт захвата нейтрона ядром. Числа ν и η не совпадают между собой, так как нейтроны в среде вызывают не только деление ядер, но также захватываются последними, не производя деления, но испуская γ-кванты, т. е. вступают в реакцию *радиационного захвата* (n,γ). (Упругое рассеяние нейтронов на ядрах в размножении числа нейтронов непосредственной роли не играет.) Пусть $\sigma_{\mathrm{n}f}$ — сечение деления, а $\sigma_{\mathrm{n}\gamma}$ — сечение радиационного захвата. Тогда средняя доля ядер, способных к делению и фактически испытавших последнее, будет $\sigma_{\mathrm{n}f}/(\sigma_{\mathrm{n}f}+\sigma_{\mathrm{n}\gamma})$. Среднее число испущенных вторичных нейтронов, приходящееся на один акт захвата нейтрона способным к делению ядром, будет

$$\eta = \nu \frac{\sigma_{\mathrm{n}f}}{\sigma_{\mathrm{n}f}+\sigma_{\mathrm{n}\gamma}}. \qquad (95.2)$$

Все величины, входящие в эту формулу, зависят от энергии нейтронов, так что сама формула относится только к моноэнергетическим нейтронам.

В табл. 16 приведены значения ν и η для ядер, делящихся тепловыми и быстрыми нейтронами. В области не слишком больших возбуждений ядра ($\mathscr{E} \lesssim 10$ МэВ) при возрастании энергии возбуждения на 1 МэВ число ν увеличивается примерно на 0,11.

2. Из табл. 16 видно, что $\eta > 1$ (и притом заметно больше единицы). Это есть необходимое (но недостаточное) условие для осуществления *цепной ядерной реакции деления атомных ядер.* Термин «цепная реакция» заимствован из химии. Под цепной понимают такую реакцию, в которой воспроизводится, и притом в большем количестве, одно из

ществляется на так называемом *обогащенном уране.* Как уже было сказано, основную массу естественного урана составляет изотоп ^{238}U (99,3 %), а изотоп ^{235}U содержится в нем в количестве примерно 0,7 %. В обогащенном уране содержание изотопа ^{235}U доводится до 2–5 %. Обогащение естественного урана производится путем *разделения изотопов.* Это очень длительный и дорогостоящий процесс. Его практически невозможно осуществить химическими методами, так как химические свойства изотопов ^{235}U и ^{238}U почти одинаковы. Приходится применять различные физические методы, использующие небольшое различие масс атомов подлежащих разделению изотопов. К таким методам в случае тяжелых элементов относятся главным образом *газовая диффузия* через пористые перегородки, *термодиффузия*, *центрифугирование*, *электромагнитный метод разделения изотопов.* Изотоп ^{235}U, добавляемый к естественному урану, может быть заменен плутонием-239 или ураном-233, свойства которых по отношению к реакции деления аналогичны свойствам урана-235.

Для смеси изотопов, заполняющей активную зону, сохраняет смысл введенный выше коэффициент η, как среднее число вторичных нейтронов деления, приходящееся на один нейтрон, поглощенный смесью. Но формула (95.2), разумеется, в этом случае неприменима. Если делящимся является только один изотоп в смеси, то она должна быть заменена на

$$\eta = \frac{\nu\sigma_{\mathrm{n}f}}{\sigma_{\mathrm{n}t}}, \tag{95.3}$$

где $\sigma_{\mathrm{n}f}$ — сечение деления, а $\sigma_{\mathrm{n}t}$ — полное сечение захвата нейтрона всеми компонентами смеси. Так, сечение деления урана-235 на тепловых нейтронах $\sigma_{\mathrm{n}f}^{(235)} = 582$ барн, сечения радиационного захвата $\sigma_{\mathrm{n}\gamma}^{(235)} = 100$ барн, $\sigma_{\mathrm{n}\gamma}^{(238)} = 2{,}73$ барн. Так как в естественном уране на один атом урана-235 приходится 140 атомов урана-238, то полное сечение для естественного урана на тепловых нейтронах будет

$$\sigma_{\mathrm{n}t} = \sigma_{\mathrm{n}f}^{(235)} + \sigma_{\mathrm{n}\gamma}^{(235)} + 140 \cdot \sigma_{\mathrm{n}\gamma}^{(238)} = 1064 \text{ барн}.$$

Для тепловых нейтронов $\nu = 2{,}42$ (см. табл. 16). Поэтому для естественного урана и тепловых нейтронов

$$\eta = \frac{\nu\sigma_{\mathrm{n}f}^{(235)}}{\sigma_{\mathrm{n}t}} = 1{,}32.$$

Важнейшей величиной, характеризующей активную зону реактора, является *коэффициент размножения* нейтронов k. Это есть отношение общего числа нейтронов в каком-либо поколении к породившему их общему числу нейтронов в предыдущем поколении. Таким образом, если в первом поколении было N_1 нейтронов, то их число в n-м поколении будет $N_n = N_1 k^n$.

При $k = 1$ реакция деления *стационарна*, т. е. число нейтронов во всех поколениях одинаково — размножения нейтронов нет. Соответствующее состояние реактора называется *критическим.* При $k > 1$

число N_n растет с увеличением n — такое состояние называется *надкритическим.* При $k < 1$ число N_n убывает с увеличением n — такое состояние называют *подкритическим.* В предельном идеальном случае, когда реактор состоит из одной только активной зоны, причем последняя бесконечно велика, коэффициент размножения нейтронов снабжают индексом ∞, т. е. обозначают через k_∞. Очевидно, коэффициент k_∞ является характеристикой только среды, заполняющей активную зону, и не имеет никакого отношения к устройству реактора.

От каких же параметров зависит коэффициент размножения нейтронов? Если бы активная зона состояла только из одного делящегося изотопа (например, урана-235), то было бы $k_\infty = \eta$. В случае реального реактора это не так. Для конкретности мы будем иметь в виду реактор на естественном или слабо обогащенном уране. Основная часть вторичных нейтронов появляется при делении ядер урана-235 тепловыми нейтронами. Но при делении образуются главным образом быстрые нейтроны. Чтобы сделать их более эффективными, используют различные *замедлители*, при рассеянии на ядрах которых нейтроны и замедляются до тепловых скоростей. Поэтому рассматриваемый реактор является *реактором на тепловых нейтронах.* Приводимая ниже формула (95.4) относится именно к таким реакторам.

В процессе замедления нейтрон может поглотиться ядрами ^{238}U, ядрами замедлителя и ядрами других конструктивных элементов реактора. Ясно, что коэффициент размножения пропорционален вероятности p того, что нейтрон не будет поглощен такими ядрами. Конечно, некоторые ядра урана-238 разделятся быстрыми нейтронами, еще не успевшими замедлиться. Это обстоятельство учитывается введением множителя ε — *коэффициента размножения на быстрых нейтронах.* Коэффициент размножения на быстрых нейтронах есть отношение числа нейтронов, возникших при делении на быстрых и тепловых нейтронах к числу нейтронов, возникших при делении только на тепловых нейтронах. Введем еще коэффициент теплового использования f — вероятность того, что замедлившийся нейтрон будет поглощен ядрами изотопов урана, а не ядрами замедлителя и других конструктивных материалов реактора. Тогда

$$k_\infty = \eta p f \varepsilon. \tag{95.4}$$

Поскольку размеры активной зоны конечны, часть нейтронов будет уходить из нее наружу. Поэтому можно положить

$$k = k_\infty P, \tag{95.5}$$

где P — вероятность того, что нейтрон не уйдет из активной зоны. Эта формула справедлива для реакторов любого вида, независимо от конкретного выражения для k_∞. Она сохраняет смысл и для атомной бомбы, активная зона которой состоит из чистого урана-235 или плутония-239. Величина P зависит от состава, размеров и формы активной зоны, а также от среды, окружающей активную зону. Если

такая среда отражает нейтроны обратно в активную зону, то она увеличивает P. В этом случае она называется *отражателем* и состоит из легких атомов, слабо поглощающих нейтроны (графит, бериллий).

Как уже указывалось выше, для работы реактора необходимо, чтобы $k \equiv k_\infty P \geqslant 1$. В пределе, когда $k_\infty P = 1$, цепная реакция, происходящая в реакторе, является самоподдерживающейся, стационарной. Размеры и массу активной зоны, а также вероятность P, когда в реакторе становится возможной самоподдерживающаяся цепная реакция, т. е. когда

$$P = P_{\text{крит}} = 1/k_\infty, \tag{95.6}$$

называют *критическими*. Эти величины, конечно, зависят от состава и формы активной зоны, а также от устройства реактора.

Для уменьшения утечки нейтронов активной зоне придают сферическую или близкую к сферической форму, например форму цилиндра с высотой порядка диаметра или форму куба, так как для тел такой формы отношение поверхности к объему тела минимально или близко к минимальному. Для чистого ^{235}U критическая масса приблизительно равна 0,8 кг, для ^{239}Pu — 0,5 кг, для ^{251}Cf — 10 г.

Если масса активной зоны значительно превышает критическую, то цепная реакция приобретает характер взрыва. На этом принципе основано действие атомной бомбы. Последняя состоит из двух или больших частей, содержащих уран-235 или плутоний-239, масса каждого из которых значительно меньше критической. Когда эти части разведены, то и масса всех частей вместе меньше критической. В таком состоянии рассматриваемое ядерное устройство не взрывается. Но если с помощью химического взрыва (при котором развивается давление порядка миллиона атмосфер) части быстро сблизить, то масса полученной активной зоны станет заметно больше критической, и произойдет ядерный взрыв.

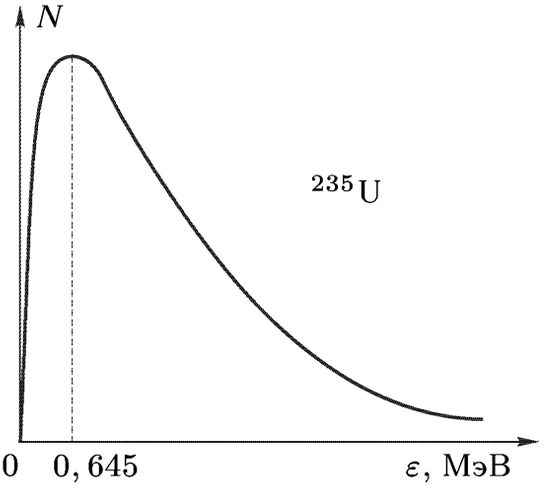


Рис. 165

4. Выясним теперь более подробно, почему в реакторах на природном или слабообогащенном уране требуется замедлитель нейтронов. Примерный энергетический спектр нейтронов, возникающих при делении урана-235, представлен на рис. 165. По оси абсцисс отложена кинетическая энергия нейтронов $\mathscr{E}$ в системе отсчета, в которой исходное делящееся ядро покоится. По оси ординат в произвольных единицах отложено число нейтронов деления $N(\mathscr{E})$, приходящееся на единичный интервал энергии. Таким образом, $N(\mathscr{E})\,d\mathscr{E}$ будет числом нейтронов, энергия которых заключена в интервале между $\mathscr{E}$ и $\mathscr{E} + d\mathscr{E}$. Кривая имеет максимум примерно при $\mathscr{E} = 0{,}645$ МэВ.

Как видно из рис. 165, при делении ^{235}U образуются преимущественно быстрые нейтроны. Ядра урана-238 делятся только под действием быстрых нейтронов с энергией $\mathscr{E} \approx 1$ МэВ. Но эффективное сечение такого деления очень мало — около 0,3 барна. Напротив, ядра урана-235 делятся под действием нейтронов любых энергий, причем эффективное сечение деления здесь быстро возрастает с уменьшением энергии. Для тепловых нейтронов оно достигает значения 582 барна. Поэтому было бы выгодно замедлить нейтроны до тепловых энергий.

Конечно, при упругих столкновениях с ядрами урана-238 и урана-235 нейтроны замедляются, но этот процесс идет очень медленно ввиду большой массы ядер урана. Главным же препятствием для замедления нейтронов в естественном уране является радиационный захват нейтронов ядрами урана-238. Особенно интенсивно процесс радиационного захвата идет в резонансной области энергий. На рис. 166 представлена

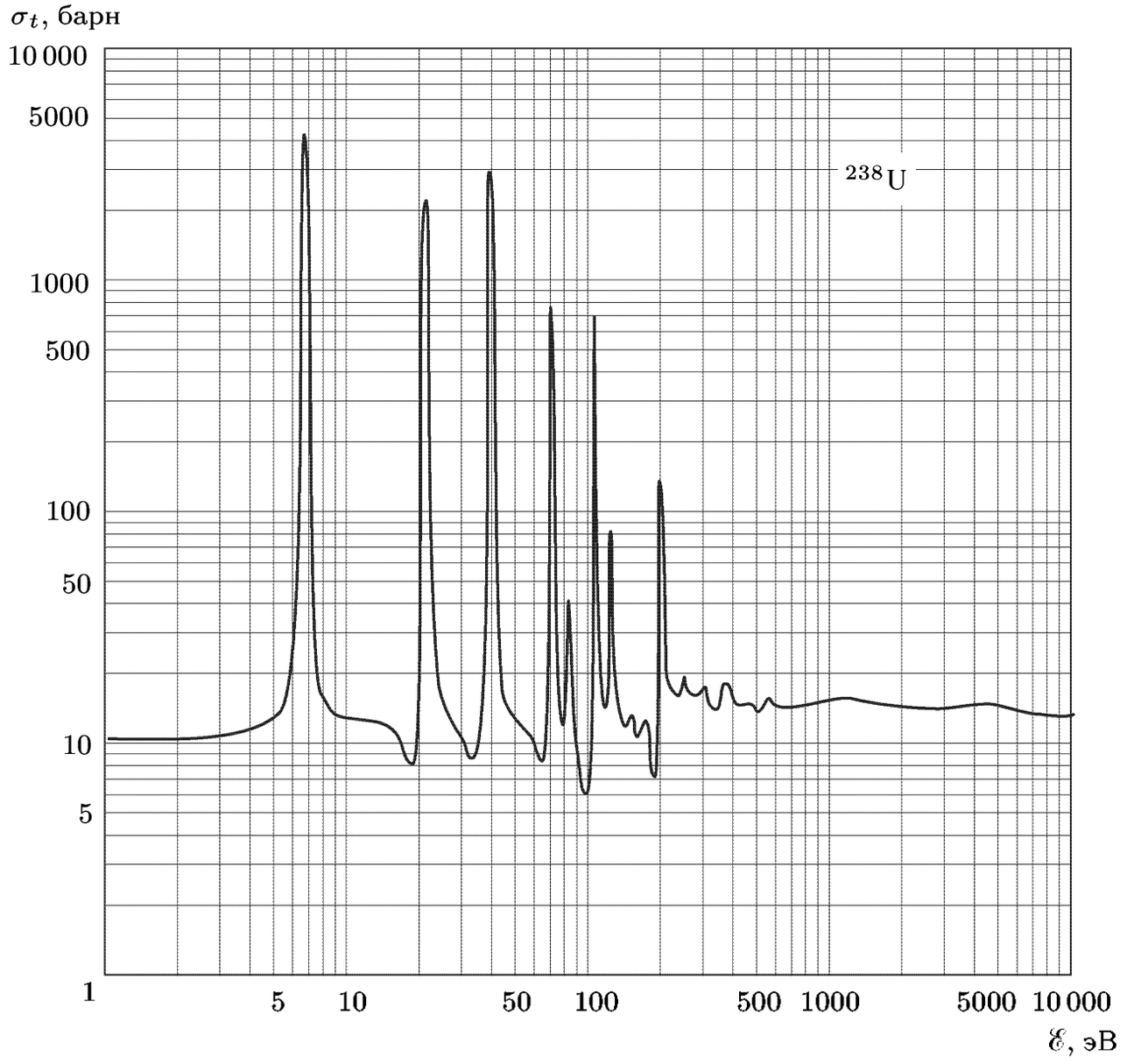


Рис. 166

зависимость полного эффективного сечения захвата σ_t нейтронов ядрами урана-238 в зависимости от энергии нейтрона. В резонансной

области энергий наблюдаются частые и резкие пики, в которых σ_t достигает нескольких тысяч барн, причем $\sigma_t \approx \sigma_{n\gamma}$, так как $\sigma_t = \sigma_{nn} + \sigma_{n\gamma}$, но $\sigma_{nn} \ll \sigma_{n\gamma}$. При замедлении нейтроны должны пройти через резонансную область, где они поглощаются и поэтому не достигают области тепловых энергий. Кроме того, известная доля нейтронов уходит за пределы активной зоны. Для того чтобы как можно более значительная доля нейтронов замедлилась до тепловых скоростей, применяется замедлитель, т. е. вещество, состоящее из легких ядер, слабо захватывающих нейтроны. В качестве замедлителя часто используется графит. Применяются также тяжелая и обычная вода (обычная вода в реакторах на природном уране не применяется из-за большого поглощения нейтронов ядрами водорода), бериллий и его окись, а также некоторые органические вещества.

Если активная зона ядерного реактора содержит замедлитель нейтронов, то основная часть ядерных делений происходит под действием тепловых нейтронов. Реакторы с замедлителем бывают *гомогенными* и *гетерогенными*. В гомогенном реакторе ядерное топливо и замедлитель нейтронов представляют собой однородную смесь (раствор или суспензию). В гетерогенном реакторе ядерное топливо в активной зоне располагается дискретно в виде вертикальных стержней, называемых *тепловыделяющими элементами* (ТВЭЛами). Обычно ТВЭЛы образуют правильную решетку, а между ними располагается замедлитель. В гомогенном реакторе на естественном уране цепная реакция не идет, так как в результате поглощения нейтронов в резонансной области тепловых нейтронов остается недостаточно и коэффициент размножения нейтронов k получается меньше единицы. Напротив, при наличии достаточно большого количества замедлителя в гетерогенном реакторе можно получить $k > 1$ даже для естественного урана, так что развитие цепной реакции становится возможным. Это объясняется тем, что большая часть нейтронов, получающихся при делении, попадает из ТВЭЛов в замедлитель, а после замедления уже в виде тепловых нейтронов возвращается опять в ТВЭЛы, где вызывает новые ядерные деления.

Первый ядерный реактор был построен под руководством Ферми в декабре 1942 г. в Чикаго. Это был гетерогенный реактор, собранный из 45 т естественного урана и 450 т графита ($k = 1{,}0006$). Его первоначальная мощность была 200 Вт. Позднее этот реактор был демонтирован и собран вновь, а его мощность была доведена до 100 кВт. Сходными параметрами обладал и первый в Европе ядерный реактор, запущенный в СССР в декабре 1946 г. под руководством И. В. Курчатова. Он содержал 50 т естественного урана и 500 т графита. Его стандартная мощность была 10 кВт.

Если замедлителя в активной зоне нет, то основная часть ядерных делений вызывается быстрыми нейтронами с энергией $\mathscr{E} \gtrsim 1$ МэВ. Возможны также ядерные реакторы на промежуточных нейтронах с энергиями 1–1000 кэВ.

5. Время жизни τ одного поколения мгновенных нейтронов сильно зависит от среды, заполняющей активную зону реактора, и меняется в пределах примерно от 10^{-3} до 10^{-8} с. Малость этого времени

вынуждает при работе с большой точностью поддерживать равенство $k = 1$. Допустим, например, что $k = 1{,}01$, и возьмем наиболее благоприятный случай, когда $\tau = 10^{-3}$ с. Тогда за одну секунду произойдет смена $1/10^{-3} = 1000$ поколений нейтронов, а их число увеличится в $1{,}01^{1000} \approx 20\,000$ раз. Это вызвало бы колоссальный взрыв реактора.

В действительности управление реактором сильно упрощается *благодаря запаздывающим нейтронам.* Коэффициент размножения нейтронов, о котором до сих пор шла речь, мы теперь обозначим через $k_{\text{м}}$, и будем называть *коэффициентом размножения мгновенных нейтронов.* Этот коэффициент не учитывает дополнительное размножение из-за появления запаздывающих нейтронов. Введем, далее, несущественное для понимания сути дела упрощающее предположение, что имеется только один тип запаздывающих нейтронов, появляющихся через время T после деления. Пусть в начальный момент общее число нейтронов было N_0. Через время $n\tau$, когда запаздывающие нейтроны еще не появлялись, число мгновенных нейтронов возрастет до $N_0 k_{\text{м}}^n$, где n — число поколений нейтронов. В момент $T = n\tau$ появятся запаздывающие нейтроны, возникшие в результате деления ядер в начальный момент. Число таких нейтронов мы обозначим через βN_0, где β — малая дробь ($\beta \approx 0{,}64\,\%$). Общее число всех нейтронов в момент T будет $N = N_0 k_{\text{м}}^n + \beta N_0$. Отношение N/N_0 можно назвать *полным коэффициентом размножения* и обозначить через k. Таким образом,

$$k = k_{\text{м}}^n + \beta. \tag{95.7}$$

При этом предполагается, что смена поколений нейтронов периодически повторяется через промежуток времени T.

Допустим, что $k = 1 + \beta \approx 1{,}0064$. Тогда $k_{\text{м}} = 1$. Это значит, что процесс размножения на мгновенных нейтронах стационарен. Размножение нейтронов будет происходить только из-за появления *запаздывающих нейтронов*, т. е. периодически повторяться через промежуток времени T, причем каждый раз число нейтронов будет возрастать в $k \approx 1{,}0064$ раза. Число нейтронов через время $10T$ увеличится в $k^{10} \approx 1{,}066$ раза. Средний период полураспада осколков деления, порождающих запаздывающие нейтроны, как легко вычислить, используя табл. 15, равен примерно 9 с. Поэтому среднее время жизни одного поколения нейтронов будет $T \approx 9/\ln 2 \approx 13$ с. Таким образом, количество нейтронов через 13 с увеличится только в 1,0064 раза, а через 130 с — в 1,066 раз. Еще медленнее увеличение количества нейтронов будет происходить, когда $1 < k < 1{,}0064$. Столь медленное размножение нейтронов существенно упрощает процесс управления работой реактора. Катастрофическое увеличение числа нейтронов наступит лишь тогда, когда k заметно начнет превышать 1,0064, так как в этом случае размножение будет идти уже на мгновенных нейтронах. (Конечно, в случае реальных реакторов на значение $k = 1{,}0064$ надо смотреть как на грубо ориентировочное, так как оно получено для сильно упрощенной модели, предполагавшей, что имеется только один вид запаздывающих нейтронов, вызывающих деление ядер через строго постоянный промежуток времени T.

Регулирование цепной реакции в ядерном реакторе на тепловых нейтронах обычно осуществляется дистанционно с пульта управления путем передвижения в активной зоне вертикальных регулирующих стержней. Такие стержни изготовляются из кадмия, карбида бора и других веществ, сильно поглощающих нейтроны. Поглощение нейтронов происходит в основном на изотопах кадмия ^{113}Cd и бора ^{10}B. Сечения поглощения тепловых нейтронов на них равны соответственно $2{\cdot}10^4$ и $4{\cdot}10^3$ барн. При увеличении глубины погружения регулирующих стержней в активную зону число поглощаемых нейтронов увеличивается, вследствие чего цепная реакция ослабевает и даже может совсем затухнуть. Наоборот, при выдвижении регулирующих стержней из активной зоны поглощение нейтронов уменьшается, а цепная реакция усиливается. Таким путем можно менять мощность реактора и поддерживать ее на требуемом уровне.

Быстрые нейтроны сравнительно слабо поглощаются в регулирующих стержнях. Поэтому для регулирования реакторов малых размеров на быстрых нейтронах применяют приближение и удаление отражателя от активной зоны реактора.

6. В результате деления ядер в активной зоне выделяется энергия в виде тепла. Количество тепловой энергии, выделяющейся в единицу времени, называется *мощностью реактора.* Так как при каждом акте деления выделяется энергия около 200 МэВ $= 3{,}2 \cdot 10^{-11}$ Дж, то мощности 1 МВт соответствует $10^6 : (3{,}2{\cdot}10^{-11}) \approx 3{\cdot}10^{16}$ актов деления в секунду. Отвод тепла из активной зоны осуществляется *теплоносителем*, который в энергетическом реакторе должен с наименьшими потерями передать его в установку, вырабатывающую электроэнергию. При не чрезмерно больших тепловых потоках в реакторах на тепловых нейтронах в качестве теплоносителя используются вода, водяной пар, воздух, азот, углекислый газ и т. д. В мощных реакторах, где активная зона прогревается до температуры 300 °С, использование воды затрудняется ее закипанием. Для избежания закипания приходится сильно повышать давление в системе теплоотвода. В реакторах на быстрых нейтронах, где энерговыделение громадно (около 0,5 кВт на 1 см^3), как правило, теплоносителем служит жидкий натрий, обладающий хорошими теплоотводящими свойствами.

7. При работе реактора мощностью 1 МВт из активной зоны исходит мощный поток нейтронов (свыше 10^{13} нейтронов на 1 см^2 в секунду, а в импульсных реакторах с длительностью импульса 0,1 с, когда мгновенная мощность реактора составляет 10^5 МВт, плотность нейтронного потока достигает 10^{18} нейтронов на 1 см^2 в секунду). Излучается также поток γ-квантов примерно такой же мощности, возникающий при делении и β-распадах радиоактивных ядер. Это примерно в 100 миллиардов раз превышает допустимую плотность потока. Защита должна сильно ослабить оба потока, чтобы они не превышали приемлемую плотность потока. Наилучшими материалами для защиты от γ-квантов являются материалы с большими атомными номерами Z. Для защиты от нейтронов, помимо хороших поглотителей, нужны также эффективные замедлители, так как проникающая способность особенно велика для

быстрых нейтронов. В качестве замедлителей используются легкие элементы, а в качестве поглотителей — бор и тяжелые элементы (кадмий, железо, свинец и др.). Должна быть предусмотрена защита и от довольно жестких вторичных γ-квантов, возникающих при радиационных захватах (n, γ) ядрами защиты. Хорошей и дешевой защитой как от нейтронов, так и от вторичных γ-квантов является бетон с железным заполнителем и соединениями бора.

8. На рис. 167 схематически представлен разрез части активной зоны гетерогенного ядерного реактора (*1* — замедлитель, *2* — отражатель, *3* — ТВЭЛ, *4* — канал охлаждения, *5* — регулирующие стержни, *6* — защита).

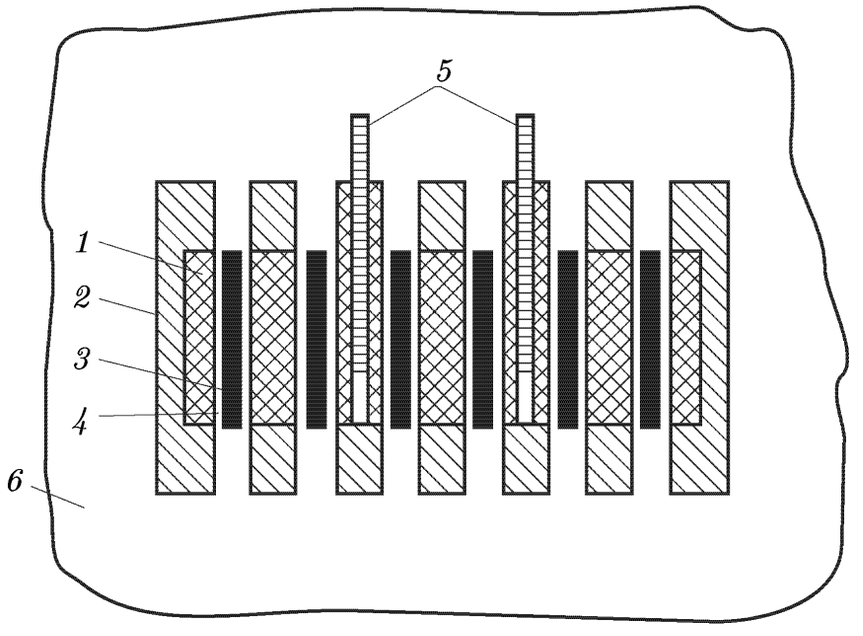


Рис. 167

В качестве примера приведем некоторые данные об активной зоне гетерогенного ядерного реактора первой в мире атомной электростанции (АЭС), вступившей в строй в Обнинске в 1954 г. Активная зона реактора имеет форму цилиндра размерами $1{,}5 \times 1{,}7$ м. В графитовый замедлитель введено 128 ТВЭЛов, содержащих около 550 кг обогащенного до 5% урана. Теплоносителем служит вода. Мощность электростанции составляет 5 МВт. Это — небольшая мощность; в мощных электростанциях количество ядерного топлива больше в сотни раз.

9. При работе ядерного реактора происходит накопление продуктов деления и образование трансурановых элементов, главным образом плутония. Накопление радиоактивных продуктов называется *отравлением реактора*, а накопление стабильных продуктов — *зашлаковыванием*. Отравление вызывается преимущественно изотопом ксенона ^{135}Xe, обладающим наибольшим сечением поглощения нейтронов ($2{,}6 \cdot 10^6$ барн). Основная часть изотопа ^{135}Xe образуется в результате β-распада изотопа иода ^{135}I ($T_{1/2} = 6{,}8$ ч). Изотоп ^{135}Xe в свою очередь претерпевает β-распад с периодом $T_{1/2} = 9{,}2$ ч и превращается в практически стабильный изотоп цезия ^{135}Cs. Отравление и зашлаковывание

приводят к уменьшению коэффициента размножения k и, следовательно, к уменьшению тепловой мощности реактора. Для увеличения срока действия ТВЭЛов, загружаемых в активную зону, регулирующие стержни сначала погружают глубоко. Затем по мере выгорания ядерного топлива их постепенно выдвигают. Замена ТВЭЛов, являющихся сильными излучателями, производится дистанционно.

10. По назначению и мощности различают следующие виды ядерных реакторов.

1) *Экспериментальные реакторы.* Их мощность не превышает нескольких киловатт. Эти реакторы предназначены для изучения и измерения различных физических величин (ν, η, ε и т. д.), знание которых необходимо для проектирования и эксплуатации других ядерных реакторов.

2) *Исследовательские реакторы.* Их мощность не превышает 100 МВт, а выделяющаяся энергия, как правило, не используется. Потоки нейтронов и γ-квантов, исходящие из активной зоны, используются для исследований в ядерной физике, физике твердого тела, радиационной химии, биологии, для испытания материалов в нейтронных потоках и т. д. Импульсные реакторы также относятся к исследовательским реакторам.

3) Реакторы, предназначенные для получения *новых делящихся трансуранов и радиоактивных изотопов.*

4) *Энергетические реакторы.* Они используются для выработки электроэнергии, теплофикации, в силовых установках на кораблях, для опреснения морской воды и т. д. Тепловая мощность современных энергетических реакторов достигает 3–5 ГВт.

Ядерные реакторы различаются также по виду ядерного топлива (естественный и обогащенный уран), по его химическому составу (металлический U, UO_2, UC и т. д.), по виду замедлителя (C, H_2O, D_2O, Be, BeO, гидриды металлов, отсутствие замедлителя), по виду теплоносителя (H_2O, газ, D_2O, органические жидкости, расплавленные металлы). Наиболее распространены гетерогенные ядерные реакторы на тепловых нейтронах с замедлителями из C, H_2O, D_2O и теплоносителями из H_2O, газа, D_2O.

11. Важная роль в ядерной энергетике в настоящее время отводится *реакторам-размножителям* (иначе называемым *бридерами*). В будущем роль таких реакторов должна становиться все большей и большей и в конце концов сделаться определяющей. Реактором-размножителем называется такой реактор, в котором происходит не только выработка энергии, но и *расширенное воспроизводство* делящегося материала. Циклы воспроизводства основаны на ядерных реакциях (94.2) и (95.1). В первой реакции не делящееся медленными нейтронами ядро ^{238}U превращается в ядро ^{239}Pu. Во второй реакции ядро ^{232}Th, также не делящееся медленными нейтронами, превращается в ядро ^{233}U. Ядра же ^{239}Pu и ^{233}U в отношении деления не хуже, а даже лучше ядра ^{235}U. Существенно заметить, что выделение образовавшегося ^{239}Pu из ^{238}U, а также образовавшегося ^{233}U из ^{232}Th производится химически, а это

несравненно легче и быстрее, чем в трудоемких и медленно идущих процессах разделения изотопов.

Коэффициентом воспроизводства $K_в$ реактора-размножителя называется отношение скорости образования делящихся ядер к скорости уничтожения исходных делящихся ядер. Для получения $K_в > 1$ необходимо, чтобы на одно поглощение нейтрона исходным делящимся ядром приходилось больше двух ($\eta > 2$) рождающихся нейтронов, так как один из этих нейтронов должен расходоваться на продолжение цепной реакции деления. Фактически из-за поглощения нейтронов в конструкционных материалах и продуктах деления необходимо $\eta > 2{,}2$–$2{,}3$. На изотопах ^{235}U и ^{239}Pu расширенное воспроизводство ядерного горючего практически возможно только в реакторах на быстрых нейтронах, так как для медленных нейтронов величина η слишком мало отличается от 2 ($\eta = 2{,}07$ для ^{235}U и $\eta = 2{,}11$ для ^{239}Pu). На быстрых же нейтронах η заметно превышает 2 (2,3 и 2,7 соответственно). У урана ^{233}U коэффициент η достаточно велик (2,28) уже для тепловых нейтронов и мало повышается (до 2,45) при переходе к быстрым нейтронам. Поэтому расширенное воспроизводство на ^{233}U можно было бы вести и в тепловых реакторах.

В реакторах-размножителях пока применяется обогащенный уран с высоким содержанием ^{235}U (до 30 %), но в будущем, по мере накопления запасов ^{239}Pu, ядерным топливом будет служить смесь естественного урана с плутонием. Торий ^{232}Th как сырьевой материал для получения делящихся ядер ^{233}U пока не нашел применения. Этот элемент не образует богатых месторождений, и технология его извлечения из руд сложнее технологии извлечения урана. Но в перспективе торий должен будет использован в ядерной энергетике широко, так как его запасы на Земле в десятки раз больше, чем урана.

В реакторах-размножителях активная зона окружена слоем воспроизводящего вещества, называемым *зоной воспроизводства.* Через реактор в качестве теплоносителя прокачивается жидкий натрий, который практически не замедляет нейтроны, но хорошо отводит тепло.

Ядерные реакторы на тепловых нейтронах могут «сжечь» 0,5–1 % урана. Использование же реакторов-размножителей увеличивает это число в десятки раз.

12. В заключение остановимся на мотивах, которые диктуют необходимость развития ядерной энергетики. В начале нашего века потребление энергии в мире удваивалось приблизительно за 50 лет. В середине века это происходило уже за 30 лет, а сейчас — за 10–15 лет. В настоящее время около 70 % энергопотребления в мире обеспечивается нефтью и газом. Однако мировые запасы нефти и газа ограничены. С учетом роста энергопотребления они могут быть исчерпаны в течение 20, максимум — 50 лет. Кроме того, нефть и газ необходимо было бы сохранить на более длительное время как сырье для получения ряда химических продуктов, пластмасс, для производства белка за счет микробиологического синтеза и т. д. Каковы же пути преодоления надвигающегося тотального энергетического кризиса?

Необходимо перестроить энергетику на новой основе. Прежде всего надо шире использовать уголь, так как запасы угля на Земле значительно превосходят запасы других природных горючих. В дальнейшем должна быть широко использована ядерная энергия. Сравнительно дешевого урана, пригодного для энергетических целей, на Земле по оценкам не так уж много — около 4 млн тонн. Сейчас на атомных электростанциях потребляется почти исключительно уран-235 и лишь около 1 % урана-238 включается в топливный цикл, который реализуется в существующих реакторах. Этих ресурсов хватит примерно на такое же время, что нефти и газа. Использование реакторов-размножителей меняет дело. В таких реакторах каждый килограмм природного урана может отдавать энергии в 30–40 раз больше, чем в обычных ядерных реакторах на тепловых нейтронах. Поэтому можно использовать не только дешевый, но и дорогой уран, а также торий. В результате энергетическими ресурсами станут десятки миллионов тонн урана, которые находятся, например, в океанской воде, в бедных урановых рудах, горных породах, а также сотни миллиардов тонн тория. А таких ядерных ресурсов хватит на тысячи лет. Еще более кардинальное решение энергетического кризиса дало бы использование термоядерной энергии, но эта проблема пока еще не решена. О ней говорится в § 98.

Существенным недостатком атомных электростанций является накопление радиоактивных отходов, которые необходимо весьма надежно и длительно хранить, чтобы предотвратить опасное загрязнение внешней среды. Это трудная инженерная проблема, которая, однако, может быть надежно решена при любых масштабах производства энергии. Более того, исследования показали, что вред, наносимый окружающей среде электростанциями на угольном топливе, существенно больше, чем вред от атомных электростанций.

§ 96. Природный ядерный реактор в Окло

1. В 1972 г. во время масс-спектрографического анализа на заводе Пьерлатте (Франция), где проводилось обогащение ядерного топлива из гексафторида урана (UF_6), было обнаружено, что в этом материале из общего числа атомов урана содержится 0,717 % атомов урана-235, тогда как в земных породах, метеоритах и образцах лунного грунта содержание урана-235 больше, а именно 0,72 %. Урановая руда для переработки доставлялась из рудника Окло, расположенного в Габоне вблизи города Франсфиля (Африка). Как выяснилось позднее, уже в более раннее время (в 1970–1972 гг.) на переработку поступала руда, содержащая временами до 20 % урана, обедненного до 0,64 % изотопом ^{235}U. Так как руда при добыче перемешивалась, то обеднение ураном-235 в отдельных образцах могло быть еще сильнее. Всего обедненного урана, использованного в цепной реакции деления, было добыто более 700 т, причем дефицит урана-235 (не замеченный первоначально) составил примерно 200 кг. Изотопный анализ урановых образцов, значительно обедненных ураном-235, обнаружил также заметные отклонения от

природного распределения именно тех редкоземельных изотопов, которые являются продуктами деления ядер урана. Каковы же возможные причины этих отклонений?

Урановое месторождение Окло, исследованное по методу содержания свинца в урановой руде, сформировалось примерно $1{,}8 \cdot 10^9$ лет тому назад. Как полагают геологи, в дельте древней реки образовался осадочный слой богатого ураном песчаника толщиной 4–10 м и шириной 600–900 м. Под ним находилась базальтовая порода, которая под действием тектонических процессов опустилась на глубину в несколько километров. При таком опускании урановая жила растрескалась, и в нее проникли грунтовые воды. Десятки миллионов лет тому назад месторождение поднялось вверх до современного уровня. В руде со средней весовой концентрацией урана 0,5 % были обнаружены 6 глинистых линз, т. е. образований размером от 10 до 20 м и толщиной порядка метра, в которых концентрация урана доходит до 20–40 % и больше. Образование линз со сверхвысоким содержанием урана произошло, по-видимому, под действием фильтрационных вод, хотя детальная картина этого процесса и не ясна.

2. Урановый рудник в Окло в какой-то мере напоминает гетерогенный ядерный реактор. Роль ТВЭЛов выполняют линзы, содержащие уран, роль замедлителя — грунтовая вода между ними. Цепная реакция в таком руднике была бы невозможна, если бы изотопный состав урана в нем был бы таким же, как и изотопный состав современного естественного урана. И действительно, в настоящее время цепная реакция в руднике не идет. Однако уран-238 распадается медленнее, чем уран-235. Их периоды полураспада равны соответственно $4{,}5 \cdot 10^9$ и $7{,}1 \cdot 10^8$ лет. Исходя из этих данных, нетрудно подсчитать, что $1{,}8 \cdot 10^9$ лет тому назад содержание атомов изотопа ^{235}U должно было составлять примерно 3,1 % от общего количества атомов урана. А для такого обогащенного урана в присутствии воды (в качестве замедлителя) коэффициент размножения k_∞ мог достигнуть критического значения $k_\infty = 1$ и даже несколько превзойти его. При таких условиях в далекие времена могла возникнуть самопроизвольная цепная ядерная реакция, что, по-видимому, и произошло на самом деле.

Продолжительность работы естественного ядерного реактора в Окло была оценена по количеству образовавшегося плутония ^{239}Pu. Таким путем было найдено, что она составляет около 0,6–0,8 млн лет. Эта оценка не очень надежна хотя бы потому, что доля делений ^{239}Pu относительно полного числа делений составляет всего лишь несколько процентов и сравнима с долей делений ^{238}U на быстрых нейтронах.

Общее количество энергии, выработанной реактором в Окло, оценивается приблизительно в $1{,}5 \cdot 10^4$ МВт · лет. Такое же количество энергии вырабатывают два блока Ленинградской АЭС за 2,3 года при полной нагрузке. Если принять среднюю продолжительность работы реактора в Окло равной 0,6 млн лет, то средняя мощность выделяющейся в нем энергии составляет всего лишь 25 кВт. Это соответствует среднему потоку тепловых нейтронов в реакторе не выше 10^8 с$^{-1}\cdot$см$^{-2}$.

Не следует думать, что цепная реакция, шедшая в Окло, носила бурный характер — с гулом, шипением пара и пр. На самом деле реакция была медленной — она скорее «тлела». Температура в руднике никогда не превышала 300–600 °C, так как природный реактор был саморегулирующейся системой. При повышении температуры из зоны реакции удалялась вода, а это вело к уменьшению количества медленных нейтронов, к затуханию и даже полному угасанию реакции. Затем рудник начинал остывать и в него снова набиралась вода, замедляющая нейтроны. Это создавало благоприятные условия для возобновления цепной реакции, а следовательно, и для повышения температуры. Вода снова удалялась из реактора, и реакция затухала. Затем опять начиналось накопление воды, увеличение количества тепловых нейтронов, возобновление цепной реакции и т. д. Таким образом, периодически происходила смена возбуждения и затухания реакции.

3. Приведенные факты дают достаточно убедительное доказательство существования в отдаленные времена природного ядерного реактора. Однако ясны далеко не все детали, относящиеся к этому вопросу. Например, среди продуктов деления в природном ядерном реакторе был обнаружен хотя и небольшой, но несомненно существующий избыток изотопов ксенона — ^{131}Xe, ^{132}Xe и ^{134}Xe — по сравнению с изотопным составом ксенона, который возникает при делении ^{235}U тепловыми нейтронами. Причина такого расхождения еще не выяснена.

§ 97. Использование антинейтрино для контроля ядерного реактора

1. В опытах Рейнеса и Коуэна (см. § 74) было найдено, что сечение реакции обратного β-распада (74.12) для антинейтрино, исходящих из ядерного реактора (т. е. антинейтрино низких энергий), составляет примерно $\sigma = 10^{-43}$ см^2. Чтобы более наглядно представить этот результат, допустим, что плоскопараллельный пучок антинейтрино распространяется в железе, причем поглощение антинейтрино происходит исключительно за счет реакции (74.12). В одном моле содержится примерно $6 \cdot 10^{23}$ атомов, т. е. $26 \cdot 6 \cdot 10^{23}$ протонов (26 — атомный номер железа). Так как атомная масса железа 56, а плотность 7,8 г/см^3, то в 1 см^3 железа будет $n = (7{,}8/56) \cdot 26 \cdot 6 \cdot 10^{23} = 2{,}2 \cdot 10^{24}$ протонов. Если среда (железо), в которой распространяются антинейтрино, однородна, то формулу (65.3) можно представить в виде

$$\frac{dN}{N} = -n\sigma l,$$

где l — путь, проходимый антинейтрино в железе. Определим путь l, на котором из пучка выбывает миллионная доля всех антинейтрино. Для этого в предыдущей формуле положим $|dN/N| = 10^{-6}$ и найдем

$$l = \frac{1}{n\sigma}\left|\frac{dN}{N}\right| = 45 \cdot 10^{11} \text{ см} = 45 \cdot 10^6 \text{ км}.$$

Таким образом, при прохождении пучка из миллиона антинейтрино через слой железа толщиной в 45 млн км (эта величина примерно только в три раза меньше расстояния от Земли до Солнца) в среднем только одно антинейтрино поглотится и выйдет из пучка. Если же толщину слоя железа довести до $45 \cdot 10^{12}$ км (приблизительно 5 световых лет), то на этой толщине интенсивность пучка антинейтрино уменьшится всего в e раз. Эти факты, кажется, не оставляют возможности для практических применений нейтрино и антинейтрино.

И все же такая возможность существует, так как ничтожное поглощение антинейтрино в веществе может быть компенсировано исключительно большой мощностью пучков антинейтрино, исходящих из крупных ядерных реакторов. Допустим в качестве примера, что детектор антинейтрино имеет форму куба со стороной 1 м и что вещество этого детектора поглощает антинейтрино так же, как и железо (это делается только для того, чтобы не повторять вычисления, аналогичные приведенным выше). Пусть в детектор попадает 10^{21} антинейтрино в сутки (это может обеспечить современный большой ядерный реактор). Толщина поглотителя по сравнению с предыдущим случаем уменьшена в $45 \cdot 10^9$ раз, а число нейтрино увеличено в $10^{21} : 10^6 = 10^{15}$ раз. Поэтому число актов обратного β-распада, зафиксированных детектором в сутки, будет

$$1 \cdot \frac{10^{15}}{45 \cdot 10^9} \approx 2 \cdot 10^4.$$

Приведенная оценка показывает, что поток антинейтрино можно регистрировать и гораздо меньшим детектором.

2. В СССР на Ровенской АЭС с 1982 г. функционирует специализированная нейтринная лаборатория. Основным помещением лаборатории является экспериментальный зал площадью $6{,}5 \times 6{,}5$ м и высотой 3,5 м. Этот зал расположен прямо под ядерным реактором на расстоянии 19,2 м от центра активной зоны. От потоков частиц из реактора зал огражден мощной защитой из нескольких слоев специального так называемого тяжелого бетона, стали, воды. Такая защита по своим поглощающим свойствам эквивалентна слою воды толщиной около 30 м. Детектором служит резервуар объемом 235,5 л, заполненный органическим сцинтиллятором — декалином ($C_{10}H_8$), в состав которого входит большое количество водорода и который содержит небольшую примесь гадолиния, поглощающего нейтроны.

В осколках деления ядерной зоны реактора в процессе β-распада нейтроны превращаются в протоны с испусканием электронов и антинейтрино ($n \to p + e^- + \overline{\nu}$). Антинейтрино и должны регистрироваться. Принцип регистрации по существу не отличается от того, какой применялся в опытах Рейнеса и Коуэна (см. § 74). Антинейтрино, взаимодействуя с протоном в обратной реакции β-распада ($\overline{\nu} + p \to n + e^+$), порождает в сцинтилляторе нейтрон и позитрон. Позитрон сразу же аннигилирует с электроном, в результате чего возникает сцинтилляционная вспышка. Нейтрон же через время порядка 10^{-8} с захватывается ядром гадолиния с испусканием γ-кванта, что дает вторую вспышку.

Световые вспышки улавливаются фотоумножителями (всего их 24), сигналы с которых поступают на электронную схему. Если реакция действительно вызывается антинейтрино, то в сцинтилляторе детектора должны появиться одна за другой две вспышки через определенное короткое время.

Всего за сутки детектор регистрирует до тысячи таких пар вспышек. Трудность опыта состоит в том, что за то же время на детектор обрушивается громадное число нейтронов и γ-квантов и почти каждый из них вызывает в нем соответствующий сигнал. Задача заключается в том, чтобы выделить относительно слабые сигналы от нейтрино на громадном фоне этих не имеющих отношения к делу сигналов. Задача эта успешно решена. Не останавливаясь на этом вопросе, укажем только, что это делается с помощью электрической схемы совпадений, позволяющей регистрировать только нужные пары следующих друг за другом сцинтилляционных вспышек. Таким путем достигается, что полезный сигнал не только не теряется на фоне посторонних событий, но даже значительно превышает его.

3. Как уже говорилось в § 74, реакция обратного β-распада ($\overline{\nu} + \mathrm{p} \to \mathrm{n} + \mathrm{e}^+$) эндотермична. Действительно, сумма масс нейтрона и позитрона (939,5731 + 0,5110034 МэВ) превышает массу протона (938,2796 МэВ) на 1,8045 МэВ. Недостаток массы у протона должен быть компенсирован за счет энергии антинейтрино $\overline{\nu}$. Реакция может идти только тогда, когда энергия антинейтрино превышает 1,8 МэВ. В нейтринном спектрометре как раз и измеряется энергия позитрона, равная энергии антинейтрино за вычетом 1,8 МэВ. Это позволяет восстановить энергетический спектр антинейтринного излучения, падающего из активной зоны реактора на прибор.

Количество вылетающих антинейтрино пропорционально числу разделившихся ядер и поэтому позволяет точно судить о выделении энергии в реакторе. Кроме того, антинейтринное излучение осколков деления урана и плутония существенно различно по своему спектральному составу. Это открывает возможность по спектру нейтринного излучения контролировать накопление плутония непосредственно в процессе его образования. В начале работы реактора его антинейтринный спектр полностью определяется осколками деления изотопа $^{235}\mathrm{U}$. В дальнейшем же, по мере накопления плутония, пропорционально растет и его вклад в спектр антинейтринного излучения.

В качестве иллюстрации приведем некоторые результаты измерений, произведенных на Ровенской АЭС. В течение примерно недели непрерывно регистрировался антинейтринный поток из реактора. Одновременно тепловым методом определялась средняя мощность реактора. За это время всего было выработано энергии 9,79 ГВт · сут по измерениям нейтринным способом, тогда как тепловой метод дал близкое значение 9,64 ГВт · сут (израсходованное топливо соответственно равно 9,41 и 9,27 кг). Принципиальное достоинство нейтринного метода заключается в том, что он не требует непосредственного контакта с делящимся веществом и может производиться дистанционно. Весьма

вероятно, что нейтринный метод окажется в будущем основным источником информации о цепных реакциях, идущих в активной зоне ядерного реактора.

§ 98. Термоядерная проблема

1. Энергия связи ядра, приходящаяся на один нуклон, если отвлечься от ее нерегулярных колебаний, с возрастанием атомного номера сначала систематически возрастает, достигает максимума вблизи ядра железа, а затем начинает систематически убывать (см. рис. 120). Иными словами, слияние (или *синтез*) легких ядер и деление тяжелых приводят, как правило, к более прочной связи между нуклонами. Отсюда следует, что при делении тяжелых ядер и при синтезе легких должна освобождаться энергия. О первом явлении и его использовании в ядерной энергетике уже говорилось в § 94, 95. При делении ядра $^{235}\mathrm{U}$ освобождается энергия, составляющая около 0,85 МэВ на нуклон.

Из реакций синтеза легких ядер особый интерес представляют две реакции:

$$\mathrm{d} + \mathrm{d} \begin{cases} \nearrow {}^3\mathrm{He} + \mathrm{n} + 3{,}27 \text{ МэВ}, \\ \searrow \mathrm{t} + \mathrm{p} + 4{,}03 \text{ МэВ}, \end{cases} \tag{98.1}$$

$$\mathrm{d} + \mathrm{t} \to {}^4\mathrm{He} + \mathrm{n} + 17{,}59 \text{ МэВ} \tag{98.2}$$

(см. § 64, задача 2). Существует и много других экзотермических ядерных реакций синтеза, например

$${}^3\mathrm{He} + \mathrm{d} \to {}^4\mathrm{He} + \mathrm{p} + 18{,}3 \text{ МэВ}, \tag{98.3}$$

$${}^6\mathrm{Li} + \mathrm{p} \to {}^4\mathrm{He} + {}^3\mathrm{He} + 4 \text{ МэВ}, \tag{98.4}$$

которые в перспективе могут быть использованы для решения обсуждаемой здесь термоядерной проблемы. Однако в настоящее время основное внимание уделяется реакции (98.2), а затем реакции (98.1), так как в них участвуют самые легкие ядра, благодаря чему облегчается получение и использование выделяющейся энергии.

На один нуклон в верхней ветви реакции dd (98.1) выделяется энергия (3,27/4) МэВ, в нижней — (4,3/4) МэВ. Обе ветви реакции dd идут практически с одинаковой вероятностью. Поэтому в среднем в реакции dd на один нуклон выделяется энергия $(1/2)(3{,}27/4 + 4{,}03/2) \approx 0{,}9$ МэВ, т. е. примерно такая же, что и при одном акте деления тяжелого ядра. В реакции dt (98.2) выделяющаяся энергия на один нуклон составляет $17{,}59/5 \approx 3{,}52$ МэВ. Однако в этой реакции только 3,52 МэВ приходятся на заряженные частицы $^4\mathrm{He}$, а 14,07 МэВ уносятся нейтронами. Энергия же заряженных частиц может быть преобразована в электрическую со значительно большим КПД, чем энергия нейтронов (см. п. 5).

В реакции (98.3) на один нуклон освобождается энергия 3,66 МэВ, а в реакции (98.4) — 0,57 МэВ. Продуктами этих реакций являются только нерадиоактивные заряженные частицы, что в энергетическом

отношении должно рассматриваться как их достоинство (см. п. 5). Однако эти реакции труднее осуществить, а потому в дальнейшем они не будут рассматриваться.

Таким образом, как и при делении тяжелых ядер, в реакциях синтеза легких освобождается энергия, в миллионы раз превосходящая тепло, получающееся при сжигании химического топлива (уголь, нефть и пр.). Однако получение этой энергии в макроскопических количествах, к величайшему сожалению, удалось пока только для военных целей — в водородной бомбе, где реакции (98.1) и (98.2) осуществляются с огромной скоростью и сопровождаются чудовищным взрывом. По этой причине они совершенно неуправляемы. Для использования энергии этих реакций в мирных целях необходимо придать им спокойный управляемый характер. Соответствующая проблема называется проблемой *управляемого термоядерного синтеза* (УТС). Она усиленно разрабатывается в СССР, США и других странах, начиная примерно с 1951 г. В СССР работы по УТС в основном ведутся в Институте атомной энергии им. И. В. Курчатова. Они были начаты под руководством Л. А. Арцимовича (1909–1973) и М. А. Леонтовича (1903–1981).

Управляемый термоядерный синтез еще не осуществлен, хотя ученые и надеются получить его к концу настоящего или в начале следующего столетия. Детальное рассмотрение этой проблемы далеко выходит за рамки этой книги. Мы можем дать о ней только общее и поверхностное представление.

2. Обе реакции (93.1) и (93.2) происходят с положительно заряженными ядрами, между которыми действуют кулоновские силы отталкивания. Для возможности проведения обеих реакций взаимодействующим ядрам должна быть сообщена кинетическая энергия, достаточная для сближения их на расстояние порядка 10^{-11} см. Только после этого с заметной вероятностью начнется процесс слияния ядер за счет *туннельного эффекта.* С увеличением зарядового числа Z кулоновский барьер между сближающимися ядрами возрастает пропорционально произведению зарядов обоих ядер. Осуществить слияние ядер тем легче, чем меньше Z. Вот почему выбираются ядра дейтерия и трития, для которых число $Z = 1$, т. е. минимально.

Можно осуществить реакции слияния легких ядер, ускоряя на ускорителе ядра одного изотопа и бомбардируя ими мишени из изотопа того же или другого элемента. Однако такой метод применим и действительно применяется для изучения ядерных реакций, а также для получения нейтронов. Но он совершенно не годится для получения ядерной энергии в промышленных масштабах, если даже воспользоваться весьма сильноточным ускорителем. Дело в том, что эффективные сечения ядерных реакций, хотя и быстро возрастают с увеличением относительной скорости сталкивающихся ядер, но даже в оптимальных условиях несравненно меньше эффективных сечений атомных столкновений. При столкновениях кинетическая энергия ускоренных ядер быстро растрачивается на ионизацию и возбуждение атомов мишени. Оставшейся энергии ускоренного ядра в подавляющем большинстве случаев

совершенно недостаточно для его слияния с ядром мишени. В результате почти все столкновения ядер не будут завершаться ядерными реакциями. Полученная таким путем ядерная энергия будет ничтожна по сравнению с энергией, которую надо затратить для осуществления ускорения ядер с помощью ускорителей.

Поэтому процессы ионизации и возбуждения атомов мишени не должны происходить. А это имеет место только в уже *полностью ионизованной плазме*, нагретой до весьма высокой температуры — порядка 10 кэВ (10^8 К). Вот почему проблема получения энергии в промышленных масштабах в управляемых реакциях синтеза ядер легких элементов получила название *термоядерной проблемы.* В своей основе она прежде всего опирается на физику плазмы — на получение, изучение и управление этим состоянием вещества. Физика же плазмы ныне составляет самостоятельный и весьма обширный отдел физики. Его изложение не входит в цели настоящего руководства. По этой причине мы можем дать только общее представление о термоядерной проблеме.

3. Дейтерий ^{2}H (иначе обозначаемый через D) — стабильный изотоп водорода. Он встречается в воде в виде молекул D_2O и HDO. В естественной воде по числу атомов содержится примерно 0,015 % дейтерия. Значит, в 18 г воды содержится $N_A \cdot 0{,}00015 \sim 9 \cdot 10^{18}$, а в 1 г около $5 \cdot 10^{17}$ атомов дейтерия, содержащих около 10^{18} нуклонов. Так как на один нуклон в реакции dd освобождается около 0,9 МэВ энергии, то за счет этой реакции в 1 г воды может выделиться энергия $10^{18} \cdot 0{,}9 \approx 9 \cdot 10^{17}$ МэВ $\approx 1{,}5 \cdot 10^5$ Дж. По сравнению с энергией химического топлива это громадная величина. Например, удельная теплота сгорания каменного угля составляет около $3{,}5 \cdot 10^7$ Дж/кг (не учитывая кислорода). Таким образом, 250 г воды в энергетическом отношении эквивалентны примерно 1 кг каменного угля. Масса воды в океанах Земли составляет примерно $1{,}45 \cdot 10^{24}$ кг. По содержанию энергии это эквивалентно $6 \cdot 10^{21}$ кг $= 6 \cdot 10^{18}$ т каменного угля. Это составляет 10^{-3} массы всей Земли (около $6 \cdot 10^{21}$ т). Таким образом, дейтерий, содержащийся в океанах, является *практически неисчерпаемым* источником энергии, которой хватило бы на сотни миллионов лет.

Тритий ^{3}H (иное обозначение — T) — радиоактивный изотоп водорода с периодом полураспада $T_{1/2} = 12{,}3$ года. Тритий образуется в небольших количествах в атмосфере под действием космических лучей (примерно 1000–2000 атомов T на 1 м^2 поверхности Земли в 1 с). Получающийся таким образом тритий соединяется с кислородом воздуха, образуя сверхтяжелую воду, но из-за распада ^{3}H такая вода не накапливается. В природной воде содержится всего около 10^{-18} атомов трития по отношению к числу атомов водорода.

Таким образом, можно сказать, что из-за радиоактивности тритий в естественном состоянии в природе не встречается. Поэтому для использования в термоядерных установках тритий должен создаваться искусственно. Он получается в реакции

$$^6\mathrm{Li} + \mathrm{n} \to {}^3\mathrm{H} + {}^4\mathrm{He} + 4{,}8\ \text{МэВ} \qquad (98.5)$$

путем облучения изотопа лития ^{6}Li нейтронами от ядерного реактора. Поэтому управляемый термоядерный синтез на основе реакции (98.2) должен предусмотреть получение и воспроизводство трития. Воспроизводство трития может быть осуществлено в самом термоядерном реакторе, поскольку реакция dt сопровождается выделением нейтронов. Для этого в соответствии с реакцией (98.5) рабочую зону реактора следует окружить слоем легкого изотопа лития ^{6}Li. Такой слой называется бланкетом. Поэтому количество трития, которым можно располагать в УТС в конце концов определяется запасом дейтерия и ^{6}Li в природе (в естественном литии изотоп ^{6}Li составляет около 7,52 %, а изотоп ^{7}Li^2 около 92,48 %). Запасы этих элементов на Земле настолько велики, что при прогнозируемых темпах потребления их энергии должно хватить на сотни тысяч лет.

Заметим, что в перспективе бланкет может быть использован и для создания *комбинированных ядерно-термоядерных реакторов.* Термоядерный реактор на dd следует окружить бланкетом, содержащим ^{238}U, делящийся быстрыми нейтронами. Быстрые нейтроны, возникающие в реакции dd будут вызывать деление ^{238}U и производить ядерное горючее — плутоний, используемый в ядерном реакторе. Реактор на dt для этой цели не годится, так как получающиеся нейтроны необходимы для воспроизводства трития.

4. При использовании термоядерного топлива, как и при использовании делящихся ядер, не приходится сжигать мировые запасы кислорода или углеводородов, так что в атмосферу не попадают углекислый газ и другие продукты сгорания. В отношении загрязнения окружающей среды (т. е. с экологической точки зрения) термоядерный реактор обладает громадным преимуществом по сравнению с реактором деления. В реакциях ядерного деления образуется много долгоживущих радиоактивных отходов, захоронение которых является непростой задачей. В термоядерном реакторе эта задача не стоит, так как в реакциях синтеза (98.1) и (98.2) долгоживущих радиоактивных отходов не образуется. В этих реакциях основными продуктами будут нейтроны, нерадиоактивные ядра гелия ^{3}He и водорода, а также ядра трития. Правда, тритий радиоактивен — он испускает ионизующую радиацию в виде β-частиц с максимальной энергией 18 кэВ. Но тритий является одним из наименее токсичных радиоактивных изотопов, тогда как плутоний, служащий топливом в ядерных реакторах, — один из наиболее токсичных из известных нам радиоактивных материалов. Кроме того, в реакции dt тритий сам является термоядерным топливом, так что его можно возвратить в активную зону реактора для дальнейшего сжигания. В результате эффективное время жизни трития сводится примерно к 5 дням вместо обычных 18 лет.

5. Сформулируем общее требование, накладываемое на термоядерный реактор законом сохранения энергии, предполагая, что в реакторе идет либо реакция dd, либо реакция dt. Не приводя детального вывода, ограничимся указанием только его общего хода и окончательными результатами.

Пусть в реактор введена плазма из чистого дейтерия или равнокомпонентной смеси дейтерия и трития, уже нагретая в результате какого-либо нетермоядерного процесса. Конечно, плазма предполагается *квазинейтральной*, т. е. такой, в которой средние концентрации положительных ионов и отрицательных электронов в каждом макроскопическом объеме почти одинаковы. Основанием для такого предположения служит замечание, что даже незначительное нарушение квазинейтральности возбудило бы сильные электрические поля, которые, однако, в плазме существовать не могут из-за ее высокой электрической проводимости. В дальнейшем через n обозначается число положительных ионов (и равное ему числу электронов) в кубическом сантиметре плазмы.

Если температура T плазмы достаточно высока, что столкновения ионов друг с другом, по крайней мере частично, будут заканчиваться реакциями с выделением термоядерной энергии. Ясно, что число таких столкновений в единице объема пропорционально n^2, а коэффициент пропорциональности зависит от T. Поэтому мощность ядерной энергии, выделяющейся в единице объема активной зоны реактора, можно представить в виде $P_{\text{яд}} = f_{\text{яд}}(T)n^2$, причем функция $f_{\text{яд}}(T)$ зависит от того, используется ли в качестве топлива чистый дейтерий или равнокомпонентная смесь дейтерия и трития.

Продуктами ядерных реакций (98.1) и (98.2) будут заряженные частицы ^{3}He, t, p, ^{4}He и нейтроны, не имеющие электрического заряда. Кинетическая энергия теплового движения заряженных частиц может быть превращена в электрическую с КПД, близким к 1. Для этого можно использовать *магнитогидродинамический генератор*, в котором происходит прямое преобразование тепловой энергии в электрическую, минуя тепловой цикл. Принцип его работы основан на электромагнитной индукции: если поперек магнитного поля движется проводящая среда, то в ней возбуждается электрический ток. В качестве такой проводящей среды может быть использована высокотемпературная плазма. Ее электрическая проводимость, как показывает теория, пропорциональна $T^{3/2}$ и не зависит от концентрации частиц. При $T \approx 1{,}6 \cdot 10^7$ K ($1{,}4 \cdot 10^3$ эВ) проводимость водородной плазмы примерно равна проводимости меди при нормальной температуре, а при $T \approx 1{,}6 \cdot 10^8$ K она будет примерно в 30 раз больше. Если поток горячей плазмы поперек магнитного поля направить между пластинами конденсатора, то возникающий индукционный ток зарядит конденсатор, т. е. тепловая энергия плазмы перейдет в электрическую.

Нейтроны, поскольку они лишены электрического заряда, таким путем не могут быть использованы. Их кинетическая энергия теплового движения может быть превращена в электрическую с использованием *теплового цикла.* КПД последнего ограничен сверху теоремой Карно. С некоторой долей оптимизма примем, что этот КПД равен $1/3$.

Однако надо учесть не только тепловую энергию, освобождающуюся в ядерных реакциях. Нейтроны уходят из активной зоны реактора и поглощаются защитой. При этом выделяется энергия *нейтронного*

сродства в виде тепла. При каждом акте поглощения его выделяется примерно 8 МэВ. Это тепло может быть также использовано в тепловом цикле. Его надо прибавить к энергии нейтронов в ядерных реакциях. Таким образом, можно принять, что тепловая энергия заряженных частиц может быть преобразована в электрическую практически полностью, а тепловая энергия нейтронов (с учетом энергии нейтронного сродства) — примерно на одну треть. Формально при энергетических расчетах нейтроны можно считать *как бы* заряженными частицами, если число их уменьшить приблизительно в три раза. Предполагается, что это и делается при оценке мощности $P_{\text{яд}}$ и коэффициента $f_{\text{яд}}(T)$.

Рассмотрим сначала реакцию dt (98.2). На долю заряженной частицы ^{3}He приходится энергия 3,5 МэВ, на долю нейтрона — 14 МэВ. С учетом нейтронного сродства ее следует увеличить примерно до $14 + 8 = 22$ МэВ. Таким образом, тепловой эффект реакции dt равен $3{,}5 + 22/3 \approx 10{,}8$ МэВ, или, круглым счетом, 10 МэВ, если все продукты реакции считать как бы заряженными частицами.

Несколько сложнее обстоит дело с реакцией dd (93.1). Как уже указывалось выше, обе ветви этой реакции идут примерно с одинаковой вероятностью. Поэтому все числа, относящиеся к этим реакциям, следует уменьшить вдвое, например 3,27 МэВ заменить на 1,65 МэВ, а 4,03 МэВ — на 2 МэВ. Но главная особенность реакции dd состоит в том, что в одной из ее ветвей выделяется тритий, а он может быть использован в качестве горючего в реакторе, основанном на реакции dt. В первой ветви на долю ^{3}He приходится 0,41 МэВ, на долю нейтрона — 1,24 МэВ или, с учетом нейтронного сродства, 5,24 МэВ, что эквивалентно $5{,}24 : 3 = 1{,}7$ МэВ тепловой энергии заряженных частиц. Во второй ветви вся энергия 2 МэВ локализована в заряженных частицах t и p. К ней надо прибавить $10 : 2 = 5$ МэВ, которые выделяются в генераторе с реакцией dt. Таким образом, в пересчете на заряженные частицы тепловой эффект реакции составляет $0{,}41 + 1{,}7 + 2 + 5 = 9{,}1$ МэВ или, круглым счетом, 10 МэВ, как и в реакции dt.

Помимо выделения энергии будут происходить также ее потери: энергия будет уноситься из активной зоны реактора электромагнитным излучением плазмы и потоком уходящих частиц. Излучение в основном возникает при торможении электронов при их столкновениях с ядрами. Поэтому его мощность из единицы объема также пропорциональна n^2 и может быть представлена в виде $P_{\text{изл}} = f_{\text{изл}}(T)n^2$. Вид функций $f_{\text{яд}}(T)$ и $f_{\text{изл}}(T)$ может быть найден теоретически, но на этом вопросе мы останавливаться не можем.

Для нахождения энергетической мощности уходящего из единицы объема потока частиц введем понятие *среднего времени удержания* τ плазмы в активной зоне реактора. Так называется промежуток времени, в течение которого за единицу времени из единицы объема плазмы уходит n/τ частиц каждого знака. Это единственный феноменологический параметр, характеризующий термоядерный реактор. Поэтому условие, которое мы получим ниже из закона сохранения энергии, совершенно не зависит от индивидуальных особенностей термоядерного реактора, а должно выполняться для реактора любого типа. Через

промежуток времени τ можно выразить энергетическую мощность $P_{\text{ч}}$ потока частиц, уходящих из единицы объема плазмы. Кинетическая энергия частиц в единице объема плазмы равна $2 \cdot (3/2)nkT = 3nkT$ (половина энергии приходится на ядра, а другая половина — на электроны). Поэтому $P_{\text{ч}} = 3nkT/\tau$.

Таким образом, энергетическая мощность частиц и излучения, уходящих из единицы объема плазмы, будет $3nkT/\tau + f_{\text{изл}}(T)n^2$. Эта энергия поглощается защитой и выделяется в виде тепла. Часть такого тепла может быть использована путем превращения в работу с некоторым КПД η, значение которого зависит от уровня развития теплотехники. Мощность бесполезно потерянной энергии в единице объема плазмы будет

$$(1-\eta)\left[\frac{3nkT}{\tau} + f_{\text{изл}}(T)n^2\right],$$

или

$$\frac{2nkT}{\tau} + \frac{2}{3} f_{\text{изл}}(T)n^2,$$

если принять $\eta = 1/3$, как это уже делалось выше.

Условие протекания самоподдерживающейся ядерной реакции синтеза состоит в том, чтобы выделяющаяся термоядерная энергия была не меньше энергии, уносимой из плазмы электромагнитным излучением и потоком уходящих частиц. Математически это условие выражается формулой

$$f_{\text{яд}}(T)n^2 \geqslant \frac{2nkT}{\tau} + \frac{2}{3} f_{\text{изл}}(T)n^2. \tag{98.6}$$

В случае равенства в плазме будет протекать только самоподдерживающая термоядерная реакция синтеза, но не будет генерироваться энергия, которую можно было бы использовать. Для превращения установки в генератор энергии необходимо, чтобы в формуле (98.6) соблюдалось неравенство.

Условие (98.6) может быть записано в виде

$$n\tau \geqslant L(T), \tag{98.7}$$

если ввести обозначение

$$L(T) = \frac{2kT}{f_{\text{яд}}(T) - (2/3) f_{\text{изл}}(T)}. \tag{98.8}$$

Конечно, знаменатель в последней формуле должен быть положителен, так как освобождающаяся ядерная энергия во всяком случае должна превышать энергию, уносимую излучением.

Формула (98.7) выражает *условие возникновения самоподдерживающейся термоядерной реакции*, или *условие зажигания термоядерной реакции*. Она была получена в 1957 г. английским физиком Лоусоном (р. 1923) и называется *критерием Лоусона*. Функция температуры $L(T)$ зависит от вида применяемого термоядерного топлива: чистого дейтерия или равнокомпонентной смеси дейтерия и трития. Ее нахождение сводится к определению функций $f_{\text{яд}}(T)$ и $f_{\text{изл}}(T)$, что может

быть сделано теоретически (об этом уже было сказано выше). Выражение (98.8) получено в предположении, что КПД $\eta = 1/3$. Нетрудно обобщить это выражение и для других значений η; формула (98.7) при этом не меняет своего вида, изменяется только функция $L(T)$.

6. Минимальная температура, при которой начинает выполняться критерий (98.7), может рассматриваться как *оптимальная температура* для осуществления самоподдерживающейся термоядерной реакции. Она может быть найдена из условия $dL/dT = 0$. На рис. 168 приведены графики функций $L(T)$ для реакций dt и dd при $\eta = 1/3$. Из них видно, что для реакции dt оптимальная температура порядка $2 \cdot 10^8$ K, а для реакции dd — порядка 10^9 K. Если значение $n\tau$ таково, что точка $(T, n\tau)$ лежит на кривой $L = L(T)$, то в системе может происходить стационарная термоядерная реакция, но без генерации полезной энергии. Если же такая точка лежит выше указанной кривой, то система начинает работать как генератор энергии.

Далее, из графиков видно, что при $\eta = 1/3$ критерий Лоусона сводится к выполнению условий:

для реакции dt: $n\tau \geqslant 10^{14}$ см$^{-3} \cdot$ с, $T \approx 2 \cdot 10^8$ K;

для реакции dd: $n\tau \geqslant 10^{15}$ см$^{-3} \cdot$ с, $T \approx 10^9$ K.

Отсюда видно, что осуществление термоядерной реакции dt легче, чем реакции dd. Это связано с тем, что сечение реакции dt в области температур 10^8–10^9 K примерно на два порядка превышает сечение реакции dd. Это в свою очередь объясняется тем, что в указанном интервале температур реакция dt носит *резонансный* характер. Вот почему в настоящее время усилия физиков и инженеров направлены на осуществление термоядерной реакции dt. Но в перспективе с развитием термоядерной энергетики должен произойти переход к реакции dd, так как запасы ^{6}Li в природе все же ограничены.

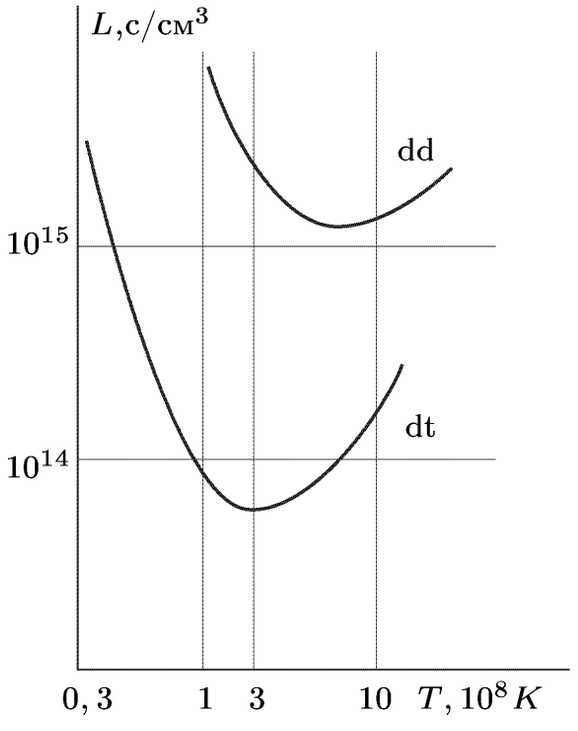


Рис. 168

Таким образом, чтобы начались термоядерные реакции, необходимо нагреть дейтерий-тритиевую плазму до температуры порядка $2 \cdot 10^8$ K, т. е. порядка 10 кэВ. При термоядерных же реакциях освобождается энергия в несколько мегаэлектронвольт на каждый акт реакции. Эта энергия в сотни раз превосходит энергию, которую надо затратить на нагревание самой плазмы. Это обстоятельство и открывает принципиальную возможность получения полезной энергии при термоядерных реакциях.

7. Для нагревания дейтерий-тритиевой плазмы до температур порядка 10^8 K могут быть использованы следующие способы.

1) *Выделение джоулева тепла* при пропускании электрического тока через плазму. Этот способ применим на начальной стадии нагревания и становится неэффективным при $T \geqslant 10^7$ К, так как проводимость плазмы быстро возрастает с температурой (приблизительно пропорционально $T^{3/2}$). Для дальнейшего нагревания плазмы требуются дополнительные способы, из которых наиболее перспективными являются способы 2) и 3), приводимые ниже.

2) *Инжекция в плазму пучков нейтральных частиц достаточно высокой энергии.* Пучки быстрых атомов формируются с помощью инжекторов — ускорителей заряженных частиц, например ядер дейтерия — дейтронов. Ускоренные дейтроны проходят через специальный слой нейтрального газа и превращаются в быстрые нейтральные атомы дейтерия, которые беспрепятственно проникают в плазму под любым углом к магнитному полю, если последнее применяется для удержания плазмы. Уже существуют инжекторы с мощностью пучка 2 МВт при энергии атомов 20–40 кэВ. Использование инжекторов привело к значительному повышению температуры плазмы. Так, на токамаке PLT (США) этим способом удалось достигнуть температуры $7 \cdot 10^7$ К.

3) *Нагревание высокочастотным электромагнитным полем.* Высокочастотное электромагнитное поле в плазме возбуждается петлей с током высокой частоты, помещаемой вблизи рабочей камеры, в которой находится плазма. Если частота поля совпадает с ионной или электронной циклотронными частотами, то происходит резонансное поглощение электромагнитных волн и соответствующее нагревание ионов или электронов плазмы. Высокая эффективность такого метода была впервые продемонстрирована в Институте атомной энергии имени И. В. Курчатова и впоследствии подтверждена на других установках. На установке PLT (США) таким методом была достигнута температура ионов до $4 \cdot 10^7$ К.

4) *Сжатие плазмы* при прохождении через нее электрического тона из-за притяжения параллельно идущих нитей тока (*пинч-эффект*). Нагревание происходит вследствие адиабатического сжатия плазмы, а также выделения тепла за счет ударных волн и турбулентных процессов.

5) *Нагревание лазерным излучением большой мощности.*

6) *Нагревание интенсивными электронными пучками.*

7) *Нагревание пучками ускоренных тяжелых ионов.*

Последние три метода относятся к установкам с плотной плазмой ($n \sim 10^{23}$ см$^{-3}$).

Указанные методы нагрева плазмы до термоядерных температур и выше являются не только принципиальными, но и технически осуществимыми. Основная трудность УТС — проблема *удержания высокотемпературной и достаточно плотной плазмы в рабочем объеме реактора.*

8. Отвлекаясь на время от способов удержания плазмы, посмотрим, какие ограничения накладываются на ее концентрацию n. Обозначим

через $\mathscr{P}_0$ давление плазмы до реакции. Во избежание разрушения стенок реактора $\mathscr{P}_0$ не может быть слишком большим, скажем, не должно превышать одной атмосферы (порядка 10^6 дин/см^2). Но $\mathscr{P}_0 \sim nkT$. При $T \approx 10^8$ К находим отсюда $n \sim \mathscr{P}_0/kT \approx 10^{14}$ см$^{-3}$. В результате выделения термоядерной энергии температура плазмы, а с ней и давление повышаются в несколько сот раз. Такое повышение давления технически допустимо; оно должно восприниматься удерживающими устройствами. При повышении температуры в сто раз выделяющаяся термоядерная энергия составляла бы $100nkT \approx 10$ Дж/см^3. В действительности энерговыделение будет больше, так как температура плазмы повышается в несколько сот раз. Это выделение энергии происходит за время удержания τ. Из критерия Лоусона (98.9) находим, что должно быть $\tau \gtrsim 1$ с. Полагая $\tau = 1$ с, видим, что мощность выделения энергии будет порядка нескольких десятков Вт/(см$^3 \cdot$ с). Для получения мощности порядка 1000 МВт необходимо, чтобы рабочий объем реактора был во много раз больше, т. е. порядка $1000 \cdot 10^6 : 10 \approx 10^8$ см$^3 = 100$ м^3. При таких объемах термоядерный реактор достигнет мощности современных крупных электростанций.

9. Следует различать два направления, по которым в настоящее время проводятся исследования по управляемому термоядерному синтезу: в одном направлении исследования ведутся с плазмой малой концентрации ($n \sim 10^{14}$ см$^{-3}$), в другом — с плазмой большой концентрации ($n \sim 10^{23}$ см$^{-3}$).

При малых концентрациях плазмы основной трудностью является достижение необходимого времени удержания (порядка секунды). Не годятся никакие вещества, из которых можно было бы сделать сосуд для удержания плазмы в течение такого времени: при соприкосновении со стенкой сосуда горячая плазма еще раньше охладилась бы и притом испарила бы стенку. Для удержания плазмы и предотвращения ее от соприкосновения со стенками сосуда используются магнитные поля различной напряженности и конфигурации. Такие поля принято называть *магнитными ловушками*. В т. III (гл. V) уже рассматривалось поведение отдельной заряженной частицы в магнитных ловушках.

В постоянном однородном магнитном поле частица движется по спирали, навивающейся на магнитную силовую линию. Ее движение ограничено только поперек магнитного поля, а параллельно полю может происходить беспредельно. Создадим однородное поле внутри ограниченного цилиндра и усилим его на концах («магнитными пробками»). Тогда получится магнитная ловушка и частица будет удерживаться в ограниченной области пространства, если только направление ее скорости не лежит внутри «конуса потерь» (см. т. III, § 88).

Ограничения области движения частицы можно также достигнуть, свернув цилиндр в тор. Однако в этом случае возникнет дрейф, приводящий к столкновению частицы со стенками тороидальной камеры (см. там же).

10. Переход от исследования движения отдельных заряженных частиц к изучению движения *всей плазмы* осложняется взаимодей-

ствием между частицами. Приближенно это взаимодействие учитывается в гидродинамической модели плазмы. В простейшей из этих моделей считается, что плазма ведет себя подобно квазинейтральной сплошной среде — жидкости, движущейся в магнитном поле. Это — *одножидкостная магнитогидродинамическая модель.* В более сложной *двухжидкостной магнитогидродинамической модели* считается, что плазма состоит как бы из двух жидкостей: одна описывает движение *электронов*, другая — *ионов.*

Можно показать, что обмен энергиями между одноименными частицами плазмы, т. е. между электронами или между ионами, происходит много быстрее, чем между разноименными, т. е. между легкими электронами и тяжелыми ионами. Иными словами, обмен теплом между различными жидкостями плазмы есть процесс более медленный, чем установление равновесного состояния каждой из этих жидкостей в отдельности. Поэтому пока не наступило полное термодинамическое равновесие плазмы, электронной и ионной компонентам можно приближенно приписать определенные температуры, вообще говоря, не совпадающие между собой.

Не входя ни в какие подробности, ограничимся простейшим примером. Пусть плазма заполняет бесконечно длинный цилиндр, параллельно оси которого в ней течет электрический ток J (такая конфигурация называется *плазменным шнуром*). Последний возбуждает магнитное поле $\mathbf{H}$, которому соответствует магнитное давление $H^2/8\pi$ (см. т. III, § 72). Если на поверхности цилиндра оно превосходит газокинетическое давление $2nkT$ (мы применим одножидкостную модель), то плазменный шнур будет сжат — это явление называется *пинч-эффектом.*

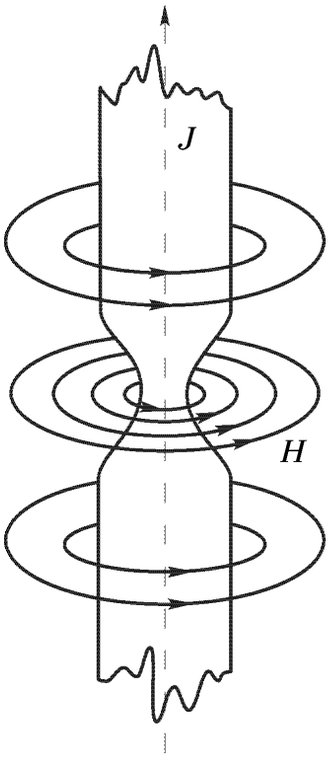


Рис. 169

Для равновесия плазменного шнура необходимо, чтобы на его поверхности магнитное давление $H^2/8\pi$ уравновешивалось газокинетическим давлением $2nkT$. Однако такое равновесие будет крайне неустойчиво. Действительно, допустим, что по случайным причинам на шнуре возникло небольшое сужение (перетяжка). В силу большой проводимости высокотемпературной плазмы магнитный поток через поперечное сечение шнура сохранится (см. т. III, § 71), так что магнитное давление на суженную часть шнура увеличится. Правда, в суженной части возрастет и газокинетическое давление. Но это вызовет только перетекание части плазмы в более широкие участки плазменного шнура. Поэтому под действием возросшего магнитного давления перетяжка на шнуре усилится, т. е. равновесие будет неустойчиво по отношению к такого рода возмущениям (рис. 169). Но равновесие плазменного шнура неустойчиво и по отношению к его малым изгибаниям. Это

видно из рис. 170 — на вогнутой стороне шнура магнитное поле, а с ним и магнитное давление больше, чем на выпуклой.

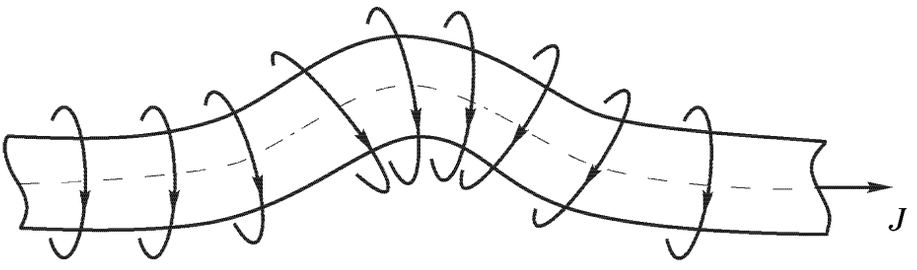


Рис. 170

В плазме существует и много других типов неустойчивостей. Природа многих из них может быть понята на основе гидродинамических моделей, но есть и такие неустойчивости, для понимания которых требуется кинетическое рассмотрение. На всех этих специальных вопросах, далеко выходящих за пределы настоящего курса, мы останавливаться не можем. Заметим лишь, что крайняя неустойчивость плазмы во всех магнитных полях простой геометрической конфигурации создает *главную трудность*, которую надо преодолеть на пути осуществления управляемого термоядерного синтеза.

Для преодоления указанной трудности идут по пути усложнения конфигурации и напряженности магнитных полей. Магнитные поля создаются не только токами, текущими в плазме, а главным образом *внешними источниками.* Не входя в детали, заметим, что удерживающее и стабилизирующее действие магнитных полей основано, в частности, на высокой электрической проводимости плазмы.

Из-за высокой проводимости магнитный поток через любой замкнутый контур, состоящий из одних и тех же частиц, остается практически постоянным при любом движении плазмы (см. т. III, § 79). Это утверждение называется *вмороженностью магнитных силовых линий в плазму.* Кроме того, вдоль магнитных силовых линий существует *натяжение* (см. там же § 72); силовые линии ведут себя подобно натянутым струнам. Оба эти фактора способствуют удержанию и стабилизации плазмы. Особенно наглядно это проявляется при устранении случайно возникших изгибов плазменного шнура.

11. Из различных магнитных ловушек в настоящее время специалисты считают наиболее перспективной ловушку, называемую *токамаком* (сокращение от «тороидальная камера с магнитными катушками»). Она была предложена и наиболее полно изучена экспериментально и теоретически в Институте атомной энергии в Москве группой физиков, возглавлявшейся Л. А. Арцимовичем и М. А. Леонтовичем, а позднее Б. Б. Кадомцевым (р. 1928). И в других странах, ведущих исследования по управляемому термоядерному синтезу (США, Япония, страны Евратома), в настоящее время отдается предпочтение токамакам.

Естественно, что в курсе общей физики можно коснуться, и притом в общих чертах, только принципа действия токамака, опуская все

конструктивные детали и инженерные вопросы. Основной частью токамака является *тороидальная вакуумная камера*, в которую вводится дейтерий (а в дальнейшем, когда от экспериментальных токамаков перейдут к сооружению энергетических реакторов, будет вводиться смесь дейтерия с тритием). Иными словами, камера представляет собой трубу, свернутую в кольцо. Радиус окружности вдоль оси этого кольца R называется *большим радиусом* тороидальной камеры. Через отверстие кольца создается магнитный поток. При его изменении во времени возбуждается параллельное оси трубы вихревое электрическое поле. Происходит электрический пробой, образуется плазма, и в кольцевой камере начинает циркулировать электрический ток. Такой ток нагревает плазму и возбуждает магнитное поле H_φ, силовые линии которого отжимают плазму от внешних стенок тороидальной камеры, формируя кольцевой плазменный шнур. Радиус поперечного сечения этого шнура a называется *малым радиусом* тороидальной камеры.

Несущественно, будет ли ток в плазменном шнуре постоянным (одного направления) или переменным, направление которого периодически меняется. Поэтому описанный индукционный способ возбуждения электрического тока в токамаке не является единственно возможным. Возможен и безындукционный способ. Можно, например, ввести в плазму электромагнитные волны определенной частоты — такие, что они вызовут движение электронов вдоль тороидального магнитного поля (см. следующий абзац), создаваемого внешними источниками. Опыты, поставленные в Японии, свидетельствуют о перспективности такого способа возбуждения тока. Его преимущество, вероятно, состоит в том, что он позволит поддерживать ток в торе токамака более длительное время, чем индукционный способ.

Для удержания и стабилизации плазменного шнура используются *тороидальное* и *полоидальное* магнитные поля, создаваемые внешними источниками. Сильное тороидальное магнитное поле H_θ направлено параллельно току в тороидальной камере и создается катушками, намотанными на тор. Как показывает теория, для магнитогидродинамической устойчивости плазмы должен выполняться так называемый *критерий Шафранова* (р. 1929)–*Крускала* (р. 1925):

$$H_\theta a > H_\varphi R. \tag{98.9}$$

Относительно слабое полоидальное поле $H_\perp \approx H_\varphi a/R$ перпендикулярно к центральной плоскости симметрии тороидальной камеры. Оно создается проводами, расположенными вдоль тора. Такое поле необходимо для удержания плазменного шнура в равновесии и удаления продуктов реакции из плазмы.

Плазма в токамаке, как уже было сказано, нагревается протекающим по ней током. Для дополнительного нагревания используются переменные электромагнитные поля, а также инжекция быстрых нейтральных атомов.

Важным параметром токамака является отношение газокинетического давления плазмы $\mathscr{P}$ к магнитному давлению $H^2/8\pi$. Его принято обозначать через β: $\beta = 8\pi\mathscr{P}/H^2$. Теория показывает, что β не

должно быть большим, так как в противном случае плазма становится неустойчивой. С другой стороны, для экономически оправданного энергетического реактора значение β должно составлять не менее 5 %. К настоящему времени в Институте атомной энергии им. И. В. Курчатова на токамаке Т-11 достигнуто значение $\beta = 3\,\%$, а в США на токамаке «Дублет-III» с эллиптическим сечением плазменного шнура — значение $\beta = 4{,}5\,\%$.

12. Для уменьшения энергетических потерь на создание сильного магнитного поля H (50–60 кГс и выше) предполагается использовать обмотки из *сверхпроводящих сплавов* с высокими значениями критического поля H, при котором начинается разрушение сверхпроводимости (см. т. III, § 80). Этот метод был впервые испытан в СССР на установке Т-7, в которой применялся сплав из ниобия с танталом при гелиевых температурах. В установке Т-15, сооружаемой в нашей стране, предполагается использовать сплав ниобия с оловом.

Решение проблемы высокотемпературной сверхпроводимости произвело бы революцию в криогенной технике, в частности при создании сверхсильных магнитных полей. В этом направлении ведутся интенсивные исследования в СССР, США, Японии и других странах. К концу 1986 г. появились сообщения, что получены материалы, становящиеся сверхпроводящими при температуре не только жидкого азота, но и при температурах, близких к комнатной. Сейчас же заметим, что получение сверхсильных магнитных полей предъявляет очень высокие требования к механической прочности сверхпроводящих катушек (уже при $H = 50$ кГс магнитное давление $H^2/8\pi$ будет порядка 10^2 атм, а при $H = 150$ кГс — порядка 10^3 атм).

13. В общем курсе физики нет смысла приводить подробные научно-технические сведения о состоянии проблемы УТС. Всякие сведения такого рода могут устареть уже к моменту выхода настоящей книги. Ограничимся поэтому только некоторыми произвольно выбранными данными о токамаках и результатах, полученных на них.

Самым крупным из действующих отечественных токамаков является токамак Т-10, запущенный в 1975 г. в Институте атомной энергии (Москва). Тороидальное магнитное поле в нем $H_\theta \approx 50$ кГс, объем плазмы — 5 м^3 ($R = 150$ см, $a = 39$ см), ток в плазме около 800 кА, $T \approx 1{,}2 \cdot 10^7$ К, $\tau \approx 0{,}07$ с, $n \approx 8 \cdot 10^{13}$ см$^{-3}$, $n\tau \approx 6 \cdot 10^{12}$ см$^{-3} \cdot$ с). В 1979 г. там же был запущен токамак Т-7 с параметрами, близкими к параметрам токамака Т-10, но со сверхпроводящими обмотками. Там же при экспериментальных исследованиях на токамаках было выяснено, что время удержания τ быстро возрастает с увеличением объема плазмы — приблизительно пропорционально aR^2.

На токамаке TFTR (США), запущенном в 1983 г., достигнуто рекордное значение $\tau = 0{,}19$ с ($R = 248$ см, $a = 85$ см, $H_\theta = 52$ кГс, ток в плазме — 2,5 МА, мощность дополнительного нагрева инжекцией быстрых атомов — 33 МВт, стоимость установки — 314 млн долларов). В конце 1986 г. на этой установке была получена рекордная температура

$2{,}2 \cdot 10^8$ К. При увеличении $n\tau$ в 5–10 раз при такой температуре токамак на смеси дейтерия с тритием стал бы работать уже в критическом режиме, когда выделяющаяся термоядерная энергия равна энергии, затрачиваемой на нагрев и удержание плазмы. На оси плазменного шнура на той же установке достигнута ионная температура $3{,}3 \cdot 10^8$ К (29 кэВ).

Токамаки — весьма сложные и дорогостоящие установки. Поэтому Советский Союз внес предложение в Международное агентство по атомной энергии (МАГАТЭ) объединить усилия стран, в которых ведутся работы по УТС, для создания опытного термоядерного реактора. Это предложение было принято. Установка получила название ИТЭР (Интернациональный термоядерный экспериментальный реактор). Определились участники проекта ИТЭР: СССР, США, Япония и европейские страны — члены Евратома, включая Швецию и Швейцарию, имеющие соглашения с Евратомом в области термоядерного синтеза. Проект должен быть разработан к концу 1990 г. Проектная группа будет постоянно работать в Институте физики плазмы имени Макса Планка в Гархинге вблизи Мюнхена. Основная задача ИТЭР — техническая демонстрация управляемого термоядерного синтеза. Эта установка еще не будет экономически выгодной, но она должна доказать возможность получения и использования в промышленных целях термоядерной энергии синтеза легких элементов. Ее назначение — стать прообразом будущего реактора термоядерной электростанции.

14. В заключение рассмотрим вопрос о принципиальной возможности термоядерного реактора, в котором не применяются никакие приспособления для удержания плазмы. Первоначально в вакуумную камеру реактора вводится смесь дейтерия и трития в твердом (замороженном) состоянии, которая практически мгновенно нагревается до температуры порядка 10^8 К. Ее пребывание в первоначальном ограниченном объеме определяется временем разлета плазмы порядка l/v, где l — линейный размер рассматриваемого объема, а v — средняя скорость частиц плазмы после ее нагревания. Это время и может быть принято за время удержания плазмы в рассматриваемом объеме, которое входит в критерий Лоусона (98.7). Таким образом, для зажигания плазмы должно быть $nl/v > L$, откуда

$$l > \frac{Lv}{n}. \tag{98.10}$$

Приняв для дейтериево-тритиевой плазмы $L = 10^{14}$ см$^{-3} \cdot$ с и полагая для такой плазмы в твердом состоянии $n = 4{,}5 \cdot 10^{22}$ см$^{-3}$, $v \approx 10^8$ см/с, получим

$$l_{\text{мин}} \approx \frac{10^{14} \cdot 10^8}{4{,}5 \cdot 10^{22}} \approx 0{,}2 \text{ см} = 2 \text{ мм}.$$

Давление плазмы непосредственно после ее нагревания будет $\mathscr{P} = 2nkT \approx 2 \cdot 4{,}45 \cdot 10^{22} \cdot 1{,}38 \cdot 10^{-16} \approx 10^{15}$ дин/см^2 $\approx 10^6$ атм, если не учитывать сжатия плазмы в процессе нагревания. После того как совершится термоядерная реакция, оно увеличится примерно еще

в 1000 раз, т. е. достигнет приблизительно 10^9 атм. Такое повышение давления может быть названо микровзрывом, так как оно происходит в объеме $\omega \approx l^3 \approx 0{,}01\text{см}^3$. Если объем камеры $V = 1$ м^3 $= 10^6$ см^3, а микровзрыв происходит в ее центре, то максимальное давление на стенки камеры будет порядка $10^9 V/\omega \approx 10$ атм, что при достаточной толщине стенок камеры не представляет опасности. В действительности при быстром введении энергии в плазму последняя сильно сжимается. Но это сжатие повышает давление плазмы только в месте ее образования, а повышения давления на стенки камеры не произойдет, так как на них действует уже расширившаяся плазма.

Очевидно, что такой термоядерный реактор будет импульсным: в рабочую камеру реактора должно периодически вводиться и взрываться в ней термоядерное топливо в виде небольших крупинок.

Трудности создания термоядерного реактора рассмотренного вида не исчезают. Они только переносятся с проблемы удержания плазмы на проблему ее нагрева. Ведь этот нагрев должен быть осуществлен за время, не превосходящее времени разлета плазмы, т. е. быстрее чем за $l_{\text{мин}}/v \approx 2 \cdot 10^{-9}$ с.

Преодоление указанной трудности идет по двум направлениям. Во-первых, нагревание плазмы до температуры порядка 10^8 К производится концентрацией лазерного излучения на крупинках термоядерного топлива (Н. Г. Басов (р. 1922) с сотрудниками). Во-вторых, такая же концентрация производится пучками ускоренных релятивистских электронов (Е. К. Завойский (1907–1976) с сотрудниками), а также ускоренных тяжелых ионов. В обоих направлениях достигнуты заметные успехи, но мы не будем на них останавливаться.

§ 99. Нейтронная оптика

1. Волновые свойства нейтронов проявляются тем отчетливее, чем больше длина волны соответствующих им волн де Бройля, т. е. чем меньше их кинетическая энергия. Напротив, для нейтронов очень высоких энергий на первый план выступают корпускулярные свойства. В целях краткости волны де Бройля для нейтронов будем называть в дальнейшем просто *нейтронными волнами.* Некоторые характеристики нейтронов приведены в табл. 17. Самыми длинноволновыми являются *ультрахолодные нейтроны*, у которых длина волны лежит в пределах оптической области электромагнитного спектра (длинный ультрафиолет, видимый свет, инфракрасное излучение). Их скорости не превышают 10 м/с. В поле тяжести Земли свободные ультрахолодные нейтроны движутся по параболам и могут подняться на высоту всего в несколько метров. *Тепловые нейтроны* характеризуются тем, что их кинетическая энергия порядка kT, причем температура T лежит в интервале от нескольких десятков до нескольких тысяч кельвинов.

В общем случае длина нейтронной волны определяется соотношением

$$\lambda = h/p. \tag{99.1}$$

Таблица 17

Некоторые характеристики нейтронов

Нейтроны	Энергия, эВ	Скорость, м/с	Длина волны, нм	Температура, К
Ультрахолодные	$< 5 \cdot 10^{-7}$	< 10	> 40	$6 \cdot 10^{-3}$
Очень холодные	$5 \cdot 10^{-7}$–$2 \cdot 10^{-4}$	10–200	40–2	$6 \cdot 10^{-3}$–2
Холодные	$2 \cdot 10^{-4}$–$5 \cdot 10^{-3}$	200–1000	2–0,4	2–60
Тепловые	$5 \cdot 10^{-3}$–10^{-1}	1000–4500	0,4–0,09	60–1000
Резонансные	10^{-1}–10^{-4}	—	0,09–$3 \cdot 10^{-5}$	—
Промежуточные	10^{4}–10^{5}	—	$3 \cdot 10^{-5}$–10^{-4}	—
Быстрые	10^{5}–10^{8}	—	10^{-4}–$3 \cdot 10^{-6}$	—
Высокоэнергетичные	10^{8}–10^{10}	—	$3 \cdot 10^{-6}$–10^{-7}	—
Релятивистские	$> 10^{10}$	—	$< 10^{-7}$	—

Если $p \ll mc$, где m — масса нейтрона, то можно пользоваться нерелятивистской формулой $p^2 = 2m\mathscr{E}$, так что

$$\lambda = \frac{h}{\sqrt{2m\mathscr{E}}}. \tag{99.2}$$

Здесь $\mathscr{E}$ означает кинетическую энергию нейтрона. Для *быстрых нейтронов* эта формула дает еще достаточную точность. Но для *высокоэнергетических* и в особенности *релятивистских* нейтронов надо пользоваться релятивистским соотношением

$$(pc)^2 = \mathscr{E}_{\text{рел}}^2 - (mc^2)^2.$$

Если учесть, что $\mathscr{E}_{\text{рел}} = \mathscr{E} + mc^2$, то получается релятивистская формула

$$\lambda = \frac{hc}{\sqrt{\mathscr{E}^2 + 2\mathscr{E}mc^2}}. \tag{99.3}$$

Общий характер дифракции волн любой природы определяется прежде всего их длиной. В этом отношении, как показывает табл. 17, условия дифракции тепловых и коротковолновых холодных нейтронов примерно такие же, что и для жестких рентгеновских лучей. Дифракция на кристаллах может отчетливо наблюдаться только тогда, когда длина волны порядка или меньше постоянной кристаллической решетки.

Понятно, что направления на главные дифракционные максимумы, поскольку они имеют общеволновой характер, определяются одинаково для волн всех видов: рентгеновских, электронных, нейтронных. В частности, для описания дифракции нейтронных волн в кристаллах справедливо условие Брэгга–Вульфа (18.1), т. е.

$$2d \sin \varphi = m\lambda, \tag{99.4}$$

где d — межплоскостное расстояние, φ — угол скольжения, $m = 1, 2, 3, \ldots$. При выполнении этого условия в кристалле происходит так называемое *интерференционное отражение* волн, т. е. усиление волн, отраженных от параллельных атомных плоскостей кристалла (см. т. IV, § 34, 61).

2. Существенно отметить, что при $\lambda > 2d$ условие (99.4) не может выполняться ни при каком угле скольжения. Таким образом, для достаточно длинных волн интерференционное отражение от кристалла невозможно. Но, конечно, как и в световой оптике, возможно обыкновенное, или полное, отражение от граней кристалла. Для ультрахолодных нейтронов реализуется именно этот случай.

Энергия нейтрона, при которой исчезает интерференционное отражение, называется энергией *брэгговского скачка.* Она различна для различных кристаллов и по порядку величины равна 10^{-3} эВ. Явление брэгговского скачка используется для отфильтровывания холодных нейтронов путем отражения пучка нейтронов от монокристаллов.

Строение кристаллов, сплавов, жидкостей, молекул и других объектов исследуется в рентгенографии и электронографии путем дифракции рентгеновских лучей и электронов. Для исследования кристаллов применяются методы Брэгга, Лауэ и Дебая–Шерера–Хелла (см. т. IV, § 61). Такие же методы применимы и в нейтронографии, основанной на дифракции тепловых нейтронов. Метод Дебая–Шерера–Хелла (метод порошка) в нейтронографии обычно применяется в тех случаях, когда трудно или даже невозможно выращивать достаточно крупные монокристаллы. Но использование порошков дает меньшую информацию по сравнению с тем, что можно получить, используя монокристаллы и добиваясь получения отражения нейтронных волн от максимально возможного числа атомных плоскостей.

Основными методами структурного анализа объемных объектов являются рентгенография и нейтронография. Электронография из-за незначительной проникающей способности электронов применяется преимущественно для исследования поверхностных структур. Метод дифракции нейтронов по распространенности не может сравниться с рентгеновским методом из-за ограниченной доступности источников нейтронов. Однако нейтронографический метод может выявить такие особенности структуры изучаемых объектов, которые недоступны рентгеновскому методу, и наоборот. Это связано с тем, что особенности дифракционных картин определяются не только длиной волны, но и другими физическими свойствами, которые у нейтронов и рентгеновских квантов различны. Оба метода не противоречат, а взаимно дополняют друг друга.

3. В нашу задачу не входит изложение методов и результатов нейтронографии. Отметим только некоторые принципиальные особенности, которые во многих областях делают нейтронографические методы незаменимыми.

При одной и той же длине волны энергии рентгеновских квантов, электронов и нейтронов не одинаковы. Так, при $\lambda = 10^{-8}$ см, как легко вычислить, энергия рентгеновского кванта равна 12,5 кэВ, энергия

электрона 150 эВ, а энергия нейтрона всего 0,08 эВ. Малая энергия частиц, из которых состоит излучение, существенна при облучении легко разрушаемых, например биологических, объектов. В этом отношении предпочтение надо отдать нейтронам по сравнению с рентгеновскими квантами.

Рентгеновские волны рассеиваются электронными оболочками атомов. Амплитуда рассеяния (определение этого понятия см. в п. 4) рентгеновских волн на атоме пропорциональна числу электронов в оболочке, т. е. зарядовому числу Z. Поэтому эти амплитуды сильно отличаются для легких и тяжелых атомов. Рентгенография практически не позволяет различать в кристалле разные атомы, если в них содержится почти одинаковое число электронов; совсем неразличимы изотопы одного и того же элемента. Рентгенография практически не позволяет обнаружить очень легкие атомы (например, водород) на фоне тяжелых (свинец, торий, уран, редкоземельные элементы и пр.). Но все это доступно методам нейтронографии, так как на нейтроны действуют не электрические, а *ядерные силы* — эти частицы рассеиваются атомными ядрами. Поэтому дифракция нейтронов дает информацию не об электронных оболочках, а о расположении *атомных ядер* — центров самих атомов. Амплитуды рассеяния и связанные с ними сечения нейтронов на ядрах одного и того же порядка для всех ядер — от самых легких до самых тяжелых. Поэтому-то на рентгенограммах химических соединений, содержащих как легкие, так и тяжелые элементы (H_2O, PbS, ThD_2 и т. д.), дифракция на легких атомах едва заметна, тогда как на нейтронограммах она проявляется отчетливо. Все же амплитуды рассеяния и сечения нейтронов на ядрах, оставаясь одного порядка по величине, нерегулярно меняются (иногда резко) при переходе от одного ядра к соседнему и от одного изотопа к другому изотопу того же элемента (см. табл. 18). Поэтому методы нейтронографии особенно необходимы для расшифровки структуры сплавов и химических соединений из элементов с близкими зарядовыми числами (FeCo, NiMn и т. д.), компоненты которых имеют сходные электронные конфигурации, но резко различаются по ядерным сечениям.

Нейтроны не имеют электрического заряда и поэтому непосредственно не вызывают ионизации вещества. Ионизация вызывается вторичными заряженными частицами (протонами отдачи, электронами, позитронами), порождаемыми нейтронами при ядерных столкновениях. Эти заряженные частицы в нейтронографии и регистрируются детекторами. Нейтроны не действуют и на фотопластинку, так как фотографическое действие вызывается электрическими, а не ядерными силами. Чтобы для регистрации нейтронов использовать фотопластинку, фотослой покрывают фольгой из элементов, которые под действием нейтронов становятся β- или γ-активными (Gd, Dy, In). В местах, куда попадают нейтроны, они реагируют с ядрами фольги. Образуются радиоактивные ядра, излучения которых и действуют на фотопластинку. Первые нейтронограммы были получены с помощью фотографического метода на экране из индия и фотопленки в 1948 г. (см. § 18, п. 9). Созданная к настоящему времени фольга из смеси ^{6}LiF–ZnS(Ag)

в сочетании с высокочувствительной рентгеновской пленкой позволяет фотометоду успешно конкурировать с традиционными методами детектирования нейтронов.

Жесткие рентгеновские лучи рассеиваются преимущественно вперед. При тех же длинах волн рассеяние нейтронов происходит изотропно. Поэтому при переходе к большим углам рассеяния интенсивность рентгеновских пиков на рентгенограмме резко падает, тогда как на нейтронограмме интенсивность нейтронных максимумов почти не меняется.

Проникающая способность нейтронов в тысячи раз больше, чем у рентгеновских лучей той же длины волны. Это связано с тем, что благодаря электрической нейтральности нейтрон не подвергается воздействию кулоновского поля атома, а потому легко проникает внутрь его и взаимодействует с ядром посредством ядерных сил. Поэтому для исследования внутренних областей образцов нейтронографические методы более эффективны, чем рентгеновские.

Особенно большое значение нейтронографический метод получил для исследования биологических макромолекул. Химический состав последних ограничен практически четырьмя элементами — водородом, углеродом, азотом и кислородом, которые являются близкими соседями по периодической системе Менделеева. А как уже подчеркивалось выше, это обстоятельство делает их практически неразличимыми рентгенографическим методом. Сложность же структурного анализа биологических кристаллов иллюстрируется следующим примером. Элементарная ячейка миоглобина содержит 1200 атомов. Для достижения нужного разрешения 0,14–0,2 нм были произведены измерения с использованием более чем 10 000 брэгговских отражений.

Тепловые и холодные нейтроны (если они не захватываются ядрами) не возбуждают атомные ядра и даже сами атомы. Но они могут возбудить акустические колебания кристаллической решетки — фононы. Это дает метод установления одной из важнейших характеристик вещества — закона дисперсии, т. е. зависимости энергии $\mathscr{E}$ элементарного возбуждения от его импульса p. Именно таким методом была получена кривая на рис. 116, первоначально постулированная Ландау на основе термодинамических соображений в его теории сверхтекучести (см. § 61).

4. Процесс распространения нейтронных волн в среде, частным случаем которого является дифракция, детально может быть описан подобно тому, как описывается распространение световых волн в молекулярной оптике. В основу такого описания кладут элементарные акты рассеяния нейтронных волн на отдельных атомных ядрах. Пусть на ядро падает плоская монохроматическая нейтронная волна вида $\psi = e^{ikr}$. Рассеянная атомным ядром волна на далеких расстояниях в линейном приближении может быть представлена в виде

$$\psi_1 = \frac{b}{r}\,\psi = \frac{b}{r}\,e^{ikr}, \tag{99.5}$$

где постоянная величина b, имеющая размерность длины, называется *амплитудой рассеяния* (начало координат помещено в рассеивающем центре, т. е. в центре ядра атома). Она может зависеть от направления рассеяния. Но если ядро сферически-симметрично, а его размеры пренебрежимо малы по сравнению с длиной нейтронной волны (только такие волны и рассматриваются в этом параграфе), то такой зависимости нет, т. е. само рассеяние также сферически-симметрично.

Через амплитуду рассеяния можно выразить эффективное сечение σ рассеяния нейтронной волны на атомном ядре. Плотность вероятности нахождения рассеянного нейтрона в пространстве равна $|\psi_1|^2 = |b|^2/r^2$, плотность радиального потока вероятности $v|\psi_1|^2 = v|b|^2/r^2$, где v — радиальная скорость нейтрона. Вероятное число нейтронов, проходящих в одну секунду через площадку dS, нормальную к радиусу, будет $dSv|b|^2/r^2 = v|b|^2\,d\Omega$, где $d\Omega = dS/r^2$ — телесный угол, под которым видна площадка dS из центра рассеяния. Плотность потока вероятности падающей волны равна $v|e^{ikr}|^2 = v$, так что для эффективного сечения рассеяния в телесный угол $d\Omega$ получаем

$$d\sigma = \frac{v|b|^2\,d\Omega}{v} = |b|^2\,d\Omega, \tag{99.6}$$

а для полного сечения рассеяния

$$\sigma = 4\pi|b|^2. \tag{99.7}$$

5. Амплитуда рассеяния может быть введена и для рассеяния рентгеновских лучей на атомах совершенно так же, как это делается для нейтронов. В табл. 18 приведены амплитуды рассеяния f рентгеновских лучей на атомах (для угла рассеяния, равного нулю) и амплитуды рассеяния нейтронов b на атомных ядрах для некоторых элементов. Если у символа химического элемента указано массовое число, то амплитуды рассеяния относятся к определенному изотопу соответствующего элемента. Если же массовое число не указано, то имеется в виду рассеяние на естественной смеси изотопов этого элемента.

Амплитуды рассеяния рентгеновских лучей на атомах, как легко проверить по табл. 18, меняются строго пропорционально числу электронов в электронных оболочках атомов. В частности, они абсолютно одинаковы для изотопов одного и того же химического элемента. На эту закономерность уже было указано выше. Амплитуды же рассеяния нейтронов на ядрах меняются нерегулярно при переходе от одного ядра к другому. По порядку величины они в среднем лежат в интервале 10^{-13}–10^{-12} см для всех ядер — от самых легких до самых тяжелых. Этого и следовало ожидать, так как величина 10^{-24} см^2 является характерным эффективным сечением для ядерных процессов. Так как интенсивность рассеяния определяется квадратом амплитуды рассеяния, то рассеяние рентгеновских лучей атомом урана в $(25{,}9/0{,}282)^2 = 8400$ раз интенсивнее, чем атомом водорода. Для нейтронов же интенсивность рассеяния увеличивается в этом случае всего в $(0{,}82/0{,}374)^2 = 4{,}8$ раз. Важность этого факта в структурном анализе была отмечена выше.

Таблица 18

Амплитуда рассеяния рентгеновских лучей (f) и нейтронов (b) некоторыми элементами (в единицах 10^{-12} см)

Элемент	Число электронов	f	b	Элемент	Число электронов	f	b
H	1	0,282	−0,374	V	23	6,5	−0,05
D	1	0,282	0,667	Fe	26	7,3	0,95
Li	3	0,84	−0,214	Ni	28	7,9	1,03
^{6}Li	3	0,84	$0{,}18 + i\,0{,}025$	^{58}Ni	28	7,9	1,44
B	5	1,41	$0{,}54 + i\,0{,}021$	^{62}Ni	28	7,9	−0,87
^{10}B	5	1,41	$0{,}14 + i\,0{,}11$	Zr	40	11,3	0,71
O	8	2,25	0,575	Cd	48	13,6	$0{,}37 + i\,0{,}16$
S	16	4,5	0,28	^{113}Cd	48	13,6	$-1{,}5 + i\,1{,}2$
Ti	22	6,2	−0,34	U	92	25,9	0,82

6. Вследствие наличия у некоторых ядер резонансных уровней энергии, близких к энергии тепловых нейтронов, амплитуда рассеяния b для таких ядер отрицательна, а в случае поглощения ядрами нейтронов становится комплексной. В табл. 18 комплексная амплитуда b дана для нейтронов с длиной волны 0,1 нм. Если ограничиться только незатухающими нейтронными волнами, то амплитуда рассеяния должна быть вещественной: положительной или отрицательной.

Знак амплитуды рассеяния может отразиться на наблюдаемых физических явлениях. Рассмотрим, например, рассеяние нейтронной волны на кристаллической решетке. Его можно интерпретировать как результат отражения волн от параллельных атомных плоскостей и их последующей интерференции. Интерференционное усиление отраженных волн выражается условием Брэгга–Вульфа (99.4). Если атомная плоскость состоит из атомных ядер одного вида, то волны, рассеянные на отдельных ядрах, усиливают друг друга в направлении, в котором должно происходить отражение от этой плоскости. Допустим теперь, что атомная плоскость состоит из атомных ядер двух сортов, амплитуды рассеяния которых вещественны, но противоположны по знаку. Тогда от нее получатся две отраженных волны, фазы которых противоположны. Такие волны будут ослаблять друг друга. В том случае, когда амплитуды рассеяния одинаковы по величине, произойдет интерференционное гашение обеих волн. Значит, на атомной плоскости не будет отражения, не появятся и дифракционные пучки.

Различие знаков амплитуд рассеяния нейтронов практически используется для приготовления так называемых *нулевых матриц*, т. е. сплавов, для которых при определенных концентрациях компонентов амплитуда рассеяния в среднем равна нулю. В таких сплавах не происходит когерентного рассеяния нейтронов. К ним относятся сплавы

Ti–Zr, сплавы на основе Mn, а также изотопическая смесь на основе Ni. Рассмотрим, например, сплав Ti–Zr. Обозначим через x процентное содержание в нем Ti (по числу атомов). Такой сплав не будет когерентно рассеивать нейтроны при выполнении условия

$$0{,}34x + 1{,}44(1 - x) = 0,$$

т. е. при $x = 80\,\%$. Впрочем, это заключение не отличается той же ясностью, какая была в случае кристаллической решетки, построенной из атомов двух сортов. Оно предполагает *когерентность волн*, рассеянных всеми атомами. А это не будет соблюдаться, если различные атомы сплава расположены в пространстве хаотически.

7. Нейтронная волна, проникая из вакуума в среду, рассеивается на атомных ядрах. Если отвлечься от всех флуктуационных процессов, то рассеянные волны будут когерентны между собой, поскольку они возбуждаются одной и той же падающей волной. Конечно, это справедливо только тогда, когда рассеянные ядра достаточно массивны — их массы должны быть велики по сравнению с массой нейтрона. В противном случае при рассеянии на ядрах будет сильно меняться длина нейтронной волны, что поведет к нарушению когерентности. В случае же тяжелых ядер, который только и рассматривается нами, этот эффект незначителен, так что рассеянные волны могут считаться практически когерентными. Разумеется, в веществе не должны происходить ядерные реакции под действием нейтронов. При выполнении этих условий нейтронные волны, рассеянные ядрами, будут интерферировать между собой и с падающей волной. В результате такой интерференции и формируется нейтронная волна в среде. Так же формируются отраженная волна и дифракционные пучки в явлениях дифракции.

8. Исходя из этих представлений, введем *показатель преломления* нейтронных волн в средах. Рассмотрим плоскопараллельный слой вещества, толщина которого Δz много меньше длины волны λ (рис. 171). Пусть на этот слой падает нейтронная волна вида e^{ikz}. Поле нейтронной волны, рассеянной одним ядром, в удаленной точке P будет be^{ikr}/r. Чтобы просуммировать такие поля, используем метод зон Френеля (см. т. IV, § 39). В оптике применимость этого метода предполагает слабое затухание волн, приходящих в точку наблюдения от последовательных зон Френеля. Такое же предположение используется и для нейтронных волн. В этом предположении полное поле излучения рассматриваемого нами слоя в точке наблюдения равно половине излучения вырезанной из него центральной зоны Френеля. Площадь центральной зоны Френеля

$$S = \pi R^2 = \pi\left(z + \frac{\lambda}{2}\right)^2 - \pi z^2 = \pi z\lambda.$$

На ней помещается $NS\Delta z$ ядер, где N — число ядер в единице объема. Если бы все волны, излучаемые ядрами, приходили в точку P в одинаковых фазах, то поле излучения в этой точке было бы

$$\frac{1}{2}\,\frac{S\Delta z N b e^{ikz}}{z} = \frac{1}{2}\,\pi N b e^{ikz}\Delta z.$$

В действительности все рассеянные волны запаздывают по фазе по сравнению с волной, исходящей из центра зоны. Для волн, исходящих из периферийных точек зоны, это запаздывание составляет π. Чтобы учесть запаздывание, применим метод векторных диаграмм. Для этого разобьем всю центральную зону на бесконечно узкие кольца одинаковой площади. Волны, излучаемые такими кольцами, будут иметь одинаковые амплитуды (если не учитывать влияние наклона лучей к плоскости зоны, что приближенно допустимо). Если бы все фазы таких волн

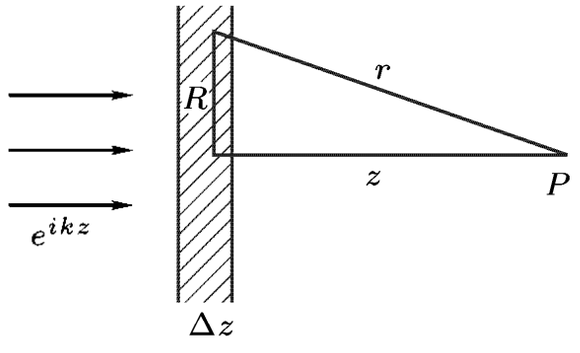


Рис. 171

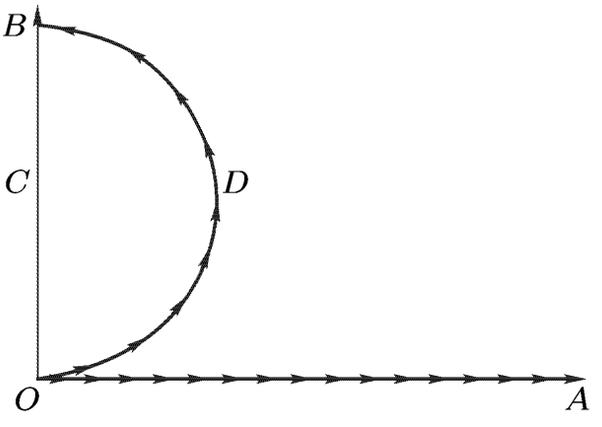


Рис. 172

были одинаковы, то результирующая волна, излучаемая центральной зоной в точке наблюдения P, изобразилась бы на векторной диаграмме прямолинейной цепочкой OA из бесконечно малых векторов, представляющих излучения таких колец (рис. 172). На самом деле из-за сдвига фаз эти векторы должны быть повернуты относительно друг друга на один и тот же угол, причем направление последнего вектора должно быть противоположно направлению первого. Иными словами, прямолинейная цепочка OA должна быть заменена ломаной ODB той же длины, образующей половину правильного многоугольника.

В пределе, когда длина каждого звена этой ломаной стремится к нулю, она переходит в полуокружность той же длины, что и длина прямолинейной цепочки OA. Вектор $\overrightarrow{OB}$, замыкающий эту полуокружность, и представляет на векторной диаграмме волну в точке наблюдения, излучаемую центральной зоной, а половина $\overrightarrow{OC}$ этого вектора — результирующую волну всего плоскопараллельного слоя. Вектор $\overrightarrow{OC}$ повернут относительно прямолинейной цепочки на угол $\pi/2$. Это значит, что результирующая волна запаздывает по фазе на $\pi/2$ относительно волны, исходящей из центра зоны. Следовательно, для получения правильной амплитуды надо предыдущее выражение умножить на $2/\pi$, а для получения правильной фазы — на $e^{-i\pi/2} = -i$. Таким образом,

рассеянная волна в точке P будет

$$-i\lambda N b e^{ikz}\Delta z.$$

Добавив сюда падающую волну e^{ikz}, получим полную волну в точке P:

$$\psi_P = (1 - i\lambda N b\Delta z)e^{ikz} = e^{i(kz-\lambda N b\Delta z)},$$

так как для применимости вывода необходимо предположить, что $\lambda N|b|\Delta z \ll 1$. С другой стороны, ту же волну можно выразить через показатель преломления n нейтронных волн, а именно

$$\psi_P = e^{ik(z-\Delta z)+ikn\Delta z} = e^{i[kz+k(n-1)\Delta z]}.$$

Сравнивая это выражение с предыдущим, получим

$$n = 1 - \frac{Nb\lambda}{k} = 1 - \frac{1}{2}\frac{Nb\lambda^2}{\pi}. \tag{99.8}$$

9. Формула (99.8) справедлива только при условии

$$N|b|\lambda^2 \ll 1. \tag{99.9}$$

Действительно, в формуле (99.5) под ψ надо понимать *полную волновую функцию*, создаваемую всеми внешними источниками в точке нахождения рассеивающего ядра. Между тем при выводе (99.8) вместо полной волновой функции была использована только *падающая* плоская волна и не были учтены волны, рассеянные всеми ядрами, внешними по отношению к рассматриваемому ядру. В частности, не учитывались эффекты отражения волн на границах слоя Δz. Это, между прочим, проявилось в том, что при прохождении через слой в рассматриваемом приближении менялась только фаза волны, но не ее амплитуда.

Можно было бы обобщить формулу (99.8) на случай, когда условие (99.9) не соблюдается, приняв во внимание рассеяние на ядрах не только первичной падающей волны, но и рассеяние всех вторичных волн, исходящих из ядер среды. Мы не будем этого делать, а предпочтем формальный, но более простой путь, использующий аналогию между волновой оптикой и классической механикой (см. т. IV, § 3, 4).

Пусть нейтрон падает из вакуума на плоскую границу среды. Так как он обладает и волновыми, и корпускулярными свойствами, то показатель преломления n можно определить двояко: либо как отношение фазовых скоростей нейтронных волн в вакууме и среде, либо как отношение скорости самого нейтрона в среде v_2 к его скорости в вакууме v_1:

$$n = v_2/v_1.$$

Траектория нейтрона испытывает преломление на границе среды потому, что на него со стороны границы действует сила отталкивания или притяжения. Это действие можно учесть, введя скачок U потенциальной энергии нейтрона при пересечении границы среды:

$$\frac{mv_1^2}{2} = \frac{mv_2^2}{2} + U,$$

где m — масса нейтрона. Следовательно,

$$n^2 = \frac{v_2^2}{v_1^2} = 1 - \frac{2U}{mv_1^2}.$$

Введя сюда длину нейтронной волны в вакууме $\lambda = h/(mv_1)$, получим

$$n^2 = 1 - \frac{2Um\lambda^2}{h^2}.$$

Для определения постоянной Um/h^2 замечаем, что при $|n-1| \ll 1$ последняя формула переходит в

$$n = 1 - \frac{Um\lambda^2}{h^2}.$$

Но в этом случае справедлива также формула (99.8). Из сравнения обеих формул находим $Um/h^2 = Nb/(2\pi)$, так что окончательно

$$n^2 = 1 - \frac{Nb\lambda^2}{\pi}. \tag{99.10}$$

10. При выполнении условия (99.9), как уже было отмечено, показатель преломления нейтронных волн весьма близок к единице. Но в зависимости от знака b он может быть как меньше (когда $b > 0$), так и больше единицы (когда $b < 0$). В этом одно из отличий нейтронных волн от рентгеновских, для которых всегда $n < 1$. Это объясняется тем, что частота рентгеновских волн всегда больше собственных частот колебаний электронов в атомах. Энергия же тепловых нейтронов может как превосходить энергии резонансных уровней ядер, так и быть меньше их. Если $n < 1$, то при достаточно малых углах скольжения нейтронная волна, падая из вакуума, испытывает полное отражение от границы среды.

Более интересен случай ультрахолодных нейтронов, когда выполняется условие

$$\frac{Nb\lambda^2}{\pi} > 1. \tag{99.11}$$

В этом случае, согласно формуле (99.10), показатель преломления n чисто мнимый. Следовательно, если нейтронная волна падает на поверхность среды из вакуума, то она может проникнуть в среду только на малую глубину, а затем полностью отразиться (см. т. IV, § 64, п. 2).

Условию (99.11) можно придать следующую форму:

$$\lambda > \lambda_{\text{гр}}, \tag{99.12}$$

где

$$\lambda_{\text{гр}} = \sqrt{\frac{\pi}{Nb}}. \tag{99.13}$$

Эта величина называется *граничной длиной волны нейтрона*, а соответствующие ей энергия и скорость

$$\mathscr{E}_{\text{гр}} = \frac{mv_{\text{гр}}^2}{2} = \frac{1}{2}\,\frac{Nbh^2}{\pi m}, \tag{99.14}$$

$$v_{\text{гр}} = \frac{h}{m\lambda_{\text{гр}}} \tag{99.15}$$

— *граничными энергией* и *скоростью нейтрона*. Эти величины приведены для некоторых материалов в табл. 19.

Таблица 19

Граничные энергия, длина волны и скорость нейтронов для некоторых материалов

Материал	$\mathscr{E}_{\text{гр}}$, 10^{-7} эВ	$\lambda_{\text{гр}}$, нм	$v_{\text{гр}}$, м/с
Алюминий	0,55	122	3,2
Магний	0,605	116	3,36
Медь	1,72	69	5,7
Углерод	1,94	65	6,1
Бериллий	2,40	58	6,8

11. Для получения ультрахолодных нейтронов используются ядерные реакторы. При делении ядер урана или плутония в ядерных реакторах получаются нейтроны со средней энергией в несколько мегаэлектронвольт. Они замедляются до тепловых энергий в веществе, окружающем ТВЭЛы с делящимся веществом. При этом возникают и ультрахолодные нейтроны. Однако их число составляет примерно только 10^{-13} часть от общего числа нейтронов (если температура замедлителя близка к комнатной). Ультрахолодные нейтроны выводятся из реактора по сильно изогнутой трубе — *нейтроноводу*. Все нейтроны, за исключением ультрахолодных, не в состоянии преодолеть резкие изгибы нейтроновода и поглощаются его стенками и окружающим веществом. Ультрахолодные же нейтроны испытывают полное отражение на границе вакуум — среда. Обычно вблизи активной зоны реактора в нейтроноводе помещается дополнительный замедлитель, называемый *конвертором*. Его полезно охлаждать для увеличения доли ультрахолодных нейтронов. Часто нейтроновод делают наклонным и поднимающимся вверх. Тогда нейтроны будут испытывать дополнительное замедление под действием силы тяжести.

Другой способ получения ультрахолодных нейтронов использует отражение нейтронов от движущегося («убегающего») магнитного зеркала (принцип работы последнего указан ниже). Удобство этого способа состоит в том, что в нем все приборы располагаются вне защиты реактора.

Указанные способы позволяют получать потоки ультрахолодных нейтронов с интенсивностью в сотни нейтронов в секунду с каждого

квадратного сантиметра поперечного сечения нейтроновода. Реально получаемые плотности ультрахолодных нейтронов не превосходят $1\ \text{см}^{-3}$.

12. На полном отражении ультрахолодных нейтронов, удовлетворяющих условию (99.12), основано устройство *ловушек ультрахолодных нейтронов*, принцип действия которых был указан Я. Б. Зельдовичем (1914–1987) в 1959 г. Ловушка представляет собой полый сосуд, в котором создан вакуум. Ультрахолодные нейтроны, удовлетворяющие условию (99.11) и находящиеся в таком сосуде, падая на его стенки, полностью отражаются обратно, каков бы ни был угол падения. Так действовала бы идеальная ловушка с идеально чистыми и гладкими стенками. В реальной ловушке стенки шероховаты, загрязнены, совершают колебания и т. д. Взаимодействие нейтронов со стенками приводит к их поглощению и нагреванию нейтронного газа. Особенно велико влияние пленки загрязнений, состоящей из водородсодержащих веществ, в частности воды. Все это понижает время хранения ультрахолодных нейтронов в сосуде. Все же к настоящему времени удалось довести это время до 1000 с, что практически совпадает со средним временем жизни нейтрона до β-распада: (917 ± 14) с.

Ультрахолодные нейтроны могут быть использованы в опытах по обнаружению у нейтрона возможного электрического заряда Q_{n} и электрического дипольного момента d_{n}, а также для изучения других фундаментальных свойств нейтрона. Опыты показали, что $Q_{\text{n}} < 2 \cdot 10^{-21} e$ (e — элементарный заряд), а $d_{\text{n}} < 2 \cdot 10^{-25} e$ (см).

13. Нейтроны обладают магнитным моментом $\mu = 6{,}03 \times \times 10^{-12}$ эВ/Гс. Поэтому нейтронные волны могут рассеиваться не только на ядрах, но и на самих атомах, если их электронные оболочки также обладают магнитными моментами. (Магнитные моменты ядер в тысячи раз меньше и поэтому роли не играют.) Это рассеяние, таким образом, обусловлено не ядерными, а *электромагнитными* силами. Оно может быть описано аналогично рассеянию на ядрах, а именно с помощью *амплитуды магнитного рассеяния нейтронов*. Для магнитных веществ (железо, кобальт, никель и пр.) ядерная и магнитная амплитуды рассеяния — одинакового порядка. Если магнитные моменты атомов в веществе ориентированы хаотически, то магнитное рассеяние атомов не когерентно. Но если их направления упорядочены, то получается когерентное рассеяние с дифракционными максимумами, положение которых определяется магнитной структурой вещества. В качестве примера на применение магнитной нейтронографии укажем, что именно она позволила установить магнитную структуру антиферромагнетиков.

Благодаря наличию магнитного момента ультрахолодные нейтроны могут испытывать полное отражение от областей сильного магнитного поля (магнитных зеркал). Для этого магнитный момент нейтрона должен быть ориентирован так, чтобы нейтрон выталкивался из магнитного поля и чтобы его кинетическая энергия $\mathscr{E}$ не превосходила μH. Если $\mathscr{E} = 10^{-7}$ эВ, то должно быть $H > \mathscr{E}/\mu = 1{,}7 \cdot 10^4$ Гс.

Полное отражение нейтронов от магнитных зеркал в магнитном поле позволяет получать интенсивные пучки *полностью поляризованных* нейтронов. Методы магнитной нейтронографии дают информацию об ориентации магнитных моментов каждого из компонентов в бинарных магнитных сплавах.

Созданы нейтронные ловушки — *нейтронные накопительные кольца*, использующие иной принцип удержания ультрахолодных нейтронов, чем ловушки, описанные выше. В таких кольцах нейтроны с энергией меньше $2 \cdot 10^{-6}$ эВ удерживаются на орбите с диаметром порядка 1 м неоднородным магнитным полем. Время удержания нейтрона в нейтронных накопительных кольцах сейчас уже превышает время жизни нейтрона до β-распада. Принцип удержания ультрахолодных нейтронов с помощью неоднородных магнитных полей был предложен в 1960 г. В. В. Владимирским (р. 1915).

Глава XV

НЕКОТОРЫЕ ВОПРОСЫ АСТРОФИЗИКИ

§ 100. Источники энергии звезд

1. Звезды излучают энергию за счет происходящих внутри них *термоядерных реакций.* Хотя эта мысль в общей форме давно высказывалась некоторыми учеными, но основанная на ней подробная количественная теория источников звездной энергии была развита Бете (р. 1906) только в 1939 г.

По современным представлениям звезды рождаются из протяженных газово-пылевых комплексов, состоящих преимущественно из водорода. Из-за гравитационной неустойчивости газово-пылевой комплекс распадается на множество более мелких частей — облаков. Каждое из этих облаков еще не является звездой. Но облако может превратиться в звезду, если масса его достаточно велика. Поэтому его называют *протозвездой.* В результате гравитационного сжатия протозвезда разогревается. Когда внутри протозвезды начинают происходить протон-протонные термоядерные реакции и дальнейшее гравитационное сжатие ее останавливается силами возросшего газово-кинетического давления, протозвезда и становится *звездой.*

2. Оценим среднюю температуру звезды к моменту ее образования из газово-пылевого облака. Очевидно, для этого достаточно знать среднюю кинетическую энергию теплового движения частиц звезды. Для простоты будем предполагать, что звезда состоит из водорода, который при высоких температурах в недрах звезды *полностью ионизован*, т. е. состоит из «голых» атомных ядер (протонов) и электронов. Энергию теплового движения эти частицы получают за счет гравитационной энергии, освобождающейся при сжатии звезды. Однако не вся освобождающаяся гравитационная энергия идет на нагревание звезды. Значительная часть ее тратится на излучение. Поэтому мы воспользуемся не законом сохранения энергии, а классической *теоремой вириала.*

Теорема вириала относится к поведению механической системы частиц, совершающей финитное движение. Если $\mathbf{r}_i$ — радиус-вектор i-й частицы, m_i — ее масса, а $\mathbf{F}_i$ — действующая на нее сила, то

$$\frac{d}{dt}\,(m_i\mathbf{r}_i\dot{\mathbf{r}}_i) = m_i\dot{\mathbf{r}}_i^2 + \mathbf{r}_i m_i\ddot{\mathbf{r}}_i = 2K_i + \mathbf{r}_i\mathbf{F}_i.$$

Просуммируем это соотношение по всем частицам системы и обозначим через K ее кинетическую энергию. Тогда

$$\frac{d}{dt}\sum_i m_i\mathbf{r}_i\dot{\mathbf{r}}_i = 2K + \sum_i \mathbf{r}_i\mathbf{F}_i.$$

Усреднив это равенство по физически бесконечно большому промежутку времени T, получим

$$\frac{1}{T}\left[\sum(m_i\mathbf{r}_i\dot{\mathbf{r}}_i)_{t=T}-\sum(m_i\mathbf{r}_i\dot{\mathbf{r}}_i)_{t=0}\right]=2\overline{K}+\overline{\sum\mathbf{r}_i\mathbf{F}_i},$$

где черта означает усреднение по времени. При $T\to\infty$ ввиду ограниченности пространства, в котором движется система, левая часть обращается в нуль и в результате имеем

$$2\overline{K}+\overline{\sum\mathbf{r}_i\mathbf{F}_i}=0. \tag{100.1}$$

Это равенство и выражает теорему вириала (вириалом называется выражение $(1/2)\overline{\sum\mathbf{r}_i\mathbf{F}_i}$). Теорема вириала есть точное следствие ньютоновской классической механики, если только под $\mathbf{F}_i$ понимать полную силу, действующую на i-ю частицу. Усреднение в теореме вириала понимается в смысле *усреднения по времени*, а в нашей задаче требуется *усреднение по совокупности частиц.* Однако если внешние условия в течение времени T не меняются, то средние значения в указанных двух смыслах совпадают между собой.

В нашей задаче теорему вириала следует, конечно, применять не к протозвезде, а к образовавшейся из нее *звезде.* Протозвезда подвергается гравитационному сжатию и, следовательно, находится в нестационарном состоянии. Для нее не имеет смысла говорить о средних величинах, о которых идет речь в теореме вириала. Только тогда, когда гравитационное сжатие будет остановлено возросшими силами газово-кинетического давления, т. е. когда протозвезда станет звездой, наступает стационарное состояние, в котором средние значения кинетической (тепловой) энергии беспорядочного движения частиц и потенциальной энергии их гравитационного притяжения, а также другие величины принимают определенные значения.

3. Вычислим теперь вириал для звезды, состоящей из равного числа протонов и электронов. Между этими частицами действуют кулоновские электрические силы. Однако при вычислении вириала эти силы учитывать не надо, так как звезда в целом *электрически нейтральна.* Действительно, рассмотрим какую-либо пару частиц i и j. В сумму $\sum\limits_i\mathbf{r}_i\mathbf{F}_i$ эта пара вносит слагаемое

$$\mathbf{r}_i\mathbf{F}_i+\mathbf{r}_j\mathbf{F}_j=(\mathbf{r}_i-\mathbf{r}_j)\mathbf{F}_{ij}=\mathbf{r}_{ij}\mathbf{F}_{ij},$$

где $\mathbf{F}_{ij}$ — сила, с которой частица j действует на частицу i, а $\mathbf{r}_{ij}$ — радиус-вектор, проведенный от частицы j к частице i. Если частицы одноименно заряжены, то $\mathbf{F}_{ij}$ будет силой отталкивания, а произведение $\mathbf{r}_{ij}\mathbf{F}_{ij}$ — величиной положительной. Напротив, для разноименно заряженных частиц произведение $\mathbf{r}_{ij}\mathbf{F}_{ij}$ отрицательно. Пусть теперь звезда содержит n протонов и n электронов. Число пар протонов и пар электронов, очевидно, равно $n(n-1)/2\approx n^2/2$ — всего n^2 пар одноименно заряженных частиц. Но таково же будет и число пар разноименно заряженных частиц. Поэтому полная сумма $\sum\mathbf{r}_i\mathbf{F}_i$, относящаяся к силам кулоновского взаимодействия, в среднем обратится в нуль.

Не надо учитывать и магнитные силы, если таковые имеются. Сила $\mathbf{F}$, действующая на частицу в магнитном поле $\mathbf{H}$, пропорциональна $[\mathbf{vH}]$, где $\mathbf{v}$ — скорость частицы. Для одной и той же частицы она с одинаковой вероятностью может быть направлена как в одну, так и в прямо противоположную сторону. Поэтому среднее значение скалярного произведения $(\mathbf{r}[\mathbf{vH}])$ равно нулю.

Единственными существенными силами, определяющими значение вириала звезды, являются *силы тяготения.* Это силы притяжения, а потому для них сумма $\sum \mathbf{r}_i\mathbf{F}_i$ будет отрицательна. Силы тяготения потенциальны. Сумма $\sum \mathbf{r}_i\mathbf{F}_i$ может быть выражена через потенциальную энергию гравитационного взаимодействия частиц системы $U = \sum U_{ij}$, где

$$U_{ij} = -\frac{G}{2}\frac{m_i m_j}{r_{ij}}. \tag{100.2}$$

Здесь m_i и m_j — массы частиц i и j, а r_{ij} — расстояние между ними. Выделим какие-либо две частицы с номерами i и j. Они вносят в сумму $\sum \mathbf{r}_i\mathbf{F}_i$ слагаемое $\mathbf{r}_i\mathbf{F}_{ij} + \mathbf{r}_j\mathbf{F}_{ji}$, где $\mathbf{F}_{ij}$, — сила, с которой частица j действует на частицу i. В силу равенства действия и противодействия $\mathbf{F}_{ji} = -\mathbf{F}_{ij}$, так что рассматриваемое слагаемое можно переписать в виде

$$\mathbf{r}_i\mathbf{F}_{ij} - \mathbf{r}_j\mathbf{F}_{ij} = (\mathbf{r}_i - \mathbf{r}_j)\mathbf{F}_{ij} = \mathbf{r}_{ij}\mathbf{F}_{ij}.$$

Но, очевидно, $\mathbf{F}_{ij} = -\operatorname{grad}_i U_{ij}$, где градиент берется по координатам частицы i в предположении, что частица j остается неподвижной. Следовательно,

$$\mathbf{r}_i\mathbf{F}_{ij} + \mathbf{r}_j\mathbf{F}_{ji} = -\mathbf{r}_{ij}\operatorname{grad}_i U_{ij} = -r_{ij}\frac{\partial U_{ij}}{\partial r_{ij}}.$$

Здесь у U_{ij} индексы можно опустить, если частную производную по r_{ij} брать в предположении, что меняется расстояние только между частицами i и j, а все остальные междучастичные расстояния остаются неизменными. Таким образом,

$$\sum \mathbf{r}_i\mathbf{F}_i = -\sum_{i,j} r_{ij}\frac{\partial U}{\partial r_{ij}}.$$

Но U есть однородная функция междучастичных расстояний r_{ij} степени -1. Поэтому в силу известной теоремы Эйлера об однородных функциях

$$\sum_{i,j} r_{ij}\frac{\partial U}{\partial r_{ij}} = -U.$$

В результате соотношение (100.1) переходит в

$$2\overline{K} + \overline{U} = 0. \tag{100.3}$$

С частным случаем этого соотношения мы уже встречались в механике (см. т. I, § 58). Это случай движения планеты вокруг Солнца (или искусственного спутника вокруг Земли) по круговой орбите. Только

в этом случае усреднения не требуется, поскольку сами величины K и U постоянны.

Сопоставим соотношение (100.3) с законом сохранения энергии. Гравитационная энергия, освобождающаяся в процессе сжатия протозвезды, расходуется не только на увеличение кинетической (тепловой) энергии K последней, но и тратится на электромагнитное и нейтринное излучения. Обозначим через $\mathscr{E}_{\text{изл}}$ полную энергию, унесенную излучением. Тогда

$$K + U + \mathscr{E}_{\text{изл}} = 0.$$

Усреднив это соотношение и вычитая его из (100.3), получим

$$\overline{K} = \overline{\mathscr{E}}_{\text{изл}}. \tag{100.4}$$

Таким образом, *половина гравитационной энергии, освобожденной при гравитационном сжатии протозвезды к моменту превращения ее в звезду, идет на увеличение кинетической (тепловой) энергии звезды, а другая половина уносится излучением.* Этот вывод имеет общее значение и не связан со специальным предположением, что звезда состоит только из водорода. Когда начнутся термоядерные реакции и наступит стационарное состояние, величины $\overline{K}$ и $\overline{U}$ будут оставаться неизменными. Тогда вся энергия, освобождающаяся при термоядерных реакциях, будет уноситься излучением.

4. Теперь мы подготовлены к тому, чтобы оценить среднюю температуру звезды $\overline{T}$. С этой целью обозначим через $m(r)$ массу звездного вещества внутри сферы радиусом r, центр которой совпадает с центром звезды. При падении на эту сферу из бесконечности массы dm выделяется гравитационная энергия $Gm\,dm/r$. Полная гравитационная энергия, освободившаяся при образовании звезды, выражается интегралом

$$G\int_0^M \frac{m}{r}\,dm,$$

где M — масса образовавшейся звезды. Как доказано выше, половина этой энергии идет на нагревание звезды. В дальнейшем, когда гравитационное сжатие прекратится, внутри звезды должна выделяться энергия в результате термоядерных реакций, чтобы поддержать температуру и излучение звезды на неизменном уровне. В результате тепловая энергия звезды K будет оставаться неизменной и выражаться половиной написанного выше интеграла. Этот интеграл можно было бы вычислить точно, если бы была известна плотность звездного вещества $\rho = \rho(r)$. Из-за незнания функции $\rho(r)$ точное вычисление мы вынуждены заменить оценкой. Очевидно,

$$\overline{K} = \frac{G}{2}\int_0^M \frac{m}{r}\,dm = \frac{GM^2}{4}\left\langle\frac{1}{r}\right\rangle,$$

или

$$\overline{K} = \frac{GM^2}{4R}\left\langle \frac{R}{r} \right\rangle, \tag{100.5}$$

где R — радиус звезды, $\langle R/r \rangle$ означает усредненное определенным образом значение R/r, а именно

$$\left\langle \frac{R}{r} \right\rangle = \frac{2}{M^2}\int\limits_0^M \frac{R}{r}\, m\, dm. \tag{100.6}$$

Мы занимаемся оценкой средней температуры не звезды вообще, а звезды, только что образовавшейся из газово-пылевого облака, состоящего практически только из полностью ионизованного водорода. К этому времени водород еще не успел «выгореть» в результате термоядерных реакций. Из-за высокой температуры к нему применима классическая статистика Больцмана, которая и используется в дальнейшем.

Средняя энергия теплового движения протона равна $(3/2)k\overline{T}$, где k — постоянная Больцмана. Такова же и средняя энергия электрона. Число протонов (а также электронов) в звезде составляет $M/m_{\text{р}}$, где $m_{\text{р}}$ — масса протона. Поэтому тепловая энергия всей звезды равна $3Mk\overline{T}/m_{\text{р}}$. Приравняв ее выражению (100.5), получим

$$\overline{T} = \frac{GMm_{\text{р}}}{12kR}\left\langle \frac{R}{r} \right\rangle. \tag{100.7}$$

Точное вычисление по формуле (100.7) требует знания плотности вещества звезды ρ в зависимости от расстояния r до ее центра. Только тогда можно найти среднее значение отношения R/r. Но так как $R/r > 1$, то во всяком случае должно быть

$$\overline{T} > \frac{GMm_{\text{р}}}{12kR}. \tag{100.8}$$

5. Можно указать и более точную оценку нижней границы для $\overline{T}$. Температура, стоящая в правой части формулы (100.8), получена в предположении, что $R/r = 1$. Такую температуру звезда получила бы, если бы звездное вещество конденсировалось только на ее поверхность. Эта температура заведомо ниже действительной температуры звезды, так как при дальнейшем перемещении вещества к ее центру производится дополнительная работа гравитационных сил, идущая на дальнейшее нагревание звезды. Дополнительную работу можно частично учесть, если предположить, что конденсация ограничивается образованием звезды постоянной плотности ρ. В таком случае $m = (4\pi/3)\rho r^3$, $dm = 4\pi\rho r^2\, dr$ и формула (100.6) дает $R/r = 6/5$. В результате получается более точная, но все еще заниженная оценка средней температуры звезды

$$\overline{T} > \frac{GMm_{\text{р}}}{10kR}. \tag{100.9}$$

Применим полученную оценку к Солнцу ($M_\odot = 2 \cdot 10^{33}$ г, $R = 7 \times \times 10^{10}$ см), точнее — к водородной звезде с такими же значениями массы и радиуса. Получим

$$\overline{T}_\odot > \frac{6{,}67 \cdot 10^{-8} \cdot 2 \cdot 10^{33} \cdot 1{,}67 \cdot 10^{-24}}{10 \cdot 1{,}38 \cdot 10^{-16} \cdot 7 \cdot 10^{10}} = 2{,}3 \cdot 10^6 \text{ K}.$$

Этот результат по порядку величины дает правильное, хотя и значительно заниженное значение средней температуры Солнца.

Оптическим методам доступна температура только поверхности Солнца. Она составляет около 6000 K. Однако в современных моделях Солнца масса наружной оболочки, в которой температура меньше 10^6 K, составляет всего около 1 % общей массы Солнца. Поэтому оболочка практически не сказывается на средней температуре Солнца.

6. Для точного вычисления температуры, как уже указывалось выше, надо знать плотность вещества в недрах Солнца. Но она также подлежит определению.

Вообще, точное вычисление температуры в недрах Солнца и звезд — не изолированная, а сложная комплексная задача. В нашем изложении речь шла не о точном вычислении, а о грубой оценке температуры. При строгой постановке должна быть определена не только температура, но и вся совокупность взаимно связанных параметров, характеризующих состояние звезды: давление, плотность, температура, химический состав, светимость звезды и пр. В частности, необходимо, чтобы при этих параметрах получилось равновесное состояние звезды. Все это находится в результате громоздкого численного интегрирования. При этом вместе с измеренными значениями массы и размеров Солнца используются уравнения сохранения и переноса энергии, уравнения гидродинамического, лучистого и конвективного равновесия, закон Стефана–Больцмана и пр. В основу расчета кладутся определенные модели Солнца. Согласованность результатов позволяет выбрать правдоподобную модель Солнца. Все расчеты теперь выполняются на ЭВМ методом проб и ошибок. В теории эволюции звезд такие же расчеты выполняются для моделей звезд с различными параметрами.

Не останавливаясь на этих вопросах, приведем данные, характеризующие Солнце на современном этапе его эволюции. Солнце состоит из водорода H, гелия He и остальных элементов. Относительные содержания их по массам в астрофизике принято обозначать соответственно через X, Y, Z. Для внешних слоев Солнца путем усреднения по различным моделям получено $X = 0{,}71$, $Y = 0{,}265$, $Z = 0{,}025$. Такие данные характерны и для всего Солнца на начальном этапе его эволюции. Но вблизи центра C Солнца в настоящее время $X_\text{C} = 0{,}38$. Температура, давление и плотность в центре Солнца равны соответственно $T_\text{C} = = 15 \cdot 10^6$ K, $\mathscr{P}_\text{C} = 3{,}4 \cdot 10^{17}$ дин/см^2, $\rho_\text{C} = 160$ г/см^3.

7. В проблеме источников энергии звезд основной интерес представляет не средняя температура, а температура в глубоких недрах звезд, так как именно там происходят термоядерные реакции.

Приведем одну из возможных оценок температуры в центре звезды, хотя она и обладает теми же принципиальными недостатками, что и приведенная выше оценка средней температуры звезды. Будем предполагать, что в звезде *нет конвекции*. Для этого необходимо, чтобы температурный градиент dT/dr был не меньше так называемого адиабатического температурного градиента (см. т. II, § 121). В реальной звезде из-за местного перегревания ее при термоядерных реакциях конвективное перемешивание, конечно, происходит и притом в некоторых звездах весьма интенсивно. Но мы рассматриваем идеальный случай, когда влияние этого перемешивания на отвод выделяющегося тепла из звезды не очень существенно. В пределе перемешивание исчезает как раз тогда, когда температурный градиент становится адиабатическим. Поэтому мы и принимаем, что в звезде устанавливается адиабатическое распределение температуры. Если еще звездное вещество считать идеальным газом, то должно быть

$$\frac{dT}{dr} = -\frac{g}{c_{\mathscr{P}}}, \tag{100.10}$$

где $c_{\mathscr{P}}$ — удельная теплоемкость при постоянном давлении, а g — ускорение свободного падения (см. т. II, § 121). Интегрирование этого уравнения дает

$$T(r) = T_{\mathrm{C}} - \int_0^r \frac{g}{c_{\mathscr{P}}}\, dr,$$

где T_{C} — температура в центре звезды. Температуру $T(R)$ на поверхности звезды можно принять равной нулю, так как она пренебрежимо мала по сравнению с T_{C}. В этом предположении

$$T_{\mathrm{C}} = \int_0^R \frac{g(r)}{c_{\mathscr{P}}}\, dr. \tag{100.11}$$

Если ввести массу вещества m внутрь сферы радиусом r, то $g = Gm/r^2$. Удельную теплоемкость $c_{\mathscr{P}}$ оценим в предположении, что звезда состоит из полностью ионизованного водорода. На каждую частицу (протон и электрон) приходятся средняя кинетическая энергия $(3/2)kT$ и теплоемкость при постоянном объеме $(3/2)k$, а при постоянном давлении $(5/2)k$. Число частиц (протонов + электронов) в единице массы равно $2 \cdot (1/m_{\mathrm{p}})$. Поэтому $c_{\mathscr{P}} = 5k/m_{\mathscr{P}}$. В результате

$$T_{\mathrm{C}} = \frac{Gm_{\mathrm{p}}}{5k} \int_0^R \frac{m(r)}{r^2}\, dr. \tag{100.12}$$

Если бы плотность ρ внутри звезды была постоянной, то $m(r) = (4\pi/3)\rho r^3$. В этом случае интеграл в (100.12) легко вычисляется.

Получается

$$T_{\text{С}} = \frac{GMm_{\text{р}}}{10kR},$$

что в точности равно средней температуре звезды, вычисленной в тех же предположениях. Тем не менее приведенная оценка температуры в центре звезды не лишена смысла. Дело в том, что при истинной зависимости плотности вещества ρ от радиуса r подынтегральное выражение в формуле (100.11) значительно быстрее возрастает к центру звезды, чем соответствующее подынтегральное выражение в (100.6). Следствием этого является интуитивно очевидное утверждение, что температура в центре звезды выше ее средней температуры.

Насколько существенно распределение плотности вещества в звезде влияет на температуру в ее центре, показывает следующий пример. По современной модели Солнца в сфере радиусом $r = R/2$ сосредоточено около 94 % полной массы. Если массой наружной оболочки пренебречь, то можно воспользоваться предыдущей формулой, заменив в ней радиус R вдвое меньшей величиной. Тогда получилось бы

$$T_{\text{С}} = \frac{GMm_{\text{р}}}{5kR},$$

что вдвое больше оценки, полученной ранее. На самом деле концентрация вещества во внутренних зонах Солнца приводит к еще большему повышению температуры $T_{\text{С}}$.

8. Итак, гравитационное сжатие разогревает внутренние недра звезды до температур порядка десяти миллионов кельвинов (1 кэВ) и выше. Этого достаточно, чтобы в недрах звезды начался синтез более тяжелых ядер из менее легких. Такой синтез и является источником энергии, излучаемой звездами. В основном это синтез более тяжелых элементов (преимущественно гелия) из водорода, так как по современным спектроскопическим данным Вселенная состоит на 70 % из водорода (по массе), 30 % из гелия и 1 % из остальных элементов (углерода, кислорода и пр.). В протозвезде начавшийся синтез идет недостаточно интенсивно, так что потери энергии на излучение в основном компенсируются гравитационным сжатием протозвезды. Когда же энергия синтеза достигает величины, достаточной для компенсации потерь энергии на излучение, гравитационное сжатие протозвезды прекращается. С этого момента протозвезда и становится звездой. В звезде гравитационные силы уравновешиваются возросшим газово-кинетическим и отчасти световым давлением.

Как показывают приведенные выше оценки, температура в недрах звезды при заданных размерах приблизительно пропорциональна ее массе M. Светимость же звезды L, т. е. полная излучаемая ею энергия в единицу времени, согласно теории, пропорциональна примерно M^3. Теоретические оценки показывают, что при $M \lesssim 0{,}1M_{\odot}$ ($M_{\odot}$ — масса Солнца) гравитационное сжатие недостаточно для достижения термоядерных температур. Вот почему процесс гравитационного сжатия всех планет Солнечной системы (включая Юпитер) не привел к образованию звезд.

9. В космических масштабах гравитация снимает основные трудности, которые надо преодолеть, чтобы практически осуществить управляемый термоядерной синтез. Громадное давление, создаваемое гравитацией, удерживает термоядерную плазму в недрах звезд. Слой же вещества громадной толщины, отделяющий горячую плазму в центральных областях звезды от холодной периферии, надежно обеспечивает ее термоизоляцию. Термоядерная энергия, освободившаяся в глубоких недрах звезды, переносится к ее периферии в основном посредством лучеиспускания. На своем пути излученная энергия поглощается и снова переизлучается с изменением спектрального состава. Это переизлучение происходит практически изотропно-равномерно во все стороны. Перенос излучения к периферии звезды напоминает диффузию и происходит сравнительно медленно. Как показывают расчеты, тепло, выделившееся в центре звезды, доходит до ее периферии за времена порядка миллиона лет.

10. *Основным процессом, в котором происходит освобождение термоядерной энергии в нормальных звездах, является превращение водорода в гелий.* При этом масса вещества уменьшается примерно на 0,7 % и освобождается энергия в соответствии с соотношением Эйнштейна $\mathscr{E} = mc^2$. Если бы Солнце состояло только из водорода и весь водород затем превратился в гелий, то масса Солнца уменьшилась бы примерно на $\Delta M = 0{,}007 M_\odot = 0{,}007 \cdot 2 \cdot 10^{33} = 1{,}4 \cdot 10^{31}$ г. При этом освободилась бы энергия $\Delta M c^2 \approx 1{,}26 \cdot 10^{52}$ эрг. При настоящем темпе излучения Солнца излучаемая им энергия составляет примерно $L_\odot = 3{,}83 \cdot 10^{33}$ эрг/с. Если бы этот темп сохранился в дальнейшем, то всей энергии выгоревшего водорода хватило бы на $(1{,}26 \cdot 10^{52}) : (3{,}83 \cdot 10^{33}) = 3{,}3 \cdot 10^{18}$ с $\approx 10^{11}$ лет.

Средняя интенсивность энерговыделения ε при термоядерных реакциях в типичных звездах по земным масштабам исключительно мала. Так, для Солнца $\varepsilon = L_\odot / M_\odot = (3{,}83 \cdot 10^{33}) : (2 \cdot 10^{33}) \approx 2$ эрг/(с · г). В результате жизнедеятельности человеческого организма выделяется в сутки примерно 3000 ккал $= 3 \cdot 10^6$ кал $= 12{,}5 \cdot 10^{13}$ эрг. Приняв массу человека равной 60 кг $= 6 \cdot 10^4$ г, найдем, что скорость выделения энергии в человеческом организме составляет около $2{,}4 \cdot 10^4$ эрг/(с · г). Это примерно в десять тысяч раз больше, чем для Солнца. Малость величины ε позволила выше оценить температуру в недрах звезды, полностью отвлекаясь от энерговыделения при ядерных реакциях. Однако благодаря громадной массе Солнца излучаемая им мощность очень велика ($3{,}83 \cdot 10^{33}$ эрг/с $= 3{,}83 \cdot 10^{26}$ Вт). Из-за излучения масса Солнца уменьшается примерно на 4 млн т в секунду.

11. Превращение водорода в гелий идет не непосредственно, а через ряд промежуточных реакций. Оно может выполняться двумя путями: 1) в *протонно-протонной* (pp) цепочке реакций, или *водородном цикле*; 2) в *углеродно-азотном* или *углеродном цикле.*

Водородный цикл начинается с реакции между двумя протонами, в результате которой образуются дейтрон, позитрон и нейтрино

(табл. 20). Эта реакция вызывается слабыми взаимодействиями, а потому идет чрезвычайно медленно; в земных условиях она непосредственно не наблюдалась. В недрах звезд кинетическая энергия сталкивающихся протонов недостаточна, чтобы преодолеть кулоновский потенциальный барьер между ними. Как правило, все столкновения между протонами происходят упруго. Только примерно одна стомил-

Таблица 20

Водородный цикл

Реакция	Энерговыделение, МэВ	Среднее время реакции
$p + p \to d + e^+ + \nu_e$	$2 \cdot 0{,}164 + (2 \cdot 0{,}257)$	$1{,}4 \cdot 10^{10}$ лет
$e^+ + e^- \to 2\gamma$	$2 \cdot 1{,}02$	—
$p + d \to {}^3He + \gamma$	$2 \cdot 5{,}49$	5,7 с
${}^3He + {}^3He \to {}^4He + 2p$	12,85	10^6 лет
Итого $4p \to {}^4He + 2e^+ + 2\nu$	$26{,}21 + (0{,}514)$	

лионная доля столкновений завершается реакцией туннельным способом. При этом за время столкновения (порядка 10^{-21} с) один протон должен превратиться в нейтрон с испусканием позитрона и нейтрино. Позитрон немедленно аннигилирует с электроном, а образовавшийся из n и p дейтрон очень быстро (в течение нескольких секунд) вступает в реакцию с одним из ближайших протонов с образованием ядра ^{3}He. В дальнейшем возможны три ветви ядерных реакций.

Первая ветвь — это реакция между двумя ядрами ^{3}He. Но так как в первых трех реакциях ядро ^{3}He получается только один раз, то в рассматриваемый вариант полного водородного цикла эта реакция должна входить дважды, что и отмечено множителем 2 во втором столбце таблицы. В итоге цикла четыре протона превращаются в ядро ^{4}He, два позитрона и два нейтрино. В таблице приведено энерговыделение в соответствующих реакциях, а также примерное среднее время каждой реакции, рассчитанное для условий в центре Солнца. В скобках указана доля выделяющейся энергии, безвозвратно уносимая нейтрино.

При достаточно больших концентрациях ^{4}He и температурах $T > (10\text{–}15) \cdot 10^6$ K в полном энерговыделении начинает преобладать вторая ветвь водородного цикла. В этом варианте первые три реакции такие же, как и в предыдущем, но не повторяются дважды, а реакция $^3\mathrm{He} + {}^3\mathrm{He}$ заменяется на цепочку

$$^3\mathrm{He} + {}^4\mathrm{He} \to {}^7\mathrm{Be} + \gamma,$$

$$^7\mathrm{Be} + e^- \to {}^7\mathrm{Li} + \nu,$$

$$p + {}^7\mathrm{Li} \to 2\,{}^4\mathrm{He}.$$

При еще более высоких температурах преобладающим в энерговыделении является водородный цикл с завершающей цепочкой

$$^3\mathrm{He} + {}^4\mathrm{He} \to {}^7\mathrm{Be} + \gamma, \quad \mathrm{p} + {}^7\mathrm{Be} \to {}^8\mathrm{B} + \gamma,$$

$$^8\mathrm{B} \to {}^8\mathrm{Be} + \mathrm{e}^+ + \nu, \quad {}^8\mathrm{Be} \to 2\,{}^4\mathrm{He}.$$

В обоих случаях основным итогом снова является превращение четырех протонов в ядро ^{4}He.

12. В табл. 21 приведен углеродный (C–N) цикл. В нем, как и в водородном цикле, освобождается энергия 26,7 МэВ, причем около 6,8 % этой энергии уносится нейтрино. Характерной особенностью углеродного цикла является воспроизводство углерода ^{12}C, и притом в таком же количестве, какое было использовано в начале цикла. Углерод ^{12}C не затрачивается, а выполняет роль *катализатора*, обеспечивающего превращение водорода в гелий. Для Солнца и менее ярких звезд преобладающим является водородный, а для более ярких — углеродный цикл.

Сказанное выше в основном относится к нормальным звездам, или к звездам главной последовательности, к которым относится и Солнце. В этих звездах энерговыделение происходит главным образом за счет превращения водорода в гелий.

Таблица 21

Углеродный цикл

Реакция	Энерговыделение, МэВ	Среднее время реакции
$\mathrm{p} + {}^{12}\mathrm{C} \to {}^{13}\mathrm{N} + \gamma$	1,95	$1{,}3 \cdot 10^7$ лет
$^{13}\mathrm{N} \to {}^{13}\mathrm{C} + \mathrm{e}^+ + \nu_\mathrm{e}$	1,50 + (0,72)	7,0 мин
$\mathrm{p} + {}^{13}\mathrm{C} \to {}^{14}\mathrm{N} + \gamma$	7,54	$2{,}7 \cdot 10^6$ лет
$\mathrm{p} + {}^{14}\mathrm{N} \to {}^{15}\mathrm{O} + \gamma$	7,35	$3{,}3 \cdot 10^8$ лет
$^{15}\mathrm{O} \to {}^{15}\mathrm{N} + \mathrm{e}^+ + \nu_\mathrm{e}$	1,73 + (0,98)	82 с
$\mathrm{p} + {}^{15}\mathrm{N} \to {}^{12}\mathrm{C} + {}^4\mathrm{He}$	4,96	$1{,}1 \cdot 10^5$ лет
Итого $4\mathrm{p} \to {}^4\mathrm{He} + 2\mathrm{e}^+ + 2\nu_\mathrm{e}$	25,03 + (1,70)	

13. Для звезд-гигантов (см. следующий параграф) с плотными «выгоревшими» (не содержащими водорода) ядрами существенны *гелиевый* и *неоновый* циклы, протекающие при значительно более высоких температурах и плотностях, чем водородный и углеродный циклы. Основной реакцией гелиевого цикла, идущей начиная с температуры $200 \cdot 10^6$ K, является реакция

$$3\,{}^4\mathrm{He} \to {}^{12}\mathrm{C} + \gamma_1 + \gamma_2 + 7{,}3\ \text{МэВ}.$$

(Далее могут следовать реакции $^{12}\mathrm{C} + {}^4\mathrm{He} \to {}^{16}\mathrm{O} + \gamma$, $^{16}\mathrm{O} + {}^4\mathrm{He} \to {}^{20}\mathrm{Ne} + \gamma$.)

Если продукты реакций гелиевого цикла вступят в контакт с H, то осуществится неоновый (Ne–Na) цикл. В нем ядро ^{20}Ne выполняет

роль катализатора в процессе превращения H в He. Последовательность реакций здесь вполне аналогична углеродному (C–N) циклу (табл. 21), только ядра ^{12}C, ^{13}N, ^{13}C, ^{14}N, ^{15}O, ^{15}N заменяются соответственно ядрами ^{20}Ne, ^{21}Na, ^{21}Ne, ^{22}Na, ^{23}Na, ^{23}Mg. Мощность этого цикла как источника энергии относительно невелика.

ЗАДАЧА

Оценить нижний предел гравитационного давления $\mathscr{P}$ в центре звезды. Чему было бы равно это давление, если бы звезда была однородна по плотности? Провести численный расчет для Солнца.

Р е ш е н и е. Из уравнения гидростатики

$$\frac{\partial \mathscr{P}}{\partial r} = -\rho g$$

получаем

$$\mathscr{P}(r) = \mathscr{P}_{\mathrm{C}} - \int_0^r \rho g \, dr,$$

где $\mathscr{P}$ — давление в центре звезды. Оно найдется из условия, что на поверхности звезды $\mathscr{P}(R) = 0$. Это дает

$$\mathscr{P}_{\mathrm{C}} = \int_0^R \rho g \, dr = \frac{G}{4\pi} \int_0^M \frac{m \, dm}{r^4},$$

или

$$\mathscr{P}_{\mathrm{C}} = \frac{GM^2}{8\pi} \left\langle \frac{1}{r^4} \right\rangle, \tag{100.13}$$

где угловые скобки означают надлежащим образом выполненное усреднение. Из (100.13) следует

$$\mathscr{P}_{\mathrm{C}} > \frac{GM^2}{8\pi R^4}. \tag{100.14}$$

Для Солнца $\mathscr{P}_{\mathrm{C}} > 4{,}4 \cdot 10^{14}$ дин/см^2 $\approx 4{,}4 \cdot 10^8$ атм. Если бы Солнце было однородно, то $\mathscr{P}_{\mathrm{C}}$ было бы втрое больше, т. е. около $13{,}2 \cdot 10^8$ атм. В общем случае для звезды

$$\mathscr{P}_{\mathrm{C}} = \frac{\gamma M^2}{R^4}, \tag{100.15}$$

где γ — безразмерный коэффициент, зависящий только от закона изменения плотности ρ вдоль радиуса звезды.

Насколько существенно возрастание плотности к центру звезды, показывает следующий пример. Вообразим, что наружная оболочка Солнца с $r > R/2$ удалена, а масса оставшегося вещества не изменилась. Согласно современной модели Солнца оставшаяся масса равна $0{,}94 M_\odot$. Тогда нижний предел давления в центре не уменьшится, а увеличится в

$$\left(\frac{0{,}94 M_\odot}{M_\odot}\right)^2 \left(\frac{R}{R/2}\right)^4 \approx 14 \text{ раз}.$$

По давлению $\mathscr{P}_C$ можно было бы вычислить температуру T_C в центре звезды, пользуясь равенством $\mathscr{P}_C = 2n_C kT$, где $n_C = \rho_C/m_p$ — число протонов в единице объема в центре звезды. Однако это вычисление требует знания плотности вещества ρ_C в центре звезды.

§ 101. Некоторые сведения из астрономии

1. Рассмотрению вопроса об эволюции звезд следует предпослать некоторые сведения из астрономии.

В астрономии *светимостью звезды* L называют полное количество энергии, излучаемое ею за единицу времени. *Звездная величина* в астрономии определяет *блеск* звезды (а не геометрические ее размеры). Блеск звезды зависит от ее светимости, расстояния до звезды и от спектральной чувствительности прибора, применяемого для наблюдения звезды. Числовое значение звездной величины считается тем больше, чем слабее звезда. Так, звезды 1-й величины ярче звезд 2-й величины, звезды 2-й величины ярче звезд 3-й величины и т. д.

Блеск звезды принято выражать в логарифмической шкале звездных величин. Эта шкала строится так, чтобы при увеличении звездной величины m на 5 единиц соответствующая интенсивность наблюдаемого излучения I уменьшалась в 100 раз. Таким образом, если звездная величина m увеличивается в арифметической прогрессии, то соответствующая интенсивность наблюдаемого излучения уменьшается в геометрической прогрессии. Если в качестве знаменателя геометрической прогрессии взять число 10, то

$$I = I_0 \cdot 10^{-0,4m},$$

где I_0 — интенсивность наблюдаемого излучения звезды нулевой величины. Действительно, при $m = 5$ эта формула дает $I/I_0 = 10^{-2} = 1/100$, т. е. при увеличении звездной величины на 5 единиц интенсивность убывает в 100 раз, как это и требуется при построении шкалы звездных величии. Преобразуя (101.1), можем написать

$$I = I_0 \cdot (2{,}5119)^{-m}, \qquad (101.1)$$

т. е. при увеличении звездной величины на единицу интенсивность излучения уменьшается приблизительно в 2,5 раза.

Символ a^m означает, что звездная величина рассматриваемого небесного светила равна a. Звездам высокой светимости соответствуют отрицательные значения звездной величины. Например, 5^m соответствует звезде 5-й величины, символ -4^m — звезде величины -4 и т. п. Символ $-2^m{,}5$ означает, что звездная величина астрономического светила равна $-2{,}5$.

Самым крупным 6-метровым телескопом (СССР) на пределе возможностей можно сфотографировать небесные объекты звездной величины (порядка 24^m). Визуальному наблюдению с помощью того же телескопа доступны объекты примерно в 10^4 раз слабее. Впрочем, предельно доступная звездная величина сильно зависит от атмосферных условий.

Такие яркие звезды, как Альдебаран и Альтаир принадлежат к звездам 1-й величины; звездная величина Капеллы $-0^m,2$; Сириуса $-1^m,6$; Солнца $-26^m,74$. Так как различные звезды находятся на различных расстояниях от Земли, то видимые звездные величины не характеризуют их светимость. Для сравнения светимостей звезд вместо видимых вводят *абсолютные звездные величины.* Абсолютная звездная величина есть такая величина, которую получила бы звезда при наблюдении со стандартного расстояния 10 пк (1 пк = 3,2616 св. лет = $3{,}0857 \cdot 10^{18}$ см), если бы не было поглощения света в межзвездной среде. Абсолютная звездная величина Солнца составляет $+4^m,75$. На таком стандартном расстоянии Солнце казалось бы совсем слабенькой звездочкой, хотя и видимой невооруженным глазом. Сириус казался бы в 14 раз слабее, чем сейчас, и уже больше не выделялся бы на ночном небе как самая яркая звезда. Напротив, Ригель (β Ориона), светимость которого в 20 000 раз превосходит светимость Солнца, затмил бы на ночном небе все остальные объекты, за исключением Луны.

2. Приемник излучения обладает различной чувствительностью к различным участкам спектра. Например, красная звезда может казаться яркой для глаза, но слабой для фотопластинки. Так называемая *визуальная звездная величина* соответствует спектральной чувствительности человеческого глаза. Она практически совпадает с *фотовизуальной звездной величиной*, определяемой фотометрированием изображения, полученного на ортохроматической пластинке с применением желтого светофильтра. *Фотографической* или *синей* называется звездная величина, определяемая фотометрированием изображения на обычной фотопластинке или фотоэлектрическим способом с применением синего фильтра. Конечно, технология при изготовлении фотопластинок и светофильтров должна быть точно указана и строго выдержана. Разность между фотографической и визуальной звездными величинами называется *показателем цвета.*

Изложенный метод определения показателя цвета звезд и классификация звездных спектров, приводимая ниже, были разработаны в Гарвардской астрономической обсерватории (США). Позднее в эти методы в других обсерваториях мира были введены существенные усовершенствования, но для наших целей нет необходимости их рассматривать.

Температура, химический состав, степень ионизации атомов и прочие параметры наружных слоев звездных атмосфер, от которых доходит до нас световое излучение, проявляются в особенностях спектров звезд. В Гарвардской обсерватории эмпирически спектры звезд были подразделены на семь классов, которые обозначаются буквами O, B, A, F, G, K, M. Для указания более детальных подробностей, выявляемых фотопластинкой, эти спектральные классы подразделяются на подклассы, спектры которых являются промежуточными между спектрами соседних классов. Спектральные подклассы отмечаются номерами от 0 до 9, прибавляемыми к букве, обозначающей класс.

Спектры поглощения звездных атмосфер, типичные для семи указанных классов, приведены на рис. 173 (*а* — λ Цефея, O6; *б* — τ Скор-

пиона, B0; *в* — α Северной Короны, A0; *г* — δ Орла, F0; *д* — Солнце, G2; *е* — ε Лебедя, K0; *ж* — α Геркулеса *A*, M5). Даем для примера очень краткую характеристику некоторых спектральных классов.

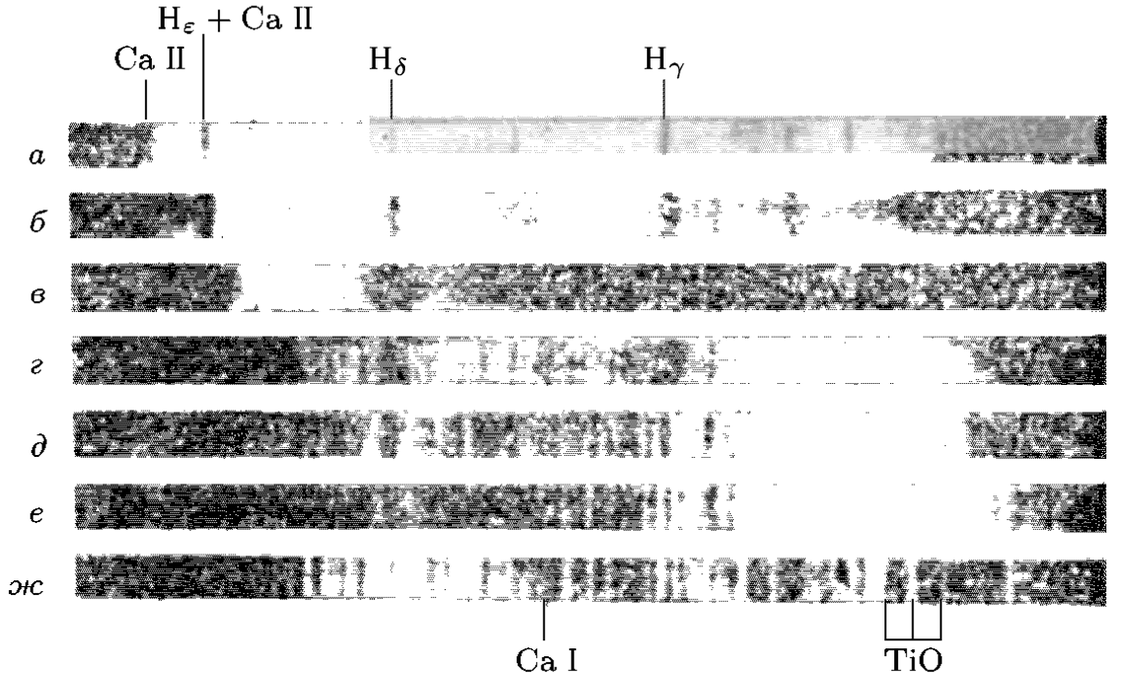


Рис. 173

Класс O. Температура $T \approx 50\,000$ K. Наблюдаются линии высокоионизованных атомов: He II, Si IV, N III и т. п. (Римские цифры I, II, III, ... относятся соответственно к нейтральному атому, к однократно ионизованному атому, к двукратно ионизованному и т. д.) Линии водорода слабы, иногда наблюдаются слабые эмиссионные линии He II и N III.

Класс B0. $T \approx 25\,000$ K. Линии He II отсутствуют; линии Mg II сильны (максимум у звезд B2); присутствуют линии Si III и O II; линии H сильнее, чем в классе O.

Класс A0. $T \approx 11\,000$ K. Линии водорода H достигают максимума; линии He I отсутствуют; линии Mg II и Si II сильны; линии Fe II, Cr II и другие слабы.

Класс F0. $T \approx 7600$ K. Линии H значительно слабее, чем в классе A0; линии Ca II сильны; линии Fe II, Cr II и другие примерно той же интенсивности, что и линии Fe I, Cr I и пр.

Класс G0. $T \approx 6000$ K. Яркость линий Ca II достигает максимума; атомный спектр очень богат благодаря многочисленным линиям Fe I и других нейтральных элементов; присутствуют молекулярные полосы CH.

Класс K0. $T \approx 5100$ K. Линии H относительно слабы; очень сильны линии нейтральных металлов; усиливаются молекулярные полосы.

Класс M0. $T \approx 3600$ K. Очень сильны линии нейтральных металлов; полосы TiO присутствуют, но довольно слабые.

Приведенная спектральная классификация используется для расположения звезд по цвету. Сначала идут горячие (голубые) звезды,

а в конце располагаются значительно более холодные красные звезды класса M. Например, яркие голубые звезды в созвездии Ориона относятся к классу B. Сириус — беловатая звезда класса A0. Солнце (желтая звезда) принадлежит к спектральному классу G2, Капелла — желтая звезда класса G0, Арктур — яркая оранжевая звезда класса K0, Бетельгейзе и Антарес — красные звезды класса M.

3. На большом статистическом материале датский астроном Герцшпрунг и американский астроном Рессел независимо друг от друга в начале нашего столетия эмпирически установили более или менее четко выраженную связь между спектральным классом (или показателем цвета) звезды и ее абсолютной величиной (или светимостью). Эта связь представляется графически на так называемой *диаграмме Герцшпрунга–Рессела* (рис. 174). Каждой точке на горизонтальной оси

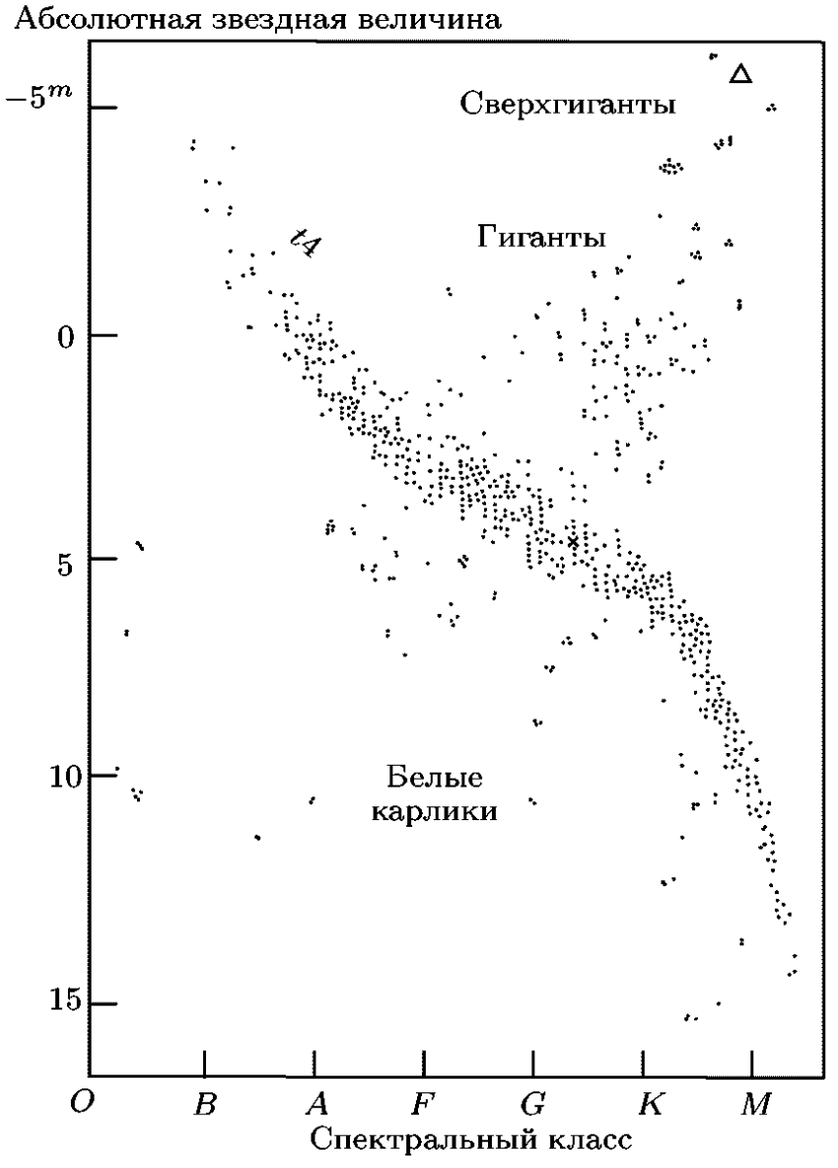


Рис. 174

этой диаграммы соответствует определенный спектральный класс (или показатель цвета). По существу, на горизонтальной оси отложена какая-то величина, характеризующая поверхностную температуру звезды. На вертикальной оси отложены соответствующие абсолютные величины

(или светимости) звезд. Звезда изображается точкой в плоскости диаграммы Герцшпрунга–Рессела.

Большинство звезд располагаются в полосе, идущей из верхнего левого угла в нижний правый угол диаграммы Герцшпрунга–Рессела. Эта полоса называется *главной последовательностью*, а соответствующие ей звезды — *звездами главной последовательности.* В верхней части главной последовательности располагаются наиболее массивные и горячие голубые звезды большой светимости, принадлежащие классам O и B, внизу — наименее массивные (*красные карлики*). Солнце (желтая звезда) — средняя звезда, расположенная где-то в центральной части главной последовательности. На диаграмме положение Солнца отмечено крестиком.

Справа вверху от главной последовательности расположены *красные гиганты* и *красные сверхгиганты.* Гиганты с поверхностной температурой порядка 4000 K примерно на 10 звездных величин ярче звезд главной последовательности с той же поверхностной температурой. Над гигантами располагаются очень редкие звезды, называемые сверхгигантами. Примером сверхгиганта может служить Бетельгейзе (α Ориона) — переменная звезда класса M2–M3. Ее видимая визуальная величина меняется от $+0^m,4$ до $-1^m,3$, а абсолютная звездная величина составляет около -3^m. Она примерно на 16 звездных величин ярче, а ее радиус примерно в 1000 раз больше звезды класса M3 главной последовательности. По своему радиусу сверхгигант может быть даже в 10 раз больше гиганта и иметь в 100 раз большую светимость.

Слева внизу от главной последовательности располагаются *белые карлики* — слабые голубые или желтые звезды, которые лежат на диаграмме Герцшпрунга–Рессела на 9–10 звездных величин ниже главной последовательности. По массе белые карлики сравнимы с Солнцем, а по размерам — с Землей. Следовательно, плотность белых карликов чрезвычайно велика — порядка 10^5–10^6 г/см^3. Соответственно очень велико и ускорение свободного падения на поверхности этих звезд, а толщина атмосферы составляет всего несколько десятков сантиметров.

Для полноты укажем еще, что между гигантами и главной последовательностью параллельно ей идет полоса звезд с промежуточными параметрами, называемых *субгигантами.* Аналогично, ниже главной последовательности также параллельно ей располагаются *субкарлики* — звезды с промежуточными свойствами между белыми карликами и звездами главной последовательности.

Массы звезд заключены в пределах от 0,03 до $60M_\odot$. Светимость стационарных звезд лежит в интервале от 10^{-4} до $10^5 L_\odot$, а радиусы — от 10 км (нейтронные звезды) до $10^3 R_\odot$ (сверхгиганты).

§ 102. Краткие сведения об эволюции звезд

1. По двум причинам вопрос об эволюции звезд будет затронут весьма кратко и схематично. Во-первых, он относится к астрономии и астрофизике. Во-вторых, теория эволюции звезд весьма сложна и очень далека от завершения. Многие вопросы этой теории совсем

неясны или ясны недостаточно. Достаточно подробное и доступное изложение вопросов эволюции звезд можно найти, например, в книге И. С. Шкловского «Звезды, их рождение, жизнь и смерть» (М.: Наука, 1984).

2. В процессе гравитационного сжатия протозвезды температура ее недр непрерывно повышается. Когда она поднимется до миллиона Кельвинов, начинаются первые термоядерные реакции с участием протонов и легких ядер с низким кулоновским барьером (дейтерий, литий, бериллий, бор):

$$\mathrm{d} + \mathrm{p} \to {}^3\mathrm{He}, \quad {}^6\mathrm{Li} + \mathrm{p} \to {}^3\mathrm{He} + {}^4\mathrm{He},$$

$${}^7\mathrm{Li} + \mathrm{p} \to 2\,{}^4\mathrm{He}, \quad {}^{10}\mathrm{Be} + 2\mathrm{p} \to 3\,{}^4\mathrm{He}.$$

Но из-за малого содержания в протозвезде этих элементов продолжительность указанных реакций по звездным масштабам невелика — всего порядка миллиона лет.

В результате освобождения термоядерной энергии в указанных реакциях температура и давление в протозвезде повышаются. Возросшее давление замедляет гравитационное сжатие протозвезды. Когда легкие ядра с низким кулоновским барьером будут израсходованы, гравитационное сжатие протозвезды усилится, что поведет к дальнейшему повышению ее температуры. Наконец, когда температура достигнет порядка 10 млн кельвинов, вступят в игру протон-протонный, а за ним углеродно-азотный циклы, являющиеся основными источниками выделения термоядерной энергии. Возросшее давление звездной плазмы стабилизирует звезду — ее гравитационное сжатие приостанавливается. С этого момента протозвезда и становится звездой, вступающей на главную последовательность. Место вступления ее на главную последовательность определяется первоначальной массой протозвезды. Чем больше эта масса, тем выше оказывается протозвезда на главной последовательности.

Ввиду того что водорода в звезде много — она почти целиком состоит из водорода, — «выгорание» водорода охватывает несравненно более длительное время, чем «выгорание» дейтерия, лития, бериллия и бора. Звезды в верхней части главной последовательности более массивны и обладают более высокими температурами. Поэтому в них процесс «выгорания» водорода происходит более быстро и охватывает время порядка ста миллионов лет. Напротив, у менее массивных и менее горячих звезд, как Солнце, этот процесс длится 10–15 млрд лет. Еще более медленно происходит эволюция звезд, находящихся в нижней части главной последовательности, — красных карликов.

Выйдя на главную последовательность, звезда длительное время остается там, почти не меняя своих свойств. Поэтому основная часть звезд наблюдается именно на главной последовательности. Однако по мере «выгорания» водорода состояние звезды все же неуклонно, но очень медленно изменяется, так что изображающая звезду точка на диаграмме Герцшпрунга–Рессела перемещается и может уйти с главной последовательности.

3. Характер эволюции звезды существенно зависит от того, происходит ли в ее недрах перемешивание вещества или нет. Вычисления на моделях показывают, что если бы эволюция звезды сопровождалась полным перемешиванием (вследствие чего ее химический состав оставался бы однородным по всему объему звезды), то изображающая ее точка на диаграмме Герцшпрунга–Рессела перемещалась бы от главной последовательности влево. Наоборот, при полном отсутствии перемешивания изображающая точка перемещалась бы вправо от главной последовательности. Беглый взгляд на диаграмму Герцшпрунга–Рессела показывает, что на ней нет звезд, расположенных вблизи главной последовательности вверху и слева. Зато имеется очень много звезд справа от нее — это красные гиганты и субгиганты. Следовательно, перемешивание вещества в недрах звезды в процессе смещения ее из главной последовательности играет лишь второстепенную роль. В массивных звездах, масса которых превышает примерно $3M_\odot$, перенос тепла в основном осуществляется лучеиспусканием. В звездах с массами, меньшими примерно $M_\odot$, он осуществляется главным образом конвекцией, т. е. перемешиванием вещества.

Эволюция звезд, масса которых меньше приблизительно массы Солнца, происходит настолько медленно, что время пребывания таких звезд в пределах главной последовательности превышает возраст Галактики (около $1{,}2 \cdot 10^{10}$ лет). В звездах с малыми массами (меньшими $0{,}3M_\odot$) преобладающим механизмом теплопередачи является конвекция. У таких звезд температура в центре недостаточно высока, чтобы водородный цикл работал полностью. Он обрывается на образовании изотопа гелия ^{3}He, а изотоп ^{4}He уже не синтезируется. При несколько больших массах (но меньших примерно $3M_\odot$) эволюция звезды заканчивается на образовании изотопа ^{4}He. Основной интерес представляет эволюция звезд, первоначальная масса которых превышает примерно $3M_\odot$. На ней мы и сосредоточим внимание.

4. Энерговыделение в звезде после ее образования и выхода на главную последовательность длительное время происходит преимущественно в центральной ее части за счет «выгорания» водорода. После выгорания водорода в центральной части звезды образуется центральное гелиевое ядро и окружающая его оболочка, богатая водородом. Звезда по своему составу становится сильно неоднородной. Ядерные реакции в ядре звезды прекращаются, но интенсивно происходят в ее оболочке. Таким образом, основное энерговыделение из центральной части переносится в оболочку. В результате оболочка сильно раздувается. Звезда сходит с главной последовательности и превращается в красный гигант или сверхгигант. Размеры звезды сильно увеличиваются (см. § 101, п. 3). Однако массы красных гигантов могут и не намного превышать массу Солнца. Благодаря большой поверхности светимость красного гиганта очень велика, несмотря на пониженную температуру его поверхности.

По мере «выгорания» водорода в оболочке звезды масса центрального гелиевого ядра возрастает, а его гравитационное сжатие усиливается. Это приводит к возрастанию плотности и температуры гелие-

вого ядра звезды. Когда эти величины достигают значений примерно 10^6 г/см^3 и 10^8 К соответственно, начинает эффективно идти реакция «сгорания» гелия. На короткое мгновение (порядка 10^{-16} с) создается неустойчивое ядро бериллия:

$$^4\mathrm{He} + {}^4\mathrm{He} \rightleftarrows {}^8\mathrm{Be} - 95 \text{ кэВ}. \tag{102.1}$$

Это ядро может тут же распасться на две α-частицы, но может успеть захватить еще одну α-частицу, в результате чего образуется ядро углерода в возбужденном состоянии:

$$^8\mathrm{Be} + {}^4\mathrm{He} + 363 \text{ кэВ} \to {}^{12}\mathrm{C}^*. \tag{102.2}$$

Возбуждение снимается γ-излучением:

$$^{12}\mathrm{C}^* \to {}^{12}\mathrm{C} + \gamma + 7{,}68 \text{ МэВ}. \tag{102.3}$$

В результате синтеза ядра ^{12}C из трех ядер ^{4}He освобождается энергия 7,22 МэВ. Скорость генерации энергии в звезде при этой реакции пропорциональна примерно $(T/10^8)^{30}$. Значит, «сгорание» гелия в этой реакции эффективно лишь при температуре, превышающей 10^8 К.

Когда накопится достаточное количество изотопа углерода ^{12}C, гелий начинает «выгорать» также в реакциях

$$\begin{gathered} {}^{12}\mathrm{C} + {}^4\mathrm{He} \to {}^{16}\mathrm{O} + \gamma, \\ {}^{16}\mathrm{O} + {}^4\mathrm{He} \to {}^{20}\mathrm{Ne} + \gamma, \\ {}^{20}\mathrm{Ne} + {}^4\mathrm{He} \to {}^{24}\mathrm{Mg} + \gamma, \end{gathered} \tag{102.4}$$

сопровождающихся дальнейшим и, наконец, полным исчерпанием гелия. Эти реакции требуют все более и более высокой температуры порядка 10^8–$1{,}5\cdot10^8$ К. В результате завершения этих реакций получается ядро звезды, состоящее из изотопов углерода ^{12}C, кислорода ^{16}O, неона ^{20}Ne и магния ^{24}Mg.

5. Все ядерные реакции в недрах Солнца и звезд недоступны непосредственному наблюдению. Существует лишь единственный косвенный метод наблюдения реакций, происходящих при участии слабого взаимодействия. В таких реакциях образуются нейтрино, а они практически не поглощаются, пронизывая всю толщу звездного вещества. Как указывалось в § 74, для регистрации солнечных нейтрино Б. М. Понтекорво предложил воспользоваться реакцией

$$^{37}_{17}\mathrm{Cl} + \nu_\mathrm{e} \to {}^{37}_{18}\mathrm{Ar} + \mathrm{e}^-. \tag{102.5}$$

Эта реакция имеет порог 0,81 МэВ и в принципе может быть использована для регистрации солнечных нейтрино почти всех энергий. Однако реально в реакции (102.5) ядро $^{37}_{18}$Ar с подавляющей вероятностью получается в возбужденном состоянии с энергией возбуждения примерно 5 МэВ. Это увеличивает эффективный порог реакции до 5,8 МэВ. Поэтому реакция (102.5) может быть использована практически для регистрации быстрых нейтрино. Такие нейтрино на Солнце получаются

только в одной из завершающих реакций водородного цикла, а именно

$$^8\mathrm{B} \to {}^8\mathrm{Be} + \mathrm{e}^+ + \nu_\mathrm{e} \quad (\mathscr{E}_\nu \leqslant 14{,}1 \text{ МэВ}) \tag{102.6}$$

(см. предыдущий параграф, п. 10). Как указывалось в § 74, такие «борные» нейтрино были зафиксированы в опытах Дэвиса. Однако их число оказалось примерно в 3–4 раза меньше наинизшего теоретического предела, найденного на основе современных моделей Солнца. Причина этого расхождения еще не ясна, хотя гипотезы для его объяснения и предлагались.

6. Перейдем к рассмотрению завершающих стадий эволюции звезды. Как уже было сказано, в начальной стадии гравитационное сжатие звезды прекращается в результате освобождения энергии при ядерных реакциях. Возникает вопрос, существуют ли силы, способные приостановить гравитационное сжатие, когда источники ядерной энергии будут исчерпаны? Единственными силами, противостоящими гравитационным, тогда окажутся силы давления звездного вещества.

При высоких плотностях звездного вещества последнее существует только в форме «голых» атомных ядер и электронов. Давление газа создается движением электронов и атомных ядер. Но это есть не только тепловое движение, исчезающее при абсолютном нуле температуры. На него накладывается еще *квантовое движение*, не прекращающееся и при абсолютном нуле. Будем рассматривать звездную плазму как идеальный газ, состоящий из электронов и положительно заряженных ядер, компенсирующих заряд электронов (газ из одних только электронов был бы вообще неустойчив). Отвлечемся от атомных ядер и сосредоточим внимание только на электронах.

Обозначим через n число электронов в единице объема. Пусть газ находится при абсолютном нуле температуры. Движение электронов квантовано. Число возможных квантовых состояний с импульсом меньше $p_{\text{макс}}$ найдется, если объем октанта импульсного пространства $(1/8)(4\pi/3)p_{\text{макс}}^3$ разделить на объем квантовой фазовой ячейки h^3, т. е. это число пропорционально $p_{\text{макс}}^3$. Максимальному импульсу соответствует максимальный уровень энергии, называемый *уровнем Ферми.* Согласно принципу Паули в каждом квантовом состоянии может находиться не более двух электронов с противоположно направленными спинами. При абсолютном нуле температуры электроны заполнят все энергетические уровни ниже уровня Ферми. Уровни, лежащие выше уровня Ферми, окажутся свободными. Таким образом, число электронов в единице объема будет пропорционально $p_{\text{макс}}^3$ или $\overline{p^3}$. Давление электронного газа $\mathscr{P}_{\text{газ}}$ пропорционально $\overline{nvp} \propto \overline{vp^4}$, где v — скорость электрона (см. т. II, § 59). Если газ нерелятивистский, то $v \propto p$, и, следовательно,

$$\mathscr{P}_{\text{газ}} \propto \overline{p^5} \propto \overline{n^{5/3}} \propto \rho^{5/3} \quad (\text{нерелятив.}), \tag{102.7}$$

где ρ — плотность электронного газа.

Если же газ ультрарелятивистский, то скорость v можно считать постоянной и равной скорости света c. В этом случае $\mathscr{P}_{\text{газ}} \propto \rho^{4/3}$, т. е.

$$\mathscr{P}_{\text{газ}} \propto \overline{n^{4/3}} \propto \rho^{4/3} \quad \text{(ультрарелятив.)}, \tag{102.8}$$

Электронный газ (как и газ любых ферми-частиц) при абсолютном нуле температуры называется *вырожденным*. Чем больше n, тем больше высоких энергетических уровней будет заполнено электронами. Пусть теперь температура T отлична от нуля. Тогда появится кинетическая энергия теплового движения электронов. Если она будет мала по сравнению с кинетической энергией квантового движения, то тепловым движением можно пренебречь и пользоваться формулами для вырожденного газа. Чем больше плотность газа, тем ближе он к состоянию вырождения. Так, плотность газа свободных электронов в металлах настолько велика, что при комнатных температурах и даже при температурах порядка 10 000 К он может считаться вырожденным.

7. Возникает сомнение в применимости к плотному вырожденному газу формул для идеального газа. Однако *вырожденный электронный газ тем ближе к идеальному, чем больше его плотность*. Это парадоксальное заключение легко понять. Идеальный газ — это такой газ, кинетическая энергия частиц которого настолько велика, что силами взаимодействия между ними можно пренебречь. Количественно это условие означает, что кинетическая энергия частиц газа очень велика по сравнению с потенциальной энергией их взаимодействия. Но это как раз и имеет место для электронного газа при высоких плотностях. Действительно, плотность кинетической энергии частиц нерелятивистского газа пропорциональна $n\overline{p^2} \propto \rho^{5/2}$. Плотность потенциальной энергии кулоновского отталкивания электронов пропорциональна ne^2/r, где r — среднее расстояние между электронами. Иными словами, потенциальная энергия пропорциональна $\rho^{4/2}$. Таким образом, с возрастанием ρ кинетическая энергия растет быстрее, чем потенциальная. Может показаться, что это не справедливо для ультрарелятивистского газа, так как плотность кинетической энергии частиц такого газа пропорциональна $\rho^{4/3}$, т. е. меняется с ρ так же, как и плотность потенциальной энергии отталкивания между электронами. Но это неверно, так как в приведенном рассмотрении учтено кулоновское взаимодействие только между электронами, но не приняты во внимание атомные ядра, наличие которых необходимо, чтобы сделать плазму квазинейтральной. Отталкивание между электронами в среднем компенсируется притяжением их к атомным ядрам. Поэтому с большой степенью приближения электронный газ в звездной плазме может рассматриваться как газ невзаимодействующих частиц.

Выше ничего не было сказано о давлении газа, состоящего из атомных ядер. Однако из-за больших масс скорости атомных ядер, а с ними и давление, производимое ядрами, малы по сравнению с теми же величинами для электронов. В звездах главной последовательности газ атомных ядер еще не вырожден. Только при экстремально высоких температурах и плотностях, когда атомные ядра релятивистские, их

давление становится сравнимым с электронным давлением. А в случае ультрарелятивистских ядер давление вообще не зависит от массы частиц, т. е. для атомных ядер становится таким же, как и для электронов.

Количественное рассмотрение вырожденного газа ферми-частиц дается в дополнении к этому параграфу.

8. Итак, при высоких плотностях развиваются громадные силы давления вырожденного электронного газа. Способно ли такое давление приостановить гравитационное сжатие звезды? Гравитационное давление в центре звезды

$$\mathscr{P}_g \sim \frac{M^2}{R^4} \sim M^{2/3}\rho^{4/3} \tag{102.9}$$

(см. задачу к предыдущему параграфу). Давление вырожденного нерелятивистского электронного газа возрастает с плотностью ρ быстрее, чем $\mathscr{P}_g$, а именно по формуле (102.7). Если бы электронный газ оставался нерелятивистским, то соответствующим выбором плотности ρ всегда можно было бы добиться, чтобы давление газа $\mathscr{P}_{\text{газ}}$ превышало гравитационное давление $\mathscr{P}_g$. Следовательно, в нерелятивистском случае давление электронного газа всегда было бы в состоянии стабилизировать звезду любой массы.

Но при очень высоких плотностях ($\rho \gg 2 \cdot 10^6$ г/см^3) электронный газ становится ультрарелятивистским, а в этом случае, как показывает формула (102.8), давление газа меняется с плотностью так же, как и гравитационное давление. Последнее, однако, пропорционально также $M^{2/3}$. Поэтому стабилизация звезды определяется значением ее массы. При массах, меньших некоторого критического значения $M_{\text{кр}}$, давление газа превышает гравитационное давление. В этом случае звезда может быть стабилизирована давлением вырожденного электронного газа.

Если же $M > M_{\text{кр}}$, то гравитационное давление начинает превышать давление электронного газа, и последнее уже не в состоянии приостановить гравитационное сжатие звезды.

Критическое значение массы $M_{\text{кр}}$ называется *чандрасекаровским пределом* по имени Чандрасекара (р. 1910), установившего существование такого предела. Чандрасекаровский предел определяется из условия, чтобы давление вырожденного электронного газа в центре звезды было равно гравитационному давлению. Численные расчеты показывают, что

$$M_{\text{кр}} \approx \frac{5{,}75}{\mu^2} M_\odot, \tag{102.10}$$

где μ — число нуклонов в ядре, приходящихся на один электрон атома. Если вещество состоит из относительно легких ядер (гелиевое ядро звезды состоит из ядер изотопа ^{4}He), для которых $N = Z$, то $\mu = 2$ и предыдущая формула переходит в

$$M_{\text{кр}} \approx 1{,}44 M_\odot. \tag{102.11}$$

С учетом вращения (а оно действует против тяготения), которое у белых карликов может быть очень быстрым, предельная масса повышается до $2M_\odot$.

Таким образом, при $M < M_{\text{кр}}$ звезда может быть стабилизирована давлением вырожденного электронного газа независимо от того, будет ли этот газ нерелятивистским или релятивистским. Такие звезды образуются из красных гигантов, плотные ядра которых в результате термоядерных реакций (101.6) сбрасывают свои оболочки и начинают вести самостоятельное существование. Это белые карлики. Они характеризуются малыми размерами, очень большими плотностями (порядка 10^6–10^7 г/см^3) и высокими температурами. Массы белых карликов порядка солнечной, а геометрические размеры порядка размеров Земли. Поскольку «ядерное горючее» в недрах белых карликов израсходовано, их излучение происходит за счет охлаждения. А так как поверхность белых карликов очень мала, то и светимость их также очень мала — в сотни и тысячи раз меньше солнечной. Поэтому остывание белых карликов происходит очень медленно и длится несколько миллиардов лет.

Вещество недр белых карликов отличается высокой прозрачностью и теплопроводностью. Высокая прозрачность связана с тем, что в силу принципа Паули все квантовые состояния электронов в недрах белого карлика заполнены. Поглощение же квантов света связано с переходом электронов из одного квантового состояния в другое. Между заполненными квантовыми состояниями эти переходы невозможны. Невозможно и поглощение световых квантов. Исключение составляют только квантовые состояния вблизи уровня Ферми, которые заполнены электронами не целиком. Поэтому только быстрые электроны вблизи уровня Ферми могут поглощать кванты излучения. Из-за вырождения электронного газа велика и теплопроводность вещества белого карлика, аналогично тому, что имеет место в металлах при обычных температурах. Вследствие высокой прозрачности и теплопроводности в недрах белого карлика не могут возникать большие перепады температуры. Перепад температуры происходит в тонком поверхностном слое белого карлика, где электронный газ не вырожден. Толщина этого слоя порядка 1 % радиуса карлика. В нем температура возрастает от нескольких тысяч кельвинов (на границе поверхности) до десятков миллионов, а затем вплоть до центра звезды почти не меняется.

9. При достаточно высоких плотностях равновесие звезды начинает нарушаться *процессом нейтронизации* звездного вещества. Как известно, при β^--распаде ядра часть энергии уносится электроном, а остальная часть — нейтрино. Энергия каждой из этих двух частей подвержена случайным вариациям, но их сумма остается постоянной. Эта постоянная величина называется *верхней границей β^--распада.* Так вот, если энергия Ферми превышает верхнюю границу β^--распада, то становится более и более вероятным процесс, обратный β^--распаду: ядро поглощает электрон, а его зарядовое число Z уменьшается на единицу — *электронный захват* (см. § 74). Вновь образовавшееся ядро может также захватить электрон, и т. д. В результате таких процессов

концентрация электронов в звезде уменьшается, уменьшается и давление вырожденного электронного газа, поддерживающее звезду в равновесии. Это ведет к дальнейшему гравитационному сжатию звезды, а с ним и к дальнейшему повышению средней и максимальной энергии электронов вырожденного газа — вероятность захвата электронов ядрами возрастает. Такие процессы могут привести к образованию нейтроноизбыточных ядер. В конце концов нейтронов может накопиться так много, что звезда будет состоять преимущественно из нейтронов. Такие звезды называются *нейтронными*. Нейтронная звезда, конечно, не может состоять из одних только нейтронов, так как необходимо давление электронного газа, чтобы предотвратить превращение нейтронов в протоны. В нейтронной звезде имеется небольшая примесь (около 1–2 %) электронов и протонов. Благодаря тому что нейтроны не испытывают кулоновского отталкивания, средняя плотность вещества внутри нейтронной звезды очень высока — примерно такая же, как в атомных ядрах. При такой плотности радиус нейтронной звезды с массой порядка $M_\odot$ примерно в 10^5 раз меньше солнечного, т. е. составляет не более 10 км. Нейтронная звезда сильно неоднородна и имеет сложную внутреннюю структуру. Но этого вопроса мы касаться не будем. Ограничимся замечанием, что упругость вырожденного нейтронного газа в звезде достаточна для противодействия гидростатическому давлению, если только масса звезды не превосходит определенного значения. Вопрос о верхнем пределе массы не вполне выяснен. Теоретические расчеты на моделях показывают, что верхний предел массы нейтронной звезды определяется оценочной формулой

$$M_{\text{пр}} \approx (2\text{–}3)M_\odot. \qquad (102.12)$$

При массе $M \gtrsim 1{,}2M_\odot$ нейтронные звезды образуются в результате *гравитационного коллапса* (быстрого сжатия) звезд. Начальная температура в центре образовавшейся нейтронной звезды очень высока (порядка 10^{11} К). Но уже через время порядка 10–100 с температура в центре понижается до 10^9 К за счет излучения нейтрино.

10. Существование нейтронных звезд было предсказано теоретически еще в 1938 г. Р. Оппенгеймером (1904–1967) и Г. Волковым. Однако из-за их ничтожной визуальной яркости их обнаружение выходило далеко за пределы возможностей оптических телескопов, даже самых мощных (ожидаемая видимая звездная величина нейтронной звезды, находящейся даже на минимальном расстоянии — порядка 10 световых лет, составляла всего 27^m). Положение изменилось после второй мировой войны, когда для исследования неба стали применять радиотелескопы. Начиная с 1964 г. на радиотелескопе Кембриджской лаборатории Кембриджского университета, возглавляемой Хьюишем, исследовались вариации радиоизлучения от дискретных космических источников. Летом 1967 г. аспирантка Хьюиша Жаклин Белл обратила внимание на один довольно необычный источник, посылающий, как выяснилось позднее, строго периодические радиоимпульсы. После этого в той же лаборатории, а позднее и в других лабораториях были открыты и исследованы аналогичные источники радиоизлучения,

которые были названы *пульсарами.* Периоды правильных пульсаций таких источников лежат в пределах 0,00154 до 3,75 с. До 1982 г. самым быстрым пульсаром с периодом 0,033 с считался пульсар, находящийся в Крабовидной туманности (см. п. 12). В 1982 г. был открыт пульсар с периодом 0,00154 с. Всего к концу 1987 г. было открыто 4 миллисекундных пульсара с периодами 0,00154; 0,003054; 0,0054; 0,006 с.

У пульсаров, расстояние до которых известно, можно подсчитать мощность радиоизлучения. У пульсара Крабовидной туманности она достигает 10^{35} эрг/с, что примерно в 25 раз больше мощности излучения Солнца ($3{,}83 \cdot 10^{33}$ эрг/с). Но есть и пульсары с мощностью излучения до 10^{38} эрг/с.

Спустя короткое время пульсары были отождествлены с предсказанными теоретически быстро вращающимися *нейтронными звездами.* Быстрое вращение нейтронной звезды связано с тем, что в силу закона сохранения момента импульса величина R^2/T должна сохраняться во времени. А так как из-за высокой проводимости плазменного вещества сохраняется и магнитный поток (см. т. III, § 71), то не должна меняться во времени и величина HR^2. Поэтому нейтронная звезда должна быть сильно намагничена. Например, при образовании нейтронной звезды радиусом $R = 10$ км из обычной звезды ($R \approx 10^6$ км) скорость вращения звезды и ее магнитное поле должны возрасти в 10^{10} раз. Но напряженность магнитного поля в некоторых пульсарах может достигать и 10^{12} Гс. Существенно, что направления магнитного поля и оси вращения, вообще говоря, не совпадают. С этими особенностями и связано радиоизлучение нейтронной звезды. Действительно, при несовпадении направлений оси вращения нейтронной звезды и ее магнитного момента последний должен вращаться, т. е. меняться во времени. А изменяющийся магнитный диполь излучает электромагнитные волны (см. т. III, § 141, задача 1). При этом излучение происходит преимущественно в направлении, перпендикулярном к магнитному моменту, и сосредоточено в пределах узкого конуса с углом раствора порядка 10°. Когда Земля попадает в пределы этого вращающегося конуса, обнаруживается максимум излучения. С этим и связан периодический импульсный характер излучения.

Благодаря потери энергии на излучение вращение нейтронной звезды непрерывно замедляется. Так, период пульсара, расположенного в Крабовидной туманности, регулярно увеличивается на $3{,}6 \cdot 10^{-8}$ секунд в сутки, или на $4{,}2 \cdot 10^{-13}$ секунд в секунду. У других пульсаров регулярное увеличение периода вращения заметно меньше. Но и у них относительное изменение периода вращения меньше чем 10^{-14}. Только самые лучшие кварцевые часы отличаются таким удивительным постоянством хода.

Нейтронная звезда может быть источником и рентгеновского излучения. Для этого вблизи нее должна находиться другая звезда, образующая вместе с ней двойную систему. Сплошное рентгеновское излучение возникает в результате *аккреции*, т. е. захвата вещества из другой звезды и его падения на нейтронную. Если масса нейтронной звезды

порядка солнечной, то при таком падении получается кинетическая энергия порядка 10^{20} эрг на грамм захваченного вещества (100 МэВ на нуклон). Для объяснения наблюдаемой интенсивности рентгеновского излучения (она того же порядка, что и интенсивность радиоизлучения) необходимо предположить, что масса вещества, захватываемая нейтронной звездой в год, составляет около 10^{-9} солнечных масс. По нескольким двойным системам оценена масса нейтронных звезд ($M \approx (1{,}4\text{–}3)M_\odot$).

Пульсар Крабовидной туманности является пока единственным пульсаром, от которого наблюдаются импульсы оптического излучения с периодом 0,033 с, т. е. с тем же периодом, что и импульсы рентгеновского излучения. Высокая интенсивность ультрафиолетовой части излучения указывает на то, что поверхность этой нейтронной звезды очень горяча. В спектре этой звезды отсутствуют какие бы то ни было спектральные линии как излучения, так и поглощения.

11. Если на конечной стадии эволюции звезды ее масса окажется больше предельной (102.12), то нерелятивистскими формулами пользоваться уже нельзя. В частности, нельзя пользоваться формулой (102.9), полученной на основании ньютоновской теории тяготения. В этом случае рассмотрение всех вопросов возможно только в рамках *общей теории относительности.* Поэтому мы должны довольствоваться лишь сообщением некоторых результатов.

Общая теория относительности приводит к выводу, что в рассматриваемом случае квантовомеханическое внутреннее давление вещества уже не в состоянии противостоять гравитационному давлению. Произойдет беспредельное катастрофическое сжатие звезды (гравитационный коллапс). Этот вывод не требует никаких предположений относительно конкретных свойств вещества (например, предположения, что вещество ведет себя как идеальный газ).

Для удаленного наблюдателя (т. е. наблюдателя, находящегося вне гравитационного поля звезды) радиус звезды асимптотически стремится к так называемому *гравитационному радиусу*

$$r_g = \frac{2GM}{c^2}. \tag{102.13}$$

Гравитационный радиус Солнца равен 2,95 км, Земли — 0,886 см. Сфера радиусом r_g называется *сферой Шварцшильда* (1873–1916), впервые получившего точное решение уравнений гравитации общей теории относительности для сферически-симметричного гравитационного поля. При приближении радиуса звезды к гравитационному радиусу скорость сжатия для удаленного наблюдателя бесконечно замедляется, так что звезда «застывает» в своем развитии. Излучение звезды по мере приближения ее радиуса к гравитационному становится все более слабым. В пределе звезда полностью изолируется от внешнего наблюдателя («самозамыкается»). Такая звезда называется *черной дырой.*

Никакие сигналы (свет, нейтрино и любые частицы), посланные в пределах сферы Шварцшильда, не могут выйти наружу и достигнуть внешнего наблюдателя. По этой причине сфера Шварцшильда

называется *горизонтом событий черной дыры*. Впрочем, удаленный наблюдатель никогда не заметит, что тело, падающее на черную дыру, пересечет горизонт событий, так как по его часам для этого требуется бесконечное время. По часам свободно падающего наблюдателя из-за замедления хода часов в гравитационном поле (см. т. IV, § 109) это время, конечно, определяется временем свободного падения. Дело в том, что при приближении к сфере Шварцшильда гравитационное поле стремится к бесконечности, как это показывает общая теория относительности.

Во внешнем пространстве наличие черной дыры проявляется в ее гравитационном поле, вращающем моменте и электрическом заряде, если сколлапсировавшая звезда была заряжена. На расстояниях, очень больших по сравнению с радиусом сферы Шварцшильда, гравитационное поле определяется законом всемирного тяготения Ньютона, а движение других тел, взаимодействующих с черной дырой, подчиняется законам ньютоновской механики.

Отметим одну особенность *вращающихся черных дыр*, которая может быть понята также на основе общей теории относительности. У вращающейся черной дыры вне горизонта существует область, называемая *экзосферой*, попав в которую вещество вовлекается во вращение вокруг черной дыры, что замедляет вращение и приводит к потере вращательной энергии черной дыры. Это может произойти, например, тогда, когда тело, попавшее в экзосферу, распадается на две части, одна из которых продолжает падать на черную дыру, а другая вылетает из экзосферы в направлении вращения. При этом, конечно, сохраняется полный вращающий момент системы, состоящей из черной дыры и влетевшего тела, но может произойти перераспределение энергии между ними. В результате часть энергии вращения черной дыры может быть унесена вылетевшей частью распавшегося тела.

Но из-за *квантовых эффектов* потеря энергии происходит и у невращающихся черных дыр, на что обратил внимание Хоукинг в 1974 г. Не имея возможности остановиться на этом вопросе, заметим, что Хоукинг показал, что невращающаяся черная дыра массой M излучает частицы (фотоны, нейтрино, гравитоны) с такой скоростью, как если бы она была черным телом, нагретым до температуры

$$T = \frac{\hbar c^3}{8\pi k G M} = \frac{6{,}2}{M/M_\odot}\, 10^{-8}\ \text{K}. \qquad (102.14)$$

Так, при $M = 3M_\odot$ эта формула дает $T \approx 2 \cdot 10^{-8}$ K. Таким образом, существует возможность уменьшения массы черной дыры путем «квантового испарения». Однако температуры (102.14) для черных дыр, возникающих в процессе коллапса звезд, слишком низки, и для таких черных дыр «квантовое испарение» не имеет значения. Время жизни этих черных дыр более чем на 50 порядков превышает возраст Вселенной. Для черных дыр с массой около 10^{15} г время жизни около 10^{10} лет. Но черные дыры столь малых масс при коллапсе звезд не образуются. Возможно, малые черные дыры возникают на *ранней стадии* эволюции Вселенной из первоначальных неоднородностей крайне

плотного расширяющегося вещества. Масса таких черных дыр по теории $M \approx c^3/\sqrt{\rho G^3}$, где ρ — плотность вещества. Квантовое испарение черных дыр существенно только для малых черных дыр, масса которых меньше или порядка 10^{15} г. Заключительный этап эволюции черной дыры такой массы заканчивается взрывом, при котором из дыры массой m порядка $3 \cdot 10^9$ г за последние 0,1 с выделяется громадная энергия — порядка $mc^2 \approx 10^{30}$ эрг!

12. Черные дыры с достоверностью еще не обнаружены. Если они существуют, то их наблюдение осложняется тем обстоятельством, что они могут проявлять себя *только своим гравитационным полем.* Изложим одну из идей, предложенных для обнаружения черных дыр. Если черная дыра образует двойную систему с обычной (видимой) звездой, то она — источник рентгеновского излучения, которое возникает при аккреции вещества обычной звезды на черную дыру. Но источниками рентгеновского излучения являются также белые карлики и нейтронные звезды, образующие двойные системы с обычными звездами. Поиски источников рентгеновского излучения нельзя проводить наземными приборами из-за поглощения рентгеновских лучей земной атмосферой. Для серьезных поисков должны применяться только рентгеновские телескопы, установленные на борту искусственных спутников Земли. Масса обычной (видимой) звезды может быть оценена по ее спектру, а доплеровское смещение спектральных линий при обращении видимой звезды вокруг центра масс ее и невидимого компонента позволяет определить период обращения. По этим данным, как показывается в астрономии, можно оценить и массу невидимой компоненты двойной звезды. Если эта масса окажется меньше $1{,}44M_\odot$, то невидимой компонентой будет белый карлик; если она больше этой величины, но меньше $3M_\odot$, то это — нейтронная звезда; если же она превосходит $3M_\odot$, то только черная дыра.

Рентгеновский телескоп был впервые установлен на борту итало-американского спутника «Ухуру», запущенного 12 декабря 1970 г. К весне 1972 г. спутник накопил информацию, достаточную для составления детального каталога 125 рентгеновских источников. Внимание астрономов привлек рентгеновский источник, расположенный в созвездии Лебедя и получивший название «Лебедь X-1». В результате исследований на спутнике «Ухуру» и с помощью рентгеновских телескопов на высотных баллонах удалось в 1971 г. с большой точностью установить положение источника. В том же году были зарегистрированы изменения рентгеновской светимости источника и наблюдалось резкое возрастание излучения радиоисточника, расположенного в том же месте неба. Естественно, что оба источника были отождествлены. Тем самым положение Лебедя X-1 было определено с точностью до угловой секунды. В том же месте была обнаружена горячая голубая спектрально-двойная звезда с периодом 5,6 сут, расположенная на расстоянии 2000 пк от Солнца. Оказалось, что и рентгеновское излучение Лебедя X-1 содержит компоненту с тем же периодом. На этом

основании было сделано заключение, что указанная звезда и рентгеновский источник образуют двойную систему. Масса звезды больше $20M_\odot$, а невидимой компоненты больше $8M_\odot$. Поэтому естественно предположить, что рентгеновским источником является черная дыра, хотя к этому предположению следует относиться с осторожностью.

13. Как показывают наблюдения, нейтронные звезды (пульсары), вероятно, чаще всего возникают при вспышках *сверхновых*. Вспышка сверхновой — это гигантский по своим масштабам взрыв звезды. В таком взрыве выбрасывается вещество с массой порядка солнечной и много больше. Средняя скорость выброшенного вещества порядка 10^9 см/с. В момент вспышки сверхновой светимость звезды возрастает в миллиарды раз, так что на короткое время (порядка месяца) звезда по своей яркости становится сравнимой с целой галактикой, состоящей из миллиардов звезд.

В течение нескольких месяцев взорвавшаяся звезда излучает столько света, сколько излучает Солнце за миллиард лет (около 10^{50} эрг). Но электромагнитное излучение уносит лишь малую часть полной энергии, освобожденной сверхновой. Кинетическая энергия разлетающегося вещества примерно в 10 раз больше. Еще большую энергию (примерно в 100 раз — по сравнению с электромагнитным излучением) уносят нейтрино, которые в основном испускаются во вспышке, длящейся около секунды. После взрыва большая часть массы звезды оказывается рассеянной и наблюдается в виде туманности. В центре звезды остается слабая звездочка — пульсар, которая в некоторых случаях может превратиться в черную дыру.

В каждой галактике сверхновые вспыхивают редко. За прошедшие 1000 лет в нашей Галактике наблюдались по крайней мере шесть сверхновых: в 1006, 1054, 1181, 1572, 1604 и 1667 гг. Особую роль в астрономии сыграла сверхновая 1054 г., появление которой было зафиксировано в китайских летописях. На месте взорвавшейся звезды в настоящее время наблюдается Крабовидная туманность, имеющая форму довольно правильного эллипса, угловые размеры которого приблизительно равны $4'' \times 6''$. В центре туманности наблюдается маленькая звездочка 16-й видимой звездной величины — пульсар, о котором уже говорилось в п. 8. Видимая звездная величина самой туманности $8^m,5$, т. е. она примерно в 1000 раз ярче, чем эта звездочка. В ближайшей к нам галактике — Большом Магеллановом облаке — последняя сверхновая вспыхнула в феврале 1987 г.

Если бы вспышки сверхновых наблюдались только «близко», т. е. в пределах нашей Галактики, то мы знали *бы* о сверхновых очень мало. Однако светимость сверхновых настолько велика, что они видны даже в удаленных галактиках. Так, в 1885 г. астроном Гартвиг на астрономической обсерватории в Тарту (Эстония) наблюдал вспышку сверхновой в крупнейшей из галактик — туманности Андромеды. Звезда имела видимый блеск $6^m,5$, т. е. люди с хорошим зрением могли видеть ее даже невооруженным глазом (а туманность Андромеды удалена от Земли на расстояние 670 тыс. парсек, т. е. свыше 2 млн световых лет!). Видимый блеск туманности Андромеды равен $4^m,5$, т. е. яркость сверхновой была

только в 6,25 раз меньше всей туманности Андромеды. В настоящее время астрономы открывают более 10 сверхновых в год.

Откуда берется энергия, необходимая для взрыва звезды при вспышке сверхновой? Считается, что одним из источников этой энергии является «выгорание» углерода в первой реакции (102.4). В этой реакции выделяется энергия около 1 МэВ $= 1{,}6 \cdot 10^{-6}$ эрг на нуклон, или 10^{18} эрг на грамм. Чтобы за счет углеродной реакции выделилась энергия 10^{50} эрг, необходимо «сжечь» взрывным образом массу углерода ${}^{12}_{6}\mathrm{C}$, равную $10^{50}/10^{18} = 10^{32}$ г, а если учесть еще кинетическую энергию разлетающегося вещества и энергию, уносимую нейтрино, то потребуется величина, примерно в 100 раз большая. А эта величина в несколько раз превышает массу Солнца.

Другим и притом более мощным источником энергии является гравитационное сжатие звезды. Гравитационная энергия, освобождающаяся при гравитационном сжатии звезды до размеров сферы Шварцшильда, составляет mc^2, т. е. 938 МэВ на нуклон. При сжатии же до размеров нейтронной звезды эта энергия равна 100 МэВ на нуклон. (Напомним, что энергия связи нуклона в ядре равна 8 МэВ.) Таким образом, при сжатии нейтронной звезды до размеров сферы Шварцшильда освобождается гравитационная энергия 800 МэВ на нуклон. Понятно, что этой энергии недостаточно, чтобы отбросить все вещество, упавшее на сферу Шварцшильда. Может быть сброшена только часть этого вещества.

Механизм взрывного освобождения ядерной и гравитационной энергии, а также механизм самого взрыва еще недостаточно ясны. Эти вопросы усиленно изучаются численными методами на моделях звезд с широким использованием ЭВМ. Интересующихся мы отсылаем к статье Бете и Брауна (В мире науки. — 1985, № 7).

14. В заключение не можем удержаться от искушения сделать следующее замечание. Еще в 1795 г. Лаплас (1749–1827) допускал существование излучающих, но невидимых звезд. Он исходил из ньютоновской корпускулярной гипотезы о природе света. Ньютоновская световая корпускула подчиняется законам ньютоновской механики и подвержена действию гравитационных сил. Обозначим через v скорость корпускулы, меняющуюся во время ее движения. В гравитационном поле звезды движение корпускулы будет финитным, если полная энергия ее отрицательна:

$$\frac{Pmv^2}{2} - \frac{GMm}{r} < 0.$$

Такая корпускула не дойдет до бесконечно удаленного наблюдателя, и звезда будет невидимой. Допустим, далее, что в момент испускания скорость корпускулы всегда равна одной и той же величине c. Испускание света происходит с поверхности звезды. Обозначим через r_0 ее радиус. Заменяя в предыдущем неравенстве r на r_0, а v — на c, получим

$$r_0 < \frac{2GM}{c^2}.$$

В предельном случае равенства эта формула совпадает с (102.13). Таким образом, если радиус звезды меньше гравитационного, то по Лапласу звезда для удаленного наблюдателя не будет видна.

Этому рассуждению, конечно, нельзя придавать серьезного значения, так как его исходные положения неверны. А сфера радиуса r_0 по своим свойствам в ньютоновской механике не имеет ничего общего со сферой того же радиуса (сферой Шварцшильда) в общей теории относительности. В общей теории относительности сфера Шварцшильда есть *горизонт событий*, и луч света не может выйти из нее наружу. В ньютоновской механике сфера радиуса r_0 этим свойством не обладает: луч света может выйти из этой сферы и дойти до наблюдателя, находящегося на конечном расстоянии, только он не может дойти до бесконечности. Далее, при падении частицы массы m из бесконечности на сферу Шварцшильда гравитационные силы в общей теории относительности совершают работу mc^2. В ньютоновской механике, как легко подсчитать, эта работа равна $mc^2/2$, т. е. вдвое меньше. Тем не менее, помимо исторического интереса, приведенное рассуждение заслуживает внимания еще и потому, что оно дает наглядный способ запоминания формулы для гравитационного радиуса.

Добавление

Идеальный газ Ферми при абсолютном нуле температуры

Согласно классическим представлениям при абсолютном нуле температуры движение частиц тела прекращается. Однако вблизи абсолютного нуля классическая физика неприменима. Там существенно проявляются *квантовые свойства вещества.* Поведение электронного и всякого газа из фермионов подчиняется квантовой статистике Ферми–Дирака. Рассмотрим поведение идеального ферми-газа в простейшем случае, когда абсолютная температура равна нулю. Нас будет интересовать в первую очередь электронный газ.

Пусть n — общее число частиц в единице объема газа. Найдем число частиц dn в единице объема с импульсами между p и $p+dp$. Число квантовых состояний в указанном интервале равно объему $4\pi p^2 dp$, приходящемуся на этот интервал импульсного пространства, деленному на объем одной квантовой ячейки h^3. При абсолютном нуле заполняются все низшие квантовые состояния, а все высшие остаются свободными. При этом, согласно принципу Паули, в каждом заполненном состоянии могут находиться два фермиона с противоположно направленными спинами. Таким образом, распределение числа частиц по заполненному пространству импульсов определяется формулой

$$dn = 2\,\frac{4\pi p^2 dp}{h^3}. \tag{102.15}$$

Согласно условию нормировки

$$\int dn = \frac{8\pi p_F^3}{3h^3} = n, \tag{102.16}$$

где p_F — максимальное значение импульса, которое может принимать частица при заданном n. Это значение называется *границей Ферми.* Таким

образом,

$$p_F = \sqrt[3]{\frac{3}{8\pi}}\, hn^{1/3} = 3{,}28 \cdot 10^{-27} n^{1/3}. \tag{102.17}$$

Уравнение (102.15) и дает распределение ферми-частиц в импульсном пространстве. В него не входит температура. Это и понятно, так как оно описывает распределение частиц в импульсном пространстве при абсолютном нуле температуры и в его ближайшей окрестности. Соответствующее состояние ферми-газа называется *вырожденным.* Таким образом, в вырожденном состоянии ферми-газа существует только один независимый макроскопический параметр, например плотность, которым определяются и все остальные параметры.

Граничная энергия $\mathscr{E}_F$, т. е. энергия, соответствующая импульсу $p = p_F$, связана с концентрацией газа соотношением

$$\begin{aligned} \mathscr{E}_F &= \frac{p_F^2}{2m} = \left(\sqrt[3]{\frac{3}{8\pi}}\right)^2 \frac{h^2}{2m}\, n^{2/3} \quad \text{при } \mathscr{E}_F \ll mc^2, \\ \mathscr{E}_F &= cp_F = \sqrt[3]{\frac{3}{8\pi}}\, hcn^{1/3} \qquad\quad \text{при } \mathscr{E}_F \gg mc^2. \end{aligned} \tag{102.18}$$

Для электронного газа

$$\begin{aligned} \mathscr{E}_F &= 0{,}584 \cdot 10^{-26} n^{2/3} \text{ эрг} = 3{,}65 \cdot 10^{-21} n^{2/3} \text{ МэВ} \quad \text{при } \mathscr{E}_F \ll mc^2, \\ \mathscr{E}_F &= 9{,}78 \cdot 10^{-17} n^{1/3} \text{ эрг} = 0{,}610 \cdot 10^{-10} n^{1/3} \text{ МэВ} \quad \text{при } \mathscr{E}_F \gg mc^2. \end{aligned} \tag{102.19}$$

В ультрарелятивистском случае, когда $mc^2 \ll \mathscr{E}_F$, а потому $\mathscr{E} = pc$, энергия Ферми $\mathscr{E}_F$ выражается формулами (102.18) и (102.19) одинаково для всех частиц, независимо от их массы. То же относится и ко всем другим соотношениям в ультрарелятивистском случае. В нерелятивистском приближении это не так. Так, в этом случае $\mathscr{E}_F$ обратно пропорциональна массе частицы m.

Газ может считаться вырожденным, когда его температура ниже так называемой температуры вырождения $T_{\text{выр}}$. Последняя определяется соотношением

$$kT_{\text{выр}} = \mathscr{E}_F. \tag{102.20}$$

Для электронов в нерелятивистском случае из формулы (102.19) получаем

$$T_{\text{выр}} = 0{,}423 n^{2/3} \cdot 10^{-10} \text{ К (электроны, } \mathscr{E}_F \ll mc^2). \tag{102.21}$$

В ультрарелятивистском случае для всех ферми-частиц (102.19) дает

$$T_{\text{выр}} = 0{,}708 n^{1/3} \text{ К } (\mathscr{E}_F \gg mc^2). \tag{102.22}$$

Для свободных электронов в металлах при обычных плотностях $n \sim 10^{22}$ температура вырождения $T_{\text{выр}} \approx 20\,000$ К. При плотностях $\rho \sim 10^6$ г/см^3 $T_{\text{выр}} \approx 2 \cdot 10^8$ К. В случае ультрарелятивистских частиц той же плотности $T_{\text{выр}} \approx 10^9$ К. Приведенные оценки показывают, что при температуре белых карликов электронный газ действительно может считаться вырожденным.

Рассчитаем теперь давление $\mathscr{P}_{\text{газ}}$ ферми-газа. В нерелятивистском случае

$$\mathscr{P}_{\text{газ}} = \frac{1}{3}\, nm\overline{v^2} = \frac{n\overline{p^2}}{3m} = \frac{1}{3m}\int\limits_0^{p_F} p^2\, dm = \frac{h^2}{15m}\left(\frac{3}{8\pi}\right)^{5/3} n^{5/3} \quad (\mathscr{E}_F \ll mc^2). \tag{102.23}$$

Здесь давление газа при заданной концентрации частиц n обратно пропорционально массе частицы m. Поэтому давлением газа нуклонов можно пренебречь по сравнению с давлением электронного газа. В релятивистском случае точное вычисление давления газа потребовало бы знания функции распределения не только при $p \ll p_F$, но и во всем интервале изменения импульсов. Такое вычисление довольно громоздко, и мы его приводить не будем Ограничимся ультрарелятивистским случаем, где все значительно упрощается. В пренебрежении энергией покоя связь между энергией и импульсом частицы принимает вид $\mathscr{E} = pc$, т. е. такой же, как и для фотона. Поэтому можно воспользоваться результатом вычисления давления для фотонного газа (см. т. II, § 61). Когда $\mathscr{E} \gg mc^2$, можно принять приближенно, что все частицы ультрарелятивистские. Тогда

$$\mathscr{P}_{\text{газ}} \approx \frac{1}{3}\, n\mathscr{E}_F = \frac{1}{3}\, ncp_F = \sqrt[3]{\frac{1}{72\pi}}\, hcn^{4/3} \quad (\mathscr{E}_F \geqslant mc^2). \tag{102.24}$$

Плотность газа $\rho = n_{\text{N}} m_{\text{N}}$, где n_{N} и m_{N} — концентрация и масса нуклонов, так как массой электронов можно пренебречь. Если μ — число нуклонов, приходящееся на один электрон, то $n_{\text{N}} = \mu n_{\text{e}}$, где n_{e} — концентрация электронов. Таким образом,

$$n_{\text{e}} \approx \frac{\rho}{\mu m_{\text{N}}}. \tag{102.25}$$

Для водородной плазмы $\mu = 1$. Если же вещество состоит из относительно легких ядер, то $N = Z$, а потому $\mu = 2$. Для вырожденного электронного газа легко получить

$$\begin{aligned} &\mathscr{P}_{\text{газ}} = 10^7 (\rho/\mu)^{5/3} \text{ дин/см}^2 && (\mathscr{E}_F \ll mc^2), \\ &\mathscr{P}_{\text{газ}} = 1{,}2 \cdot 10^9 (\rho/\mu)^{4/3} \text{ дин/см}^2 && (\mathscr{E}_F \gg mc^2). \end{aligned} \tag{102.26}$$

При $\mathscr{E}_F = mc^2$ плотность вырожденного электронного газа $\rho \approx 2 \times$ $\times\, 10^6$ г/см^3. Для ориентировки укажем, что при $\rho \approx 10^7$ г/см^3 давление вырожденного электронного газа $\mathscr{P}_{\text{газ}} \approx 10^{12}$ атм ($\mu \approx 2$).

§ 103. Космические лучи

1. Начало изучения космических лучей относится к августу 1912 г., когда австрийский физик Гесс (1883–1964) поднялся с электроскопами на воздушном шаре для исследования ионизации воздуха. О степени ионизации он судил по скорости разряда электроскопов. Чтобы чувствительность электроскопов не менялась с высотой, они были герметизированы, так что плотность воздуха в них оставалась постоянной. Еще до Гесса аналогичные опыты производились Гоккелем, поднимавшимся на высоту 4000 м над уровнем моря. Но результаты его опытов были

сомнительны и подверглись критике, так как он применял прибор, давление воздуха внутри которого изменялось при изменении наружного давления, а поправок на это он не вводил. Для того чтобы выяснить, насколько обоснованной была эта критика, Гесс, наряду с герметизированными электроскопами, захватил и негерметизированный электроскоп.

Еще к 1902 г. было установлено, что сухой воздух в герметически закрытом сосуде всегда слабо ионизован, даже после того, как совершенно распались газообразные радиоактивные вещества, содержащиеся в нем в ничтожных количествах. Если окружить сосуд свинцовой оболочкой с толщиной стенок около 2–3 см для защиты его от внешнего ионизующего излучения, то ионизация значительно уменьшается, но не исчезает полностью. Это заставляет предположить, что источники излучения, вызывающего ионизацию, находятся вне сосуда, в котором заключен воздух.

Наряду с другими предположениями было высказано и предположение, что ионизация вызывается излучением радиоактивных веществ, всегда имеющихся в небольших количествах в почве, горных породах, воде и пр. Если бы это было так, то ионизация воздуха убывала бы с высотой.

Для проверки этого заключения Гесс и совершил свой полет на воздушном шаре. Он обнаружил, что при подъеме на первоначальные 600 м над уровнем моря ионизация воздуха действительно убывала, хотя и медленнее, чем это ожидалось. Но начиная с высоты 600 м она стала возрастать — сначала медленно, а затем все быстрее и быстрее. На высоте 4800 м ионизация стала примерно в четыре раза больше, чем на уровне моря. Гесс пришел к заключению, что результаты его наблюдений лучше всего объясняются предположением, что на границу земной атмосферы из мирового пространства падает излучение очень большой проникающей способности.

Кольхерстер (1887–1946), многократно поднимавшийся на воздушных шарах, и другие ученые в опытах с электроскопами и ионизационными камерами подтвердили результаты наблюдений Гесса и обнаружили дальнейшее увеличение ионизации воздуха с высотой. На высоте 8400 м ионизация оказалась приблизительно в 10 раз больше, чем над уровнем моря. Тем самым было подтверждено и предположение Гесса о внеземном происхождении ионизующего излучения. Окончательно справедливость этого предположения была к концу 1926 г. доказана Милликеном (1868–1953), осуществившим в 1923–1926 г. серию опытов по поглощению такого ионизующего излучения. Падающее на Землю проникающее излучение, приходящее из космоса, было названо *космическими лучами.*

Дальнейшие опыты с запускаемыми на высоту шарами-зондами показали, что интенсивность потока космических лучей возрастает лишь до высоты около 20 км над уровнем моря. На этой высоте она достигает максимума, а при дальнейшем подъеме снижается. Это связано с тем,

что падающее на Землю космическое излучение взаимодействует с атмосферой, образуя множество вторичных частиц, о чем будет сказано в дальнейшем.

В дальнейшем, за немногими исключениями, мы опустим исторические и экспериментальные подробности, а ограничимся весьма кратким сообщением важнейших результатов из обширнейшей области физики космических лучей.

2. При исследовании природы космических лучей громадную роль сыграли методы регистрации заряженных ионизующих частиц с помощью камеры Вильсона, счетчиков Гейгера–Мюллера и толстослойных фотографических эмульсий. Последний метод, разработанный в середине 40-х годов, аналогичен методу камеры Вильсона. Заряженные ионизующие частицы, проходя через слой фотоэмульсии, сталкиваются с зернами бромистого серебра и взаимодействуют с ними, в результате чего в эмульсии возникает «скрытое изображение». После проявления в тех местах, через которые пролетели ионизующие частицы, в зернах восстанавливается металлическое серебро. При рассматривании в микроскоп в поле зрения видны траектории отдельных частиц, отмеченные темными зернами серебра.

Отметим, что начиная с 1924 г. Д. В. Скобельцын (р. 1892) провел длинный ряд важных исследований с камерой Вильсона, помещенной в сильное магнитное поле H. Траектория заряженной частицы в магнитном поле искривляется, но если поле постоянно, то ее кинетическая энергия, а с ней и релятивистская масса остаются постоянными. Если частица движется перпендикулярно к магнитному полю, то, приравнивая центробежную силу pv/R лоренцевой силе $(Ze/c)vH$, получаем

$$cp = ZeHR, \tag{103.1}$$

где Ze — заряд частицы, p — ее импульс, а R — радиус кривизны ее траектории. Произведение HR называется *магнитной жесткостью* частицы. Для ультрарелятивистской частицы кинетическая энергия $\mathscr{E} = pc$. Поэтому с помощью формулы (103.1) можно определить энергию и знак заряда ультрарелятивисткой частицы. Этим методом в 1929 г. Скобельцын нашел в космических лучах отрицательные частицы, энергия которых оказалась значительно больше энергии обычных β-частиц.

Отметим еще идею *схемы совпадений* на электронных лампах, предложенную Росси (р. 1905) и широко использованную им и другими исследователями космических лучей. Эта схема позволяет судить, в каком направлении прошла ионизующая частица, если только она заряжена. Схема совпадений содержит два или больше расположенных вдоль одной прямой счетчиков Гейгера–Мюллера G_1, G_2, G_3, ..., которые соединены с электронными лампами L_1, L_2, L_3, ..., как показано на рис. 175. Пока в счетчиках не произошел разряд, ток от положительного полюса батареи B, питающей лампы, через сопротивление R и электронные лампы проходит в землю. Если сопротивление R очень велико, то на нем происходит большое падение напряжения, так что потенциал в точке A будет близок к потенциалу земли. Когда в один из счетчиков

попадает ионизующая частица, она вызывает в нем электрический заряд, на сетку лампы, соединенной с этим счетчиком, поступает отрицательный импульс напряжения, так что эта лампа перестает пропускать ток. Однако лампы, соединенные с остальными счетчиками, в которых разряд не произошел, продолжают пропускать ток. При достаточно большом сопротивлении R ток на землю изменится очень мало. То же самое будет происходить, если частица пройдет через несколько, но не через все счетчики. Но положение изменится, если она пройдет одновременно через все счетчики. Тогда все лампы окажутся запертыми, и ток на землю прекратится. В результате потенциал в точке A резко возрастет до потенциала на положительном полюсе батареи. Это скачкообразное изменение потенциала в точке A можно измерить вольтметром или зафиксировать каким-либо другим прибором. Срабатывание этого прибора свидетельствует о том, что ионизующая частица прошла через все счетчики установки. После срабатывания схемы заряды в счетчиках Гейгера–Мюллера затухают, положительные заряды с сеток электронных ламп стекают в землю, и схема вновь готова для регистрации частиц. В схеме Росси лампы можно заменить транзисторами.

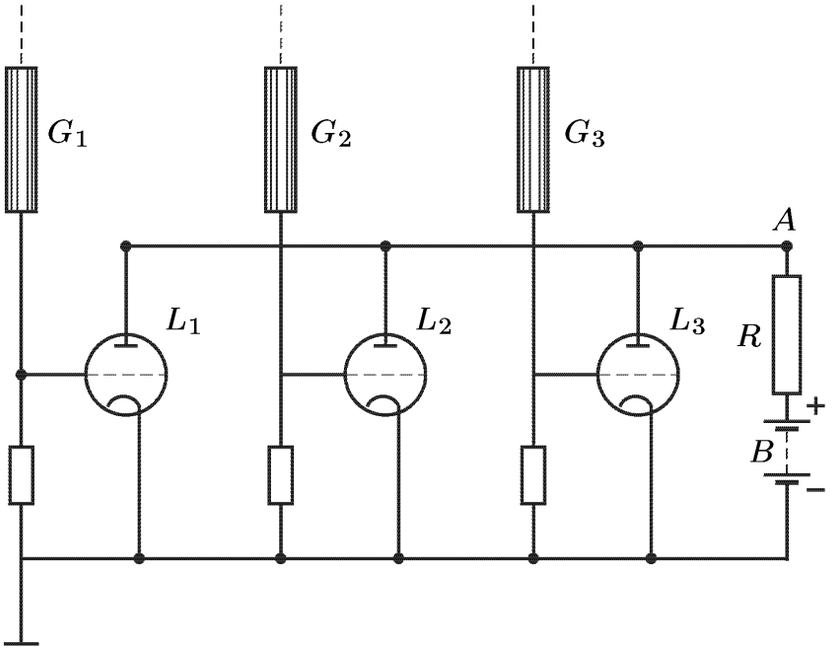


Рис. 175

Конечно, схема Росси и аналогичные ей схемы могут регистрировать не только события, вызываемые одной частицей, проходящей через все счетчики, но и такие события, которые вызываются частицами, проходящими друг за другом через слишком короткие промежутки времени. Чтобы этого не произошло, эти промежутки должны быть больше *разрешающего времени схемы*, определяющегося главным образом временем зажигания разряда в счетчиках Гейгера–Мюллера. Разрешающее время схемы Росси порядка микросекунды. Разработаны схемы совпадений, аналогичные схеме Росси, характеризующиеся меньшими разрешающими временами.

3. Космические лучи разделяются на *первичные* и *вторичные*. Первичные космические лучи падают на Землю из космоса. Проходя через толстый слой земной атмосферы (около 1000 г/см^2), они претерпевают сложную цепь превращений, в результате чего возникает новое излучение, которое и называется вторичными космическими лучами. По своему составу вторичные космические лучи не имеют ничего общего с первичными.

В околоземном космическом пространстве первичное космическое излучение состоит из *галактического космического излучения* и *солнечного космического излучения*. Первое приходит из удаленных объектов нашей звездной системы — Галактики — и других галактик, второе — от Солнца. Рассмотрим сначала первичное галактическое излучение.

Природа первичного космического излучения длительное время оставалась неясной. Только начиная с 40-х годов окончательно было установлено, что первичные космические лучи состоят из ядер тех же химических элементов, которые входят и в состав Земли, главным образом из ядер водорода. Частицы космических лучей обладают громадными кинетическими энергиями по сравнению с частицами обычного вещества, даже нагретого до миллиардов кельвинов. Энергии самых быстрых частиц космических лучей столь велики, что они в этом отношении долго будут оставаться вне конкуренции с самыми быстрыми частицами, ускоряемыми на современных ускорителях. Основная масса первичных космических лучей имеет энергии от 10^9 эВ (1 ГэВ) и выше, что в десятки тысяч раз превышает тепловую энергию частиц в самых горячих частях Вселенной. Отсюда следует, что энергия частиц космического излучения имеет нетепловое происхождение.

4. То, что первичное космическое излучение, по крайней мере частично, состоит из заряженных частиц, подтверждается *геомагнитными явлениями*. К ним относится, во-первых, *широтный эффект* — зависимость интенсивности I космических лучей от геомагнитной широты места. Отвлечемся ради простоты от действия магнитного поля Солнца (оно будет учтено в п. 9) и примем во внимание только магнитное поле Земли $\mathbf{H}$. Сила Лоренца $(e/c)[\mathbf{vH}]$ перпендикулярна к магнитному полю и к скорости частицы. Если космическая частица приближается к геомагнитному полюсу, двигаясь вдоль $\mathbf{H}$, то сила Лоренца обращается в нуль. В этом случае Земли могут достигнуть частицы любой энергии. Если же частица приближается к Земле в плоскости геомагнитного экватора, то сила Лоренца максимальна. Она изгибает траекторию частицы. Земли могут достигнуть только частицы, энергии которых больше некоторой определенной величины. Если же энергия меньше, то частица не достигнет Земли, а отразится ее магнитным полем. Достаточно медленные частицы будут отражаться и на других геомагнитных широтах. Количественно широтный эффект характеризуется величиной $(I_{90^\circ}-I_{0^\circ})/I_{90^\circ}$. Измерения показали, что эта величина на уровне моря составляет около 10 %, а на высоте 10 км — около 36 %.

Во-вторых, к геомагнитным явлениям относится *долготный эффект*. Так называют изменение интенсивности космических лучей при

смещении вдоль одной и той же параллели. Магнитное поле Земли в первом приближении можно моделировать с помощью магнитного диполя с магнитным моментом $8{,}1 \cdot 10^{25}$ Гс $\cdot$ см^3, наклоненным под углом $11{,}5°$ к земной оси. Однако такой диполь надо поместить не в центре Земли, а сместить от него перпендикулярно к земной оси приблизительно на 342 км. Вследствие этого, а также небольшого отличия магнитного поля Земли от поля диполя это поле немного меняется вдоль параллели. С этим и связан долготный эффект. На уровне моря интенсивность космических лучей на одной и той же параллели отклоняется от среднего значения на несколько процентов.

Наконец, в-третьих, наблюдается *асимметрия в направлении космических лучей*, приходящих на Землю. Она проявляется в зависимости интенсивности космических лучей от ориентации регистрирующего прибора относительно стран света. С запада приходит больше космических частиц, чем с востока. Количественно восточно-западная асимметрия характеризуется величиной $2(I_з - I_в)/(I_з + I_в)$, т. е. отношением разности интенсивностей космических лучей, приходящих соответственно с запада и востока, к их полусумме. На больших высотах эта величина составляет около 0,5.

Опуская строгое и сложное рассмотрение вопроса, для уяснения причины восточно-западной асимметрии ограничимся частным случаем, когда заряженная космическая частица движется к Земле в плоскости ее геомагнитного экватора. На рис. 176 представлен этот случай для положительно ($а$) и отрицательно ($б$) заряженных частиц. Предполагается, что магнитное поле перпендикулярно к плоскости рисунка и направлено к читателю. Траектория положительной частицы загибается магнитным полем к востоку, т. е. к скорости такой частицы добавляется слагающая, направленная на восток. Если, как это установлено, вдали от Солнечной системы интенсивность космических лучей изотропна в пространстве, то большая часть первичных положительных частиц, достигающих Земли, будет двигаться в направлении с запада на восток, а не в противоположном направлении. Для отрицательных частиц положение будет обратным. Из фактически наблюдаемой восточно-западной асимметрии был сделан вывод, что в *первичных космических лучах большая часть частиц заряжена положительно.*

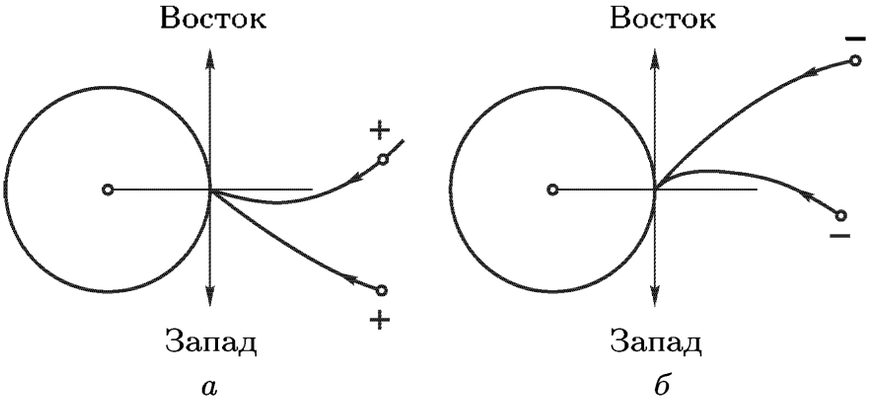


Рис. 176

Существуют и другие геомагнитные явления, о которых будет сказано ниже.

5. Исследования космических лучей производились с разнообразными детекторами в сочетании с фильтрами (из свинца, железа, углерода, парафина и пр.). Они производились на поверхности и в грунте Земли (от поверхности моря до глубин около 3 км), в верхних слоях атмосферы, в космическом пространстве около Земли и далеко за ее пределами.

Наиболее подробные данные о составе первичных космических лучей получены для частиц с энергией, большей 2,5 ГэВ на нуклон. Они получены главным образом из измерений в ядерных эмульсиях, поднимавшихся на шарах-зондах в верхние слои атмосферы. Эти данные приведены в табл. 22. В ней указаны интенсивности первичных космических лучей для различных частиц с энергией, большей 2,5 ГэВ на нуклон, т. е. число частиц заданных энергий, проходящих через единичную площадку нормально к ней, отнесенное к единице телесного угла. Если космические лучи распределены изотропно в пространстве, то для получения полного потока первичных космических частиц указанных энергий через единичную площадку приведенные данные надо умножить на π. Таким путем находим, что поток частиц около Земли у границы атмосферы равен 0,45 $с^{-1} \cdot см^{-2}$. Если учесть также все частицы с меньшими энергиями, то для полного потока первичных частиц в том же месте получится 1 $с^{-1} \cdot см^{-2}$.

Из табл. 22 видно, что в первичных космических лучах представлено больше всего протонов (92 %), затем идут α-частицы (6,6 %). Более тяжелых ядер (главным образом C, N, O) около 0,8 %. Одним из аргументов в пользу того, что тяжелые ядра принадлежат к первичным, а не к вторичным космическим лучам, может служить то обстоятельство, что заряды тяжелых ядер больше зарядов ядер атмосферных газов. Поэтому тяжелые ядра не могут возникать под действием первичных космических лучей. Средняя энергия, приходящаяся на один нуклон, приблизительно одинакова для всех ядер. Поэтому многозарядные ядра (т. е. ядра с $Z > 1$) несут заметную долю (около $1/2$) полной энергии первичных космических лучей.

Наряду с протонами и ядрами в космических лучах содержится незначительное количество электронов и позитронов (около 1 %) и γ-квантов (около 0,01 % при энергиях, больших 100 МэВ). Несомненно, имеются и нейтрино, но при настоящем уровне науки нет способов их детектировать в космических лучах. Присутствие мезонов и нейтронов в первичных космических лучах исключено, поскольку эти частицы нестабильны и должны распадаться на пути к Земле.

6. В последних двух столбцах табл. 22 сравнивается среднее относительное содержание различных частиц в космических лучах с естественным содержанием их во Вселенной, полученным из данных о составе земной коры, метеоритов, атмосферы Солнца и звезд. В космических лучах значительно больше «очень тяжелых» и «самых тяжелых» ядер, чем в среднем во Вселенной. В особенности такое преобладание наблюдается для ядер группы L, в которую входят литий, бериллий

Таблица 22

Состав космических лучей с энергиями, большими 2,5 Гэв на нуклон

Группа	Частицы, входящие в группу	Заряд ядра	Средняя атомная масса	$\frac{\text{Число частиц}}{\text{м}^2 \cdot \text{с} \cdot \text{ср}}$	Число ядер на 10 тыс. протонов	
					в космических лучах	в среднем во Вселенной
p	Протоны	1	1	1300	10 000	10 000
α	Ядра гелия	2	4	94	720	1600
L	«Легкие» ядра	3–5	10	2,0	15	10^{-4}
M	«Средние» ядра	6–9	14	6,7	52	14
H	«Тяжелые» ядра	$\geqslant 10$	31	2,0	15	6
VH	«Очень тяжелые» ядра	$\geqslant 20$	51	0,5	4	0,06
SH	«Самые тяжелые» ядра	> 30	100	$\sim 10^{-4}$	10^{-3}	$7 \cdot 10^{-5}$
e	Электроны и позитроны	1	$\frac{1}{1836}$	13	100	10 000
				1418	10 900	21 620

и бор. В космических лучах среднее относительное содержание ядер этой группы примерно в 100 тысяч раз больше, чем в среднем во Вселенной.

Аномально высокое содержание в космических лучах ядер группы L по сравнению с их средним содержанием во Вселенной делает невероятным предположение, что эти ядра имеются в достаточных количествах уже в самих источниках космических лучей. Более вероятна гипотеза, что они образуются при столкновениях тяжелых космических частиц с ядрами межзвездного газа. При таких столкновениях тяжелые ядра могут расщепляться на части (процесс *фрагментации*), в результате чего и образуются ядра группы L. Можно предположить, что это основной и даже практически единственный способ образования в космических лучах указанных ядер. Если эта гипотеза верна, то можно оценить, какой средний путь проходит космическая частица,

порожденная источником, от места ее образования до Земли — вопрос, очень важный для решения проблемы происхождения первичных космических лучей.

Для решения этого вопроса рассмотрим в космическом газе однородный параллельный пучок тяжелых космических частиц, т. е. частиц, масса которых больше массы ядер группы L. Направление распространения этого пучка примем за ось X. При выходе из источника первоначально пучок состоит только из тяжелых частиц. Из-за столкновения их с ядрами галактического газа образуются легкие ядра группы L, благодаря чему интенсивность I рассматриваемого пучка тяжелых частиц будет убывать. Допустим, что это единственный механизм убывания интенсивности пучка. Обозначим через σ среднее эффективное сечение столкновения тяжелой частицы с атомным ядром галактического газа и притом такое, при котором образуются легкие ядра группы L. Тогда процесс ослабления интенсивности $I_т$ потока тяжелых ядер будет описываться уравнением

$$\frac{dI_т}{dx} = -\sigma N I_т,$$

где N — число атомных ядер галактического газа в единице объема. Считая для простоты это число постоянным, получим

$$I_т = I_{т0} e^{-\sigma N x}. \tag{103.2}$$

Введем упрощающее предположение, что при каждом исчезновении тяжелой частицы рождается только одна легкая частица группы L. Если бы легкие частицы двигались также в направлении пучка, то образовался бы поток легких частиц с интенсивностью

$$I_л = I_{т0} - I_т = I_т(e^{\sigma N x} - 1).$$

Тогда отношение концентраций легких и тяжелых частиц в космических лучах было бы равно

$$\frac{n_л}{n_т} = \frac{I_л}{I_т} = e^{\sigma N x} - 1. \tag{103.3}$$

В действительности полученное выражение справедливо независимо от введенного ранее предположения относительно направления. Только под x следует понимать длину пути, действительно проходимого космической частицей от места ее образования до места наблюдения. А этот путь может значительно отличаться от прямолинейного.

Как видно из табл. 22, $n_л/n_т = 15/(52 + 15 + 4) = 0{,}21 \approx 1/5$. Благодаря малости этого отношения можно ограничиться линейным приближением

$$\frac{n_л}{n_т} = \sigma N x, \tag{103.4}$$

$$x = \frac{1}{\sigma N}\,\frac{n_л}{n_т} = \frac{1}{5}\,\frac{1}{\sigma N}. \tag{103.5}$$

Допустим, что источник космических лучей находится в Галактике. Для средней концентрации газа в Галактике можно принять $N \approx 1$ см$^{-3}$, а для среднего эффективного сечения $\sigma \approx 10^{-26}$ см^2. При таких значениях получается

$$x \approx 2 \cdot 10^{25} \text{ см}.$$

Скорость космической частицы незначительно отличается от скорости света c. Поэтому время движения частицы от источника до Земли $\tau \approx (2 \cdot 10^{25}) : (3 \cdot 10^{10}) \approx 0{,}7 \cdot 10^{15}$ с $\approx 2 \cdot 10^7$ лет.

7. По-видимому, найденные значения x и τ несколько занижены. Одна из причин этого состоит в том, что Галактика имеет форму диска диаметром $D_{\text{Гал}} \approx 25$ кпс $= 7{,}7 \cdot 10^{22}$ см и толщиной 2 кпс, который окружен гало, т. е. заполненной космическим веществом сферой диаметром 30 кпс. Плотность галактического газа в гало примерно в 10 раз меньше, чем в галактическом диске. Наличие гало несколько снижает величину N и завышает значения x и τ. Наиболее вероятными считаются значения $x \approx 10^{26}$ см, $\tau \approx 10^8$ лет. Значение x примерно в 1000 раз превышает диаметр галактического диска. Это не противоречит использованному нами (но не доказанному) предположению, что источник космических лучей находится в пределах нашей Галактики. Дело в том, что космическая частица при своем движении подвергается действию хаотически меняющихся в пространстве галактических магнитных полей, искривляющих ее траекторию. Последняя не прямолинейна, а имеет довольно запутанный вид: блуждание космической частицы в Галактике напоминает броуновское движение в гигантских масштабах.

В процессе длительного блуждания космических частиц в хаотически меняющихся галактических магнитных полях стирается всякая пространственная и временная выделенность источников галактических космических лучей по отношению к Земле. С этим, по-видимому, и связана наблюдаемая высокая изотропия галактических космических лучей, т. е. однородность распределения направлений их движения в пространстве, установленная с точностью до 0,1 %. С этим же связано и высокое постоянство их интенсивности во времени. Конечно, все это справедливо, если исключить искажающее влияние магнитных полей Земли и межпланетного пространства.

Наиболее простой и надежный метод измерения магнитного поля $\mathbf{H}$ в Галактике основан на исследовании зеемановского расщепления спектральной линии водорода $\lambda = 21$ см (см. § 66, п. 4). Таким и другими способами найдено, что средняя напряженность магнитного поля в межзвездном пространстве Галактики равна около $7 \cdot 10^{-6}$ Гс. Этому соответствует средняя плотность магнитной энергии в том же пространстве $H^2/8\pi \approx 1{,}5 \cdot 10^{-12}$ эрг/см$^3 \approx 1{,}2$ эВ/см^3.

8. Распределение космических лучей по энергиям характеризуется *энергетическим спектром*, который может быть *интегральным* и *дифференциальным*. Интегральным энергетическим спектром $I(\mathscr{E})$ называется поток космических частиц, кинетическая энергия которых

превышает $\mathscr{E}$. Его производная по энергии $dI/d\mathscr{E}$, взятая по абсолютной величине, определяет дифференциальный энергетический спектр. Очевидно, это есть поток космических частиц, отнесенный к единичному интервалу энергии.

В интервале энергий 10–10^6 ГэВ интегральный энергетический спектр хорошо аппроксимируется выражением

$$I(\mathscr{E}) \propto \mathscr{E}^{-\gamma}, \tag{103.6}$$

где γ в среднем равно приблизительно 1,6 — точнее, слабо меняется от 1,4 до 1,8, как показано ниже:

Интервал изменения $\mathscr{E}$, ГэВ	10–10^4	10^4–10^6	10^6–10^9
γ	1,4–1,6	1,6–1,8	1,8–2,2

При энергиях порядка $3 \cdot 10^6$ ГэВ спектр $I(\mathscr{E})$ опускается вниз более круто, претерпевая сравнительно небольшой и плавный излом. Это, возможно, связано с тем, что частицы столь высоких энергий перестают эффективно удерживаться магнитным полем Галактики и начинают заметно уходить в метагалактическое пространство. При $\mathscr{E} \approx 3 \cdot 10^{10}$ ГэВ спектр $I(\mathscr{E})$, по-видимому, становится опять более пологим.

Частиц с энергией $\mathscr{E} \gtrsim 10^6$ ГэВ очень мало, а потому они уже не могут регистрироваться с помощью счетчиков или фотоэмульсий. Например, на площадь 1 м^2 частица с энергией $\mathscr{E} > 10^7$ ГэВ попадает в среднем только один раз в год. Такие частицы обнаруживаются только по широким атмосферным ливням (см. п. 12). Зарегистрированы частицы с энергией порядка 10^{20} эВ $= 10^{11}$ ГэВ $= 12$ Дж. Энергии такой частицы достаточно, чтобы поднять груз массой 1 кг на высоту 1 м. Однако такие частицы встречаются очень редко — на поверхность в 10 км^2 за год в среднем падает не более одной частицы с энергией 10^{20} эВ.

При уменьшении энергии $\mathscr{E}$, начиная примерно с 10 ГэВ, рост интенсивности $I(\mathscr{E})$ замедляется и практически совсем прекращается при $\mathscr{E} \lesssim 2$ ГэВ. Магнитное поле Земли приводит к тому, что при таких относительно малых энергиях протоны могут достигать поверхности Земли лишь в высоких широтах (в окрестности полюсов). Но даже и в высоких широтах столь малоэнергичные частицы часто не достигают земной поверхности из-за потерь их энергии в атмосфере. Интенсивность космических лучей при $\mathscr{E} \lesssim 2$ ГэВ нерегулярно меняется во времени, так как они подвержены сильному влиянию магнитного поля, переносимого сгустками солнечной плазмы (так называемым *солнечным ветром*). Эти космические лучи заметно повышают нормальный уровень ионизации воздуха в верхних слоях атмосферы, что делает их непроницаемыми для радиошумов, приходящих из Галактики. Они приводят также к прерыванию радиосвязи на коротких волнах в полярной зоне (см. т. IV, § 87).

9. По энергетическому спектру можно вычислить среднюю плотность энергии космических лучей в межзвездном пространстве нашей Галактики и сравнить ее с тем, что дают другие источники энергии. В плотность энергии вносят свой вклад:

Космические лучи	$1{,}6 \cdot 10^{-12}$ эрг/см^3 = 1,3 эВ/см^3,
Турбулентное движение газа	$0{,}5 \cdot 10^{-12}$ эрг/см^3 = 0,4 эВ/см^3,
Магнитное поле	$1{,}5 \cdot 10^{-12}$ эрг/см^3 = 1,25 эВ/см^3,
Полное излучение звезд	$0{,}7 \cdot 10^{-12}$ эрг/см^3 = 0,6 эВ/см^3.

Таким образом, в пределах нашей звездной системы — Галактики — полная энергия всех космических лучей сравнима с энергией звездного излучения, энергией теплового и турбулентного движения межзвездного газа, а также с энергией магнитного поля Галактики. По теореме вириала (см. § 100) средняя кинетическая энергия движения межзвездного газа равна половине его потенциальной энергии, взятой со знаком минус. Поэтому в динамике Вселенной, помимо сил тяготения, газового и светового давления, должны приниматься во внимание и сравнимые с ними силы, обусловленные движением частиц космических лучей и создаваемых ими магнитных полей.

10. Подавляющая часть первичных космических лучей приходит к Земле из окружающего галактического пространства. Но небольшая часть космических лучей, преимущественно низких энергий ($\mathscr{E} < 1$ ГэВ), приходит к Земле и от Солнца. Энергетический спектр солнечных космических лучей круто падает в области высоких энергий, так что они вносят заметный вклад только в первичные космические лучи низких энергий. Средний по времени поток энергии солнечных космических лучей у поверхности Земли составляет около $3{,}5 \cdot 10^{-4}$ эрг/(с · см^2), что в 20 раз меньше потока энергии всех космических лучей $7 \cdot 10^{-3}$ эрг/(с · см^2). Однако во время кратковременных вспышек на Солнце поток солнечных космических лучей может превысить его среднее значение в тысячи раз.

Например, 23 февраля 1956 г. на поверхности Солнца произошла сильная хромосферная вспышка, зарегистрированная на Земле. Спустя несколько минут скорость счета детекторов космических лучей на специальных станциях, расположенных в различных местах Земли, стала быстро возрастать. В течение 15–20 мин скорость счета достигла максимума, затем начала спадать, а через несколько часов снизилась до нормального уровня. Возрастание скорости счета было зарегистрировано не только в высоких широтах, но и вблизи экватора. На этом основании можно заключить, что в пришедших солнечных космических лучах должны содержаться хотя бы в относительно небольшом количестве частицы с энергиями больше 10–20 ГэВ, чтобы преодолеть вблизи экватора отталкивающее действие земного магнитного поля. В последующие годы — примерно раз в год — наблюдалось еще несколько подобных явлений, хотя и менее крупного масштаба.

Солнечные космические лучи сравнительно легко отделить от галактических, так как они появляются эпизодически — во время сильных хромосферных вспышек на Солнце. От галактических они отличаются еще составом — в них практически нет ядер легких атомов Li, Be, B, а также ядер тяжелых элементов. В целом химический состав солнечных космических лучей близок к составу атмосферы Солнца.

Солнечная активность проявляется также *в вариации интенсивности* и галактических космических лучей, приходящих на Землю. В результате «магнитных бурь» на Земле, связанных с солнечной активностью, меняется земное магнитное поле, а с ним и интенсивность космических лучей, наблюдаемых на Земле. Это явление наблюдается не только в полярных, но и в экваториальных областях Земли. Например, в период максимальной солнечной активности в 1957–1958 гг. интенсивность космических лучей с энергиями, превышающими 1 ГэВ, была почти втрое меньше, чем в последующий период минимума активности в 1964–1965 гг.

11. Затронем вопрос о происхождении космических лучей, хотя он и очень далек от окончательного решения. Солнце испускает частицы с энергиями в сотни мегаэлектронвольт. Во время солнечных вспышек в межпланетное пространство выбрасываются большие массы ионизованных газов и порождаются, хотя и в малых количествах, частицы с энергиями 10–20 ГэВ. Нельзя ли предположить, что в результате каких-то еще не известных процессов возникают в нужном количестве и частицы более высоких энергий, наблюдаемые в космических лучах? От такой гипотезы следует отказаться. Основным возражением против нее является постоянство интенсивности и изотропия направлений космических лучей в любое время дня и ночи. Главную долю наблюдаемых космических лучей составляют частицы с энергией порядка 10 ГэВ. Если бы такие частицы возникали на Солнце, то они должны были бы приходить на Землю по направлению от него и, следовательно, наблюдались бы только днем, а не ночью. Тем более это заключение относится к частицам больших энергий. Возьмем для примера протон с энергией 10^{15} эВ. Магнитная жесткость HR, вычисленная по формуле (103.1), для такого протона составит $3 \cdot 10^{12}$ Гс $\cdot$ см. Магнитное поле в межпланетном пространстве — порядка нескольких единиц на 10^{-5} Гс. Возьмем ориентировочно $3 \cdot 10^{-5}$ Гс. Тогда радиус кривизны траектории протона будет порядка 10^{17} см, что примерно в 6500 раз больше радиуса земной орбиты. Напряженность земного магнитного поля равна 0,5 Гс. В нем радиус кривизны меньше $6 \cdot 10^{12}$ см, но и эта величина все же в 10 000 раз больше радиуса Земли.

Можно предположить, что источниками космических лучей являются звезды. Однако такие источники не могли бы создать необходимую энергию космических лучей в Галактике, если бы они были обыкновенными нормальными звездами. Действительно, объем Галактики 10^{67} см^3, средняя плотность энергии космических лучей в Галактике $1{,}6 \cdot 10^{-12}$ эрг/см^3, а полная энергия космических лучей в ней 10^{55} эрг.

Мощность излучения солнечных космических лучей 10^{24} эрг/с. Если предположить, что со времени образования Солнца ($5 \cdot 10^9$ лет) оно излучало космические лучи равномерно, а это излучение целиком удерживалось магнитными полями в Галактике, то энергия солнечных космических лучей в объеме Галактики составляла бы 10^{40} эрг. Если бы все звезды Галактики в среднем излучали энергию космических лучей так же, как Солнце, то полная энергия космических лучей в Галактике была бы равна 10^{51} эрг, что в 10 000 раз меньше требуемой.

Поэтому естественно предположить, что основная масса космических лучей возникает при взрывах сверхновых. При взрыве сверхновой освобождается энергия, сравнимая с собственной энергией Солнца:

$$M_\odot c^2 = 2 \cdot 10^{33} \cdot (3 \cdot 10^{10})^2 = 1{,}8 \cdot 10^{53} \text{ эрг}.$$

Известная часть этой энергии уносится космическими лучами. Взрывы сверхновых, по-видимому, в состоянии обеспечить необходимую плотность энергии космических лучей в Галактике.

Считается маловероятным, что основная часть космических лучей имеет метагалактическое происхождение, т. е. приходит в Солнечную систему от удаленных галактик, лежащих за пределами нашей Галактики. Космические лучи «блуждают» в Галактике, отклоняясь галактическими магнитными полями. Их движение в Галактике имеет характер диффузии. За время существования Вселенной (10^{10} лет) за пределы Галактики продиффундирует или может вступить в нее извне лишь сравнительно небольшая часть космических частиц. Это могут сделать только частицы очень высоких энергий, которые почти не отклоняются галактическими магнитными полями. Но таких частиц в Галактике и за ее пределами относительно мало. Для космических частиц, энергии которых заполняют практически весь энергетический спектр, указанные эффекты не играют существенной роли. Подавляющая часть света (за исключением света от Солнца, планет и их спутников) приходит к Земле от звезд нашей Галактики. На долю Метагалактики приходится лишь незначительная часть. По-видимому, это справедливо и для космических лучей. Если это так, то космические лучи, приходящие на Землю, в основном должны возникать в пределах нашей Галактики.

Если даже справедлива гипотеза, что источниками космических лучей являются сверхновые нашей Галактики, то еще остается открытым вопрос о механизме ускорения заряженных частиц до тех громадных энергий, которые характерны для космических лучей. Вопрос этот не может считаться решенным. Одна из гипотез была предложена Ферми, который предположил, что при взрывах сверхновых образуются протяженные намагниченные облака плазмы. Не только плазма в облаках совершает сложное турбулентное движение, но и сами облака движутся с громадными скоростями. Заряженные частицы могут сталкиваться с такими облаками и отражаться от них. Если частица движется навстречу облаку, то при отражении ее энергия увеличивается; если она движется в ту же сторону, что и облако, то энергия уменьшается. При

хаотическом движении облаков преобладают встречные столкновения, так что в среднем происходит увеличение кинетической энергии частиц.

12. До сих пор речь шла о первичных космических лучах, главным образом галактических. Теперь коротко рассмотрим *вторичные космические лучи.* Они возникают в земной атмосфере в длинной и разветвленной цепи превращений, вызываемых первичными космическими лучами. Исходным звеном такой цепи является взаимодействие космической частицы с ядрами атомов воздуха. Ядро-мишень при этом, как правило, распадается на протоны и нейтроны. Распадается и первичная частица, если она была сложной. Образующиеся при этом протоны в дальнейшем ведут себя так же, как и первичные протоны.

Количественно прохождение частицы в веществе удобно характеризовать средним пробегом ее λ до взаимодействия с ядром среды. Средний пробег λ принято выражать количеством вещества в граммах в столбе вещества площадью в 1 см^2 и высотой, равной длине пути, проходимой частицей до взаимодействия. В тех же массовых единицах удобно выражать вообще путь частицы в веществе. Ослабление интенсивности I пучка частиц в веществе при прохождении пути x выражается формулой

$$I = I_0 e^{-x/\lambda}. \tag{103.7}$$

Так, толщина земной атмосферы в этих единицах составляет около 1000 г/см^2. Для протонов $\lambda = 70$ г/см^2, для α-частиц $\lambda = 25$ г/см^2, для более тяжелых ядер λ еще меньше. Отсюда видно, что вероятность достижения частицей земной поверхности исключительно мала. Так, для протона она равна $e^{-1000/70} \sim 10^{-7}$. Из 10 миллионов первичных протонов в среднем только один достигнет земли. Небольшое количество первичных протонов можно наблюдать лишь на высоких горах. Для наблюдения более тяжелых первичных ядер уже необходимо поднимать приборы в верхние слои атмосферы или в космос.

При столкновении с ядром протон теряет в среднем 30–50 % первоначальной энергии. Основная часть первичных космических лучей со средней энергией 10 ГэВ сначала теряет энергию на рождение новых частиц при столкновении с ядрами, а затем энергия идет на ионизацию атомов воздуха. Лишь протоны с первоначальной энергией в несколько сот гигаэлектронвольт и выше могут достигнуть поверхности земли. Когда энергия первичной частицы уменьшится примерно до 100 МэВ, дальнейшее рождение новых частиц становится маловероятным.

Процесс столкновения протона с атомным ядром схематически изображен на рис. 177. В ядерном столкновении образуются заряженные пионы $\pi^{\pm}$ и нейтральный пион π^0. Образуются также пары протон-антипротон и нестабильные частицы: K-мезоны и гипероны, но основную роль в дальнейших процессах играют пионы.

Заряженные пионы, распадающиеся сравнительно медленно (за время в среднем $2{,}6 \cdot 10^{-8}$ с), успевают частично провзаимодействовать с другими ядрами воздуха и вызвать такие же ядерные процессы, как и первичные частицы. Другая часть заряженных пионов, не успев провзаимодействовать, распадается на мюоны ($\mu^{\pm}$) и нейтрино. Нейтрино,

как очень слабо взаимодействующие частицы, уходят, пронизывая не только атмосферу, но и весь земной шар. Положительный мюон μ^+ распадается по схеме $\mu^+ \to e^+ + \nu_e + \overline{\nu}_\mu$, отрицательный — по схеме $\mu^- \to e^- + \overline{\nu}_e + \nu_\mu$. Время жизни мюонов порядка $2{,}2 \cdot 10^{-6}$ с. За это время успевает распасться лишь небольшая часть мюонов. Поэтому они сравнительно легко проходят через атмосферу и поглощаются глубоко под землей. Здесь существенную роль играет релятивистский эффект замедления хода движущихся часов (см. т. IV, § 106, п. 4). На уровне моря мюоны составляют большую часть «проникающей компоненты» космических лучей.

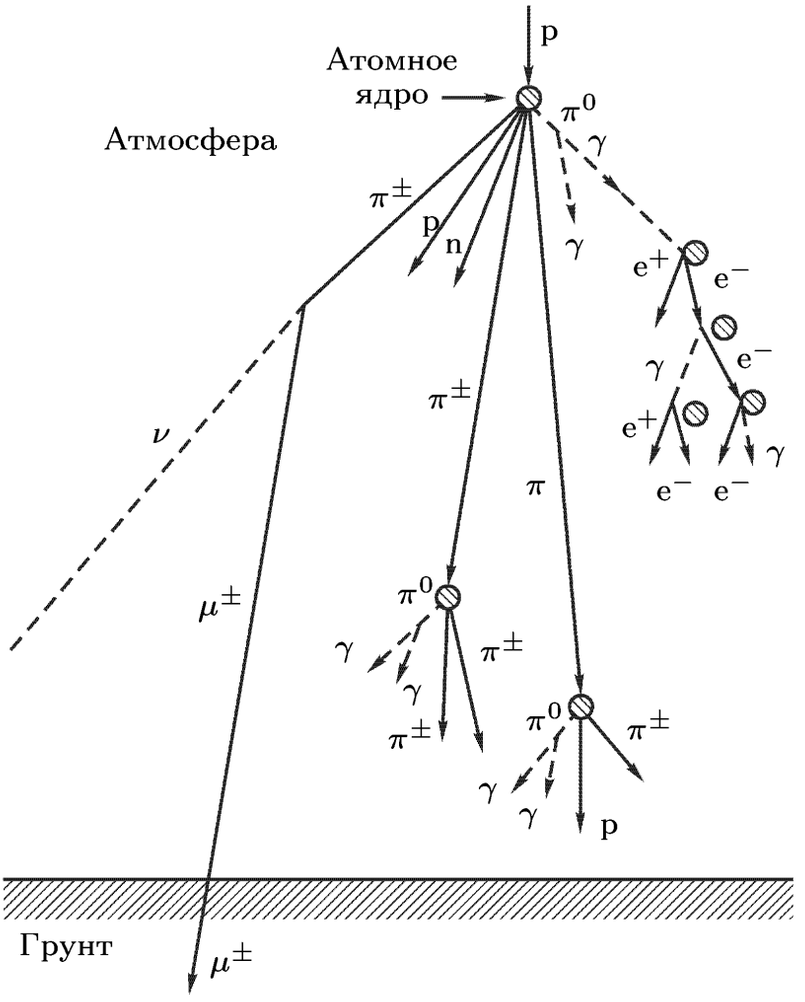


Рис. 177

Теперь проследим судьбу нейтрального пиона π^0. Среднее время жизни его очень короткое ($1{,}8 \cdot 10^{-16}$ с). Поэтому он успевает уйти лишь на ничтожное расстояние от места своего рождения и, как правило, не провзаимодействовав с ядрами атомов воздуха, распадается на два фотона высокой энергии: $\pi^0 \to \gamma + \gamma$. В поле ядра эти фотоны порождают электрон–позитронные пары. Средний пробег γ-фотона в воздухе до рождения электрон–позитронной пары $\lambda \approx 35$ г/см^2. Образовавшиеся электрон и позитрон в процессе тормозного излучения в веществе излучают новые фотоны. Энергия последних еще достаточно высока, чтобы породить новые электрон–позитронные пары. В итоге распад нейтрального пиона высокой энергии приводит к образованию

в атмосфере *электронно-фотонного каскадного ливня.* Когда энергия становится меньше 72 МэВ на частицу, главную роль начинают играть не тормозное излучение и рождение пар фотонами, а потери энергии электронов и позитронов на ионизацию, а фотонов — на комптоновское рассеяние. На этом развитие электронно-фотонного каскадного ливня прекращается.

В развитие электронно-фотонного каскадного ливня вносят некоторый вклад нейтральные пионы, образующиеся при распаде других нестабильных частиц — K-мезонов и гиперонов, а также электроны, образующиеся при распаде мюонов.

При энергиях первичной частицы 10^{14}–10^{15} ГэВ и выше каскад, сопровождающийся развитием электронно-фотонных лавин, получил название *широкого атмосферного ливня.* Развитие такого ливня начинается на высоте 20–25 км над уровнем моря. Общее число частиц в широком атмосферном ливне достигает многих миллионов. На уровне моря частицы ливня покрывают площадь до нескольких квадратных километров. Огромное число частиц в ливне позволяет его обнаруживать и изучать с помощью сравнительно небольшого количества счетчиков, размещенных на большой площади. Одновременное срабатывание счетчиков свидетельствует о прохождении широкого атмосферного ливня вторичных частиц, но лишь малая доля их фиксируется счетчиками. В первом приближении можно считать, что на каждую заряженную частицу (в основном электрон и позитрон) приходится 2–3 ГэВ энергии первичной частицы. Подсчет общего числа заряженных частиц в ливне позволяет достаточно точно оценить энергию первичной частицы, вызвавшей ливень. Никакими другими способами это сделать невозможно.

Глава XVI

ЭЛЕМЕНТАРНЫЕ ЧАСТИЦЫ

§ 104. Что такое элементарные частицы

1. В общем курсе о физике элементарных частиц можно дать только предварительные представления. Это касается как экспериментальных методов, так в особенности и теоретических представлений. Этому разделу науки посвящены специальные курсы, в которых излагается как предмет в целом, так и его отдельные части. Наше весьма неполное изложение носит описательный информационный характер и, как правило, не затрагивает теоретические основы изучаемых закономерностей.

Отметим с самого начала, что в физике элементарных частиц вводится ряд понятий с весьма экзотическими названиями: *странность*, *очарование*, *красота* (или *прелесть*), *цвет* и пр. Все эти термины не имеют никакого отношения к тому, в каком смысле они употребляются в обыденной жизни. Это просто какие-то *квантовые числа*, вводимые для описания характеристик элементарных частиц.

2. При введении понятия *элементарных частиц* первоначально предполагалось, что это есть *первичные*, далее *неделимые частицы*, из которых состоит вся материя. Таковыми вплоть до начала XX века считались атомы (слово «атом» в переводе с греческого означает «неделимый»). После того как была установлена сложная структура атомов, они перестали считаться элементарными частицами в указанном смысле слова. Такая же судьба постигла ядро, а затем протон и нейтрон, у которых была установлена внутренняя структура. Открывались новые и новые объекты (*мюоны*, *пионы*, *нейтрино* и пр.), которые могли претендовать на роль элементарных частиц. Для большинства из них эти претензии были отклонены очень быстро. Но и в настоящее время мы с достоверностью не знаем, какие частицы являются действительно элементарными и есть ли вообще элементарные частицы в первоначальном смысле этого слова. *Элементарными частицами сейчас условно называют большую группу мельчайших микрочастиц, не являющихся атомами или атомными ядрами* (за исключением протонов — ядер атома водорода). Общее, что роднит все элементарные частицы, состоит в том, что все они являются специфическими формами материи, не ассоциированной в атомы и атомные ядра.

В настоящее время к «истинно» элементарным принято относить следующие частицы (и их античастицы): 1) *лептоны* (e, μ, τ и соответствующие им *нейтрино*); 2) *кварки*; 3) *фотоны* и *промежуточные бозоны* $W^{\pm}$, Z^0. Необходимо, однако, заметить, что существуют гипотезы (основанные на наблюдаемой на опыте симметрии между кварками

и лептонами в электромагнитных взаимодействиях, а также на идеях Великого объединения сил) о том, что кварки и лептоны сами состоят из более фундаментальных частиц — «преонов». То же относится к $W^{\pm}$-, Z^0-бозонам (у которых предполагается существование дискретных возбужденных состояний). Если разности энергетических уровней велики по сравнению с энергиями, воздействующими на систему, то последняя ведет себя как целая — как элементарная частица.

3. Естественным источником различных частиц высоких энергий являются космические лучи. Не случайно поэтому, что до начала 50-х годов физика элементарных частиц была тесно связана с изучением космических лучей. В настоящее время источниками заряженных частиц высоких энергий являются в основном ускорители. Они в комбинации с детекторами позволяют исследовать процессы, в которых образуются и взаимодействуют различные элементарные частицы в лучших контролируемых условиях. Понятно, почему физику элементарных частиц называют также *физикой высоких энергий.*

Для осуществления взаимодействия на очень малых расстояниях частицы должны обладать очень высокими энергиями. Но в ультрарелятивистской области полная энергия частицы связана с ее импульсом соотношением $\mathscr{E} = pc$, так что соотношение неопределенностей $\Delta x \cdot \Delta p \approx \hbar$ принимает вид

$$\Delta\mathscr{E} \cdot \Delta x \gtrsim \hbar c/2, \tag{104.1}$$

или

$$\Delta\mathscr{E} \gtrsim 10^{-14}/\Delta x, \tag{104.2}$$

если энергию выражать в гигаэлектронвольтах, а расстояние Δx — в сантиметрах. Из формулы видно, что для проникновения на расстояние 10^{-16} см требуется энергия, превышающая примерно 100 ГэВ. Именно энергия такого порядка потребовалась, например, для получения промежуточных векторных бозонов $W^{\pm}$ и Z^0.

4. Наиболее характерным свойством элементарных частиц, связанным по современным представлениям с корпускулярно-волновым дуализмом, является их *способность рождаться и взаимопревращаться друг в друга* при столкновениях. При этом полная (релятивистская) энергия сохраняется, т. е. полная энергия всех частиц до столкновения равна полной энергии всех частиц после столкновения. Например, при столкновении движущегося протона с неподвижным протоном может происходить реакция

$$p + p \to p + p + \pi^0,$$

в которой возникает нейтральный пион. Каково необходимое (но недостаточное) условие возможности такой реакции? Масса пиона равна 135 МэВ. Поэтому для осуществления этой реакции кинетическая энергия налетающего протона во всяком случае должна быть не меньше 135 МэВ. На самом деле она должна быть не меньше 290 МэВ, так как в соответствии с законом сохранения импульса часть энергии уносится

системой в ее поступательном движении и не участвует в реакции (см. задачу 3 к § 108).

Вторая характерная черта элементарных частиц состоит в том, что подавляющая часть их *нестабильна.* Частицы самопроизвольно распадаются. Среднее время жизни τ частицы в свободном состоянии меняется в широких пределах — от 10^{-24} с до бесконечности (для протона, например, экспериментально установлено, что $\tau > 10^{32}$ лет).

Распад элементарных частиц и их взаимопревращаемость друг в друга ставят под сомнение сам смысл вопроса: «Из чего состоит частица?» Например, при распаде нейтрона ($n \to p + e^- + \overline{\nu}_e$) возникают новые частицы: протон, электрон и антинейтрино. Но это не означает, что нейтрон состоит из таких частиц. На опыте констатируется только, что эти частицы *рождаются* в результате распада нейтрона. Нейтрон в той же мере элементарен, что и протон. Новые частицы могут рождаться и при распадах и взаимодействиях других частиц. Частицу можно считать сложной, построенной из нескольких других частиц в том случае, когда энергия связи каждой из составляющих частиц много меньше ее энергии покоя. Именно в таком смысле надо понимать утверждение, что атомное ядро состоит из протонов и нейтронов.

5. В § 63 (п. 9) уже говорилось, что все процессы и явления в природе осуществляются в результате четырех взаимодействий: *сильных*, *электромагнитных*, *слабых* и *гравитационных.* Сравнение интенсивностей этих взаимодействий приобретает определенный смысл только тогда, когда точно указаны условия, при которых происходит сравнение. В § 63 проводилось сравнение при условии, что средние кинетические энергии сталкивающихся частиц порядка 1 ГэВ.

Сильное взаимодействие вызывают процессы, протекающие наиболее быстро по сравнению с другими процессами. Оно обеспечивает и самую сильную связь элементарных частиц. В частности, связь между нуклонами в атомных ядрах обусловлена сильным взаимодействием. Им объясняется исключительная прочность атомных ядер, лежащая в основе стабильности вещества в земных условиях.

Электромагнитное взаимодействие сводится к взаимодействию электрических зарядов (и магнитных моментов) частиц с электромагнитным полем. Процессы, связанные с электромагнитным взаимодействием, протекают значительно менее быстро, чем процессы, вызываемые сильным взаимодействием. Электромагнитное взаимодействие обеспечивает связь электронов в атомах, ионов в кристаллах, атомов в молекулах. Электромагнитное взаимодействие (наряду с тяготением) играет основную роль в окружающем нас макроскопическом мире. Это связано с тем, что радиус действия сильных взаимодействий порядка 10^{-13} см и на больших расстояниях сильное взаимодействие фактически исчезает. Электромагнитное же взаимодействие (и тяготение) характеризуется бесконечным радиусом действия.

Слабое взаимодействие, как показывает само название его, вызывает очень медленно протекающие процессы с элементарными частицами. Хорошей иллюстрацией этого может служить исключительная слабость взаимодействия нейтрино низких энергий с веществом (см. § 74).

Ведь нейтрино свойственно одно только слабое взаимодействие. Со слабым взаимодействием связана также относительная медленность распада квазистабильных частиц. Время жизни большинства этих частиц лежит в диапазоне 10^{-8}–10^{-13} с, тогда как время жизни сильно взаимодействующих частиц составляет 10^{-23}–10^{-24} с. Интенсивность слабого взаимодействия растет с энергией (в системе центра масс). При $\mathscr{E} \sim M_{\mathrm{W}}$ слабое взаимодействие сравнивается с электромагнитным.

Гравитационное взаимодействие доминирует в случае больших макроскопических масс (планет, звезд). Но в мире элементарных частиц, ввиду малости их масс, даже на самых малых характерных для них расстояниях порядка 10^{-13} см это взаимодействие ничтожно. В физике элементарных частиц при современном ее состоянии гравитационное взаимодействие не учитывается. Оно, возможно, существенно лишь на расстояниях порядка 10^{-33} см.

§ 105. Классификация элементарных частиц

1. Все частицы (в том числе и неэлементарные частицы и квазичастицы) разделяются на *бозоны* и *фермионы*. Бозонами (или *бозе-частицами*) называются частицы или квазичастицы, обладающие нулевым или целочисленным спином. Бозоны подчиняются статистике Бозе–Эйнштейна (отсюда и происходит их название). К бозонам относятся: гипотетический *гравитон* (спин 2), *фотон* (спин 1), *промежуточные векторные бозоны* (спин 1), *глюоны* (спин 1), *мезоны* и *мезонные резонансы*, а также *античастицы* всех перечисленных частиц. Частицы или квазичастицы с полуцелым спином называются *фермионами* (или *ферми-частицами*). Для них справедлив принцип Паули, и они подчиняются статистике Ферми–Дирака (отсюда и происходит их название). К фермионам относятся: *лептоны*, все *барионы* и *барионные резонансы*, *кварки* (спин 1/2), а также соответствующие *античастицы*.

2. По времени жизни τ различают *стабильные*, *квазистабильные* и *резонансные частицы*. Последние для краткости называют просто *резонансами*. Резонансными называют частицы, распадающиеся за счет сильного взаимодействия с временем жизни 10^{-23} с. Нестабильные частицы, время жизни которых превышает 10^{-20} с, распадаются за счет электромагнитного или слабого, но не за счет сильного взаимодействия. Эти частицы относят к квазистационарным. Время 10^{-20} с, ничтожное в обыденных масштабах, должно считаться большим, если его сравнивать с ядерным временем. Последнее есть время, требующееся свету на прохождение диаметра ядра (10^{-13} см). Даже за время 10^{-20} с свет успевает пробежать расстояние в 10^3–10^4 нуклонных диаметров. За это время может совершиться еще много внутринуклонных процессов. Вот почему частицы, названные нами квазистабильными, в справочниках называют просто стабильными. Впрочем, абсолютно стабильными частицами являются, по-видимому, только фотон γ, электрон e^-, протон $\mathrm{p}(?)$, электронное ν_{e}, мюонное ν_μ и таонное ν_τ нейтрино и их античастицы — распад всех этих частиц на опыте не зарегистрирован.

Распады могут происходить по сильному, электромагнитному и слабому взаимодействиям. Наиболее быстро происходят распады по сильному взаимодействию — таковы распады резонансов. Квазистабильные частицы распадаются за счет слабого или электромагнитного взаимодействия. Они стали бы абсолютно устойчивыми, если бы можно было мысленно «выключить» эти взаимодействия, оставив только сильное. Наиболее стабильными резонансами являются частицы J/ψ и Υ, для которых $\tau \approx 10^{-20}$ с. Они относятся к резонансам потому, что у них имеются каналы распада, обусловленные сильным взаимодействием, но эти распады подавлены законами сохранения *очарования* и *красоты* при сильных взаимодействиях (см. § 110, пп. 4, 5).

В силу малости времени жизни τ, резонансы не обладают определенной массой. Это видно из соотношения неопределенностей $\Delta\mathscr{E} \cdot \tau \approx \hbar$. Резонансы описываются непрерывным спектром масс. Положение максимума этого спектра и называется массой резонанса. Ширина спектра Γ определяется обычным соотношением $\Gamma \approx \hbar/\tau$. При очень малых временах жизни она бывает сравнима со значением самой массы резонанса. Именно ширина Γ (а не τ) обычно и приводится в таблицах в качестве меры нестабильности резонанса. Так, при $\tau \approx 10^{-23}$ с получаем $\Gamma \approx 100$ МэВ. Поэтому резонансами можно назвать частицы с большой шириной спектра масс $\Gamma \approx 100$ МэВ.

3. Особую группу элементарных частиц составляют фотоны, являющиеся *переносчиками электромагнитного взаимодействия*, и родственные им $W^{\pm}$-, Z^0-*бозоны*, являющиеся *переносчиками слабого взаимодействия*. Эти четыре частицы образуют группу так называемых *переносчиков взаимодействия*. К переносчикам взаимодействия относятся и *глюоны*, а также гипотетические *гравитоны*. Все остальные частицы разделяются на *лептоны* и *адроны*.

Лептонами называются частицы, не участвующие в сильных взаимодействиях и имеющие спин $1/2$. В настоящее время установлено существование шести заряженных лептонов: *электрон* e^-, *позитрон* e^+, *мюоны* $\mu^{\pm}$, *тяжелые лептоны* $\tau^{\pm}$ (*таоны*), и соответствующих им шести нейтральных частиц: *электронное нейтрино* ν_e и *антинейтрино* $\overline{\nu}_e$, *мюонное нейтрино* ν_μ и *антинейтрино* $\overline{\nu}_\mu$, таонное нейтрино ν_τ и *антинейтрино* $\overline{\nu}_\tau$. Нейтральные лептоны (нейтрино) не участвуют и в электромагнитных взаимодействиях.

Все лептоны, на современном уровне знания, можно назвать *истинно элементарными частицами*, так как у них в отличие от адронов не обнаружена внутренняя структура. В этом смысле лептоны называются *точечными частицами*.

Мюоны были открыты в космических лучах Андерсоном вместе с Неддермайером в 1937 г. Наличие у мюонов собственного (мюонного) нейтрино было установлено позже — только в начале 60-х годов. τ-лептоны были открыты в 1975 г. в Стэнфорде (США) группой экспериментаторов во главе с Перлом (р. 1927) в опытах со встречными электрон–позитронными пучками. Тау-лептон получается в результате аннигиляции электрона и позитрона ($e^+ + e^- \to \tau^+ + \tau^-$). Масса мюона

$m_\mu = 105{,}7$ МэВ, время жизни $\tau = 2{,}2 \cdot 10^{-6}$ с, масса таона $m_\tau \sim$ $\sim 1{,}8$ ГэВ, время жизни $\tau_\tau \sim 5 \cdot 10^{-13}$ с.

Наши сведения о нейтрино очень неполные (см. § 74, п. 11). Особенно это касается μ- и τ-нейтрино. Даже в отношении электронного нейтрино нельзя категорически утверждать, равна ли масса этой частицы нулю или только очень мала.

4. Адронами называются элементарные частицы, участвующие в сильных взаимодействиях. Они, как правило, участвуют также и во всех других взаимодействиях — электромагнитном и слабом.

Эти частицы, в основном резонансы, составляют наиболее многочисленную группу элементарных частиц — их насчитывается около 400. Адроны подразделяются на *стабильные* и *квазистабильные адроны* и *резонансы.* В свою очередь стабильные адроны подразделяются на *мезоны* и *барионы.* Теоретические мотивы такого подразделения выяснятся в § 110 (в кварковой модели). В группу резонансов входят мезонные и барионные резонансы.

Мезонами называются нестабильные заряженные или нейтральные адроны, обладающие нулевым или целочисленным спином, а потому принадлежащие к классу бозонов. Сюда относятся π^0- и $\pi^\pm$-мезоны, $K^\pm$-мезоны. Эти мезоны были открыты раньше других. Масса их — промежуточная между массами электрона и протона (отсюда и происходит их название — от греческого слова mesos, что означает «средний, промежуточный»). Позднее были открыты более тяжелые $\mathrm{D}^\pm$-, D^0-, $\mathrm{F}^\pm$-мезоны, масса которых больше массы протона. Было открыто также много мезонных резонансов, т. е. мезонов с временами жизни порядка 10^{-23} с. Масса некоторых из них также превосходит массу протона. Мюоны μ первоначально назывались μ-мезонами, но они не относятся к классу мезонов, так как имеют спин $1/2$ и не участвуют в сильных взаимодействиях.

Барионами и барионными резонансами называются адроны с полуцелым спином и массами, не меньшими массы протона. К ним относятся *нуклоны* (*протоны* и *нейтроны*), *гипероны* и др. Протон и нейтрон — самые легкие барионы. Протон — единственный стабильный барион (см., впрочем, § 108, п. 4), все остальные барионные резонансы нестабильны и путем последовательных распадов превращаются в нуклоны и легкие частицы: *π-мезоны*, *электроны*, *нейтрино*, *γ-кванты.* (Нейтрон в свободном состоянии — нестабильная частица со временем жизни $\sim$ 16 мин, но в связанном состоянии внутри ядра он стабилен, если ${}^A_Z M < {}_{Z+1}{}^A M + m_\mathrm{e}$, т. е. когда не происходит β^--распада. Если же ${}^A_Z M > {}_{Z-1}{}^A M + m_\mathrm{e}$, то нестабилен протон и происходит позитронный β^+-распад: $\mathrm{p} \to \mathrm{n} + \mathrm{e}^+ + \nu_\mathrm{e}$.)

Нестабильные барионы с массами, бóльшими массы нуклона (протона и нейтрона), и бóльшим временем жизни по сравнению с ядерным временем (порядка 10^{-23} с) называются *гиперонами.* Первые гипероны (Λ) были открыты в космических лучах. Детальное изучение их стало возможным после того, как их стали получать на ускорителях заряженных частиц высоких энергий при столкновениях быстрых нуклонов, π-

и K-мезонов с нуклонами атомных ядер. Известно несколько типов гиперонов: *лямбда* (Λ^0), *сигма* ($\Sigma^-, \Sigma^0, \Sigma^+$), *кси* ($\Xi^-, \Xi^0$), *омега* ($Q^-$), Λ_c. Все гипероны имеют спин $1/2$, за исключением гиперона Ω^-, спин которого равен $3/2$. Таким образом, гипероны, как и все барионы, являются фермионами. Время жизни гиперонов $\tau \sim 10^{-10}$ с (за исключением Σ^0 и Λ^0 и Λ_c, для которых τ равно 10^{-19} и 10^{-13} с соответственно). За это время они распадаются на нуклоны и легкие частицы (π-мезоны, электроны, нейтрино, γ-кванты).

В 70-х годах на больших ускорителях были созданы пучки заряженных и нейтральных гиперонов высоких энергий (20–100 ГэВ). Это позволило проверить формулу для релятивистского замедления времени в лучших условиях по сравнению с тем, как это делалось раньше (см. т. IV, § 106, пп. 4, 5). Если бы не было релятивистского замедления времени, то гипероны от своего рождения до распада пробегали бы путь порядка $c\tau$, т. е. порядка сантиметра или десятков сантиметров. На самом деле этот путь достигает нескольких метров.

Детализация классификации элементарных частиц будет произведена в следующих параграфах по мере надобности. Отличительные признаки лептонов, мезонов и барионов станут более ясными, если при классификации частиц исходить из кварковой модели (см. § 110).

§ 106. Античастицы

1. В микромире каждой *частице* соответствует *античастица.* В некоторых случаях частица совпадает со своей античастицей, т. е. все свойства частицы и античастицы тождественны. В таком случае элементарные частицы называют *истинно нейтральными частицами.* К ним относятся *фотон* γ, π^0-мезон, η^0-мезон, J/ψ-мезон, *ипсилон-частица* Υ. Если же частица и античастица не совпадают, то массы, спины, изотопические спины, времена жизни у частицы и античастицы одинаковы, а прочие характеристики (электрический заряд, магнитный момент, лептонные и барионные заряды, странность, очарование, красота) одинаковы по абсолютной величине, но противоположны по знаку. (Встречающиеся здесь термины будут определены в следующих параграфах.) Так, электрон и протон отличаются от позитрона (антиэлектрона) и антипротона прежде всего знаком электрического заряда. Нейтрон отличается от антинейтрона знаком магнитного момента. Лептонные заряды у лептонов и антилептонов, барионные заряды у барионов и антибарионов противоположны по знаку.

Понятия частицы и античастицы относительны. Что назвать частицей и что античастицей — это вопрос соглашения. Электрон считают частицей, а позитрон — античастицей только потому, что в нашей Вселенной преобладают именно электроны, а позитроны являются более экзотическими объектами. Но, в принципе, с равным успехом частицей можно было бы назвать позитрон, а электрон — античастицей.

2. Первая античастица — *позитрон* (антиэлектрон) была предсказана теоретически Дираком в 1931 г. и обнаружена в 1932 г. Андерсоном (р. 1905). Дирак исходил из предложенного им релятивистского волнового уравнения и принципа Паули. Рассмотрение его теории, по существу, далеко выходит за рамки нашего курса, и мы не будем ее касаться. Отметим только, что теория Дирака в вопросе о частицах и античастицах, во всяком случае, недостаточна. Это видно уже из того, что она построена для электрона, а он имеет спин $1/2$ и относится к фермионам. Античастицами же обладают не только фермионы, но и бозоны. Существование античастиц является более фундаментальным фактом, чем думали сначала. Современная квантовая теория поля решает этот вопрос, но на этом мы останавливаться не можем.

Андерсон открыл позитрон в составе космических лучей, фотографируя следы космических частиц в камере Вильсона. След позитрона был похож на след электрона, но в магнитном поле он загибался в противоположную сторону. Это свидетельствовало о положительном знаке заряда наблюдаемой частицы. О направлении полета частицы можно было судить по увеличению кривизны следа при ее движении. Для того чтобы усилить этот эффект, Андерсон ставил на пути позитрона свинцовую пластинку, проходя через которую позитрон тормозился, и уменьшение радиуса кривизны его следа становилось более значительным. По кривизне следа Андерсон вычислил энергию частицы. Если бы это был протон, то его пробег при установленном Андерсоном значении энергии был бы примерно в 10 раз меньше наблюдаемого в действительности. Это означало, что масса открытой положительно заряженной частицы была меньше массы протона.

В 1933 г. вскоре после открытия Андерсона Блэккет (1897–1974) и Оккиалини (р. 1907) открыли *электронно-позитронные ливни* в космических лучах. Они пользовались камерой Вильсона, управляемой посредством счетчиков. Камера помещалась между двумя счетчиками и срабатывала только тогда, когда через оба счетчика одновременно пролетала ионизующая заряженная частица. В этот момент и производилось фотографирование. Измеряя кривизну треков, исследователи также пришли к заключению о существовании позитрона.

3. В вакууме при отсутствии вещества *позитрон столь же стабилен, что и электрон.* Однако при встрече электрона с позитроном эти частицы «аннигилируют», т. е. превращаются в два, три или несколько квантов излучения. Один γ-квант излучиться не может, так как в этом случае нарушился бы закон сохранения импульса. Это очевидно, если рассмотреть процесс в системе центра масс и принять во внимание, что число излученных γ-квантов от выбора системы отсчета не зависит. Значит, и в любой системе отсчета один γ-квант излучиться не может.

Существует обратный процесс: γ-квант может породить *пару* $\mathrm{e}^+\mathrm{e}^-$. Для этого необходимо, чтобы энергия γ-кванта была не меньше собственной энергии пары $2m_\mathrm{e}c^2$. Этот процесс может происходить только в присутствии *третьего тела*, например атомного ядра, иначе нарушился бы закон сохранения импульса. Действительно, в системе центра масс образовавшейся пары импульс пары был бы равен нулю,

тогда как импульс породившего ее γ-кванта отличен от нуля. При наличии атомного ядра импульс γ-кванта будет восприниматься ядром. В произвольной же системе отсчета импульс исчезнувшего кванта распределится между тремя частицами: атомным ядром, электроном и позитроном. Нарушения закона сохранения импульса не произойдет.

Позитроны могут появляться также в позитронной радиоактивности и в других процессах превращения ядерных и элементарных частиц. Образование пар e^+e^- γ-квантами высоких энергий и позитронная радиоактивность служат основными процессами для получения позитронов.

При столкновениях медленных позитронов с атомами вещества позитроны могут захватывать электроны из атомной оболочки. В результате образуется связанная система из электрона и позитрона, называемая *позитронием.* Различают *ортопозитроний* (когда спины e^+ и e^- направлены параллельно) и *парапозитроний* (когда спины направлены противоположно). Позитроний — нестабильная частица, так как позитрон и электрон быстро аннигилируют с образованием γ-квантов. Парапозитроний аннигилирует в два γ-кванта за время $1{,}25 \cdot 10^{-10}$ с, а ортопозитроний — в три γ-кванта за время $1{,}4 \cdot 10^{-7}$ с. Позитроний — простейшая система, связанная чисто электромагнитными силами (без участия сильных взаимодействий). Поэтому изучение позитрония представляет особый интерес для проверки квантовой электродинамики.

4. В 1955 г. на ускорителе протонов в Беркли (максимальная энергия 6,3 ГэВ) группой американских физиков были открыты *антипротоны.* Согласно закону сохранения барионного заряда (см. § 108) антипротон может образоваться только в паре с протоном (или нейтроном, если позволяет закон сохранения электрического заряда). Антипротоны получались при столкновениях ускоренных протонов с протонами же, входящими в состав ядер медной мишени. Пороговая энергия протонов (в лабораторной системе отсчета, где мишень покоится) равна 5,6 ГэВ (см. § 107, п. 4). Система отклоняющих магнитов отбирала отрицательно заряженные частицы, подавляющее большинство которых составляли π^--мезоны. Выделение антипротонов на подавляющем фоне π^--мезонов и представляло главную трудность эксперимента (например, при энергии 6,2 ГэВ на 62 000 π^--мезонов приходился один антипротон). Массы частиц определялись измерением их импульса (по отклонению в магнитном поле) и скорости. Последняя определялась двумя независимыми способами: по времени пролета между двумя люминесцентными счетчиками и счетчиками Черенкова. Черенковский счетчик был отъюстирован так, что он регистрировал только частицы со скоростями, равными скорости антипротонов, а более быстрые π-мезоны не регистрировал. На современных ускорителях получаются пучки антипротонов, содержащие до 10^6 частиц в пучке.

Через год после открытия антипротона был получен и *антинейтрон.* Антинейтроны получались перезарядкой антипротонов, т. е. в результате реакции

$$\overline{p} + p \to n + \overline{n}. \qquad (106.1)$$

Появление антинейтрона обнаруживалось наблюдением его аннигиляции с нуклонами. Будучи нейтральным, антинейтрон не ионизует атомы вещества, через которое он пролетает. По этой причине антинейтрон не оставляет следа в детекторе. Однако при аннигиляции антинейтрона возникает несколько заряженных частиц, следы которых выходят из одной точки (звезда).

5. Поскольку позитрон и антипротон так же стабильны, как и соответствующие им частицы, наряду с обычным веществом физика допускает существование и *антивещества.* Ядра атомов антивещества построены из антипротонов и антинейтронов. Их оболочки состоят из позитронов. Первое *антиядро — антидейтрон* d — было получено в 1965 г. в ускорителе группой американских физиков под руководством Ледермана (р. 1922). В 1969 г. на ускорителе в Серпухове (76 ГэВ) под руководством Ю. Д. Прокошкина (р. 1929) было зарегистрировано ядро антигелия $^3\overline{\text{He}}$, состоявшее из двух антипротонов и одного антинейтрона. В 1974 г. там же было получено ядро *антитрития* $^3\overline{\text{H}}$. Оно состоит из одного антипротона и двух антинейтронов. Во Вселенной антивещество астрономами не обнаружено. Не исключено, что во Вселенной антивещества и нет. (В первичных космических лучах число регистрируемых антипротонов примерно в 10^3–10^4 раз меньше числа протонов.) Причину такой асимметрии, если она существует, надо искать в происхождении и эволюции Вселенной.

Аннигилируют не только электрон с позитроном, но и всякая частица со своей античастицей. Однако при аннигиляции тяжелых частиц и античастиц возникают не столько γ-кванты, сколько другие легкие частицы. Причина этого в том, что аннигиляция позитрона с электроном происходит в результате *электромагнитного взаимодействия*, тогда как аннигиляция более тяжелых частиц и античастиц, являющихся адронами, вызывается *сильным взаимодействием.* Так, при аннигиляции протона с антипротоном доля появляющихся γ-квантов ничтожна по сравнению с долей π-мезонов. Аннигиляция же с испусканием нескольких (больше двух) π-мезонов более вероятна, чем двухмезонная. Среднее число всех π-мезонов (заряженных и нейтральных), появляющихся при аннигиляции одного медленного протона с одним антипротоном, равно приблизительно 4,8, а одних заряженных — 3,1.

§ 107. Законы сохранения энергии и импульса и их приложения

1. Законы сохранения в физике элементарных частиц играют значительно бóльшую роль, чем в макроскопической физике. Нам известны точные динамические законы, управляющие макроскопическими явлениями более детально, чем законы сохранения. Правда, и в макроскопической физике часто можно более быстро получить ответ с помощью одних только законов сохранения, не вникая в механизм явления. Но сами законы сохранения в макроскопической физике являются следствиями

динамических законов, так что здесь в принципе можно обойтись и без законов сохранения.

Иное положение в физике элементарных частиц. Здесь не существует сколько-нибудь законченной теории, тогда как законы сохранения хорошо соблюдаются. Некоторые из них встречаются и в макроскопической физике, но большинство являются новыми. Кроме того, в микромире законы сохранения приобретают новую особенность, не свойственную аналогичным законам в макромире. В макроскопической области явление может и не происходить, если оно даже удовлетворяет всем законам сохранения. Например, если на пути шара, катящегося по горизонтальной плоскости, поставить достаточно высокий барьер, для преодоления которого энергии шара недостаточно, то по классическим законам шар не может оказаться по другую сторону барьера, хотя это и не противоречит закону сохранения энергии и другим законам сохранения. Подобных барьерных запретов не существует в области микромира, поскольку там действуют квантовые законы (см. § 28, с. 163). *В микромире все явления должны происходить обязательно, если только они удовлетворяют всем законам сохранения.* Вероятность явления может быть очень мала, но оно рано или поздно произойдет, если только при этом будут соблюдены все законы сохранения.

2. Как можно считать сейчас установленным, каждый закон сохранения связан с какой-либо *симметрией законов природы*, хотя и не для всех законов эта симметрия выяснена. Так, в основе законов сохранения энергии $\mathscr{E}$, импульса $\mathbf{p}$ и момента импульса $\mathbf{M}$ лежат соответственно *однородность времени*, *однородность* и *изотропия пространства.* Разумеется, сами по себе пространство и время еще не включают понятий о различных физических величинах. Об этом уже говорилось в томе I. Сейчас же, предполагая, что изучающий ядерную физику уже успел ознакомиться с аналитической механикой, добавим, что перечисленные свойства пространства и времени в классической механике надо понимать в смысле *инвариантности функции Лагранжа* (или *Гамильтона*) относительно изменения начала отсчета времени, переноса начала координат и поворота координатных осей.

Аналогично обстоит дело и в квантовой механике, но на этом вопросе мы не будем останавливаться.

К точным законам сохранения, выполняющимся при любых взаимодействиях, относятся законы *сохранения энергии*, *импульса*, *момента импульса*, *электрического заряда*, *барионного заряда* и трех *лептонных зарядов.* Остальные законы сохранения: *странности*, *очарования*, *красоты*, *изотопического спина* и некоторые другие являются приближенными и выполняются не при всех взаимодействиях. Впрочем, следует заметить, что сохранение барионного и лептонного зарядов является эмпирическим законом и не имеет столь глубоких оснований, как сохранение $\mathscr{E}$, $\mathbf{p}$, $\mathbf{M}$. Поиски явлений с нарушением барионного и лептонного зарядов — одно из важнейших направлений современных экспериментальных исследований (см. § 108, пп. 3, 4).

3. Рассмотрим в этом параграфе важнейшие законы сохранения энергии и импульса и некоторые их применения.

В физике элементарных частиц помимо энергии покоя энергия встречается только в двух формах: *кинетической* и *потенциальной*. Полная энергия $\mathscr{E}$ равна их сумме и связана с релятивистской массой системы соотношением $\mathscr{E} = m_{\text{рел}}c^2$. В этом разделе физики принято пользоваться системой единиц, в которой скорость света c принимается за единицу (постоянная $\hbar$ также принимается равной единице). Тогда полная энергия частицы $\mathscr{E} = m_{\text{рел}}c^2 = m_{\text{рел}}$, а квадрат четырехмерного вектора энергии — импульса принимает вид

$$\mathscr{E}^2 - P^2 = m^2, \tag{107.1}$$

где под m понимается *масса покоя* частицы, часто называемая также *инвариантной массой* (обозначение m_0 не применяется; полная масса обозначена через $m_{\text{рел}}$). Таким образом, энергия, масса и импульс в указанной системе единиц имеют *одинаковую размерность*. Их единицей обычно служит гигаэлектронвольт.

4. С помощью законов сохранения энергии и импульса можно определять *энергетический порог* той или иной реакции между частицами. Рассчитаем, например, порог рождения антипротона в реакции столкновения двух протонов:

$$\mathrm{p} + \mathrm{p} \to \mathrm{p} + \mathrm{p} + \mathrm{p} + \overline{\mathrm{p}}. \tag{107.2}$$

Один из протонов (мишень) покоится, другой налетает на него. Требуется определить минимальную энергию налетающего протона, чтобы эта реакция стала возможной. Таким образом, реакция рассматривается в лабораторной системе отсчета.

Перейдем на мгновение в систему центра масс. Тогда протоны перед столкновением будут двигаться навстречу друг другу, так что их суммарный импульс будет равен нулю. Очевидно, необходимая минимальная энергия получится тогда, когда все четыре частицы после реакции будут находиться в состоянии покоя.

Теперь можно вернуться в лабораторную систему отсчета. В ней все четыре частицы будут двигаться с одной и той же скоростью, а следовательно, и с одним и тем же импульсом (так как массы частицы и античастицы одинаковы). Пусть кинетическая энергия налетающего протона равна $\mathscr{E}$, а, следовательно, полная энергия системы до столкновения равна $\mathscr{E} + 2m_{\mathrm{p}}$. Импульс налетающего протона $\mathbf{P}$ равен импульсу всей системы до, а следовательно, и после столкновения. Но после столкновения получается система четырех частиц, полная инвариантная масса которой равна $4m_{\mathrm{p}}$. Следовательно, на основании формулы (107.1)

$$(\mathscr{E} + 2m_{\mathrm{p}})^2 - P^2 = \text{Инв} = (4m_{\mathrm{p}})^2.$$

На основании той же формулы для налетающего протона

$$(\mathscr{E} + m_{\mathrm{p}})^2 - P^2 = \text{Инв} = m_{\mathrm{p}}^2,$$

откуда $P^2 = \mathscr{E}^2 + 2\mathscr{E}m_{\mathrm{p}}$. Исключив P и сократив на m_{p}, получим

$$\mathscr{E} = 6m_{\mathrm{p}} = 5{,}63 \text{ ГэВ}. \tag{107.3}$$

5. Допустим, что в реакцию вступают частицы с одинаковыми скоростями (по модулю и направлению). Тогда из законов сохранения энергии и импульса следует, что сумма масс получающихся частиц — продуктов реакции — не может превосходить сумму масс исходных частиц. Подчеркнем, что здесь речь идет об *инвариантных массах* (массах покоя). Допустим, например, что частицы a_1 и a_2 вступили в реакцию $a_1 + a_2 \to b_1 + b_2 + \ldots$ Так как частицы a_1 и a_2 движутся с одинаковыми скоростями, то можно рассмотреть реакцию в системе центра масс, в которой полный импульс системы равен нулю, а полная энергия равна сумме масс $m_{a_1} + m_{a_2}$. В конечном состоянии полный импульс системы, конечно, по-прежнему будет равен нулю. Но отдельные частицы, вообще говоря, могут приобрести скорости в разных направлениях. Полная энергия системы, разумеется, не изменится. Но она может быть представлена также суммой масс образовавшихся частиц. Однако это будут уже не массы покоя (инвариантные массы), а *релятивистские массы*, а они больше масс покоя. Для масс покоя (инвариантных масс) всех частиц можно поэтому написать

$$m_{a_1} + m_{a_2} \geqslant m_{b_1} + m_{b_2} + \ldots \tag{107.4}$$

Доказанное утверждение справедливо и для одной частицы, распадающейся на несколько других.

6. Законы сохранения энергии и импульса используются и для определения массы m и времени жизни τ нейтральных частиц. Примером может служить определение m и τ для нейтральных частиц Λ^0 и K^0.

Нейтральная частица, пролетая через атомы среды, не «обдирает» их электронные оболочки и поэтому не оставляет после себя следов в регистрирующих приборах (фотоэмульсия, пузырьковая камера, камера Вильсона и др.). Только в редких случаях нейтральная частица сталкивается с атомным ядром и вызывает взрыв последнего. Если при таком взрыве возникают заряженные частицы, то от места взрыва они разлетаются в разные стороны и оставляют следы (треки), образующие так называемую «звезду».

По треку заряженной частицы можно судить о ее энергии и импульсе. Чем длиннее трек частицы (например, в пузырьковой камере), тем больше ее энергия. Чем толще трек, тем меньше ее скорость в соответствующей точке трека. (Это справедливо для нерелятивистских частиц. В релятивистской области ионизация приблизительно постоянна.) Импульс частицы можно измерить по кривизне трека в магнитном поле: $P = eRH$. Направление же импульса $\mathbf{P}$ совпадает с направлением трека.

После этих отступлений вернемся к определению параметров K^0-мезона и Λ^0-гиперона. Уже в конце 40-х — начале 50-х годов при фотографировании треков космических лучей в камере Вильсона на фотопластинках были замечены следы пар заряженных частиц, исходящие из одной точки. Такие пары следов стали называть *вилками*. Очень часто вершины вилок находились вблизи звезд. Было высказано предположение, что при образовании звезды наряду с заряженными

частицами возникали и нейтральные, не оставляющие следов в камере. Исследования показали, что зубцы вилок в одних случаях — это следы π^+- и π^--мезонов, в других случаях — следы протонов p и π^--мезонов. Гипотетическую нейтральную частицу, при распаде которой возникают эти частицы, обозначили через V^0. Таким образом, предполагали, что эта частица распадается по одному из каналов:

$$\mathrm{V}^0 \to \pi^+ + \pi^-, \quad \mathrm{V}^0 \to \mathrm{p} + \pi^-. \tag{107.5}$$

Чтобы проверить это предположение, исследовали разные вилки и в каждом случае находили энергии и импульсы заряженных частиц, образующих вилку. Складывая энергии обеих частиц вилки, а по правилу параллелограмма — их импульсы, находили суммарную полную энергию $\mathscr{E}$ и импульс $\mathbf{P}$ частиц вилки. Эти величины должны быть равны энергии и импульсу незаряженной гипотетической частицы. По ним находили инвариантную массу этой частицы

$$m = \sqrt{\mathscr{E}^2 - P^2}. \tag{107.6}$$

Оказалось, что вилки, состоящие из $\pi^\pm$-мезонов, дают для массы гипотетической частицы около $0,500$ ГэВ, а вилки из протонов и π^--мезонов — около 1,11 ГэВ. Тем самым было доказано, что при образовании звезды возникали нейтральные частицы двух сортов. Более легкую назвали K^0-мезоном, а более тяжелую — Λ^0-гипероном. Они распадаются по схемам

$$\mathrm{K}^0 \to \pi^+ + \pi^-, \quad \Lambda^0 \to \mathrm{p} + \pi^-. \tag{107.7}$$

Зная импульсы и массы частиц K^0 и Λ^0, можно было вычислить их скорости. Измерив же расстояние от звезды до вершины вилки и разделив его на скорость частицы, можно было вычислить время ее жизни в лабораторной системе отсчета. А поскольку скорость частицы известна, можно было пересчитать это время и к системе отсчета, в которой частица покоится. При таком пересчете скорость частицы на пути от звезды к вершине вилки можно считать постоянной ввиду малости этого пути, а главное потому, что частица нейтральная и по этой причине слабо взаимодействует с окружающей средой.

7. Время жизни Λ^0-гиперона $12{,}6 \cdot 10^{-10}$ с. Такого же порядка время жизни K^0-мезона. Если считать, что скорость частицы порядка скорости света, то за это время она проходит расстояние около 3 см, которое легко измерить. Но не так обстоит дело с нейтральными частицами, время жизни которых порядка 10^{-16} с и меньше. К ним относится прежде всего π^0-мезон. Масса и время жизни π^0-мезона также были измерены с использованием законов сохранения энергии и импульса. Наиболее точно эти величины были найдены в результате изучения реакции распада K^+-мезона

$$\mathrm{K}^+ \to \pi^+ + \pi^0, \tag{107.8}$$

за которым следует распад π^0-мезона по схеме [1])

$$\pi^0 \to \gamma + e^+ + e^-. \qquad (107.9)$$

Так как частицы K^+ и π^+ заряженные, то энергии и импульсы этих частиц можно найти, изучая оставляемые ими треки. После этого из реакции (107.8) можно вычислить энергию (полную), а значит и скорость π^0-мезона. Затем надо измерить расстояние от конца трека K^+-мезона до точки, из которой исходят частицы пары e^+, e^-. Это расстояние и есть пробег π^0-мезона за время его жизни. Правда, измерение этого расстояния, составляющего доли микрометра, лежит на границе возможного. Так как скорость π^0-мезона известна, то по пробегу находится и время его жизни в лабораторной системе отсчета. Затем оно может быть пересчитано и к системе, в которой π^0-мезон покоится. По современным данным

$$m_{\pi^0} = (264{,}113 \pm 0{,}008) m_e, \quad \tau_{\pi^0} = (0{,}828 \pm 0{,}057) \cdot 10^{-16} \text{ с}.$$

8. При меньших временах жизни пробег нейтральной (а следовательно, невидимой) частицы не поддается прямому измерению. Рождение и распад нейтральной частицы происходят в столь малой области, что прямыми методами ее невозможно отличить от точки. Такое положение имеет место при рождении и распаде резонансных частиц (резонансов), времена жизни которых 10^{-23} с (пробег 10^{-13} см, т. е. порядка диаметра ядра). Обнаружение таких частиц и определение их масс и времен жизни возможны только косвенными методами. Они используют те же законы сохранения энергии и импульса в сочетании со статистической обработкой данных с помощью ЭВМ. Поясним это на примере.

В 1961 г. в Беркли (США) группа физиков открыла нейтральную частицу — так называемый резонанс ω^0. На пути пучка антипротонов $\bar{p}$, вышедших из ускорителя с кинетической энергией $\mathscr{E}_{\text{кин}} \sim 1{,}61$ ГэВ, была поставлена большая водородная пузырьковая камера. При столкновении антипротонов с протонами (ядрами водорода) образовывались $\pi^\pm$- и π^0-мезоны. Понятно, что суммарный электрический заряд всех образовавшихся частиц должен быть равен нулю. Легко оценить верхний предел для числа N образующихся π-мезонов.

Так как в пучке движутся только антипротоны, а сталкивающихся частиц $\bar{p}$ и p две, то полная энергия их будет $\mathscr{E}_{\text{кин}} + 2m_p$. Импульс системы до и после столкновения, очевидно, равен импульсу антипротона $\mathbf{P}$. Величина $(\mathscr{E}_{\text{кин}} + 2m_p)^2 - P^2$ есть инвариант и равна квадрату инвариантной массы системы. При столкновении эта величина не меняется. Но после столкновения инвариантную массу можно представить в виде Nm_π, если пренебречь разницей масс заряженного и нейтрального мезонов. В самом деле, в системе центра масс максимальное число π-мезонов N получится тогда, когда все π-мезоны получатся в состоянии

[1]) Этот распад в 80 раз менее вероятен распада $\pi^0 \to 2\gamma$, но для нашей цели важен именно этот распад.

покоя. В этом случае полная инвариантная масса всех мезонов будет Nm_π, а она во всех системах отсчета одинакова. Итак,

$$(\mathscr{E}_{\text{кин}} + 2m_\text{p})^2 - P^2 = (Nm_\pi)^2. \qquad (107.10)$$

Но для налетающего антипротона величина $(\mathscr{E}_{\text{кин}} + m_\text{p})^2 - P^2$ также инвариантна и равна m_p^2. Отсюда находим $P^2 = \mathscr{E}_{\text{кин}}^2 + 2m_\text{p}\mathscr{E}_{\text{кин}}$. Исключение P из этого и предыдущего уравнений дает

$$N^2 = 2m_\text{p}(\mathscr{E}_{\text{кин}} + 2m_\text{p})/m_\pi^2. \qquad (107.11)$$

Подставляя сюда $m_\text{p} = 0{,}938$ ГэВ, $m_\pi = 0{,}140$ ГэВ, $\mathscr{E}_{\text{кин}} = 1{,}61$ ГэВ, получаем $N \sim 18$. Это дает верхний предел для N. На самом деле при рассматриваемых энергиях он никогда не достигается. В таких случаях среднее $N \sim 5$–6. Были изучены фотоснимки с зарегистрированными на них звездами, образованными заряженными π-мезонами. Исследователи прежде всего отобрали 2500 «четырехлучевых звезд», т. е. таких, из которых исходили четыре трека. Исследуя треки каждой звезды, можно было вычислить полную инвариантную массу системы всех четырех π-мезонов. Если бы только они образовывались из исходной системы $\bar{\text{p}}\text{p}$, то инвариантная масса каждой звезды была бы равна исходной массе $2m_\text{p}$. На самом деле это случалось редко. Та же картина получалась бы, если бы звезда образовывалась от распада на четыре π-мезона одной нейтральной частицы, образующейся при столкновении антипротона с протоном. Это тоже встречалось редко.

Поэтому было высказано предположение, что в каждой звезде, помимо четырех заряженных π-мезонов, должны присутствовать какие-то нейтральные частицы. Можно предположить, что звезды содержали только одну нейтральную гипотетическую частицу. Эту гипотетическую частицу мы обозначим через ω^0. Полные энергия $\mathscr{E}$ и импульс P системы при столкновении не меняются. Эти величины известны, поскольку антипротоны получаются от ускорителя с заранее известной энергией. Поэтому энергию и импульс частицы ω^0 можно вычислить по формулам

$$\mathscr{E}_{\omega^0} = \mathscr{E} - \mathscr{E}_1 - \mathscr{E}_2 - \mathscr{E}_3 - \mathscr{E}_4, \quad P_{\omega^0} = P - P_1 - P_2 - P_3 - P_4,$$

а ее инвариантную массу — по формуле $m_{\omega^0}^2 = \mathscr{E}_{\omega^0}^2 - P_{\omega^0}^2$. Здесь цифрами обозначены энергии и импульсы отдельных заряженных π-мезонов. Оказалось, что $m_{\omega^0} \approx 0{,}135$ ГэВ — величина, характерная для π^0-мезонов. Из всех 2500 четырехлучевых звезд было отобрано 800 звезд, в каждой из которых содержался один π^0-мезон.

По трекам можно было определить энергию и импульс каждого заряженного π-мезона. Эти же величины для нейтрального π^0-мезона можно было вычислить по формулам, приведенным выше.

При исследовании четверок и троек заряженных π-мезонов получались различные соответствующие им инвариантные массы. Если по оси абсцисс откладывать инвариантные массы троек заряженных π-мезонов, а по оси ординат — число случаев, в которых они появляются на различных небольших интервалах энергии определенной величины,

то получится плавная кривая без сколько-нибудь заметных максимумов и минимумов. Это указывает на статистический, некоррелированный характер появления соответствующих масс в каждой четверке или тройке заряженных π-мезонов.

Если же брать тройки π-мезонов, в каждой из которых содержится один нейтральный π-мезон, то картина резко меняется. На определенном месте на кривой появляется высокий узкий максимум, свидетельствующий о том, что частицы каждой тройки рассматриваемого вида не независимы, а коррелируют между собой. Это связано с тем, что все мезоны таких троек получаются в результате распада *одной и той же* нейтральной частицы, возникшей при столкновении антипротона с протоном. Такую частицу мы обозначили через ω^0. Изученная реакция идет через появление промежуточной нейтральной частицы ω^0:

$$\begin{aligned} \overline{\mathrm{p}} + \mathrm{p} \to \pi^+ + \pi^- + \omega^0 \\ \hookrightarrow \pi^+ + \pi^- + \pi^0. \end{aligned} \tag{107.12}$$

Вершине соответствующей кривой соответствует определенное значение массы, которая и принимается за массу частицы ω^0. Становится понятным, почему частица ω^0 названа *резонансной частицей*, или *резонансом.* Сам резонанс характеризуется определенной шириной (полушириной) Г. По этой ширине и определяется время жизни частицы $\tau \sim \hbar/\Г$. Для ω^0-резонанса $m_{\omega^0} = (782{,}6 \pm 0{,}3)$ ГэВ, $\Г = 10{,}1$ МэВ ($\tau = 6{,}52 \cdot 10^{-23}$ с).

9. Теперь скажем несколько слов о законе сохранения момента импульса. Момент импульса определяется одинаково в нерелятивистской и релятивистской классической механике, т. е. формулой $\mathbf{M} = [\mathbf{rP}]$. Той же формулой, но в операторной форме орбитальный момент $\mathbf{M}$ определяется в квантовой механике. Собственный (или спиновый) момент частицы определяется иначе. Об этом вскользь было сказано в § 36 (п. 5). Подробное рассмотрение этого вопроса выходит за рамки общего курса физики. В квантовой механике момент импульса квантуется и, кроме того, частица может обладать внутренним моментом — спином. Сохраняется полный момент: орбитальный плюс спиновый. Например, в системе центра масс в распаде $\rho^0 \to \pi^+ + \pi^-$ спиновый момент равен 1, а пионы возникают в p-состоянии; в распаде $\Lambda^0 \to \mathrm{p} + \pi^-$ спин равен $1/2$, а p и π^- могут быть в s- и p-состояниях.

§ 108. Законы сохранения электрического, лептонных и барионного зарядов

1. Названные законы выполняются для всех взаимодействий (сильного, электромагнитного, слабого; для гравитационного взаимодействия элементарных частиц они, разумеется, не проверены). Их происхождение неясно, в частности не выяснено, с какой симметрией они связаны. Существуют веские основания считать, что закон сохранения электрического заряда — точный (масса фотона нулевая, на опыте

установлено, что $m_\Phi < 10^{-21} m_e$). Законы сохранения лептонных и барионного зарядов — чисто эмпирические, и не исключено, что они могут нарушаться (см. пп. 3, 4).

Электрический заряд Q частицы играет двоякую роль. Во-первых, он является *источником силового поля.* Во-вторых, это есть *квантовое число*, характеризующее частицу. Лептонные и барионные заряды, насколько известно в настоящее время, не являются источниками каких бы то ни было дальнодействующих силовых полей. Это просто квантовые числа, характеризующие внутренние свойства элементарных частиц.

2. Различают три вида лептонных зарядов: *электронный* L_e (для e и ν_e), *мюонный* L_μ, (для μ и ν_μ), *таонный* L_τ (для τ и ν_τ). Барионный заряд один и обозначается через B.

С помощью лептонных зарядов простейшим образом интерпретируется установленный на опыте закон, согласно которому в замкнутой системе при любых процессах остается постоянной разность между числом лептонов и антилептонов каждого вида. То же относится и к барионам. Условились считать, что лептонный заряд L_e равен $+1$ (для e^- и ν_e), $L_\mu = +1$ (для μ^- и ν_μ), $L_\tau = +1$ (для τ_+ и $\overline{\nu}_\tau$) и -1 для всех антилептонов (e^+, $\overline{\nu}_e$, μ^+, $\overline{\nu}_\mu$, τ^+, $\overline{\nu}_\tau$). Для всех остальных элементарных частиц лептонные заряды принимаются равными нулю. Барионный заряд принимается равным $+1$ для всех барионов и барионных резонансов и -1 для их античастиц. Все заряды аддитивны. Это значит, что для сложной системы заряд каждого вида равен сумме зарядов того же вида всех входящих в нее элементарных частиц. Например, барионный заряд атомного ядра равен сумме всех барионных зарядов нуклонов, из которых построено ядро. Иными словами, барионный заряд ядра равен массовому числу A.

3. Ни одного случая нарушения закона сохранения лептонного заряда не обнаружено. Однако точность, с которой проверен этот закон, невелика. Не исключено, что закон сохранения лептонного заряда приближенный и в определенных условиях может нарушаться. Если бы это было так, то в вакууме могли бы осуществляться превращения нейтрино в антинейтрино и наоборот, а также превращения одного типа нейтрино в другие. Поиски этих явлений представляют большой интерес.

Закон сохранения лептонного заряда требует, например, чтобы при электронном β-распаде ($n \to p + e^- + \overline{\nu}_e$) вместе с электроном обязательно рождалось электронное антинейтрино, так как суммарный лептонный заряд этих двух частиц равен нулю. При позитронном β-распаде ($p \to n + e^+ + \nu_e$), наоборот, вместе с позитроном должно появляться электронное нейтрино.

Законом сохранения лептонного заряда объясняется отсутствие таких процессов, как

$$\nu_e + p \nrightarrow e^+ + n, \quad \nu_\mu + p \nrightarrow \mu^+ + n, \tag{108.1}$$

хотя другими законами сохранения они разрешены. Напротив, процессы

$$\overline{\nu}_e + p \to e^+ + n, \quad \overline{\nu}_\mu + p \to \mu^+ + n, \tag{108.2}$$

удовлетворяющие закону сохранения лептонного заряда, наблюдаются экспериментально. Первый процесс наблюдается в ядерных реакторах при взаимодействии электронных антинейтрино с протонами вещества. Второй процесс наблюдался на ускорителях высоких энергий, где получались мюонные нейтрино в реакциях распада π^-- и K^--мезонов; $\pi^- \to \mu^- + \overline{\nu}_\mu$ и $K^- \to \mu^- + \overline{\nu}_\mu$. Эти антинейтрино и взаимодействовали с протонами вещества.

Примеры (108.1) и (108.2) показывают, что нейтрино (как электронное, так и мюонное) не тождественны своим античастицам. В 1962 г. было установлено (в частности, по отсутствию реакции $\nu_\mu + n \to e^- + p$), что ν_e и ν_μ — разные частицы. После этого и были введены разные лептонные заряды L_e и L_μ. Аналогично обстояло дело с введением $\tau^\pm$-лептонов и таонных нейтрино $\nu_\tau \overline{\nu}_\tau$ и соответствующего им лептонного заряда L_τ.

4. Из закона сохранения барионного заряда следует, что протон не может превратиться, например, в позитрон и фотон, хотя при таком превращении не нарушались бы ни закон сохранения энергии, ни другие уже рассмотренные выше законы сохранения. Если бы такое превращение было возможно, то оно неизбежно привело бы к аннигиляции атомов вещества, так как образовавшиеся позитроны аннигилировали бы с электронами атомных оболочек. Из того же закона следует, что антибарион может рождаться только в паре со своим барионом. Так, антипротон рождается в реакции $p+p \to p+p+p+\overline{p}$. При столкновении двух протонов могут возникнуть и два антипротона. Но тогда появятся и два новых протона — всего получится шесть частиц, а это значительно повысит порог реакции (см. § 107, п. 4). Аналогично, среди продуктов распада бариона должен обязательно присутствовать также барион. Могут появиться и несколько барионов, но совместно с антибарионами, число которых будет на единицу меньше. Все сказанное относится и к лептонам.

Не исключено, что барионный заряд не всегда строго соблюдается. В так называемой теории великого объединения допускается возможность распада протона на более легкие частицы, например $p \to \pi^0 + e^+$. Такой распад был бы связан с изменением барионного заряда. Экспериментально распад протона пока не обнаружен. На опыте установлено, что время жизни протона $\tau > 10^{32}$ лет.

ЗАДАЧИ

1. Определить порог рождения антипротона:
а) при соударении электрона с покоящимся протоном;
б) при соударении электрона с покоящимся электроном;
в) при соударении фотона с покоящимся протоном;
г) при соударении фотона с покоящимся электроном.

Указание. Задачи решаются аналогично задаче из § 107, п. 4. При выборе соответствующей реакции необходимо руководствоваться законом сохранения барионного заряда. В частности, надо учесть, что антипротон всегда рождается в паре с протоном. В соответствии с этим надо выбрать реакции:

а) $$e^- + p \to e^- + p + p + \overline{p}, \tag{108.3}$$

б) $$e^- + e^- \to e^- + e^- + p + \overline{p}, \tag{108.4}$$

в) $$\gamma + p \to \gamma + p + p + \overline{p}, \tag{108.5}$$

г) $$\gamma + e^- \to \gamma + e^- + p + \overline{p}. \tag{108.6}$$

При рассмотрении реакций с участием γ-квантов надо использовать соотношение $P_\gamma = \mathscr{E}_\gamma$. Порог реакции соответствует случаю, когда энергия γ-кванта после реакции стремится к нулю. Пороги приведенных реакций соответственно равны:

а) $$4m_p + 2m_e = 3{,}754 \text{ ГэВ}, \tag{108.3а}$$

б) $$2m_p\,(m_p/m_e + 2) = 3446 \text{ ГэВ}, \tag{108.4а}$$

в) $$4m_p = 3{,}753 \text{ ГэВ}, \tag{108.5а}$$

г) $$2\,(m_p/m_e + 1)\,m_p = 1724 \text{ ГэВ}. \tag{108.6а}$$

2. Определить порог реакции

$$p + p \to p + p + \pi^0. \tag{108.7}$$

Ответ.

$$2m_{\pi^0} + \frac{m_{\pi^0}^2}{2m_p} = 284 \text{ МэВ}. \tag{108.7а}$$

3. K^+-мезоны распадаются по схеме $K^+ \to \mu^+ + \nu_\mu$. По мере уменьшения скорости мезона в среде след его становится все более извилистым и запутанным — частица начинает беспорядочно блуждать между атомами. Наконец, след обрывается из-за остановки частицы. Из точки остановки исходят две частицы — μ^+ и ν_μ. Заряженный мюон μ^+ оставляет след, нейтрино следа не оставляет. По следу заряженной частицы и можно судить о характере процесса. Рассчитать энергию мюона и нейтральной частицы в точке, из которой они исходят.

То же самое сделать, заменив K^+-мезон π^+-мезоном, т. е. для распада $\pi^+ \to \mu^+ + \nu_\mu$.

Решение. Полная энергия остановившегося мезона равна его массе m_{K^+}. Если $\mathscr{E}_\mu$ и $\mathscr{E}_\nu$ — полные энергии мюона и нейтрино в момент их образования, то

$$\mathscr{E}_\mu + \mathscr{E}_\nu \to m_K.$$

Для мюона

$$\mathscr{E}_\mu^2 - P_\mu^2 \to m_\mu^2.$$

Наконец, для нейтрино $P_\nu = \mathscr{E}_\nu$. Из этих трех уравнений однозначно получаем (так как $\mathscr{E}_\mu \neq \mathscr{E}_\nu$)

$$\mathscr{E}_\mu = \frac{m_K}{2} + \frac{m_\mu^2}{2m_K}, \quad \mathscr{E}_\nu = \frac{m_K}{2} - \frac{m_\mu^2}{2m_K}.$$

Итак, более легкая частица (нейтрино) получает меньшую энергию. Не противоречит ли это известному утверждению, что при распаде неподвижной

частицы большая энергия уносится меньшей частицей? Нет, не противоречит. Это утверждение относится не к полной, а к кинетической энергии. Кинетические энергии частиц будут

$$\mathscr{E}_\mu^{\text{кин}} = \frac{m_{\text{K}^+}}{2} + m_\mu \left(1 - \frac{m_\mu}{2m_{\text{K}^+}}\right) = 152 \text{ МэВ},$$

$$\mathscr{E}_\nu^{\text{кин}} \equiv \mathscr{E}_\nu = \frac{m_{\text{K}^+}}{2} - \frac{m_\mu^2}{2m_{\text{K}^+}} = 236 \text{ МэВ}.$$

В случае реакции с π^+-мезоном справедливы те же формулы, если сделать замену $m_{\text{K}^+} \to m_\pi$. Для этого случая получается $\mathscr{E}_\mu^{\text{кин}} = 4{,}36$ МэВ, $\mathscr{E}_\nu^{\text{кин}} = 29{,}8$ МэВ. Так как согласно общему правилу (107.4): $m_{\text{K}^+} > m_\mu$, $m_{\pi^+} > m_\mu$, то всегда получается $\mathscr{E}_\nu^{\text{кин}} > \mathscr{E}_\mu^{\text{кин}}$.

Найденные числа поучительны еще в одном отношении. Кинетическая энергия мюона в случае распада K^+-мезона равна 152 МэВ, тогда как в случае распада π^+-мезона она составляет всего 4,36 МэВ. Отсюда следует, что трек мюона при распаде останавливающегося K^+-мезона должен быть во много раз длиннее трека мюона при распаде π^+-мезона. Это позволяет отличить один распад от другого.

§ 109. Другие законы сохранения и квантовые числа

1. Было замечено, что гипероны интенсивно рождаются при столкновениях адронов высоких энергий. Это указывает на то, что их рождение связано с *сильными* взаимодействиями. Поэтому казалось бы, что время жизни гиперонов должно быть порядка 10^{-23} с, что характерно для процессов, связанных с сильными взаимодействиями. Это время жизни примерно в 10^{13} раз меньше найденного экспериментально для гиперонов. Это казалось странным.

Проблема указанного странного поведения частиц была решена на основе гипотезы парного рождения частиц, подтвержденной на ускорителях. Было установлено, что при столкновениях π-мезонов и нуклонов с нуклонами гипероны всегда рождаются *парами*, или *большими группами* совместно с K-мезонами или другими гиперонами. Ранее наблюдалось совместное рождение только частиц с античастицами. Здесь же парами рождались совсем другие частицы. Например, при столкновении протонов наблюдалась реакция

$$\text{p} + \text{p} \to \text{p} + \Lambda^0 + \text{K}^+, \tag{109.1}$$

причем Λ^0-гиперон появлялся только совместно с K^+-мезоном или Σ^+-гипероном, но никогда не появлялся вместе с K^--мезоном или Σ^--гипероном.

Гипероны и K-мезоны были названы *странными частицами*. Для K-мезонов также характерны относительно большие времена жизни (и даже на 1–2 порядка бóльшие, чем у гиперонов). Это время, как и для гиперонов, того же порядка, что и у заряженных пионов. Но последние распадаются в результате слабого взаимодействия главным

образом по схеме

$$\pi^+ \to \mu^+ + \nu_\mu, \quad \pi^- \to \mu^- + \overline{\nu}_\mu. \tag{109.2}$$

Это наводит на мысль, что гипероны и K-мезоны распадаются также за счет *слабого взаимодействия.* Все факты свидетельствуют, что это действительно так. В частности, относительная вероятность слабых взаимодействий в исследуемой области энергий на 12–14 порядков меньше вероятности сильных взаимодействий. Это и ведет к увеличению времени жизни примерно в то же число раз.

2. Для количественного описания парного рождения и истолкования относительно большого времени жизни странных частиц Гелл-Манн (р. 1929) и независимо от него Нишиджима (р. 1926) ввели новое квантовое число S, которое было названо *странностью.* Поведение странных частиц можно объяснить, если предположить, что частицы Λ^0, Σ^+, Σ^-, Σ^0, K^-, $\overline{\mathrm{K}}^0$ имеют странность -1, частицы $\overline{\Lambda}^0$, $\overline{\Sigma}^+$, $\overline{\Sigma}^-$, $\overline{\Sigma}^0$, $\overline{\mathrm{K}}^+$, K^0 — странность $+1$, частицы Ξ^-, Ξ^0 — странность -2, частицы $\overline{\Xi}^-$, $\overline{\Xi}^0$ — странность $+2$, частица Ω^- — странность -3, а нуклоны, пионы и η^0-мезоны лишены странности. Далее, надо предположить, что странность аддитивна, в сильных и электромагнитных взаимодействиях она сохраняется, а в слабых может меняться на ± 1.

Странные частицы рождаются в сильных взаимодействиях. Так как при этом странность не меняется, то странные частицы могут рождаться только парами частиц с противоположными странностями. Вот почему в реакции (109.1) появляются две частицы Λ^0 и K^+ с противоположными странностями, но не частицы Λ^0 и K^-, имеющие странности, знаки которых совпадают. В реакции

$$\mathrm{p} + \mathrm{p} \to \Xi^0 + \mathrm{p} + \mathrm{K}^0 + \mathrm{K}^+ \tag{109.3}$$

одновременно вместе с частицей Ξ^0, странность которой равна -2, появляются два странных мезона K^0 и K^+, суммарная странность которых равна $+1 + 1 = 2$.

В процессах распада странных частиц меняется странность продуктов распада. Это указывает на то, что эти процессы не вызываются сильными и электромагнитными взаимодействиями. В противном случае странность не менялась бы. Значит, за распад странных частиц ответственны *слабые силы.* Вот почему скорость распада уменьшается в 10^{12}–10^{14} раз по сравнению со скоростями распадов, вызываемых сильными взаимодействиями.

Сильные взаимодействия не способны вызывать распад странных частиц. Масса странных частиц (каонов, Λ^0, $\Sigma^\pm$, Ξ^0, Ξ^-, Ω^--гиперонов) такова, что распады их с сохранением странности на более мелкие частицы невозможны; таким образом, эти распады не могут происходить благодаря сильному взаимодействию, в котором странность сохраняется. Например, если бы было $M_{\Sigma^+} > M_{\mathrm{n}} + M_{\pi^+}$, то мог бы происходить распад $\Sigma^+ \to \mathrm{n} + \pi^+$. Но на самом деле $M_{\Sigma^+} < M_{\mathrm{n}} + M_{\pi^+}$, так что рассматриваемый распад невозможен.

Есть случаи, когда распад странных частиц происходит за счет сильного взаимодействия. Например,

$$K^*_{892} \to K_{494} + \pi, \quad \Sigma^*_{1382} \to \Lambda_{1232} + \pi.$$

Странность здесь сохраняется, поскольку распад идет за счет сильного взаимодействия. (Странность резонансов K^*, Σ^*, Λ, а также K равна $+1$, π-мезонов — 0. Нижний цифровой индекс означает массу соответствующей частицы в мегаэлектронвольтах. Ширины резонансов K^* и Σ^* равны соответственно 51 и 35 МэВ, так что их времена жизни порядка 10^{-23} с.)

Электромагнитные взаимодействия происходят без изменения странности, но в исключительных случаях могут вызвать распад странных частиц. Примером может служить распад $\Sigma^0 \to \Lambda^0 + \gamma$, происходящий под действием электромагнитных сил. Здесь энергии Σ^0 достаточно для рождения Λ^0-гиперона и γ-кванта, а странность не меняется. (Эта реакция не может происходить за счет сильного взаимодействия, так как $M_{\Sigma^0} < M_{\Lambda^0} + M_{\pi^0}$.) За исключением подобных редких случаев распад странных частиц вызывается слабыми силами.

Вместо странности S часто используют гиперзаряд Y, определяемый соотношением

$$Y = B + S. \tag{109.4}$$

Так как барионный заряд целочисленный, аддитивен и сохраняется, то гиперзаряд обладает теми же свойствами, что и странность, и совпадает с ней для частиц, у которых $B = 0$.

3. Аналогами квантового числа S являются также аддитивные целочисленные квантовые числа: *очарование* (*шарм*) C и *красота* (*прелесть*) b. Эти величины сохраняются не во всех, а только в сильных и электромагнитных взаимодействиях. Значения C приводятся в таблицах в конце книги. Об очаровании C и красоте b более подробно говорится в следующем параграфе.

Упомянем еще о *законе сохранения четности*, о котором подробно говорилось в § 69. Четность сохраняется во всех взаимодействиях за исключением слабого, в котором она нарушается. Заметим еще, что состояние со спином J и четностью P изображается символом J^P (например, 1^+ или $1/2^-$).

4. Существуют адроны, весьма близкие по своим физическим свойствам, объединяемые в группы, называемые *изотопическими мультиплетами* (*дуплетами*, *триплетами* и т. д.). Они одинаковым образом участвуют в сильных взаимодействиях, имеют приблизительно равные массы, одни и те же барионный заряд, спин, одинаковые внутреннюю четность, странность и отличаются друг от друга электромагнитными характеристиками (электрический заряд, магнитный момент). Если бы не было электромагнитных и слабых взаимодействий, то все свойства таких частиц были бы одинаковыми. Так, объединяются в изотопический дуплет протон и нейтрон. Эти две частицы рассматриваются как различные квантовые состояния *одной и той же частицы — нуклона.* Существуют изотопические триплеты

частиц, например (π^-, π^0, π^+), $(\Sigma^-, \Sigma^0, \Sigma^+)$, а также мультиплеты (особенно среди резонансов), состоящие из большего числа частиц. Существуют и одиночные частицы, не входящие в изотопические мультиплеты; они называются *синглетами*.

Число частиц в мультиплете по одной из оправдавших себя гипотез можно представить формулой

$$N = 2T + 1, \tag{109.5}$$

где T принимает одно из значений $T = 0, 1/2, 1, 3/2$. Эта формула аналогична формуле, определяющей число возможных проекций углового момента (спина) на избранное направление. Чтобы провести эту формальную аналогию дальше, вводят некоторое абстрактное изотропное пространство (не имеющее ничего общего с обычным пространством), которое называется *изотопическим пространством* (правильнее его надо было бы назвать *изобарическим пространством*, но по случайным причинам этот термин не привился). В этом пространстве каждому адрону соответствует некоторое направление и на нем откладывается вектор длины T. Длина этого вектора называется *изотопическим спином* или, короче, *изоспином* соответствующего адрона. В изотопическом пространстве можно произвольно выбрать некоторую ось и назвать ее *изотопической осью*. Проецированием вектора $\mathbf{T}$ на эту ось получают систему точек, отстоящих одна от другой на расстояние 1. Каждой такой проекции соответствует частица мультиплета. Проекцию вектора $\mathbf{T}$ на изотопическую ось мы будем обозначать через T_3. При заданном изоспине T проекция T_3 может принимать следующие значения:

$$T_3 = -T, -(T-1), \ldots, +(T-1), +T.$$

Так, нуклон (протон и нейтрон) имеет изоспин $T = 1/2$; протону условились приписывать проекцию $T_3 = +1/2$, а нейтрону — $T_3 = -1/2$. Трем π-мезонам соответствует изоспин $T = 1$ $(3 = 2 \cdot 1 + 1)$, и т. д.

Все изложенное носит чисто формальный характер. Реальной физической величиной пока что является только число частиц N в изотопическом мультиплете, и совсем не обязательно представлять это число формулой (109.5). Однако можно указать и утверждения *физического характера*, которые нагляднее всего формулируются с использованием понятия *изотопического спина*. Для этого надо путем определения ввести правило, по которому находится изоспин системы адронов по изоспинам частиц, из которых состоит эта система. Условились складывать векторы изоспинов частиц по тому же правилу *векторного сложения*, по которому складываются угловые моменты (спины) частиц. Существенно заметить, что изоспин системы зависит не только от изоспинов всех составляющих частиц, но и от *углов* между векторами изоспинов этих частиц. Поэтому при одном и том же составе систем они могут обладать различными изоспинами. Например, изоспин системы, состоящей из нуклона и пиона, может быть либо $1/2$ (когда изоспины этих частиц направлены противоположно), либо $3/2$ (когда они направлены одинаково).

Физическое утверждение, о котором упоминалось выше, — это закон *сохранения изотопического спина.* Этот закон заключается в том, что изотопический спин сохраняется при *сильных взаимодействиях*, но нарушается в электромагнитных и в других взаимодействиях.

Сильное взаимодействие для всех частиц, входящих в один и тот же изотопический мультиплет, одинаково, т. е. не зависит от электрических зарядов частиц. В этом проявляется так называемая *изотопическая инвариантность* элементарных частиц, присущая сильному взаимодействию. Частным случаем ее является зарядовая независимость ядерных сил. Формально математически изотопическая инвариантность может быть интерпретирована как независимость сильного взаимодействия от вращения в изотопическом пространстве.

На основании изотопической инвариантности удается предсказать существование, массу и заряд новых частиц, если известны их изотопические «партнеры». Именно так было предсказано существование и свойства π^0, Σ^0, Ξ^0 по известным $\pi^\pm$, $\Sigma^\pm$ и Ξ^-.

Приведем другой пример. Из изотопической инвариантности следует, что вероятность реакции $\mathrm{p}+\mathrm{p}\to\mathrm{d}+\pi^+$ вдвое больше вероятности реакции $\mathrm{n}+\mathrm{p}\to\mathrm{d}+\pi^0$ (обе реакции идут за счет сильного взаимодействия). Это следует из того, что в конечном состоянии обеих реакций суммарный изоспин $T=1$ ($T=1$ для π и $T=0$ для d). В начальном состоянии первой реакции всегда $T=1$ (так как $T_3=1/2+1/2=1$), а для второй реакции суммарный изоспин может быть 1 или 0 (так как $T_3=1/2-1/2=0$). Поэтому по закону сохранения изотопического спина первая реакция может идти при любых начальных состояниях, тогда как вторая — только для половины начальных состояний.

Поскольку изотопическая инвариантность имеет место только для сильных и нарушается для электромагнитных взаимодействий, точность предсказаний на ее основе по порядку величины равна отношению сил электромагнитного и сильного взаимодействий, т. е. составляет примерно 1 %.

5. Заслуживает внимания следующее правило. Допустим, например, что наблюдается реакция

$$a+b\to c+d. \tag{109.6}$$

Здесь a, b, c и d изображают какие-то частицы. Если такая реакция идет, то должны выполняться все законы сохранения, соответствующие вызывающему ее взаимодействию. Обратно, если все законы сохранения выполняются, то реакция (109.6) должна обязательно идти (см. § 107, п. 1). (Для слабых взаимодействий суммарные странность, очарование, красота не сохраняются.)

Рассмотрим в качестве примера барионный заряд B, которым обладает, скажем, частица b. Если эту частицу перенести из левой части в правую, то барионный заряд слева уменьшится, а справа увеличится на B. Но если при таком переносе частицу b одновременно заменить античастицей $\bar{b}$, то барионный заряд и справа уменьшится на B. Равенство барионных зарядов восстановится. То же относится и к другим

зарядам. Законы сохранения допускают возможность реакции

$$a \to c + d + \overline{b}. \tag{109.6а}$$

Вообще, *если какую-либо частицу или группу частиц перенести из одной части равенства в другую, заменив их античастицами, то получится соотношение, выражающее возможную новую реакцию.*

При доказательстве не использованы законы сохранения энергии, импульса и момента импульса. Но их и не надо было учитывать, так как эти величины характеризуют не сами частицы, а состояние их движения. А эти состояния могут изменяться в широких пределах, определяемых не законами сохранения, а начальными условиями. Аналогично, при переносе частиц из-за вращения изоспинов в изотопическом пространстве получатся состояния с требуемыми суммарными изоспинами.

Приведем примеры. Возьмем реакцию распада нейтрона

$$\mathrm{n} \to \mathrm{p} + \mathrm{e}^- + \overline{\nu}_\mathrm{e}. \tag{109.7}$$

Перенесем антинейтрино $\overline{\nu}_\mathrm{e}$ справа налево, заменив его на нейтрино ν_e. Получим

$$\nu_\mathrm{e} + \mathrm{n} \to \mathrm{p} + \mathrm{e}^-. \tag{109.8}$$

Мы получили реакцию взаимодействия нейтрино с нейтроном с образованием протона p и электрона e^-.

Возьмем, далее, процесс комптоновского рассеяния кванта на электроне

$$\gamma + \mathrm{e}^- \to \gamma + \mathrm{e}^-. \tag{109.9}$$

Здесь частица γ совпадает со своей античастицей. Перенесем ее слева направо, а электрон с соответствующей заменой — справа налево. Возникнет процесс

$$\mathrm{e}^+ + \mathrm{e}^- \to \gamma + \gamma, \tag{109.10}$$

выражающий аннигиляцию пары $\mathrm{e}^+\mathrm{e}^-$ с испусканием двух γ-квантов.

Более интересен следующий пример. Процесс распада

$$\Sigma^0 \to \mathrm{p} + \mathrm{e}^- + \overline{\nu}_\mathrm{e}$$

в принципе возможен, но его не наблюдают, так как идет распад $\Sigma^0 \to$ $\to \Lambda^0 + \gamma$, который на 12–13 порядков более вероятен. Однако взаимодействие, приводящее к этому процессу, можно изучать в нейтринных опытах:

$$\overline{\nu}_\mathrm{e} + \mathrm{p} \to \Sigma^0 + \mathrm{e}^+, \quad \overline{\nu}_\mu + \mathrm{p} \to \Sigma^0 + \mu^+.$$

§ 110. Кварковая модель адронов

1. Обилие уже открытых и вновь открываемых адронов (в том числе резонансов) наводит на мысль, что все они построены из каких-то других, более фундаментальных частиц. С наибольшим успехом эта идея реализована в кварковой модели адронов. Кварковая модель объясняет не только систематику, но и динамику адронов. Она приводит

к массе оправдывающихся предсказаний и в настоящее время считается общепризнанной.

В 1964 г. Гелл-Манном и независимо от него Цвейгом (р. 1937) была выдвинута гипотеза, подтвержденная дальнейшими исследованиями, что все адроны построены из более фундаментальных частиц, которые, по предложению Гелл-Манна, были названы *кварками.* На основе кварковой гипотезы не только была понята структура уже известных адронов, но и предсказано существование новых. Ниже дается понятие о кварковой модели без каких-бы то ни было претензий на полноту и обоснование полученных результатов.

К настоящему времени установлено существование пяти разновидностей (или так называемых *ароматов*) кварков: u, d, s, c, b. Их массы: $m_u \approx 5$ МэВ, $m_d \approx 7$ МэВ, $m_s \approx 150$ МэВ, $m_c \approx 1{,}3$ ГэВ, $m_b \approx 5$ ГэВ. На эти данные надо смотреть как на оценочные и грубо ориентировочные, так как кварки в свободном состоянии не наблюдались и поэтому их массы нельзя было измерить прямыми методами (см. п. 7). Неоднократно поступали предварительные сообщения об открытии t-кварка ($m_t \gtrsim 22$ ГэВ), но окончательно существование t-кварка еще не установлено. Каждому кварку соответствует свой антикварк.

Все кварки имеют спин $1/2$ и барионный заряд $1/3$. Кварки u, c, t называют *верхними*, так как они имеют дробный электрический заряд $+2/3$. Остальные кварки d, s, b с электрическим зарядом $-1/3$ принято называть *нижними.* В соответствии с этой терминологией кварки можно расположить в таблице:

$$u, c, t \text{ (заряд } 2/3); \quad d, s, b \text{ (заряд } -1/3). \tag{110.1}$$

Кварк s является носителем странности, c — очарования, b — красоты (прелести).

Протон состоит из двух u-кварков и одного d-кварка ($p \to uud$), нейтрон состоит из одного u-кварка и двух d-кварков ($n \to udd$). Их античастицы построены из антикварков: $\overline{p} \to \overline{u}\,\overline{u}\,\overline{d}$, $\overline{n} \to \overline{u}\,\overline{d}\,\overline{d}$.

2. Мезоны построены из двух частиц: кварка и антикварка. Поэтому их барионное число равно нулю. Посмотрим, какие комбинации из двух таких частиц могут быть составлены из трех самых легких кварков u, d, s и их антикварков $\overline{u}$, $\overline{d}$, $\overline{s}$. Очевидно, всего таких комбинаций будет девять:

$$\begin{array}{lll} u\overline{u}, & u\overline{d}, & u\overline{s}, \\ d\overline{u}, & d\overline{d}, & d\overline{s}, \\ s\overline{u}, & s\overline{d}, & s\overline{s}. \end{array} \tag{110.2}$$

Кварки и антикварки могут входить в эти (и всякие другие) комбинации с различными орбитальными угловыми моментами. Спины кварков и антикварков могут различно ориентироваться друг относительно друга. Общий угловой момент составленной из кварков и антикварков частицы (античастицы) найдется по правилу векторного сложения спинового и орбитального моментов.

Допустим, что все кварки не обладают орбитальными угловыми моментами; все их моменты чисто спиновые. Допустим, далее, что спины кварка и антикварка направлены противоположно. Тогда каждая комбинация представляет мезон со спином $J = 0$. Наиболее легкие заряженные мезоны представляются комбинациями $u\overline{d}$ и $d\overline{u}$. Это — π-мезоны:

$$\pi^+ = u\overline{d}, \quad \pi^- = d\overline{u}. \tag{110.3}$$

Нейтральный π^0-мезон состоит из таких же кварка и антикварка. Но он представляет собой линейную суперпозицию состояний $u\overline{u}$ и $d\overline{d}$. Он с равной вероятностью может находиться в состоянии $u\overline{u}$ и в состоянии $d\overline{d}$.

Более тяжелые K-мезоны содержат кварк s и антикварк $\overline{s}$:

$$\mathrm{K}^+ = u\overline{s}, \quad \mathrm{K}^0 = d\overline{s}, \quad \overline{\mathrm{K}} : 0 = s\overline{d}, \quad \mathrm{K}^- = s\overline{u}. \tag{110.4}$$

K^+ и K^- являются частицей и античастицей по отношению друг к другу. То же относится к K^0- и $\overline{\mathrm{K}}^0$-мезонам, которые, таким образом, не являются истинно нейтральными частицами: K^0-мезон отличается от своей античастицы $\overline{\mathrm{K}}^0$ странностью. У $\overline{\mathrm{K}}^0$-мезона странность $+1$, а у K^0-мезона странность -1. С этим связано то, что K^0 в отличие от $\overline{\mathrm{K}}^0$ слабо поглощаются веществом. Дело в том, что странность всех барионов отрицательна или равна нулю. Поэтому поглощение K^0 веществом может происходить лишь с несохранением странности, т. е. из-за слабого взаимодействия. Масса $\mathrm{K}^\pm$-мезонов равна 493,7 МэВ, нейтрального K^0-мезона — 497,7 МэВ. Частицы, содержащие s-кварк, являются странными, а сам s-кварк называется *странным кварком*. Диагональные члены $u\overline{u}$, $d\overline{d}$, $s\overline{s}$ соответствуют нейтральным мезонам. Линейная суперпозиция двух первых из них, как уже отмечалось выше, представляет π^0-мезон. Двум остальным линейным суперпозициям $u\overline{u} + d\overline{d}$ и $s\overline{s}$ соответствуют истинно нейтральные частицы: η- и η'-мезоны с массами соответственно 548,8 и 957,6 МэВ.

Допустим по-прежнему, что орбитальных угловых моментов у кварков и антикварков нет, но их спины направлены параллельно. В этом случае возможны те же девять комбинаций кварков с антикварками. Однако спин каждой комбинации будет 1. Этим комбинациям соответствует девять более тяжелых мезонов:

ρ^+ ρ^- ρ^0	$\mathrm{K}^{*+}\mathrm{K}^{*0}\mathrm{K}^{*-}\overline{\mathrm{K}}^{*0}$	ω^0	φ^0
770 МэВ	892 МэВ	783 МэВ	1020 МэВ

3. Обратимся теперь к барионам. Спин их полуцелый. Следовательно, если кварки не обладают орбитальными моментами, то барионы должны быть построены из нечетного числа частиц. Рассмотрим случай, когда такими частицами являются кварки u, d, s. Спин бариона может быть равен либо $1/2$ (когда спины двух кварков параллельны, а спин третьего направлен противоположно), либо $3/2$ (когда спины

всех кварков параллельны). В первом случае образуется октет (восьмерка) барионов:

$$p(uud), \quad n(udd), \quad \Lambda^0(uds),$$
$$\Sigma^-(dds), \quad \Sigma^0(uds), \quad \Sigma^+(uus), \quad \Xi^-(dss), \quad \Xi^0(uss).$$

Барионы со спином $3/2$ образуют декуплет (десятку) барионов:

$ddd \quad udd \quad uud \quad uuu$	$\Delta^-\Delta^0\Delta^+\Delta^{++}$	1232 МэВ
$dss \quad uds \quad uus$	$\Sigma^-\Sigma^0\Sigma^+$	1385 МэВ
$dss \quad uss$	$\Xi^-\Xi^0$	1530 МэВ
sss	Ω^-	1672 МэВ

Ω^--гиперон, находящийся в вершине этого перевернутого треугольника, был предсказан на основании кварковой модели барионов и обнаружен экспериментально в 1964 г. Были предсказаны и его свойства — масса, странность, время жизни; все это подтвердилось на опыте.

4. Триумфом кварковой модели является открытие *очарованных частиц.* Первая очарованная частица была открыта в 1974 г. двумя группами экспериментаторов независимо друг от друга: на протонном ускорителе в Брукхейвене (США) при бомбардировке протонами ядер Be и на ускорителе со встречными электронно-позитронными пучками в Стэнфорде (США). Первая группа назвала открытую частицу мезоном J, а вторая — мезоном ψ. Поэтому обнаруженный мезон и получил двойное название J/ψ. Его масса 3,096 ГэВ. Замечательная особенность вновь открытой частицы состоит в ее относительно большой долговечности. Ее ширина 60 кэВ, тогда как обычные ширины для частиц таких энергий 10–70 МэВ, что примерно на три порядка больше. Этот факт, как и в случае странных частиц, указывает на запрет по какому-то новому квантовому числу. В результате было введено квантовое число C, получившее название *очарования*, или *шарма* (по-английски charm — очарование). Ему соответствует новый кварк c. В кварковой модели очарование определяется как разность между числом кварков (c) и антикварков ($\overline{c}$). Частицы с очарованием, отличным от нуля, называются *очарованными.* Очарование подобно странности сохраняется при сильных и электромагнитных взаимодействиях, но не сохраняется при слабых. Распады очарованных адронов происходит за счет *слабого взаимодействия*, при этом очарование меняется на единицу, J/ψ-мезон построен из кварка c и антикварка $\overline{c}$ ($J/\psi = c\overline{c}$). Его очарование равно нулю и относят его к числу мезонов со *скрытым очарованием.* Скрытое очарование у частиц проявляется в том, что они легко распадаются на очарованные частицы, если распад не запрещен законами сохранения энергии и импульса, а распады на неочарованные частицы сильно подавлены (так как подавлена аннигиляция $c\overline{c}$ в более мелкие кварки), т. е. происходят с малой вероятностью.

По своей структуре квантовая система $c\bar{c}$, называемая *чармонием*, напоминает атом водорода (скорее, позитрония). Однако последний всегда называется *атомом водорода* независимо от того, в каком из возбужденных состояний он находится. Напротив, различные возбужденные состояния чармония (и любых других кварковых систем) условились считать *различными частицами* и обозначать *различными символами*. Мезон J/ψ является одним из *возбужденных состояний чармония*. После него было открыто несколько возбужденных состояний чармония. Были открыты и мезоны с *явным очарованием*:

$D^+ - c\bar{d}$	$D^0 - c\bar{u}$	$F^+ - s\bar{c}$
$D^- - d\bar{c}$	$\overline{D}^0 - u\bar{c}$	$F^- - s\bar{c}$
1869 МэВ	1865 МэВ	2021 МэВ

Были открыты также и *очарованные барионы*.

5. История открытия нового кварка b аналогична истории открытия кварка c. В 1977 г. в Батавии (США) был открыт новый мезон, обозначенный через Υ. Он возникал при бомбардировке мишени из меди и свинца пучком протонов с энергией 400 ГэВ. Этот сверхтяжелый мезон при массе m_Υ характеризовался относительно малой шириной (около 0,04 МэВ). Свойства новой частицы не укладывались в схему четырехкварковой модели, и пришлось ввести пятый кварк b, который был назван *прелестным*, или *красивым*. (Адроны, в которые входит кварк b, называют *красивыми*, или *прелестными*.) Мезон Υ является одним из возбужденных состояний связанной системы $b\bar{b}$ со спином 1. В дальнейшем мезон Υ и другие возбужденные состояния той же системы $\Upsilon', \Upsilon'', \Upsilon'''$ получались во встречных электрон–позитронных пучках, а на встречных pp-пучках в ЦЕРН (Швейцария) был также обнаружен первый, самый легкий «красивый» барион $\Lambda_b = udb$ массой 5400 МэВ. Разность между числами b-кварков и их антикварков $\bar{b}$ называется *красотой*. *Красота сохраняется при сильных и электромагнитных взаимодействиях и может нарушаться при слабых*.

Если просмотреть все квантовые числа для адронов, то легко обнаружить, что электрический заряд адрона можно вычислить по формуле

$$Q = T_3 + \frac{1}{2}(B + S + C + b). \tag{110.5}$$

В частности, эта формула применима для нуклонов. У них $S = C = b = 0$, $B = +1$, для протона $T_3 = +1/2$, для нейтрона $T_3 = -1/2$.

6. Успешная классификация адронов на основе кварковой модели явилась веским аргументом в ее пользу. Другим аргументом являются опыты по прямому просвечиванию нуклонов и других адронов высокоэнергетическими электронами. Происходящий при этом процесс называется *глубоконеупругим рассеянием*. Теоретический анализ его привел к заключению, что внутри адронов электроны рассеиваются на точечных частицах с зарядами $+2/3$ и $-1/3$ и спином $1/2$. На этих частицах электроны резко меняют свои импульсы и энергии, сообщая их кварку, подобно тому как это происходит с α-частицами при рассеянии на ядрах

в опытах Резерфорда. О конечных размерах кварков эти опыты ничего не говорят. Из них можно заключить только, что во всяком случае на расстояниях от силового центра больше 10^{-15} см кварки ведут себя еще как бесструктурные точечные частицы. Возможно, что за пятым кварком b и предполагаемым шестым t будут открыты и другие кварки. Возможно, что увеличение числа кварков будет небольшим и ограничено общими принципами или внутренними свойствами симметрии элементарных частиц. Возможно, что и бесструктурность кварков отражает лишь достигнутый уровень исследования этих материальных объектов.

7. Многочисленные поиски свободных кварков, производившиеся на ускорителях высоких энергий, в космических лучах и окружающей среде, оказались безуспешными. В частности, кварки искали при измерении заряда капель по методу Милликена (1868–1953). Обнаружение капель с дробным зарядом свидетельствовало бы о присутствии кварка, так как дробный заряд кварка не мог бы быть компенсирован целочисленным зарядом электрона или иона. Такие опыты производились в наше время с гораздо большей точностью, чем это мог делать Милликен. К отрицательному результату привел и очень точный масс-спектроскопический анализ воды, который дал для верхнего предела отношения числа свободных кварков к числу протонов величину порядка 10^{-27}. Сейчас большинство специалистов склоняется к тому, что в свободном состоянии кварки не существуют.

Итак, кварковая модель предполагает, что кварки внутри адронов существуют, но опыт вынуждает признать, что вылететь оттуда и появиться в свободном состоянии они не могут. Такое положение называют английским словом «конфайме́нт», которое означает «пленение, тюремное заключение». Кварк, получивший энергию в результате столкновения с электроном, не вылетает наружу из адрона, а растрачивает ее на образование кварк-антикварковых пар, т. е. на образование новых адронов, в основном мезонов.

Одной из причин ненаблюдаемости кварков в свободном состоянии, возможно, являются их очень большие массы. Это препятствует рождению кварков при тех энергиях, которые достигаются на современных ускорителях. Не исключена возможность, что в свободном состоянии кварки принципиально не могут появиться в силу специфики их взаимодействия. Приводились доводы экспериментального и теоретического характера, согласно которым силы взаимодействия между кварками не ослабевают с расстоянием. Если это так, то для отделения их друг от друга требуются бесконечно большие энергии, при которых такое отделение делается невозможным. Все это — только предположения, ни в какой степени не претендующие на достоверность и даже правдоподобность.

8. Волновые функции систем кварков, образующих барионы, должны быть *антисимметричными*, так как кварки имеют спин $1/2$. Между тем встречаются барионы с одинаковыми ориентациями спинов:

$\Delta^{++}(u_\uparrow u_\uparrow u_\uparrow)$, $\Omega^-(s_\uparrow s_\uparrow s_\uparrow)$, которые *симметричны* относительно перестановки кварков, если только последние не обладают какими-либо дополнительными характеристиками. Такие характеристики и пришлось ввести. Они называются *цветами.* Каждый тип (аромат) кварка характеризуется тремя цветами. Им условно присвоили названия: *красный* R (red), зеленый G (green) и голубой B (blue). Антикваркам присвоили *антицвета* $\overline{R}$, $\overline{G}$, $\overline{B}$. Каждый антицвет является *дополнительным* к своему цвету, так что комбинации $R\overline{R}$, $G\overline{G}$, $B\overline{B}$ считаются *бесцветными* (*белыми*). Точно также бесцветной будет и комбинация $R\overline{R}+G\overline{G}+$ $+\,B\overline{B}$.

По современным представлениям сильные взаимодействия осуществляются путем обмена между кварками безмассовыми частицами со спином 1 и нулевой массой покоя. Эти частицы называются *глюонами.* Предполагается существование восьми глюонов, обладающих характеристикой «цвет». Обмен глюонами между кварками меняет их цвет, но оставляет неизменными все остальные квантовые числа (электрический заряд, странность, очарование, красоту), т. е. сохраняет аромат кварка. Каждый глюон содержит цвет и антицвет. Из трех цветов и антицветов можно составить всего девять комбинаций:

$$R\overline{R},\quad R\overline{G},\quad R\overline{B},$$

$$G\overline{R},\quad G\overline{G},\quad G\overline{B},$$

$$B\overline{R},\quad B\overline{G},\quad B\overline{B}.$$

Каждой из таких комбинаций соответствует глюон. Цвет, подобно электрическому заряду, сохраняется. Поэтому шесть недиагональных явно окрашенных комбинаций не могут перемешиваться между собой. Три диагональные комбинации бесцветны, и перемешивание их между собой не меняет цвет. Каждая из диагональных комбинаций может быть получена путем линейной суперпозиции двух остальных комбинаций. Существует, следовательно, всего восемь соответствующих им глюонов.

Представим схематически взаимодействие между кварками путем обмена глюонами. Будем изображать красный, зеленый и голубой кварки любого аромата символами q_R, q_G, q_B $(q \equiv u, d, s, c, b, t)$. Тогда взаимодействие между красным и зеленым кварками запишется так:

$$q_R \to q_G + q_{R\overline{G}},\quad q_G + q_{R\overline{G}} \to q_R.$$

Первое равенство означает, что красный кварк испустил красно-зеленый глюон и превратился в зеленый кварк q_G. Второе означает, что зеленый кварк, поглотив красно-зеленый глюон, превратился в красный кварк.

Изложенное имеет весьма формальный характер, а главное в общей физике не имеет каких-либо конкретных применений. Поэтому нет смысла заниматься этим дальше. Соответствующие вопросы относятся к специальной теоретической дисциплине — *квантовой хромодинамике* (КХД).

9. В заключение заметим, что к настоящему времени Вайнберг (р. 1915), Салам (р. 1926) и Глэшоу (р. 1932) создали теорию, в которой электромагнитное и слабое взаимодействия объединены в единое *электрослабое взаимодействие.* На расстояниях от силового центра, меньших радиуса действия слабых сил (10^{-16} см), различие между электромагнитными и слабыми взаимодействиями исчезает. На больших расстояниях, однако, сохраняется различие между ними, так что старая теория остается в силе. Мы не можем входить в рассмотрение этих вопросов по существу, так как они далеко выходят за пределы общего курса физики.

Отметим только, что за последние 15–20 лет в физике элементарных частиц произошли революционные открытия.

1. Достоверно установлено существование структуры у сильновзаимодействующих частиц — адронов (в том числе протона и нейтрона). Модель кварков не только объясняет систематику адронов, но и динамику их взаимодействия. На основе модели кварков предсказаны существование и свойства большого числа частиц, обнаруженных затем экспериментально.

2. Установлена общая природа электромагнитных и слабых взаимодействий. Открыты переносчики слабых взаимодействий: $W^{\pm}$- и Z^0-бозоны.

3. Создана теория сильного взаимодействия кварков — квантовая хромодинамика. Подтверждены экспериментально свойства переносчиков этих взаимодействий — глюонов (в частности, спин 1). Предсказания КХД (в той области, где применима теория возмущений) блестяще подтверждаются на опыте.

4. Поставлен вопрос о единой природе всех сил (Великое объединение).

5. Обнаружена глубокая связь между физикой элементарных частиц и космологией.

Таблицы

I. Лептоны ($J = 1/2$, $L = 1$, $B = 0$)

Частица	L_e	L_μ	L_τ	Масса, МэВ	Время жизни, с	Основные каналы распада
e^-	$+1$	0	0	0,511003	∞ ($> 2 \cdot 10^{22}$ лет)	
ν_e	$+1$	0	0	< 46 эВ	∞	
μ^-	0	$+1$	0	105,6592	$2{,}19713 \cdot 10^{-6}$	$e^- \overline{\nu}_e \nu_\mu$ (98,6 %)
ν_μ	0	$+1$	0	$< 0{,}52$	∞	$e^- \overline{\nu}_e \nu_\mu \gamma$ (1,4 %)
τ^-	0	0	$+1$	1784	$(4{,}6 \pm 1{,}9) \cdot 10^{-13}$	$\rho^- \nu_\tau$ (22 %) $\mu^- \overline{\nu}_\mu \nu_\tau$ (18 %) $\pi^+ \pi^- \pi^- \nu_\tau$ (18 %) $e^- \overline{\nu}_e \nu_\tau$ (16 %) $\pi^- \nu_\tau$ (11 %) $\pi^- \rho^0 \nu_\tau$ (7 %)
ν_τ	0	0	$+1$	< 250	?	?

II. Кварки

Кварк	Масса, МэВ	J^P	L	B	S	C	T	T_3	Q	Участие во всех взаимодействиях
u	≈ 5	$1/2^+$	0	1/3	0	0	$+1/2$	$+1/2$	$+2/3$	—
d	≈ 7	$1/2^+$	0	1/3	0	0	1/2	$-1/2$	$-1/3$	—
c	≈ 1350	$1/2^+$	0	1/3	0	$+1$	0	0	$+2/3$	участвует
s	≈ 150	$1/2^+$	0	1/3	-1	0	0	0	$-1/3$	во всех
t?	$> 22\,000$	$1/2^+$	0	1/3	0	0	0	0	$+2/3$	взаимо-
b	≈ 4700	$1/2^+$	0	1/3	$b = +1$		0	0	$-1/3$	действиях

III. Квазистабильные мезоны [1] ($J^P = 0^{-1}$, $L = 0$, $B = 0$)

Частица	Кварковый состав	S	C	T	T_3	Масса, МэВ	Время жизни, с	Основные каналы распада
$\pi^{\pm}$	$u\overline{d}(d\overline{u})$	0	0	1	±1	139,567	$2{,}603 \cdot 10^{-8}$	$\mu^{\pm}\nu_{\mu}(\widetilde{\nu}_{\mu})$ (100 %)
π^0	$u\overline{u}(d\overline{d})$	0	0	1	0	134,963	$0{,}83 \cdot 10^{-16}$	$\gamma\gamma$ (98,8 %)
η^0	$u\overline{u}, \overline{dd}, \overline{ss}$	0	0	0	0	548,8	$0{,}7 \cdot 10^{-18}$	$\gamma\gamma$ (39 %) $\pi^0\pi^0\pi^0$ (32 %) $\pi^+\pi^-\pi^0$ (24 %)
K^+	$u\overline{s}$	+1	0	1/2	+1/2	496,67	$1{,}237 \cdot 10^{-8}$	$\mu^+\nu_{\mu}$ (63,5 %) $\pi^+\pi^0$ (22,2 %) $\pi^+\pi^+\pi^-$ (5,6 %) —
K^0_S							$0{,}89 \cdot 10^{-10}$	$\pi^+\pi^-$ (68,8 %) $\pi^0\pi^0$ (31,4 %)
K^0	$d\overline{s}$	+1	0	1/2	−1/2	497,7	—	—
K^0_L							$5{,}18 \cdot 10^{-8}$	$\pi^{\pm}e^{\mp}\nu_e(\nu_e)$ (38,7 %) $\pi^{\pm}\mu^{\mp}\overline{\nu}_{\mu}(\nu_{\mu})$ (27,1 %) $\pi^0\pi^0\pi^0$ (21,5 %) $\pi^+\pi^-\pi^0$ (12,4 %)
D^+	$c\overline{d}$	0	+1	1/2	+1/2	1869	$\approx 9 \cdot 10^{-13}$	$\overline{K}^0X + K^0X$ ($\approx$ 50 %) $e^{\pm}X$ (20 %) K^-X (15 %) $\overline{K}^0\pi^+\pi^0$ (10 %)
D^0	$c\overline{u}$	0	+1	1/2	−1/2	1865	$\approx 5 \cdot 10^{-13}$	K^-X ($\approx$ 45 %) $\overline{K}^0X + K^0X$ ($\approx$ 30 %) $K^-\pi^+\pi^0$ (10 %)
F^+	$c\overline{s}$	+1	+1	0	0	2021	$\approx 2 \cdot 10^{-13}$	?

[1] Символом X обозначены другие частицы.

IV. Стабильные и квазистабильные барионы ($J^P = 1/2^+$, $L = 0$, $B = +1$)

Частица	Кварковый состав	S	C	T	T_3	Масса, МэВ	Время жизни, с	Основные каналы распада
p	uud	0	0	1/2	+1/2	938,279	$> 10^{32}$ лет	
n	udd	0	0	1/2	−1/2	939,573	925 ± 11	$\mathrm{pe}^- \overline{\nu}_\mathrm{e}$ (100 %)
Λ^0	uds	−1	0	0	0	1115,6	$2{,}63 \cdot 10^{-10}$	$\mathrm{p}\pi^-$ (64 %)
								$\mathrm{n}\pi^0$ (36 %)
Σ^+	uus	−1	0	1	+1	1189,4	$0{,}80 \cdot 10^{-10}$	$\mathrm{p}\pi^0$ (52 %)
								$\mathrm{n}\pi^+$ (48 %)
Σ^0	uds	−1	0	1	0	1192,5	$5 \cdot 10^{-20}$	$\Lambda\gamma$ (100 %)
Σ^-	dds	−1	0	1	−1	1197,3	$1{,}48 \cdot 10^{-10}$	$\mathrm{n}\pi^-$ (100 %)
Ξ^0	uss	−2	0	1/2	+1/2	1315	$2{,}9 \cdot 10^{-10}$	$\Lambda\pi^0$ (100 %)
Ξ^-	dss	−2	0	1/2	−1/2	1321,3	$1{,}64 \cdot 10^{-10}$	$\Lambda\pi^-$ (100 %)
Ω^-	sss	−3	0	0	0	1672,5	$0{,}8 \cdot 10^{-10}$	$\Lambda\mathrm{K}^-$ (69 %)
								$\Xi^0\pi^-$ (29 %)
								$\Xi^-\pi^0$ (8 %)
Λ_c^+	udc	0	+1	0	0	2282	$(1\text{–}2)\cdot 10^{-13}$	Наблюдалось 12 типов распадов
Λ_b^0	udb	0	0			~ 5400		

Именной указатель

Аббе 122
Абрикосов 388
Авогадро 17, 57, 295, 516
Аллен 370
Алиханов 479, 480
Алиханьян 479, 480
Аллен 480
Альварец 537, 538, 540
Ампер 80, 207
Андерсон 399, 737, 740, 740
Андроникашвили 369
Арцимович 655, 665

Бак 248, 427
Бальмер 63–66, 75–77, 90, 239–241, 249, 259, 262, 263
Бардин 382
Баркла 285, 286
Барнет 213
Басов 669
Беккер 556, 602, 603
Беккерель 390
Белл 707
Бергман 204
Бете 101, 103, 266, 516, 519, 523, 523, 683, 713
Бет 221
Биберман 110
Блох 258, 358–359, 362, 383, 516
Блэкетт 54, 740
Боголюбов 375, 376, 382,
Бозе 272, 322, 375, 375, 379, 381, 736
Больцман 256, 321, 327, 353, 489, 687, 688
Бор Нильс 65–69, 71, 72, 74, 76, 78, 79, 84, 88–91 95, 131, 133, 141, 153, 154, 186, 189, 201, 207, 211, 215, 217, 228–230, 243, 251, 255, 256, 259, 276, 277, 284, 287, 291, 313, 385, 417, 419, 420, 427, 431, 496, 516, 517, 586, 587, 595, 600, 609, 618, 627
Бор Оге 509
Борн 92, 111, 326, 333
Боте 47, 48, 556, 602, 603
Браун 713
Брейт 591
Бриллюэн 326, 340, 342–344, 348
Бройль де 92–103, 107, 109–119, 121, 124, 126, 128, 162, 163, 176, 180, 216, 217, 326, 327, 341, 354, 358, 496, 578, 586, 591, 596, 669
Брэгг Генри 91
Брэгг Лоуренс 30, 100, 102, 106, 107, 670, 675
Брэккет 64

Вавилов 40, 42, 44, 49, 88, 342, 344, 377, 531, 568, 569
Вайнберг 8, 765
Валравен 381
Ван-дер-Ваальс 318, 320, 321, 380
Ван-де-Грааф 535, 536, 547, 624, 625
Ван-Лёвен 207
Вейцзеккер 405, 406, 408, 410, 412, 470, 486, 497, 501, 608, 610
Векслер 543, 546
Вигнер 591
Видеман 88
Видерое 536, 537, 539
Вильсон 30–32, 51, 54, 297, 466, 517, 531, 561, 571, 603, 718, 740, 745
Владимирский 682
Волков 707
Вольта 168
Вуд 86, 88,
Вульф 30, 100, 102, 106, 107, 670, 675
Вустер 473, 474

Гайтлер 308, 312, 523
Гамильтон 273, 308, 743
Гамов 459, 461, 463, 466
Ган 486, 606, 607
Гартвиг 712

Гаудсмит 214
Гаусс 51
Гейгер 47–48, 50, 457, 464, 466, 561, 565, 567, 602, 603, 718, 719
Гейзенберг 92, 119–123, 125, 127, 137, 161, 177, 181, 190, 390, 397, 461
Гелл-Манн 754, 759
Гёпперт-Майер 506
Герке 417
Герлах 211, 213–217, 245, 254, 380, 429
Герц Генрих 14, 132, 221, 224, 225
Герц Густав 79, 80, 82–85,
Герцшпрунг 698–700
Герштейн 8
Гесс 716, 717
Гинзбург 42, 43,
Гиорсо 623
Глезер 571
Глэшоу 765
Гоккель 716
Гольданский 443
Горбань 8
Горьков 382
Гук 260

Даламбер 373
Дебай 27, 103, 105–107, 324, 326–330, 333, 337, 341, 346, 671
Де Гааз 213
Джермер 99–101, 103,
Джонс 370
Дирак 10, 92, 214, 230, 255, 263, 272, 322, 379, 381, 391, 527, 714, 736, 740
Добрецов 417, 418, 423, 424
Добронравов 46, 47
Доплер 34–37, 40, 42, 43, 344, 489
Дьюар 372
Дэвис 482, 483, 703
Дэвиссон 99–101, 103
Дюлонг 325, 329

Ельяшевич 245

Жолио-Кюри Фредерик 277, 442, 443, 603, 626

Завойский 257, 669
Зееман 243, 245–250, 260, 263, 420, 426–429
Зельдович 681
Зоммерфельд 68, 90, 198, 230
Зундман 90

Иваненко 391, 398
Иордан 92
Иоффе 46, 47
Ирншоу 56

Йенсен 506

Кадомцев 665
Камерлинг-Оннес 366, 371
Капица 367, 372 571
Карман 326, 333
Карно 658
Кеезом А. 366, 367, 371,
Кеезом В. 371
Кеплер 52
Кирхгоф 88
Клеве 291
Клейн 58, 59, 527
Книппинг 285
Кокрофт 576, 581
Кольхерстер 717
Комптон Артур 26–27, 29, 31, 32, 34, 92, 162, 525, 527, 529
Комптон Карл 20
Кондон 459, 461
Коуэн 480, 568, 651, 652
Кристофилос 552
Крониг 214
Крускал 666
Кузьмин 483
Кулон 52, 55, 56, 231, 265, 392, 412, 460, 464, 514
Купер 382, 383
Курант 552
Курчатов 486, 626, 643, 655, 662, 667
Кэнсман 99
Кюри (Жолио) Ирен 442, 443, 603, 607
Кюри Мария 390, 620
Кюри Пьер 390, 620

Лагранж 259, 743
Лайман 63, 64, 66, 67, 75–77, 239, 261
Ландау 368, 373–379
Ланде 246
Ланжевен 208
Лаплас 151, 259, 299, 439, 713, 714
Латтес 399

Лауэ 30, 103, 108, 285, 286, 671
Ледерман 742
Лежандр 434
Лейпунский 479, 480
Ленард 14, 16, 50
Леонтович 166, 461, 655, 665
Ливингстон 552
Лондон 308, 312, 385
Лоренц 12, 37, 207, 216, 217, 256, 542, 720
Лоуренс 541, 626
Лоусон 660, 661, 663, 668
Лукирский 20, 171
Луммер 417
Лэмб 264, 265
Любимов 478
Ляпунов 357–359, 363

Майкельсон 417
Мак Вуд 367
Мак-Миллан 543
Максвелл 10, 14, 97, 109, 128, 268, 324
Мандельштам 117, 166, 326, 339, 344, 348, 461
Марсден 50
Матьё 355
Мах 377
Мейсснер 387–388
Мейтнер Лизе 606608
Менделеев 55, 276, 277, 291, 498, 623, 629, 673
Мёссбауэр 45, 450, 487, 491–493
Милликен 20, 46, 171, 717, 763
Минковский 28, 33, 94, 95
Мозли 55, 291–292
Моттельсон 509
Мюллер 561, 565, 567, 718, 719

Намбу 382
Нернст 325, 328
Неттол 457, 464, 466
Неддермайер 399, 737
Нильсон 291
Нишиджима 754
Нишина 58, 59, 527
Новикова 8
Ньютон 37, 66, 96, 128, 710

Оганесян 627, 628, 629
Оже 297, 485
Оккиалини 399, 740
Оксенфельд 387–388
Онзагер 378
Оппенгеймер 707

Пакард 258
Парселл 258
Паули 214, 270, 272, 274–277, 303, 333, 352, 361, 375, 381, 383, 386, 416–418, 474, 475, 478, 496, 497, 502, 505, 703, 706, 714, 736, 740
Паунд 45, 258, 493
Пауэлл 399
Пашен 63, 64, 66, 75–77, 248, 427
Перл 737
Перо 77, 250, 417, 418
Петржак 443, 612
Пешков 373
Пикеринг 72, 73
Планк 9–10, 12, 18, 21, 22, 37, 42, 44, 90, 91, 93, 94, 128, 177, 190, 251, 320, 325, 327, 341, 348, 419, 461, 668
Пойнтинг 58, 219–221
Понтекорво 482, 702
Престон 248
Прилежаев 20
Прокошкин 742
Пти 325, 329
Пуассон 455
Пфунд 64

Раби 253
Рамзауэр 107
Рассел 231, 243
Ребке 45, 493
Резерфорд 50–52, 54–56, 126, 277, 390, 411, 412, 460, 461, 531, 534, 562, 602, 604, 763
Рейнес 480, 651, 652
Рейнольдс 367
Рейнуотер 509
Рессел 698, 699, 700
Риги 14
Ридберг 63, 68, 70, 72, 77, 78, 91, 202, 241, 268, 293
Ризерфорд 264, 265
Ритц 56, 63, 66, 67, 203
Ричардсон 20
Росси 718, 719
Рунге 247

Савич 607

Садовский 217, 220, 221
Салам 765
Саундерс 243
Сиборг 623
Сильвера 381
Скобельцын 31, 571, 718
Снайдер 494, 552
Стефан 327, 688
Стокс 89
Столетов 14, 16
Струтинский 609, 630
Сушкин 110

Тамм 42, 58, 398, 527
Тартаковский 8, 105, 107,
Теренин 417, 418, 423, 424
Тисса 368, 373
Томсон Дж.Дж. 14, 59, 105, 106, 527
Томсон Д.П. 105
Торри 258
Третьяков 478

Уилер 609
Уленбек 214
Уолтон 576, 581

Фабри 77, 250, 417
Фабрикант 110
Фарадей 77, 103, 268
Фаулер 72, 73
Фейнман 378
Ферми 168, 169, 272, 322, 379, 381, 383, 384, 391, 474, 475, 477–479, 572, 606, 607, 619, 621, 643, 703, 706, 714, 715, 729, 736
Физо 108
Флеров 443, 612, 624, 631
Фок 393
Франк Джеймс 79, 80, 82, 84, 85
Франк Илья Михайлович 42, 43,
Фрёлих 382
Френель 39, 676
Френкель 496, 609
Фридрих 285
Фриш Отто 607, 608
Фуко 9, 37
Фурье 109, 118, 174, 176

Хальвакс 14
Хартри 393
Хансен 258
Хелл 105–107, 671
Хилл 355–357
Хиллераас 302
Холл 362
Хоукинг 710
Хофштадтер 415
Хьюиш 707

Цвейг 759

Чандрасекар 705
Чебышев 141
Чедвик 390, 473, 602, 604, 605
Черенков 40, 42, 44, 49, 88, 342, 344, 377, 531, 568, 569, 741
Чэдвик 55

Шафранов 666
Шварцшильд 259, 709, 710, 713, 714
Шевченко 8
Шерер 103, 105–107, 671
Широков 394
Шкловский 700
Шредингер 73, 92, 95, 109, 128–131, 133, 136, 138, 139, 142, 144, 147, 150, 151, 153, 155, 157–159, 161, 163, 166, 175, 176, 179, 195, 196, 211, 259, 261, 273, 274, 285, 298, 308, 314, 322, 324, 331, 352, 354, 355, 358, 359, 363, 393, 432, 433, 462, 504, 505, 507
Шриффер 382
Штарк 259–263
Штерн 107, 211, 213–217, 245, 254, 380, 429
Штрассман 606, 607
Шюлер 417

Эйлер 685
Эйнштейн 10, 18–20, 37, 45, 91, 213, 272, 322, 324, 325, 375, 379, 381, 391, 402, 493, 494, 621, 691, 736
Эпштейн 259
Эллис 473, 474
Эрмит 141

Юдин 394
Юкава 398, 399
Юнг 336
Юри 76

Яноши 110

Предметный указатель

Автоионизация 296
Автофазировка 543
Адиабатическое приближение 309
Адроны 531, 737
Азотная катастрофа 430
Активная зона реактора 638
Активность 450
Актиниды 285
Акустическая (дебаевская) ветвь частот 339
Акцепторы 354
Альфа-распад 442, 455
— необлегченный 467
— облегченный 467
Альфа-частицы длиннопробежные 466
Америций 619
Амперовы молекулярные токи 207
Амплитуда рассеяния 674, 674
— — магнитного 681
Амплитудные коэффициенты 159
Аннигиляционные потери 532
Аннигиляция 528, 740
— двухфотонная 532
Антивещество 742
Антидейтрон 742
Антинейтрино
— мюонное 476
— таонное 476
— электронное 475
Антинейтрон 741
Антипротон 741
Антицвет 764
Античастицы 739
Аромат 759
Атом гелия 298
Атомная единица массы 392
Ахондриты 633

Барионный заряд 585, 750
Барионы 585, 738, 760
— очарованные 762
Безызлучательные переходы 66
Беккерель (единица) 450
Белые карлики 699, 706
Бериллиевы лучи 603
Берклий 620
Бета-распад 442, 467
— двойной 472
— обратный 480
— позитронный 443, 467
— электронный 443, 467
— — захват 443, 467
Бетатрон 547
Бланкет 657
Блеск звезды 695
Бозе-частицы (бозоны) 272
Бозоны 736
Боровский радиус 74, 156
Брегговский скачок 671
Бустер 551

Вакуум π-мезонов 266
— физический 266
— фотонный 266
— электронно-позитронный 266
Валентный электрон 199
Вектор поляризации 578
Векторные диаграммы 189
Великое объединение 765
Верхняя граница бета-распада 473, 706
Взаимодействия гравитационные 394, 735
— сильные 394, 735
— слабые 394, 735
— электромагнитные 394, 735
Вилка 745
Вириал и теорема вириала 684
Виртуальные кванты 397
— частицы 266, 397
Вихревая нить 378
Вихревые кольца 379
Внутренний потенциал 102

Внутренняя конверсия 485
— — коэффициент 485
— — парная 486
Водородный цикл 691
Водородоподобный атом 68
Возраст Земли 454
Волна встречная 539
— попутная 539
Волновая функция 114
Волновое число спектроскопическое 61
Волновой пакет 109, 118
Волны Блоха 358
— вероятности 111
— де Бройля (фазовые) 93
— — — групповая скорость 94
— — — — показатель преломления 95, 102
Восточно-западная асимметрия космических лучей 721
Вращательные энергетические уровни 441
Время восстановления счетчика 561
— жизни 451
Встречные пучки 555
Входной канал 576
Вырождение 197
Вырожденный газ 715
Выход реакции 556, 579
Выходной канал 577

Гамильтониан 179
Гамма-излучение 483
Гармоники (обертоны) 69
Гармонический осциллятор 138
Гелиевый цикл 693
Гелий ^{3}He 366
— I 367
— II 367
Генератор Ван-дер-Граафа 535
— перезарядный (тандем-генератор) 536
Гигантский резонанс 600
Гиперзаряд 755
Гипероны 738
Гиромагнитное отношение 210
— — для электрона 215
Глубина проникновения 160
Глубоконеупругое рассеяние 762
Глюоны 399, 764
Гравитационное смещение спектральных линий 493
Гравитационный коллапс 707, 709
— радиус 709
Гравитоны 396, 398, 736
Граница серии 64
Границы Ферми 714
Граничная длина волны нейтрона 680
— энергия нейтрона 680

Давление звука 341, 347
Движение инфинитное 138
— нормальное 368
— сверхтекучее 368
— финитное 137
Двухжидкостная модель He II 368
Двухпротонный распад 442
Дейтерий 73
Дейтрон 73
Деление ядра 606
Детекторы 560
— Гейгера 565
— — Мюллера 561, 565
— импульсные ионизационные камеры 561, 563
— ионизационные 561, 561
— полупроводниковые 561, 570
— пропорциональные 561, 564
— сцинтилляционные 561, 567
— трековые 560
— черенковские 561, 567, 568
— электронные 560
Дефект массы нейтрального атома 402
— — ядра 401
Диаграмма Герцшпрунга–Рессела 698
Дипольный магнитный момент ядра 438
Дисперсионные соотношения 342
Дифракционное рассеяние 578
Диэлектрики 352
Доноры 354
Дрейфовые (пролетные) трубки 536
Дуанты 541
Дублет 228, 229
Дырки в полупроводниках 353
— — электронных оболочках 289
Дырочная проводимость 354

Естественная ширина линии 489
— — спектральной линии 126

Жолиотий 277, 626

Загадка солнечных нейтрино 482
Закон Гейгера–Неттола 457
— кубов Дебая 328
— Мозли 292
— сохранения барионного заряда 580, 743, 749
— — изотопического спина 580, 743, 757
— — импульса 579, 743, 747
— — красоты (прелести) 743
— — лептонных зарядов 477, 580, 743, 749
— — момента импульса 579, 743, 749
— — очарования (шарма) 743
— — странности 743, 754
— — четности 433, 580, 755
— — электрического заряда 579, 743, 749
— — энергии 579, 743, 747
— $1/v$ 591
Законы дисперсии 359, 374
Запаздывающие нейтроны 444
Запрет интеркомбинаций 298
Запрещенные линии 437
— переходы 234, 238
Заряд лептонный 476, 750
— мюонный 750
— таонный 750
— ядра 53
Зарядовая независимость ядерных сил 757
— симметрия 392
Зарядовое число (порядковый номер элемента) 391
Зашлаковывание реактора 646
«Звезда» 742, 745
Звездная величина 695
— — абсолютная 696
— — визуальная 696
— — фотовизуальная 696
— — фотографическая (синяя) 696
Звезды главной последовательности 699
Звук второй 373
— первый 373
Зеркальные ядра 392
Зона Бриллюэна 334, 340
— валентная 351
— воспроизводства 648
— проводимости 351
Зоны запрещенные 351
— разрешенные 351

Излучение дипольное 203
— квадрупольное 203
— октупольное 203
— рекомбинационное 72
— сплошное 290
— температурное (равновесное) 88
— характеристическое 287
Изобары 391
Изомеры 486
Изотонический спин (изоспин) 756
Изотоны 391
Изотопическая инвариантность 392, 757
— ось 756
Изотопические мультиплеты 755
Изотопический эффект 381
Изотопическое пространство 756
— смещение спектральных линий 73
Изотопы 391
Инклюзивное сечение 577
Интерполяционная формула Дебая 329
Интерференционное отражение 99
Интерферометр Фабри–Перо 77, 250, 418
Ионизационные потери 511
Истинно элементарные частицы 737
Источники нейтронов 555
ИТЭР 668

Калифорний 620
Камера Вильсона 561, 571
— диффузионная 571
— дрейфовая 561
— искровая 561, 573
— криогенная 572
— пропорциональная 561, 565
— пузырьковая 561, 571
— стримерная 561, 574
Кандолюминесценция 89
Катодолюминесценция 89
Квадрупольный момент внешний 440
— — внутренний 440
Квазиимпульс 341, 342
Квазинейтральность 658
Квант 9
— магнитного потока 385
— магнитный дипольный 436
— электрический дипольный 436
— — квадрупольный 436

Квантование 67, 133
— водородоподобного атома 153, 195
— магнитного потока 384
— осциллятора 139
— пространственное 189
— ротатора 190
Квантовая жидкость 365
— хромодинамика 764
Квантовое число 68
— — внутреннее 227
— — главное 197, 198, 199
— — магнитное 190, 198, 199
— — орбитальное 190, 198, 199
— — проекции полного момента 227
— — радиальное 198, 199
— — спиновое 200
Квантовый ансамбль 112
— вихрь 378
Кварки 759
Кварковая модель адронов 758
Квартет 229
Квинтет 229
Классический радиус электрона 59, 527
Классы звезд 696
Комптоновская длина волны 28
Конвертор 680
Конденсат 376
Конденсация Бозе–Эйнштейна 375
Контактная разность потенциалов 168
Конус потерь 663
Конфайме нт 763
Коротковолновая граница рентгеновского излучения 21
Корпускулярно-волновой дуализм 11
Корреляции 383
Космические лучи 716, 717
— — вторичные 720
— — галактические 720
— — первичные 720
— — солнечные 720
Коэффициент воспроизводства 648
— газового усиления 564
— истинного поглощения 295
— ослабления 294
— отражения 160
— поглощения атомный 295
— — массовый 295
— — молекулярный 295
— пропускания 160, 165
Коэффициент размножения нейтронов 639
— рассеяния 60, 295
— теплового использования 640
Коэффициенты Френеля 159
Край поглощения 296
Красные гиганты 699
— карлики 699
— сверхгиганты 699
Красота (прелесть) 755, 762
Кратность (степень) вырождения 198
Кристаллы ионные 332
— ковалентные 332
— металлические 332
— молекулярные 332
Критерий Лоусона 660, 661
— Шафранова–Крускала 666
Критическая масса 641
— скорость 377
— энергия 524
Куперовская пара 383
Курчатовий 626
Кюри (единица) 450
Кюрий 620

Ларморовская частота 243
Лептоны 476, 737
Ливни 532
— широкие атмосферные 533, 726, 732
— электронно-позитронные 740
— электронно-фотонные каскадные 732
Линейный коэффициент поглощения 525
Ловушки ультрахолодных нейтронов 681
Лоренцевский триплет 243
Лоуренсий 626
Лэмбовский сдвиг 263
Люминесценция 86, 88
Люминофоры 88
Лямбда (λ)-линия 367
— точка 367

Магические числа 406, 498
— ядра — см. Ядра магические
Магнетон Бора 420, 211
— ядерный 420
Магнитная жесткость 718

Магнитное 2^L-поле 436
— поле сильное 245, 249
— — слабое 245, 249
Магнитные зеркала 681
— квадрупольные линзы 552
— ловушки 663
— моменты ядер 429
— пробки 663
Магнитный момент 208
— — электрона 255
— — — аномальный 256
— — — спиновый 214
— резонанс 251
Магнитогидродинамическая модель плазмы двухжидкостная 664
— — — одножидкостная 664
Масса инвариантная (покоя) 744
Массовое число 391
Массовый коэффициент поглощения 296, 297, 525
Мезоатомы 78
Мезоны 738, 759
— с явным очарованием 762
Менделевий 623
Металлы 352
Метод Дебая–Шерера–Хелла 103
— Лауэ 103
— магнитного резонанса 430
— носителя 482
— сравнения интенсивностей 423
Механика волновая 92
— квантовая 92
— матричная 92
Микротрон 546
Многофотонные процессы 66
Множитель Ланде 244
Модель испарения 593
Молекула водорода 307
Молекулярное (релеевское) рассеяние 294
Момент импульса 181
Мощность реактора 645
Мультиплет 229
Мультиплетность уровня 227, 233
Мультиполь 436
Мюонная молекула 314
Мюоны 415, 737

Накопительные кольца 555
— — нейтронные 682
Насыщение ядерных сил 402
Нейтрино мюонное 476
Нейтрино таонное 476
— электронное 475
Нейтронные волны 669
— звезды (пульсары) 707, 708
Нейтроновод 680
Нейтроны быстрые 637, 670
— вторичные 614
— высокоэнергетичные 670
— запаздывающие 615, 616, 644
— мгновенные 615
— релятивистские 670
— тепловые 637, 669
— ультрахолодные 669
Неоновый цикл 693
Нептуний 618
Нильсборий 277, 627
Нобелий 626
Нормальные колебания 319
— координаты 318
Нуклид 391
Нуклонное самосогласованное поле 503
Нуклонный потенциал 503
Нулевая энергия 137
Нулевые колебания 137
— матрицы 675
Нуль–нуль (0–0) переходы 235

Обертоны — см. Гармоники
Область Гейгера 565
Обменная энергия 305
Обменное взаимодействие 305, 312
Оболочка 278, 498
Образование пар 29
Обратная решетка 340
Обращение линий 293
Обращенная схема уровней энергии 288
Оператор 172–181
— Гамильтона 179
— импульса 177
— инверсии 432
— линейный 173
— магнитного момента 210
— перестановки 271
— спина фотона 215
— углового момента 181
— энергии 179
Операторный метод 172
Операторы антикоммутирующие 172
— коммутирующие 172

Оптимальная температура 661
Оптическая (борновская) ветвь частот 339
Ортоводород 315
Ортогелий 298
Ортосостояние 304
Основная частота 69
Остров стабильности 630
Отравление реактора 646
Отражатель 641
Очарование (шарм) 755, 761

Параводород 315
Парагелий 298
Парамагнетизм электронный 256
— ядерный 256
Парасостояние 304
Парциальное сечение реакции 586
Переносчики взаимодействия 737
Переходные элементы 284
Период (время) полураспада 451
Периодическая система элементов 276
Пинч-эффект 662, 664
Пионы 476
Плазменный шнур 664
Плотность вероятности 114
— — радиальная 156
Плутоний 619
Поверхность Ферми 383
Подбарьерный процесс 588
Позитроний 741, 78
Показатель преломления 596
— — нейтронных волн 676, 679
— цвета 696
Поле полоидальное 666
— тороидальное 666
Полиномы Чебышева–Эрмита 140
Полупроводники 353
Полуширина линии доплеровская 489
Поляризация вакуума 268
Поляризованные частицы 578
Порог реакции 581
Порядковый номер элемента 276
Постоянная Ляпунова 357
— Планка 10
— распада 450
— Ридберга 63, 70
— спин-орбитального взаимодействия 506
Постоянная тонкой структуры 75
— экранирования 293
Потенциалы возбуждения 82
— ионизации 82
— критические 82
Потенциальные барьеры 157
— ямы 142, 148
Потенциальный барьер 460
Правила запрета (отбора) 204
— отбора 421
Правило (метод) интервалов 422
— Престона 247
— Рунге 247
— сложения угловых моментов 194
— Стокса 89
— частот Бора 65, 132
Предравновесный распад 595
Преоны 734
Приближение ближайших соседей 334
Принцип антисимметрии 272
— комбинационный Ритца 62
— неопределенностей 119
— Паули 168, 275
— — обобщенный 276
— соответствия 68
— суперпозиции состояний 115, 130
— тождественности одинаковых частиц 271
Прицельное расстояние 53
Пробег 512, 518
— максимальный 519
— средний 519
Прозрачность (пропускаемость) барьера 160
Протий 73
Протозвезда 683
Протонный распад 442
Процессы нормальные 344
— переброса 344

Работа выхода 168
Радиатор 568
Радиационная длина 523
Радиационные поправки 268
— потери 519, 522
Радиоактивность 442
— искусственная 442
— наведенная 533
— основной закон 451
Радиоактивные семейства (ряды) 446

Радиоактивные ядра 442
Радиолюминесценция 89
Радиус ядра 411
Разрешение временно́е 561, 719
— пространственное 561
Реакторы гетерогенные 643
— гомогенные 643
— исследовательские 647
— экспериментальные 647
— энергетические 647
Реакторы-размножители (бридеры) 647
Редкие земли (лантаниды) 284
Режим критический 639
— надкритический 640
— подкритический 640
— стационарный 639
Резерфорд (единица) 450
Резерфордий 277, 626
Резонанс антиферромагнитный 253, 259
— диамагнитный (циклотронный) 253, 259
— ферримагнитный 253, 259
— ферромагнитный 253, 258
— электронный парамагнитный (ЭПР) 253
— ядерный магнитный (ЯМР) 253
Резонансная линия 64
Резонансные частицы (резонансы) 736
Рентгеновские лучи 285
Рентгенолюминесценция 89
Ротатор 190
Ротон 374

Самоподдерживающаяся реакция 641
Самосогласованное поле 355
Сверхновые звезды 712
Сверхпроводимость 381
— второго рода 388
— первого рода 388
Сверхпроводящее состояние 388
Сверхтекутесть ^{3}He 379
— ^{4}He 378
— атомарного водорода 380
Сверхтонкая структура спектральных линий 417, 231
Сверхэлементы 630
Сверхядра 630
Светимость звезды 695
Связь (j, j) 234
— ионная (гетерополярная) 308
— ковалентная (гомеополярная) 308
— нормальная (Рассела–Саундерса) 231
— химическая 307
Сериальный символ 247
Серия Бальмера 63
— Бергмана (фундаментальная) 204
— Брэккета 64
— главная 204
— диффузная (первая побочная) 204
— Лаймана 63
— Пашена 63
— Пикеринга 72
— Пфунда 64
— резкая (вторая побочная) 204
— спектральная — см. Спектральная серия
— Фаулера 72
Силы Ван-дер-Ваальса (дисперсионные) 318
— молекулярные 317
— обменные 305, 312
— химические (валентные) 317
Симметрия «вперед-назад» 589
Синглет 228, 229
Синхротрон 550
Синхрофазотрон 550
Система отсчета лабораторная (ЛС) 577
— — центра масс (ЦМ) 577
Скачок теплоемкости гелия 367
Скрытые параметры 112
Смешанное состояние 388
Собственные значения 133, 178
— функции 133, 178
Солнечный ветер 726
Соотношение неопределенностей 119, 120, 125
Составное (промежуточное) ядро (компаунд-ядро) 587
Состояние дипольное 436
— квадрупольное 436
— октупольное 436
— смешанное 578
— чистое 578
Спектр водорода 67
— дискретный 133
— непрерывный (сплошной) 133
— энергетический 67
Спектральная серия 63

Спектральные закономерности 61
— серии щелочных металлов 199
Спектры атомные 61
— линейчатые 61
— молекулярные 61
— полосатые 61
Спин-орбитальное взаимодействие 228–230
Спин фотона 217
— электрона 198, 211
— ядра 416,419, 429
Спиновая степень свободы 226
Спиновые координаты 274
Спиральность нейтрино 478
Спонтанное деление 442
Среднее время удержания 659
Статистика Бозе–Эйнштейна 272
— Ферми–Дирака 272
Стационарные состояния 130
Столкновения неупругие 81
— упругие 81
Странность 754
Странные частицы 753
Стримеры 574
Субгиганты 699
Субкарлики 699
Сфера Шварцшильда (горизонт событий) 709
Схема совпадений 718

Температура вырождения 715, 375
— дебаевская 329
Температурный коэффициент расширения гелия 367
Теория возмущений 299
— сверхпроводимости 381
— сверхтекучести 373
— теплоемкостей Дебая 324
Тепловыделяющие элементы (ТВЭЛы) 643
Теплоемкость водорода 317
Теплоноситель 645
Теплопроводность He II 371
— твердых тел 344
Термоэлектронная эмиссия 169
Термоядерная проблема 654, 656
Термы спектральные 62, 66
— — щелочных металлов 202
Ток насыщения 15
Токамак 665
Толстослойные фотоэмульсии 561, 573
Томсоновское (классическое) рассеяние 527
— сечение рассеяния 59
Тонкая структура спектра α-частиц 466
— — спектральных линий 229
— — — термов 226
— — — — водорода 238
Тормозное излучение 21, 286, 522
Точка поворота 134
Трансурановые элементы (трансураны) 617
Трансураны 276
Триболюминесценция 89
Триплет 229
Тритий 73
Тритон 73
Туннельный эффект 166, 462
Тушение свечения 87

Углеродный цикл 691, 693
Угловое распределение 577
Узлы 137
Упаковочный коэффициент 402
Управляемый термоядерный синтез 655
Уравнение Хилла 355
— Шредингера для стационарных состояний 131, 179
— — общее 128, 131, 179
Уровень Ферми 703
Ускоритель Видерое 536
— линейный 536
— на встречных пучках 555
— с бегущей волной 539
— циклический 541
Условие Брэгга–Вульфа 670, 100
— зажигания термоядерной реакции 660
— нормировки 114
— радиационного равновесия 453

Фазотрон (синхроциклотрон) 543
Ферми частицы (фермионы) 272
Фермий 621
Фермионы 736
Флюоресценция 88
Фокусировка 542
— жесткая (сильная) 552
Фононный газ 343
Фононы 340
Фонтанирование гелия 370

Формула Бете 519
— Бора 516
— Брейта и Вигнера 591
— Вейцзеккера 405
— Клейна–Нишины–Тамма 59
— Резерфорда 54
— Томсона 527
— Эйнштейна 18
Фосфоресценция 88
Фотолюминесценция 89
Фотон 10, 12
— в гравитационном поле 44
— — среде 37
Фотоэлектрический эффект 14
— — внешний 14
— — внутренний 14, 24
— — многофотонный 23
— — , низкочастотная граница 18
— — нормальный 22
— — обратный 21
— — селективный 22
Фотоядерные реакции 600
Функции нечетные 432
— четные 432
Функция антисимметричная 271
— симметричная 271

Характер уровня 436
Характеристика фотоэлемента 15
Характеристическое рентгеновское излучение 22, 287
Хемилюминесценция 89
Холодная эмиссия 169
Хондры 633

Цвет 764
Центробежный барьер 465
Цепная реакция 637
Циклотрон 541
— изохронный 553

Чандрасекаровский предел 705
Чармоний 762
Частицы истинно нейтральные 739
— квазистабильные 736
— нечетные 435
— очарованные 761
— резонансные 736
— со скрытым очарованием 761
— стабильные 736
— четные 435
Чередование интенсивности спектральных линий 316
Черная дыра 709
Четность 431
— внутренняя 435
— — несохранение при слабых взаимодействиях 433

Ширина спектра масс 737

Эйнштейний 621
Экзосфера 710
Экзотермическая реакция 581
Электрическое 2^L-поле 436
— поле сильное 263
— — слабое 263
Электролюминесценция 89
Электрон Оже 297
Электронно-позитронные ливни — см. Ливни
Электронно-фотонный каскадный ливень — см. Ливни
Электронные конфигурации 234, 278
Электронный слой 278
Электроны Оже 485
Электрослабое взаимодействие 765
Электроядерные процессы 601
Элементарные возбуждения 374
— частицы 733
Элементарный заряд 391
Эндотермическая реакция 581
Энергетическая щель 386
Энергетические зоны 351
Энергетический спектр космических лучей дифференциальный 725
— — — — интегральный 725
Энергия активации 612
— ионизации 82
— нейтронного сродства 659
— обмена — см. Обменная энергия
— реакции 581
— связи ядра 400
— — — удельная 402
— Ферми 168, 383
Эффект Вавилова–Черенкова 42
— долготный 720
— Доплера аномальный 43
— — нерелятивистский 35
— — релятивистский 36
— Зеемана 420, 427
— — простой 242
— — сложный 242

Эффект Комптона 26
— — обратный 29
— Мёссбауэра 487, 491
— механотермический (механокалорический) 370
— Оже 296
— Пашена–Бака 248
— Садовского 217
— термомеханический 370
— широтный 720
— Штарка квадратичный 260
— — линейный 260
Эффективное сечение дифференциальное 53, 577
— — интегральное (полное) 577
— — полное 54
Эффективность 561
— конверсионная 568

Ядерная (атомная) энергия 636
— модель капельная 496
— — оболочечная 498
Ядерное время 411
— гиромагнитное отношение 419
— топливо 636
Ядерные модели коллективные (сильное взаимодействие) 495
— — обобщенные 495, 509
— — одночастичные (независимых частиц) 495
— реакции передачи 595
— — подхвата 594
— — прямые 594
— — срыва 594
Ядерный магнетон — см. Магнетон ядерный
— магнитный резонанс 418
— реактор 636
— — природный 649
— фотоэффект 600
Ядерных сил насыщение 305
Ядра бета-стабильные 406
— дважды магические 405
— деформированные 497
— магические 405, 498
— нечетно-нечетные 405
— радиоактивные 394
— четно-четные 404

K-захват 443
$W^{\pm}$-, Z^0-бозоны 737

Учебное издание

СИВУХИН Дмитрий Васильевич

ОБЩИЙ КУРС ФИЗИКИ

Том V

АТОМНАЯ И ЯДЕРНАЯ ФИЗИКА

Редактор *Д. А. Миртова*
Оригинал-макет: *В. В. Худяков*
Оформление переплета: *А. А. Логунов*

ЛР № 071930 от 06.07.99
Подписано в печать 16.07.02. Формат 60×90/16
Бумага офсетная № 1. Печать офсетная
Усл. печ. л. 49. Уч.-изд. л. 55,85. Тираж 5000 экз. Заказ №

Издательская фирма «Физико-математическая литература»
МАИК «Наука/Интерпериодика»
117864 Москва, Профсоюзная, 90
E-mail: fizmat@maik.ru

Издательство МФТИ
141700 Долгопрудный, Московская обл., Институтский пер., 9

При участии издательской и книготорговой фирмы «ФИЗМАТКНИГА»
141700 Долгопрудный, Московская обл., Институтский пер., 9
Тел.: (095) 408-76-81, 409-93-28

Интернет-магазин физико-математической литературы:
http://www.fizmatkniga.ru

Отпечатано с диапозитивов
в РГУП «Чебоксарская типография № 1»
428019 Чебоксары, пр. И. Яковлева, 15

ISBN 5-9221-0230-3